

Contents lists available at ScienceDirect

Drug and Alcohol Dependence



journal homepage: www.elsevier.com/locate/drugalcdep

Loci on chromosome 20 interact with rs16969968 to influence cigarettes per day in European ancestry individuals

Pamela N. Romero Villela^{a,b}, Luke M. Evans^{a,c}, Teemu Palviainen^d, Richard Border^e, Jaakko Kaprio^d, Rohan H.C. Palmer^g, Matthew C. Keller^{a,b}, Marissa A. Ehringer^{a,e,f,*}

^a Institute for Behavioral Genetics, University of Colorado, Boulder, USA

^b Department of Psychology and Neuroscience, University of Colorado, Boulder, USA

^c Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, USA

^d Institute for Molecular Medicine Finland FIMM, University of Helsinki, USA

^e Departments of Neurology and Computer Science, University of California, Los Angeles, USA

^f Department of Integrative Physiology, University of Colorado, Boulder, USA

^g Behavioral Genetics of Addiction Laboratory, Department of Psychology, Emory University, Atlanta, GA, USA

ARTICLE INFO

Keywords: rs16969968 Smoking Interaction Cigarettes per day Magma

ABSTRACT

Background: The understanding of the molecular genetic contributions to smoking is largely limited to the additive effects of individual single nucleotide polymorphisms (SNPs), but the underlying genetic risk is likely to also include dominance, epistatic, and gene-environment interactions. Methods: To begin to address this complexity, we attempted to identify genetic interactions between rs16969968, the most replicated SNP associated with smoking quantity, and all SNPs and genes across the genome. Results: Using the UK Biobank European subsample, we found one SNP, rs1892967, and two genes, PCNA and TMEM230, that showed a significant genome-wide interaction with rs16969968 for log10 CPD and raw CPD, respectively, in a sample of 116 442 individuals who self-reported currently or previously smoking. We extended these analyses to individuals of South Asian descent and meta-analyzed the combined sample of 117 212 individuals of European and South Asian ancestry. We replicated the gene findings in a meta-analysis of five Finnish samples (N=40 140): FinHealth, FINRISK, Finnish Twin Cohort, GeneRISK, and Health-2000–2011. Conclusions: To our knowledge, this represents the first reliable epistatic association between single nucleotide polymorphisms for smoking behaviors and provides a novel direction for possible future functional studies related to this interaction. Furthermore, this work demonstrates the feasibility of these analyses by pooling multiple datasets across various ancestries, which may be applied to other top SNPs for smoking and/or other phenotypes.

1. Introduction

Smoking cigarettes is the leading cause of preventable death in the United States (Alberg et al., 2014). One in five deaths in the United States can be attributed to smoking (Alberg et al., 2014). Smoking also affects the economy; smoking-related health costs are around \$300 billion per year in the United States alone (S. Federal Trade Commission (FTC), 2019). Previous work has demonstrated a substantial genetic component to smoking behaviors, and twin studies estimate the heritability of smoking quantity and nicotine dependence to be between 40% and 75% (Kaprio, 2009; Lessov-Schlaggar et al., 2006) in adults across multiple ancestries. Recent genome-wide association studies

(GWAS) have identified several hundred individual variants associated with various smoking-related behaviors (Liu et al., 2019), but the majority of the SNP heritability for smoking quantity remains to be accounted for (Evans et al., 2021; Quach et al., 2020). Moreover, the variants within these genes and their regulatory elements are likely to influence a complex trait via minute perturbations across a complex, non-linear set of physiological networks (i.e., transcriptional, neuronal, and developmental) (Kauffman, 1993). The physiological intricacy in which complex traits such as smoking develop suggests that interactions between loci or whole genes (i.e., epistasis) are likely, since there are numerous ways and stages at which these interactions could arise. Furthermore, while current evidence of epistatic effects in humans has

https://doi.org/10.1016/j.drugalcdep.2024.111126

Received 20 October 2023; Received in revised form 12 January 2024; Accepted 7 February 2024 Available online 15 February 2024 0376-8716/Published by Elsevier B.V.

^{*} Correspondence to: Institute for Behavioral Genetics, Boulder, CO 80303, USA. *E-mail address:* marissa.ehringer@colorado.edu (M.A. Ehringer).

been limited (Hill et al., 2008), work on model organisms further suggests that epistatic effects are common (Mackay, 2013) and may be particularly important for predicting an individual's genetic risk to disease such as nicotine dependence (Mackay and Moore, 2014).

To detect novel variants and improve our understanding of the biological processes involved in nicotine dependence, we investigated SNP-SNP interactions influencing nicotine use. However, testing all pairwise, genome-wide SNP-SNP interactions is computationally infeasible and hampered by the stringent multiple testing correction required for such analyses. Instead, by reducing the total number of tests by selecting a SNP or set of well-replicated SNPs of large effect (which are more likely to harbor interactions) with known functional impact as interactors, studying epistasis is possible. We therefore selected a well-replicated SNP previously associated with nicotine behaviors as our moderator, namely, rs16969968.

SNP rs16969968 in the CHRNA5/A3/B4 gene cluster of neuronal nicotinic receptor genes is the most widely replicated genetic variant associated with smoking behaviors (Chen et al., 2018; Picciotto and Kenny, 2021; Wen et al., 2016), emerging from early GWAS studies of lung cancer and smoking behaviors (Amos et al., 2008; Hung et al., 2008; Thorgeirsson et al., 2008). Nicotine is an agonist for neuronal nicotinic acetylcholine receptors (CHRN genes) and repeated nicotine use leads to their upregulation (Fowler et al., 2020). rs16969968 was the original top SNP identified in the CHRNA5 gene and has been the major focus of further study because it changes an amino acid (aspartate to asparagine; D398N) and has been shown to confer functional effects using cell culture methods in vitro (Bierut et al., 2008) and behavioral effects in a mouse genetic model (Buck et al., 2021; Koukouli et al., 2017; O'Neill et al., 2018). Absent balanced cross-over interactions, interaction effects are likely to be associated with at least some additive effects estimated in a typical GWAS single locus, additive model. If epistatic interactions do underlie any of the variation for heaviness of smoking or nicotine dependence, rs16969968 is therefore a reasonable a priori candidate locus to study as an interactor. In addition, this SNP is relatively common in some ancestral groups, with a minor allele frequency (MAF) of 0.37-0.43 in populations of European and Middle Eastern descent according to dbSNP (NCBI, 1999). Although rs16969968 is rarer in other ancestral groups such as East Asian and African, with MAF of 2% and 7% respectively (Bierut et al., 2008), this SNP has been associated with smoking behaviors in trans-ancestry analyses (Adjangba et al., 2021; Olfson et al., 2015). In sum, because rs16969968 is a highly replicated, trans-ancestral, and common signal of large effect with known functional consequences on smoking, we hypothesized that G×GWAS investigations using rs16969968 would be better powered than other SNPs in our search for epistatic effects influencing nicotine use.

Further investigation of statistically independent SNPs within the CHRNA5/A3/B4 gene cluster has previously suggested that rs16969968 moderates the effect of other SNPs on nicotine use. In a meta-analysis of smoking quantity led by Saccone et al., the authors identified at least two signals within the region that are statistically independent of rs16969968, tagged by rs578776 and rs588765 (Saccone et al., 2010). The major (risk) allele of rs578776 is in phase with the minor (risk) allele of rs16969968. In the case of rs16969968, the minor allele increases risk for nicotine dependence, but for rs578776 the minor allele is protective against it. Consequently, although the risk loci are correlated with each other, the minor alleles are out of phase, and when controlling for rs16969968, rs578776 is no longer genome-wide significant. At a second SNP, rs588765, in linkage disequilibrium (LD) with rs16969968, the minor allele is protective in single-locus models. However, when controlling for rs16969968, the minor allele is associated with an increased risk for smoking quantity (Saccone et al., 2010). In addition, a more recent study found that women who were carriers of the rs16969968 risk allele had increased odds of stopping smoking if they had the minor allele of CHRNA3 SNP rs578776 (Jones et al., 2023). In short, previous studies have demonstrated that controlling for

rs16969968 or including it in interaction models has the potential to uncover new associations with smoking behaviors and provide further nuance to previously discovered ones.

However, to date, no study has tested for interactions of rs16969968 with other genetic loci on smoking intensity using a systematic, hypothesis-free approach. Using open-sourced software, we developed a two-step approach to explore genome-wide interactions across multiple levels of analysis, namely, at the single SNP and gene-level. This approach is flexible – allowing us to pool data from multiple sources or ancestries as well as granular – the SNP-level results from step 1 can be used to pinpoint the specific region driving any significant gene-level interactions. Using this approach, we investigated genome-wide interactions with rs16969968 at the single SNP and gene-level underlying smoking quantity.

2. Material and methods

2.1. Discovery sample: UK Biobank individuals of European and South Asian ancestry

We conducted our primary analyses in the UK Biobank (Sudlow et al., 2015), a biorepository with approximately 500 000 individuals. The initial analysis was limited to individuals of European ancestry, as detailed below. Following our replication, we included individuals of the second largest ancestry group in the UKB, namely, unrelated South Asian ancestry individuals. Details about selecting unrelated individuals of South Asian ancestry and meta-analyzing across the European and South Asian subsamples can be found under Supplementary Methods.

All unrelated participants of European ancestry who reported currently or formerly smoking and had genotype data that passed quality controls were used (N_{European}= 116 442). Participants were 40 years of age or older. Around 46% of our sample of unrelated individuals who reported smoking were female. Different ancestral populations differ in their allele frequencies; these allele frequency differences can increase false positives or decrease power in GWAS (Tian et al., 2008). To minimize such confounding, we performed all analyses only on individuals of the same ancestral population, namely, European (Euro) or and South Asian (S_Asian) descent, and then meta-analyzed across them. To identify individuals of European ancestry, we performed principal component analysis and retained those whose top scores on the first four principal components fell within the range of European ancestry previously determined by the UK Biobank (field 22006). Similar details for identifying individuals of South Asian ancestry can be found under Supplementary Methods.

All data analysis and cleaning were performed using PLINK2 (Chang et al., 2015). We first removed 849 individuals whose self-reported sex differed from their chromosomal sex determination (UKB data fields 31 and 22001) due to their increased probability of being a sample mix-up, 46 people with irregularly high inbreeding coefficients ($|F_{het}| > 0.2$), and 159 individuals who requested their information be redacted from the UKB, as well as 1 029 individuals whose genetic data did not pass quality controls identified by Affymetrix (549 individuals) and the UK Biobank (480 individuals, fields 22010 and 22051). Then, we used MAFand LD-pruned array markers (plink2 command: -maf 0.01 -hwe 1×10^{-8} –indep-pairwise 50 5 0.2) to identify unrelated individuals among all individuals of a given ancestry who reported smoking using the Genome-wide Complex Trait Analysis (GCTA) software (Yang et al., 2011) (gcta command: -grm-singleton 0.05). For our analyses, we used the HRC-imputed dosage data provided by the UK Biobank's full release, which used the HRC reference panel v.1.1 (McCarthy et al., 2016) and an information score greater or equal to 0.9. We filtered MAF > 1% and tested ~ 10 M SNPs across the 22 autosomal chromosomes.

Smoking quantity was measured by CPD; we included individuals who reported previously or currently smoking (UKB field IDs 2887, 3456, and 6183; average = 18.22, median = 20, range 1-140, inclusive). Overall, most people tend to underestimate the amount they smoke, and

this is particularly pronounced in individuals who report formerly smoking in whom telescoping can partly explain why our measure of smoking quantity was right-skewed (Krall et al., 1989) (Fig. S1A). To assess whether changes to scale influence the tests of the interactions, we also investigated log10 transformed CPD (Fig. S1B).

2.2. rs16969968×SNP analyses

All models included the following covariates: sex (UKB field 31), age at time of assessment (field 21003), genotyping batch (field 22000), assessment center (field 54), and the first 10 genetic principal components generated across the UKB to control for population- stratification. We accounted for the two different SNP chips (Bioleve and Axiom) by controlling for batch, and for local geographical effects by controlling for assessment center. To calculate these 10 genetic PCs, we used flashpca (Abraham and Inouye, 2014) on common LD-pruned array markers. To reduce collinearity in our full set of covariates, we ran principal component analysis using the prcomp function in R (R Foundation for Statistical Computing, 2018) to remove the axes that explained less than 1% of the total variance. Previous research has shown that failing to include the interactions between a moderator and covariates can inflate estimates of an interaction (Keller, 2014). As such, we also included all interactions between rs16969968 and our covariates. and each additional interacting SNP and the covariates.

We used PLINK2 to run a linear regression model (plink2 command: –linear interaction) to estimate SNP_j-by-rs16969968 interaction associations with CPD. We included all rs16969968×covariate and SNP_j×covariate interactions to avoid potential confounding (Keller, 2014). Because covariate scales varied widely, all covariates and their products were standardized (plink2 command: –covar-var-iance-standardize). We used a standard GWAS threshold of 5×10^{-8} for this analysis. Our regression model took the following form:

$$CPD = \beta_{0} + \beta_{1}G + \beta_{2}Z_{j} + \beta_{3}G * Z_{j} + \sum_{p=1}^{q}\beta_{p}X_{p} + \sum_{p=1}^{q}\beta'_{p}X_{p} * G + \sum_{p=1}^{q}\beta'_{p}X_{p} * Z_{j} + \varepsilon$$
(1)

Where X_p indicates the 1.q covariates, G indicates the number of risk alleles at rs16969968, Z_j indicates the jth SNP in the G×GWAS, ε denotes environmental noise and measurement error.

2.3. rs16969968×gene analyses

To investigate rs16969968 interactions with gene level effects, we fed the resulting rs16969968-by-SNP_i p-values into the multi-marker analysis of genomic annotation (MAGMA) (de Leeuw et al., 2015) v.1.09 to test gene-level interaction associations for CPD and log10-transformed CPD. Using MAGMA, one can employ either the "SNP-wise mean" or the "SNP-wise top" model to aggregate genome-wide signals at the gene level. The SNP-wise mean model is more powerful when several SNPs within a gene show a moderate association with the outcome of interest; the SNP-wise top model, on the other hand, is more powerful when a single SNP is strongly associated with the trait (de Leeuw et al., 2018, 2016). To ensure our analyses would be sensitive to varying unknown genetic architectures, we used both MAGMA's top and mean p-value models separately (MAGMA commands -model SNP-wise top and -model SNP-wise mean, respectively). To our knowledge, this was the first time MAGMA has been used to perform G×GWAS interaction analyses. We investigated the likelihood of getting spurious results from using MAGMA in this novel fashion by simulating a random phenotype and running our rs16969968×SNP and subsequently our rs16969968×Gene analyses genome wide (See Supplementary Methods). While we did see slight deflation of the p-values in the SNP-wise top model, no genes were significant after controlling for multiple testing via a Bonferroni correction in either model (Fig. S9).

In all the MAGMA analyses, variants were annotated to genes using a 25Kb window around the start and end of a gene. SNPs were successfully mapped onto a total of 18 573 genes using genome build 37. We used the SNP×rs16969968 interaction *p*-values for each SNP from the original GWAS, which accounted for the appropriate main effects, covariates and covariate interactions as described above, and included MAGMA's default covariates in the analysis (gene size, density, inverse minor allele count, per-gene sample size, plus the log value of each). We used a Bonferroni multiple testing correction significance threshold based on the number of genes tested ($p = 0.05/18573 = 2.70 \times 10^{-6}$), which is conservative given LD structure and overlapping gene regions.

2.4. Finnish replication sample

To replicate any significant interactions, we chose five Finnish subsets with genetic and cigarette use data available as a replication sample (N = 40 140). These five subsets include: FinHealth 2017 study (Fin-Health) (National FinHealth Study - THL, n.d.), FINRISK (The National FINRISK Study, n.d.), Finnish Twin Cohort (FTC) (Kaidesoja et al., 2019; Kaprio et al., 2019), GeneRISK (Widén et al., 2022), and the Health-2000–2011 (T2000–2011) (Health-2000–2011, n.d.). These datasets varied in sample size (ranging from around 994 smoking individuals in GeneRISK up to 26 751 in FINRISK) and the granularity of the cigarette use outcome (i.e., FTC used binned CPD while the rest of the subsets used raw CPD). For more information on these samples, please see Supplementary Methods. We confirmed our Finnish sample was an appropriate replication sample by comparing the Finnish linkage disequilibrium patterns of any gene regions of interest to those in our original UKB European sample, the largest sample in our study (Fig. S8).

To replicate any significant SNP or gene signals from the UKB analysis using the Finnish samples, we defined a replication region as all SNPs within 250 kb of the lead SNP in a significant interaction from the discovery analyses. This ensured that all SNPs in common between the Finnish and UKB samples in our region of interest plus any new SNPs that were likely to be in linkage disequilibrium with our SNPs of interest would also be included.

We performed rs16969968×SNP and rs16969968×Gene interaction analyses for any replication regions in each of the five Finnish samples as described previously (see 2.2 rs16969968×SNP Analyses and 2.3 rs16969968×Gene Analyses). We meta-analyzed the results from the rs16969968×SNP analyses across only the Finnish subsets (labeled Fin_Meta-analysis), across all European samples (labeled Euro_Metaanalysis), and trans-ancestrally across the UKB and Finnish samples (labeled All_Meta-analysis) using METAL's inverse variance weighing model (Willer et al., 2010).

To determine the number of independent tests conducted in the replication analyses, we performed principal component analysis using the UKB on all the SNPs within any replication regions of interest using R. To identify the maximum number of independent signals in a replication region, we ran principal component analysis on the genotypes of all the SNPs in our region of interest and counted the number of principal components whose standard deviations were greater than 1. To determine a significance threshold for any replication SNPs, we then adjusted for multiple testing in each region by dividing 0.05 by the effective number of independent loci the PCA analysis revealed. For example, the number of independent loci in our region of interest on chromosome 20 was 3; dividing 0.05 by 3 yielded a corrected alpha of 0.017 (0.05/3) for replicating our SNPs of interest on chromosome 20.

2.5. Characterizing significant interactions

For any statistically significant genes from the gene-level MAGMA analysis ($p < 2.70 \times 10^{-6}$), we sought to understand what drove any significant rs16969968×Gene interactions. To do this, we inspected the linkage disequilibrium patterns and performed conditional and functional analyses on any SNPs of interest within any significant genes.

SNPs of interest within significant genes were defined as SNPs with a suggestive significance of $p < 1 \times 10^{-5}$, which is about one order of magnitude lower than our gene-level significance of $p < 2.70 \times 10^{-6}$. This ensured that our follow-up analyses included all SNPs that might be driving any significant interactions observed at the gene level.

We used HaploView (Barrett et al., 2005) as well as LocusZoom (Pruim et al., 2010) to visualize the linkage disequilibrium pattern of the SNPs of interest for any genes that reached statistical significance. To test whether a significant gene contained a single or multiple signals, we conducted rs16969968×SNP interaction analyses on all SNPs of interest while conditioning on the top SNP for that gene. Our multiple testing correction threshold for these conditional analyses was defined by the number of effectively independent SNPs within a significant gene (see Section 2.4 Finnish replication sample). To test the interactive effect of each SNPs in the gene with rs16969968 while conditioning on the top SNP and its rs16969968 interaction, we exported the additive coding of all SNPs in the gene within MAGMA's 25 kb window using PLINK (plink flag: –recode A), and included in the conditional model the interaction between the top SNP and rs16969968 as well as the main effect of the

top SNP and its interactions with the rest of our covariates. The conditional analysis followed the following regression model:

$$CPD = \beta_0 + \beta_1 G + \beta_2 Z_{top} + \beta_3 Z_j \text{ within gene region} + \beta_4 G * Z_{top} + \beta_5 G * Z_j \text{ within gene region} + \sum_{p=1}^q \beta_p X_p + \sum_{p=1}^q \beta_p X_p * G + \sum_{p=1}^q \beta_p X_p * Z_{top}$$
(2)

3. Results

We used the UKB European and South Asian samples as our discovery samples. Since our largest discovery subsample was of European descent, we then used the Finnish samples to replicate any significant results. For conciseness, the main text will focus on the results from the UKB European discovery subset and replications using the Finnish samples. For additional results on the South Asian subsample or the meta-analyzed European and South Asian discovery sample, please see *Supplementary Results*.



Fig. 1. rs16969968×SNP analysis influencing cigarettes per day in the UK Biobank's European subset. No SNP reached genome-wide significance ($p < 5 \times 10^{-8}$) when analyzing the UKB's European subset for rs16969968×SNP interactions for cigarettes per day. Blue line denotes suggestive ($p < 1 \times 10^{-5}$) significance.

3.1. rs16969968×SNP analyses

In the UKB European subset, no SNP reached genome-wide significance for CPD (Fig. 1), but one SNP on chromosome 11, rs1892967, was significant for log10-transformed CPD ($p < 5 \times 10^{-8}$, $p = 3.18 \times 10^{-8}$, Fig. S6A). For rs16969968 and rs1892967's allele frequencies across our European subset, please consult Table S2.

3.2. rs16969968×gene analyses

We used MAGMA to aggregate the resulting *p*-values from the rs16969968×SNP analysis by gene to detect any potential gene-level interactions with rs16969968 ($p < 2.63 \times 10^{-6}$). In both the SNP-wise Mean and Top models, we found the *PCNA* gene to significantly interact with rs16969968 for CPD in Europeans (Fig. 2A, $p = 8.02 \times 10^{-7}$; Fig. 2B, $p = 3.67 \times 10^{-7}$, respectively). No genes reached genome-wide significance for \log_{10} CPD in the European subset (Fig. S6B-C), although *PCNA* neared suggestive significance (Fig. S6B, $p = 2.71 \times 10^{-5}$; Fig. S6C, $p = 2.21 \times 10^{-5}$). Notably, genes containing or near top SNP rs1892967 on chromosome 11 were insignificant across the SNP-wise Mean (p = 0.20, Fig. S5B) and Top (p = 0.0098, Fig. S5C) models for log10 CPD.

3.3. Finnish replication

3.3.1. Finnish replication of rs16969968×SNP analyses

Using the Finnish replication samples, we additionally tested this

Manhattan Plot of Magma SNP-wise MEAN Analysis for CPD Q-Q plot of Magma SNP-wise MEAN Analysis of CPE 10 12 11 13 14 15 16 17 **B)** SNP-wise Top model for CPD: Plot of Magma SNP-wise TOP Analysis for CPD Q-Q plot of Magma SNP-wise TOP Analysis of CPD 11 12 13 15

Fig. 2. (A) Two genes, *PCNA* and *TMEM230*, reached significance ($p < 2.63 \times 10^{-6}$) for interacting with rs16969968 influencing CPD after adjusting for multiple testing when using the SNP-wise Mean model in MAGMA. (B) One gene, *PCNA*, reached significance ($p < 2.63 \times 10^{-6}$) for interacting with rs16969968 influencing CPD after adjusting for multiple testing when using the SNP-wise Top model in MAGMA. All gene analyses shown above are using only the UKB European subset.

5

region tagged by rs73586411 on chromosome 20. When meta-analyzing across the Finnish samples, for purposes of multiple testing correction, we identified four independent tests in our SNPs of interest using principal component analysis (see 2.4 Finnish Replication Sample). Across the Finnish subsets, nine SNPs were nominally significant (p < 0.05Table S1), but no SNP reached significance after adjusting for multiple comparisons ($p < 0.05/4 = 1.25 \times 10^{-2}$). We did not detect evidence of study heterogeneity across the Finnish subsets (p < 0.05, all p > 0.278, Table S1). When meta-analyzing across all Finnish and UKB European subsets, no SNP reached genome-wide significance, but the interaction between rs1696969 and rs73586411 remained suggestively significant $(p < 5 \times 10^{-8}, p = 6.5 \times 10^{-6}, \text{ Table S1})$. We did not detect evidence of study heterogeneity between the UKB and Finnish samples (p < 0.05, p = 2.42×10^{-1} , Table S1). Fig. 3A shows the estimated effect sizes for this interaction within individual samples and across all samples where rs73586411 was available. The estimated effect size for the rs16969968×rs73586411 interaction was consistently negative across the four meta-analyses and four of the seven samples (Fig. 3A), spanning UKB European ancestry, South Asian ancestry, and Finnish ancestry. For example, while underpowered, we note that the direction of the interaction effect between rs1696969 and rs73586411 for the South Asian UKB sample is consistent with the European samples (Fig. 3A) and its standard error for the interaction is smaller than in the European samples because rs73586411's effect allele is more common in South Asians than in individuals of European descent (Table S2).

A) SNP-wise Mean model for CPD:



5 5.2 Position on chr20 (Mb) 5.4

Fig. 3. (A) Estimated rs16969968 \times rs735864111effect sizes, alongside their standard error for those estimates across samples. The sample size of each sample is denoted in parentheses; samples are ordered according to decreasing sample size. (B) Locus Zoom plot for region of interest, tagged by rs73586411.

4.8

3.3.2. Finnish replication of rs16969968×gene analyses

We used the Finnish samples to replicate our significant interaction between rs16969968 and the *PCNA* gene. The strongest SNP-level interaction associations in this region were physically located within the *CDS2* gene, but were within 25KB of *PCNA* and *TMEM230*, and therefore included in MAGMA gene analyses for all three genes. We consequently included all three genes in our Finnish replication study. In our rs16969968×Gene meta-analysis of the Finnish samples, all three genes (*CDS2*, *TMEM230*, and *PCNA*) were significant after multipletesting correction ($p < 1.67 \times 10^{-2}$) across both the SNP-wise Mean and SNP-wise Top models (Table 1A and B, respectively), successfully replicating our results of the *PCNA* gene from the UKB's European subset.

3.4. Exploring the rs16969968×SNP interactions

We used LocusZoom and HaploView to visualize the pattern of associations as a function of their linkage disequilibrium (LD) with the lead SNP (rs73586411) in the PCNA gene and our significant SNP rs1892967 in chromosome 11 for log₁₀ CPD. All our suggestively significant interactions ($p < 1 \times 10^{-5}$) from the rs16969968×SNP analyses for the PCNA gene were highly correlated with one another (Fig. 3B) and aggregated in a single LD block, block 3 (Fig. S4). To confirm whether this was a single signal, we conducted rs16969968×SNP interaction analyses for the SNPs within PCNA, conditioning on the rs16969968×rs73586411 interaction, the interaction with the lowest *p*value in the PCNA gene. No SNPs were significant after controlling for multiple comparisons (p > 0.05/3 effectively independent SNPs in the region = 0.017). In addition, we explored the LD patterns around SNP rs1892967 which was the single significant hit from our rs16969968×SNP analyses for log_{10} CPD. rs1892967 lies within the *GRAMD1B* gene and is in high linkage disequilibrium (LD > 0.8) with SNP rs1892966, which neared the genome-wide significance of 5×10^{-8} $(p < 5 \times 10^{-5}, p = 6.91 \times 10^{-8}, \text{Fig. S10}).$

4. Discussion

We conducted an exploratory study of SNP and gene interactions with the SNP rs16969968 on daily cigarette consumption. In the SNP×SNP interaction analysis, no SNP reached genome-wide significance when analyzing only European individuals. However, when we meta-analyzed across European and South Asian populations, one SNP, rs1892967, reached genome-wide significance for log₁₀ CPD ($p = 3.18 \times 10^{-8}$). Nevertheless, the gene analyses for genes near this SNP were insignificant across both the MAGMA mean and top models across all discovery subsets and the discovery meta-analysis. Further analyses with increased sample sizes and greater ancestry diversity can help clarify the role of this locus and its interaction with rs16969968 on smoking quantity.

At the gene-level, one gene, *PCNA*, did achieve genome-wide significance within a single ancestry. This result was consistent with the SNP-level analysis, where some SNPs within this region (tagged by rs73586411 and including two other genes, *CDS2* and *TMEM230*) had *p*-values approaching significance. Importantly, we replicated this genelevel finding in an independent dataset of five Finnish samples (all *p* < 6.16×10^{-3}), followed by a meta-analysis of the results (6.62E-4 > p > 5.42E-8), confirming our novel finding for all three genes. The fact that all three of these genes were statistically significant in our replication analyses using the Finnish samples supports our conclusion that a region tagged by lead SNP rs73586411 and shared across these three genes significantly modulates the effect of the risk allele of rs16969968 and its effects on daily cigarette consumption.

We conducted conditional analyses to determine the number of independent signals within our region of significance. The LD structure *PCNA* gene and conditional analyses revealed that this is a single signal coming from an LD block containing 11 SNPs. Because we used a 25 kb

Table 1

Results for the gene-level interaction analyses with rs16969968 for cigarettes per day for Europeans and South Asians in the UKB, meta-analyzed Finnish subsets, and the meta-analysis across both ancestries and all datasets (UKB and Finnish). (A) *TMEM230* and *PCNA* reached significance in the UKB European subsample using the SNP-wise Mean model. All three overlapping genes, *CDS2*, *TMEM230*, and *PCNA* were significant in the meta-analysis of the Finnish subsets and in the meta-analysis combining all UKB and Finnish sub-samples. (B) Considering the most significant signal within a gene (SNP-wise Top model), *TMEM230* and *PCNA* reached significance in the UKB discovery meta-analysis of Europeans and South Asians; *PCNA* was also significant in the UKB European sub-sample. All three overlapping genes, *CDS2*, *TMEM230*, and *PCNA* were significant in the finnish and the mega meta-analysis. Values in bold indicate significance after multiple testing ($p < 2.63 \times 10^{-6}$).

A) SNP-wise Me	ean:	for Cigorol	tos nor Dev		
Study	Gene	Gene	Sample	7-score	P-value
Study	Name	ID	Size	2-30010	1 -value
UKB_Euro	CDS2	8760	116 442	3.317	4.55E- 04
UKB_S_Asian	CDS2	8760	770	0.95998	1.69E- 01
UKB_Meta- Analysis	CDS2	8760	117 212	3.602	3.15E- 04
Fin_Meta- Analysis	CDS2	8760	40 140	2.739	6.16E- 03
Euro_Meta- Analysis	CDS2	8760	156 582	4.410	1.03E- 05
All_Meta- Analysis	CDS2	8760	165 459	4.517	6.07E- 06
UKB_Euro	TMEM230	29058	116 442	2.403	7.63E- 07
UKB_S_Asian	TMEM230	29058	770	0.10492	4.58E- 01
UKB_Meta- Analysis	TMEM230	29058	117 212	2.112	3.47E- 02
Fin_Meta- Analysis	TMEM230	29058	40 140	3.143	1.67E- 03
Euro_Meta- Analysis	TMEM230	29058	156 582	3.444	5.73E- 04
All_Meta- Analysis	TMEM230	29058	165 459	3.245	6.62E- 04
UKB_Euro	PCNA	5111	116 442	4.798	8.02E- 07
UKB_S_Asian	PCNA	5111	770	0.022242	4.91E- 01
UKB_Meta- Analysis	PCNA	5111	117 212	4.264	2.01E- 05
Fin_Meta- Analysis	PCNA	5111	40 140	2.812	4.93E- 03
Euro_Meta- Analysis	PCNA	5111	156 582	5.041	4.63E- 07
All_Meta- Analysis	PCNA	5111	165 459	4.851	4.70E- 07

B) SNP-wise Top:

Top rs16969968xGene Analyses for Cigarettes per Day

Study	Gene	Gene	Sample	Z-score	P-value
Study	Name	ID	Size	Libeore	i vulue
UKB_Euro	CDS2	8760	116 442	4.411	5.16E- 06
UKB_S_Asian	CDS2	8760	770	1.5025	6.65E- 02
UKB_Meta- Analysis	CDS2	8760	117 212	3.576	3.49E- 04
Fin_Meta- Analysis	CDS2	8760	40 140	3.135	1.72E- 03
Euro_Meta- Analysis	CDS2	8760	156 582	4.615	3.92E- 06
All_Meta- Analysis	CDS2	8760	165 459	4.485	2.05E- 06
UKB_Euro	TMEM230	29058	116 442	3.845	6.03E- 05
UKB_S_Asian	TMEM230	29058	770	-0.63345	7.37E- 01
UKB_Meta- Analysis	TMEM230	29058	117 212	2.9	3.73E- 03

(continued on next page)

Table 1 (continued)

Fin_Meta- TMEM230 29058 40 140 4.548 5. Analysis 00 Euro_Meta- TMEM230 29058 156 582 4.836 1. Analysis 00 00 00 00 00 00 Analysis 00 00 00 00 00 00 00 All Meta- TMEM230 29058 165 459 4 583 1	
Euro_Meta- TMEM230 29058 156 582 4.836 1. Analysis 00 01 04 <	.41E- 6
All Meta- TMEM230 29058 165 459 4 583 1	.32E- 6
Analysis 00	.86E- 6
UKB_Euro PCNA 5111 116 442 4.952 3.	.67E- 7
UKB_S_Asian PCNA 5111 770 -0.090806 5.	.36E- 1
UKB_Meta- PCNA 5111 117 212 3.636 2. Analysis 04	.77E- 4
Fin_Meta- PCNA 5111 40 140 4.584 4. Analysis 00	.57E- 6
Euro_Meta- PCNA 5111 156 582 5.397 6. Analysis 04	.79E- 8
All_Meta- PCNA 5111 165 459 5.228 5. Analysis 04	.42E- 8

window, all these 11 nominally significant SNPs driving the interaction with PCNA also span part of the CDS2 and TMEM230 genes (Sherry et al., 2001). Only PCNA was statistically significant in our UKB Euro analyses while CDS2 and TMEM230 were additionally significant in our replication; we hypothesize that this discrepancy was due to CDS2 and TMEM230 harboring more non-significant SNPs, which diluted the signal in the CDS2 and TMEM230 genes. To illustrate, the PCNA gene boundary contained 48 SNPs, whereas the CDS2 and TMEM230 gene region boundaries contained 221 and 67, respectively. In sum, we emphasize that this interaction is due to a single signal within the PCNA, CDS2, and TMEM230 region of chromosome 20. None of the SNPs in the LD block driving our significant gene results are located within coding regions of PCNA, CDS2, or TMEM230. Most are located within intronic regions of CDS2, but there is no evidence for functional impact based on current information available for possible epigenetic areas or other known gene regulatory elements. Therefore, prioritization of possible functional SNPs could not be identified in this study.

To our knowledge, of the three genes encompassing our epistatic region of interest, PCNA is the only one previously linked to smoking behaviors. PCNA encodes for proliferating cell nuclear antigen, which is widely expressed across many tissues and involved in leading strand synthesis of DNA during replication. According to the GWAS catalog ("GWAS Catalog," 2023), height is the only phenotype with evidence of association with PCNA (Barton et al., 2021). However, animal and transcriptomic studies have linked PCNA to nicotine. For example, animal studies have linked nicotine exposure to PCNA damage in lung and kidney cell cultures in a dose-dependent fashion (Salama et al., 2014). In addition, PCNA expression levels increased in hepatic and pancreatic cells of rats exposed to both ethanol and tobacco compared to tobacco alone (Wang et al., 2014). PCNA gene expression is up-regulated in response to complex environmental mixtures, including cigarette smoke, and is clearly involved in DNA repair following exposure to hazardous chemicals (Sen et al., 2007). Cumulatively, this previous research links PCNA to smoking behaviors. On the other hand, TMEM230 and CDS2 have been associated with a variety of other traits. For example, TMEM230 has been previously associated with acute myeloid leukemia (Lv et al., 2016), hair morphology (Medland et al., 2009), and Parkinson's Disease (Wang et al., 2021). CDS2 has emerged in four GWAS reports: two studies of height (Kichaev et al., 2019; Sakaue et al., 2021), one on Ebbinghaus illusion, an inability to contextualize relative size perception (Zhu et al., 2020), and another identifying gene-gene interactions with pathological hallmarks of Alzheimer's disease (Wang et al., 2020). While all three genes (PCNA, TMEM230, and CDS2) were significant in our replication, only PCNA was significant across both our discovery and replication analyses, and therefore functional follow-up of this interaction should prioritize the PCNA gene.

There are several caveats and limitations to our study. First, since it was computationally unfeasible for us to investigate all pair-wise genetic interactions, we only investigated genetic interactions with a single SNP of interest, rs16969968. There are likely other interactions involving SNPs other than rs169969968 influencing smoking quantity that were unexplored. Second, smoking quantity is a good proxy for nicotine dependence, but participants tend to underreport how much they smoke (Gorber et al., 2009); this is especially problematic in individuals who no longer actively smoke (Soulakova et al., 2012). While most of our sample was composed of individuals who reported actively smoking, underreporting the number of cigarettes smoked per day might have been especially pronounced on individuals who no longer report actively smoking, thereby introducing non-random bias in our outcome of interest. However, this potential bias in CPD self-reporting is unlikely to be jointly correlated with the genotype at rs16969968 and with other genes across the genome; while underreporting might bias the average CPD estimate, it would not have led to false interaction signals. Third, we estimated smoking quantity through cigarettes per day since it is the most common form of nicotine consumption, thereby excluding other ways our participants might have ingested nicotine (i.e., vapes, pipes, hookahs, etc.). In our results, both the SNP and gene level interactions for log₁₀-transformed cigarettes per day were insignificant for the chromosome 20 region surrounding PCNA and tagged by rs73586411. For example, at the SNP level using log₁₀-transformed CPD in Europeans, the interaction *p*-value for rs73586411 was 4.33×10^{-4} compared to 1.79×10^{-5} for CPD. At the gene level, the interaction between rs16969968 and PCNA for log₁₀-CPD was also insignificant in Europeans in the UKB ($p = 2.71 \times 10^{-5}$ for SNP-wise mean, $p = 2.21 \times 10^{-5}$ for SNP-wise top model). In addition, the 25 kb window we chose around the start and end of each gene is arbitrary; there is no clear standard in the field for this. When using genes discovered in model organisms associated with nicotine consumption, Palmer et al. found that heritability for human nicotine consumption was enriched in genomic regions surrounding the genes compared to the protein-coding regions of these genes. In addition, after comparing 5, 10, 25, and 35 kb gene windows, they found that enrichment decreased after 10 kb (Palmer et al., 2021). These findings suggest that it is beneficial to use a gene window, though the best window size still merits further investigation and could vary across traits and across genes.

Despite these limitations, our study has multiple strengths. Our twostep approach of conducting a genome-wide interaction study and aggregating these signals within genes increased our power relative to the initial SNP×SNP analysis to detect interactions by decreasing the multiple testing burden. In addition, based on our simulations, this approach kept our type I error rate low when evaluating unlinked SNPs. Second, this approach is flexible - we pooled data from multiple sources and ancestries with low sample sizes that would normally be disregarded in interaction studies due to low power and instead "balance statistical needs with fairness" following Ben-Eghan et al.'s call to include data from minority populations despite their low sample sizes (Ben-Eghan et al., 2020). While flexible, our approach is also granular we pinpointed the specific LD block driving our significant SNP×Gene results by inspecting any regions of interest using the SNP×SNP results from step one. Third, our method is agnostic to functional characteristics. Previous extensions to MAGMA such as H-MAGMA (Sey et al., 2020) are limited to investigating interactions that have well documented chromatin interaction profiles, such as neural tissue. Similarly, transcriptome-wide association studies combine GWAS results and gene expression to elucidate probable mechanistic relationships for the SNP and genes associated with a trait of interest. Previous work has expanded the TWAS approach to investigate gene-gene interactions and demonstrated that gene interactions influencing complex traits are pervasive and therefore important to further investigate (Evans et al., 2023). One of the limitations of our approach is that it does not elucidate biological function; however, unlike TWAS, our approach is unaffected by limitations of existing gene expression datasets, including a lack of diversity, small sample sizes, and tissue bias present (Wainberg et al., 2019). The approach developed here (pooling data from multiple datasets to

increase sample size and representativeness, limiting SNP×SNP epistatic analyses to common variants, and using a 10–25 kb upstream and downstream gene window when aggregating SNP×SNP results at the gene-level) will be useful for other researchers in the field attempting to discover genome-wide interactions with a wide range of complex traits. Potential interactions discovered using our approach can then be followed-up using H-MAGMA or TWAS to elucidate biological networks or functions underlying such interactions. These results serve as a guide for others in the field as they also attempt to study epistasis at the SNP level.

In summary, this is the first study to report an interaction between rs16969968 and any genome-wide loci influencing cigarette consumption. Five of our nominally significant SNPs, such as rs73586411 and rs6053152, previously failed to reach significance for cigarettes per day in GSCAN, with sample sizes roughly 3-10 times the size used here (Liu et al., 2019). This highlights the power of interaction studies to detect novel variants that would not be found otherwise and the importance for larger studies like GSCAN to increase access to individual-level data for the feasibility of this work. Future studies could implement our pipeline to investigate interactions between other well-replicated common polymorphisms and genome-wide loci for further characterization of the genetic factors underlying complex traits. Our approach could be especially helpful in understanding how different variants within the same gene might have different or opposing effects on a trait of interest. Like the CHRNA5 receptor gene rs16969968 is located in, the CYP2A6 gene harbors multiple polymorphisms that have been previously associated with smoking quantity but with opposing effects (Pan et al., 2015). For example, while rs16969968 has consistently been shown to increase smoking quantity, our current results show that in the presence of the minor allele of tag SNP rs73586411, smoking quantity was reduced. In short, our current findings expand our understanding of how a well-characterized and long-established SNP influencing smoking quantity alters risk for smoking behaviors in conjunction with the rest of the genome and showcases a novel way for other scientists to continue more detailed characterization of other strongly associated SNPs underlying smoking behaviors.

Funding

This work was supported by the National Institutes of Health (grant numbers R01 AG046938-06 to Chandra A. Reynolds, R01 DA044283-01 to Scott I. Vrieze, and R01 MH100141-06 to Matthew C. Keller) and the University of Colorado Boulder Institute for Behavioral Genetics.

CRediT authorship contribution statement

Evans Luke M.: Formal analysis, Supervision, Writing – review & editing. **Border Richard:** Data curation, Investigation. **Palviainen Teemu:** Formal analysis. **Ehringer Marissa A.:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Keller Matthew C.:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. **Palmer Rohan H.C.:** Funding acquisition, Validation. **Kaprio Jaakko:** Validation. **Romero Villela Pamela N.:** Formal analysis, Methodology, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

Authors report no personal conflicts of interest.

Acknowledgments

This research was done using the UK-Biobank under project ref. 16651; We thank the participants and administrators of the UK Biobank and Finnish studies. This work was supported by the National Institutes of Health (grant numbers R01 AG046938–06 to Chandra A. Reynolds,

R01 DA044283–01 to Scott I. Vrieze, and R01 MH100141–06 to Matthew C. Keller) and the University of Colorado Boulder Institute for Behavioral Genetics. This work utilized the Summit supercomputer, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University. The Summit supercomputer is a joint effort of the University of Colorado Boulder and Colorado State University. Access to the Finnish samples was made possible thanks to Dr. Palmer through his NIH grant ref. DA042742. We also would like to thank Michael Stallings, Ph.D. and Chelsie Benca-Bachman, Ph.D. for their feedback on this manuscript. JK acknowledges support of the Sigrid Juselius Foundation and the Academy of Finland.

Author contributions

MCK and MAE conceived the study; MCK, MAE, and LME supervised the implementation of the study, interpreted the results, and revised the paper. PNRV conducted the analyses, interpreted the results, drafted, and revised the paper. RB conducted preliminary analyses and revised the paper. TP contributed to the analyses. JK and RHP facilitated access to the data and commented on multiple drafts of the manuscript. All authors approved the final version of this manuscript.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.drugalcdep.2024.111126.

References

- Abraham, G., Inouye, M., 2014. Fast principal component analysis of large-scale genomewide data. PLoS One 9, e93766. https://doi.org/10.1371/JOURNAL. PONE.0093766.
- Adjangba, C., Border, R., Romero Villela, P.N., Ehringer, M.A., Evans, L.M., 2021. Little evidence of modified genetic effect of rs16969968 on heavy smoking based on age of onset of smoking. Nicotine Tob. Res. 23, 1055. https://doi.org/10.1093/NTR/ NTAA229.
- Alberg, A.J., Shopland, D.R., Cummings, K.M., 2014. The 2014 Surgeon General's Report: commemorating the 50th anniversary of the 1964 Report of the Advisory Committee to the US Surgeon General and updating the evidence on the health consequences of cigarette smoking. Am. J. Epidemiol. 179, 403–412. https://doi. org/10.1093/AJE/KWT335.
- Amos, C.I., Wu, X., Broderick, P., Gorlov, I.P., Gu, J., Eisen, T., Dong, Q., Zhang, Q., Gu, X., Vijayakrishnan, J., Sullivan, K., Matakidou, A., Wang, Y., Mills, G., Doheny, K., Tsai, Y.Y., Chen, W.V., Shete, S., Spitz, M.R., Houlston, R.S., 2008. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1, 2008 405 Nat. Genet. 40, 616–622. https://doi.org/10.1038/ ng.109.
- Barrett, J.C., Fry, B., Maller, J., Daly, M.J., 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265. https://doi.org/10.1093/ bioinformatics/bth457.
- Barton, A.R., Sherman, M.A., Mukamel, R.E., Loh, P.R., 2021. Whole-exome imputation within UK Biobank powers rare coding variant association and fine-mapping analyses, 2021 538 Nat. Genet. 53, 1260–1269. https://doi.org/10.1038/s41588-021-00892-1.
- Ben-Eghan, C., Sun, R., Hleap, J.S., Diaz-Papkovich, A., Munter, H.M., Grant, A.V., Dupras, C., Gravel, S., 2020. Don't ignore genetic data from minority populations, 2021 5857824 Nature 585, 184–186. https://doi.org/10.1038/d41586-020-02547-2
- Bierut, L.J., Stitzel, J.A., Wang, J.C., Hinrichs, A.L., Grucza, R.A., Xuei, X., Saccone, N.L., Saccone, S.F., Bertelsen, S., Fox, L., Horton, W.J., Breslau, N., Budde, J., Cloninger, C.R., Dick, D.M., Foroud, T., Hatsukami, D., Hesselbrock, V., Johnson, E. O., Kramer, J., Kuperman, S., Madden, P.A.F., Mayo, K., Nurnberger, J., Pomerleau, O., Porjesz, B., Reyes, O., Schuckit, M., Swan, G., Tischfield, J.A., Edenberg, H.J., Rice, J.P., Goate, A.M., 2008. Variants in nicotinic receptors and risk for nicotine dependence. Am. J. Psychiatry 165, 1163–1171. https://doi.org/ 10.1176/appi.ajp.2008.07111711.
- Buck, J.M., O'Neill, H.C., Stitzel, J.A., 2021. The Intergenerational transmission of developmental nicotine exposure-induced neurodevelopmental disorder-like phenotypes is modulated by the Chrna5 D397N polymorphism in adolescent mice. Behav. Genet. 51, 665–684. https://doi.org/10.1007/S10519-021-10071-X/ FIGURES/9.
- Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7. https://doi.org/10.1186/S13742-015-0047-8/2707533.

Chen, L.S., Horton, A., Bierut, L., 2018. Pathways to precision medicine in smoking cessation treatments. Neurosci. Lett. 669, 83–92. https://doi.org/10.1016/J. NEULET.2016.05.033.

- de Leeuw, C.A., Mooij, J.M., Heskes, T., Posthuma, D., 2015. MAGMA: generalized geneset analysis of GWAS data. PLoS Comput. Biol. 11 https://doi.org/10.1371/journal. pcbi.1004219.
- de Leeuw, C.A., Neale, B.M., Heskes, T., Posthuma, D., Christiaan A. de Leeuw, Benjamin M.Neale, T.H, D.P, 2016. The statistical properties of gene-set analysis. Nat. Rev. Genet. 17, 353–364. https://doi.org/10.1038/nrg.2016.29.
- de Leeuw, C.A., Stringer, S., Dekkers, I.A., Heskes, T., Posthuma, D., 2018. Conditional and interaction gene-set analysis reveals novel functional pathways for blood pressure. Nat. Commun. 9, 3768. https://doi.org/10.1038/s41467-018-06022-6.
- Evans, L.M., Jang, S., Hancock, D.B., Ehringer, M.A., Otto, J.M., Vrieze, S.I., Keller, M.C., 2021. Genetic architecture of four smoking behaviors using partitioned SNP heritability. Addiction 116, 2498–2508. https://doi.org/10.1111/ADD.15450.

Evans, L.M., Arehart, C.H., Grotzinger, A.D., Mize, T.J., Brasher, M.S., Stitzel, J.A., Ehringer, M.A., Hoeffer, C.A., 2023. Transcriptome-wide gene-gene interaction associations elucidate pathways and functional enrichment of complex traits. PLOS Genet. 19, e1010693 https://doi.org/10.1371/JOURNAL.PGEN.1010693.

- Fowler, C.D., Turner, J.R., Imad Damaj, M., 2020. Molecular mechanisms associated with nicotine pharmacology and dependence. Handb. Exp. Pharmacol. 258, 373–393. https://doi.org/10.1007/164_2019_252.
- Gorber, S.C., Schofield-Hurwitz, S., Hardt, J., Levasseur, G., Tremblay, M., 2009. The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and cotinine-assessed smoking status. Nicotine Tob. Res. 11, 12–24. https://doi.org/10.1093/NTR/NTN010.

GWAS Catalog, 2023. [WWW Document]. URL: (https://www.ebi.ac.uk/gwas/home) (accessed 3.16.22).

Health-2000-2011, n.d. THL [WWW Document]. URL: (https://thl.fi/en/web/thlfi-en/res earch-and-development/research-and-projects/health-2000-2011) (Accessed 3.16.22).

Hill, W.G., Goddard, M.E., Visscher, P.M., 2008. Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 4 https://doi.org/10.1371/ JOURNAL.PGEN.1000008.

Hung, R.J., McKay, J.D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., Mates, D., Bencko, V., Foretova, L., Janout, V., Chen, C., Goodman, G., Field, J.K., Liloglou, T., Xinarianos, G., Cassidy, A., McLaughlin, J., Liu, G., Narod, S., Krokan, H.E., Skorpen, F., Elvestad, M.B., Hveem, K., Vatten, L., Linseisen, J., Clavel-Chapelon, F., Vineis, P., Bueno-de-Mesquita, H.B., Lund, E., Martinez, C., Bingham, S., Rasmuson, T., Hainaut, P., Riboli, E., Ahrens, W., Benhamou, S., Lagiou, P., Trichopoulos, D., Holcátová, I., Merletti, F., Kjaerheim, K., Agudo, A., Macfarlane, G., Talamini, R., Simonato, L., Lowry, R., Conway, D.I., Znaor, A., Healy, C., Zelenika, D., Boland, A., Delepine, M., Foglio, M., Lechner, D., Matsuda, F., Blanche, H., Gut, I., Heath, S., Lathrop, M., Brennan, P., 2008. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 452, 633–637. https://doi.org/10.1038/ nature06885.

Jones, S.K., Alberg, A.J., Wallace, K., Froeliger, B., Carpenter, M.J., Wolf, B.J., 2023. CHRNA5-A3-B4 and DRD2 genes and smoking cessation throughout adulthood: a longitudinal study of women. Nicotine Tob. Res. 25, 1164–1173. https://doi.org/ 10.1093/NTR/NTAD026.

- Kaidesoja, M., Aaltonen, S., Bogl, L.H., Heikkilä, K., Kaartinen, S., Kujala, U.M., Kärkkäinen, U., Masip, G., Mustelin, L., Palviainen, T., Pietiläinen, K.H., Rottensteiner, M., Sipilä, P.N., Rose, R.J., Keski-Rahkonen, A., Kaprio, J., 2019. FinnTwin16: a longitudinal study from age 16 of a population-based Finnish twin cohort. Twin Res. Hum. Genet. 22, 530–539. https://doi.org/10.1017/ THG.2019.106.
- Kaprio, J., 2009. Genetic epidemiology of smoking behavior and nicotine dependence. COPD J. Chronic Obstr. Pulm. Dis. 6, 304–306. https://doi.org/10.1080/ 15412550903049165.
- Kaprio, J., Bollepalli, S., Buchwald, J., Iso-Markku, P., Korhonen, T., Kovanen, V., Kujala, U., Laakkonen, E.K., Latvala, A., Leskinen, T., Lindgren, N., Ollikainen, M., Piirtola, M., Rantanen, T., Rinne, J., Rose, R.J., Sillanpää, E., Silventoinen, K., Sipilä, S., Viljanen, A., Vuoksimaa, E., Waller, K., 2019. The older Finnish twin cohort – 45 years of follow-up. Twin Res. Hum. Genet. 22, 240–254. https://doi. org/10.1017/THG.2019.54.

Kauffman, S., 1993. The Origins of Order. Oxford University Press, Oxford.

Keller, M.C., 2014. Gene × environment interaction studies have not properly controlled for potential confounders: The problem and the (simple) solution. Biol. Psychiatry 75, 18–24. https://doi.org/10.1016/j.biopsych.2013.09.006.

Kichaev, G., Bhatia, G., Loh, P.R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc, B., Price, A.L., 2019. Leveraging polygenic functional enrichment to improve GWAS power. Am. J. Hum. Genet. 104, 65–75. https://doi.org/10.1016/J. AJHG.2018.11.008.

Koukouli, F., Rooy, M., Tziotis, D., Sailor, K.A., O'Neill, H.C., Levenga, J., Witte, M., Nilges, M., Changeux, J.P., Hoeffer, C.A., Stitzel, J.A., Gutkin, B.S., Digregorio, D.A., Maskos, U., 2017. Nicotine reverses hypofrontality in animal models of addiction and schizophrenia. Nat. Med. 23, 347–354. https://doi.org/10.1038/nm.4274.

Krall, E.A., Valadian, I., Dwyer, J.T., Gardner, J., 1989. Accuracy of recalled smoking data. Public Health.

Lessov-Schlaggar, C.N., Pang, Z., Swan, G.E., Guo, Q., Wang, S., Cao, W., Unger, J.B., Johnson, C.A., Lee, L., 2006. Heritability of cigarette smoking and alcohol use in Chinese male twins: the Qingdao twin registry. Int. Epidemiol. Assoc. Int. J. Epidemiol. 35, 1278–1285. https://doi.org/10.1093/ije/dyl148. Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D.M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., Tian, C., Zhan, X., Choquet, H., Docherty, A.R., Faul, J. D., Foerster, J.R., Fritsche, L.G., Gabrielsen, M.E., Gordon, S.D., Haessler, J., Hottenga, J.-J., Huang, H., Jang, S.-K., Jansen, P.R., Ling, Y., Mägi, R., Matoba, N., McMahon, G., Mulas, A., Orrù, V., Palviainen, T., Pandit, A., Reginsson, G.W., Skogholt, A.H., Smith, J.A., Taylor, A.E., Turman, C., Willemsen, G., Young, H., Young, K.A., Zajac, G.J.M., Zhao, W., Zhou, W., Bjornsdottir, G., Boardman, J.D., Boehnke, M., Boomsma, D.I., Chen, C., Cucca, F., Davies, G.E., Eaton, C.B., Ehringer, M.A., Esko, T., Fiorillo, E., Gillespie, N.A., Gudbjartsson, D.F., Haller, T., Harris, K.M., Heath, A.C., Hewitt, J.K., Hickie, I.B., Hokanson, J.E., Hopfer, C.J., Hunter, D.J., Iacono, W.G., Johnson, E.O., Kamatani, Y., Kardia, S.L.R., Keller, M.C., Kellis, M., Kooperberg, C., Kraft, P., Krauter, K.S., Laakso, M., Lind, P.A., Loukola, A., Lutz, S.M., Madden, P.A.F., Martin, N.G., McGue, M., McQueen, M.B., Medland, S.E., Metspalu, A., Mohlke, K.L., Nielsen, J.B., Okada, Y., Peters, U., Polderman, T.J.C., Posthuma, D., Reiner, A.P., Rice, J.P., Rimm, E., Rose, R.J., Runarsdottir, V., Stallings, M.C., Stančáková, A., Stefansson, H., Thai, K.K., Tindle, H.A., Tyrfingsson, T., Wall, T.L., Weir, D.R., Weisner, C., Whitfield, J.B., Winsvold, B.S., Yin, J., Zuccolo, L., Bierut, L.J., Hveem, K., Lee, J.J., Munafò, M.R., Saccone, N.L., Willer, C.J., Cornelis, M.C., David, S.P., Hinds, D.A., Jorgenson, E., Kaprio, J., Stitzel, J.A., Stefansson, K., Thorgeirsson, T.E., Abecasis, G., Liu, D.J., Vrieze, S., 2019. 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. https://doi. 41588-018-0307-5

Lv, H., Zhang, M., Shang, Z., Li, J., Zhang, S., Lian, D., Zhang, R., Lv, H., Zhang, M., Shang, Z., Li, J., Zhang, S., Lian, D., Zhang, R., 2016. Genome-wide haplotype association study identify the FGFR2 gene as a risk gene for Acute Myeloid Leukemia. Oncotarget 8, 7891–7899. https://doi.org/10.18632/ ONCOTARGET.13631.

- Mackay, T.F.C., 2013. Epistasis and quantitative traits: using model organisms to study gene–gene interactions, 2013 151 Nat. Rev. Genet. 15, 22–33. https://doi.org/ 10.1038/nrg3627.
- Mackay, T.F.C., Moore, J.H., 2014. Why epistasis is important for tackling complex human disease genetics, 2014 66 Genome Med. 6, 1–3. https://doi.org/10.1186/ GM561.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A. Timpson, N., Koskinen, S., Vrieze, S., Scott, L.J., Zhang, H., Mahajan, A., Veldink, J., Peters, U., Pato, C., Van Duijn, C.M., Gillies, C.E., Gandin, I., Mezzavilla, M., Gilly, A., Cocca, M., Traglia, M., Angius, A., Barrett, J.C., Boomsma, D., Branham, K., Breen, G., Brummett, C.M., Busonero, F., Campbell, H., Chan, A., Chen, S., Chew, E., Collins, F.S., Corbin, L.J., Smith, G.D., Dedoussis, G., Dorr, M., Farmaki, A.E., Ferrucci, L., Forer, L., Fraser, R.M., Gabriel, S., Levy, S., Groop, L., Harrison, T., Hattersley, A., Holmen, O.L., Hveem, K., Kretzler, M., Lee, J.C., McGue, M., Meitinger, T., Melzer, D., Min, J.L., Mohlke, K.L., Vincent, J.B., Nauck, M., Nickerson, D., Palotie, A., Pato, M., Pirastu, N., McInnis, M., Richards, J.B., Sala, C., Salomaa, V., Schlessinger, D., Schoenherr, S., Slagboom, P.E., Small, K., Spector, T., Stambolian, D., Tuke, M., Tuomilehto, J., Van Den Berg, L.H., Van Rheenen, W., Volker, U., Wijmenga, C., Toniolo, D., Zeggini, E., Gasparini, P., Sampson, M.G., Wilson, J.F., Frayling, T., De Bakker, P.I.W., Swertz, M.A., McCarroll, S., Kooperberg, C., Dekker, A., Altshuler, D., Willer, C., Iacono, W., Ripatti, S., Soranzo, N., Walter, K., Swaroop, A., Cucca, F., Anderson, C.A., Myers, R.M., Boehnke, M., McCarthy, M.I., Durbin, R., Abecasis, G., Marchini, J., 2016. A reference panel of 64,976 haplotypes for genotype imputation, 2016 4810 Nat. Genet. 48, 1279–1283. https://doi.org/10.1038/ng.3643. Medland, S.E., Nyholt, D.R., Painter, J.N., McEvoy, B.P., McRae, A.F., Zhu, G., Gordon, S.

- Medland, S.E., Nyholt, D.R., Painter, J.N., McEvoy, B.P., McRae, A.F., Zhu, G., Gordon, S. D., Ferreira, M.A.R., Wright, M.J., Henders, A.K., Campbell, M.J., Duffy, D.L., Hansell, N.K., Macgregor, S., Slutske, W.S., Heath, A.C., Montgomery, G.W., Martin, N.G., 2009. Common variants in the trichohyalin gene are associated with straight hair in Europeans. Am. J. Hum. Genet. 85, 750–755 https://doi.org/10.1016/J.AJHG.2009.10.009/ATTACHMENT/DF3336F9-8BCA-4128-96C8-DEF141A61DD9/MMC1 PDF
- National FinHealth Study THL, n.d. [WWW Document]. URL: (https://thl.fi/en/web/t hlfi-en/research-and-development/research-and-projects/national-finhealth-study) (Accessed 3.16.22).
- NCBI, 1999. rs16969968 RefSNP Report dbSNP NCBI [WWW Document]. URL: https://www.ncbi.nlm.nih.gov/snp/rs16969968#variant_details)(Accessed 9.22.23).
- O'Neill, H.C., Wageman, C.R., Sherman, S.E., Grady, S.R., Marks, M.J., Stitzel, J.A., 2018. The interaction of the Chrna5 D398N variant with developmental nicotine exposure. Genes Brain Behav. 17, e12474 https://doi.org/10.1111/GBB.12474.
- Olfson, E., Saccone, N.L., Johnson, E.O., Chen, L.S., Culverhouse, R., Doheny, K., Foltz, S. M., Fox, L., Gogarten, S.M., Hartz, S., Hetrick, K., Laurie, C.C., Marosy, B., Amin, N., Arnett, D., Barr, R.G., Bartz, T.M., Bertelsen, S., Borecki, I.B., Brown, M.R., Chasman, D.I., Van Duijn, C.M., Feitosa, M.F., Fox, E.R., Franceschini, N., Franco, O. H., Grove, M.L., Guo, X., Hofman, A., Kardia, S.L.R., Morrison, A.C., Musani, S.K., Psaty, B.M., Rao, D.C., Reiner, A.P., Rice, K., Ridker, P.M., Rose, L.M., Schick, U.M., Schwander, K., Uitterlinden, A.G., Vojinovic, D., Wang, J.C., Ware, E.B., Wilson, G., Yao, J., Zhao, W., Breslau, N., Hatsukami, D., Stitzel, J.A., Rice, J., Goate, A., Bierut, L.J., 2015. Rare, low frequency and common coding variants in CHRNA5 and their contribution to nicotine dependence in European and African Americans, 2016 215 Mol. Psychiatry 21, 601–607. https://doi.org/10.1038/mp.2015.105.
- Palmer, R.H.C., Benca-Bachman, C.E., Huggett, S.B., Bubier, J.A., McGeary, J.E., Ramgiri, N., Srijeyanthan, J., Yang, Jingjing, Visscher, P.M., Yang, Jian, Knopik, V. S., Chesler, E.J., 2021. Multi-omic and multi-species meta-analyses of nicotine consumption, 2021 111 Transl. Psychiatry 11, 1–10. https://doi.org/10.1038/ s41398-021-01231-y.

- Pan, L., Yang, X., Li, S., Jia, C., 2015. Association of CYP2A6 gene polymorphisms with cigarette consumption: a meta-analysis. Drug Alcohol Depend. 149, 268–271. https://doi.org/10.1016/J.DRUGALCDEP.2015.01.032.
- Picciotto, M.R., Kenny, P.J., 2021. Mechanisms of nicotine addiction. Cold Spring Harb. Perspect. Med. 11, a039610. https://doi.org/10.1101/CSHPERSPECT.A039610.
- Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., Willer, C.J., Frishman, D., 2010. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336–2337. https://doi.org/10.1093/BIOINFORMATICS/BTO419.
- Quach, B.C., Bray, M.J., Gaddis, N.C., Liu, M., Palviainen, T., Minica, C.C., Zellers, S., Sherva, R., Aliev, F., Nothnagel, M., Young, K.A., Marks, J.A., Young, H., Carnes, M. U., Guo, Y., Waldrop, A., Sey, N.Y.A., Landi, M.T., McNeil, D.W., Drichel, D., Farrer, L.A., Markunas, C.A., Vink, J.M., Hottenga, J.J., Iacono, W.G., Kranzler, H.R., Saccone, N.L., Neale, M.C., Madden, P., Rietschel, M., Marazita, M.L., McGue, M., Won, H., Winterer, G., Grucza, R., Dick, D.M., Gelernter, J., Caporaso, N.E., Baker, T. B., Boomsma, D.I., Kaprio, J., Hokanson, J.E., Vrieze, S., Bierut, L.J., Johnson, E.O., Hancock, D.B., 2020. Expanding the genetic architecture of nicotine dependence and its shared genetics with multiple traits, 2020 111 Nat. Commun. 11, 1–13. https:// doi.org/10.1038/s41467-020-19265-z.
- R Foundation for Statistical Computing, 2018. R: A Language and Environment for Statistical Computing.
- S. Federal Trade Commission (FTC), 2019. Federal Trade Smokeless Tobacco Report for 2019.
- Saccone, N.L., Culverhouse, R.C., Schwantes-An, T.H., Cannon, D.S., Chen, X., Cichon, S., Giegling, I., Han, S., Han, Y., Keskitalo-Vuokko, K., Kong, X., Landi, M.T., Ma, J.Z., Short, S.E., Stephens, S.H., Stevens, V.L., Sun, L., Wang, Y., Wenzlaff, A.S., Aggen, S. H., Breslau, N., Broderick, P., Chatterjee, N., Chen, J., Heath, A.C., Heliövaara, M., Hoft, N.R., Hunter, D.J., Jensen, M.K., Martin, N.G., Montgomery, G.W., Niu, T., Payne, T.J., Peltonen, L., Pergadia, M.L., Rice, J.P., Sherva, R., Spitz, M.R., Sun, J., Wang, J.C., Weiss, R.B., Wheeler, W., Witt, S.H., Yang, B.Z., Caporaso, N.E., Ehringer, M.A., Eisen, T., Gapstur, S.M., Gelernter, J., Houlston, R., Kaprio, J., Kendler, K.S., Kraft, P., Leppert, M.F., Li, M.D., Madden, P.A.F., Nöthen, M.M., Pillai, S., Rietschel, M., Rujescu, D., Schwartz, A., Amos, C.I., Bierut, L.J., 2010. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a metaanalysis and comparison with lung cancer and COPD. PLoS Genet. 6 https://doi.org/ 10.1371/journal.pgen.1001053.
- Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., Narita, A., Konuma, T., Yamamoto, Kenichi, Akiyama, M., Ishigaki, K., Suzuki, A., Suzuki, K., Obara, W., Yamaji, K., Takahashi, K., Asai, S., Takahashi, Y., Suzuki, T., Shinozaki, N., Yamaguchi, H., Minami, S., Murayama, S., Yoshimori, K., Nagayama, S., Obata, D., Higashiyama, M., Masumoto, A., Koretsune, Y., Ito, K., Terao, C., Yamauchi, T., Komuro, I., Kadowaki, T., Tamiya, G., Yamamoto, M., Nakamura, Y., Kubo, M., Murakami, Y., Yamamoto, Kazuhiko, Kamatani, Y., Palotie, A., Rivas, M.A., Daly, M.J., Matsuda, K., Okada, Y., 2021. A cross-population atlas of genetic associations for 220 human phenotypes. Nat. Genet. 53, 1415–1424. https://doi.org/10.1038/S41588-021-00931-X.
- Salama, S.A., Arab, H.H., Omar, H.A., Maghrabi, I.A., Snapka, R.M., 2014. Nicotine mediates hypochlorous acid-induced nuclear protein damage in mammalian cells. Inflammation 37, 785–792. https://doi.org/10.1007/S10753-013-9797-6/ FIGURES/5.
- Sen, B., Mahadevan, B., DeMarini, D.M., 2007. Transcriptional responses to complex mixtures—A review. Mutat. Res. Mutat. Res. 636, 144–177. https://doi.org/ 10.1016/J.MRREV.2007.08.002.
- Sey, N.Y.A., Hu, B., Mah, W., Fauni, H., McAfee, J.C., Rajarajan, P., Brennand, K.J., Akbarian, S., Won, H., 2020. A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles, 2020 234 Nat. Neurosci. 23, 583–593. https://doi.org/10.1038/s41593-020-0603-0.
- Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., Sirotkin, K., 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29, 308–311. https://doi.org/10.1093/NAR/29.1.308.
- Soulakova, J.N., Hartman, A.M., Liu, B., Willis, G.B., Augustine, S., 2012. Reliability of adult self-reported smoking history: data from the tobacco use supplement to the

current population survey 2002–2003 cohort. Nicotine Tob. Res. 14, 952–960. https://doi.org/10.1093/NTR/NTR313.

- Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., Collins, R., 2015. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 12, 1001779. https://doi.org/10.1371/JOURNAL. PMED.1001779.
- The National FINRISK Study, n.d. THL [WWW Document]. URL (https://thl.fi/en/web/ thlfi-en/research-and-development/research-and-projects/the-national-finrisk-s tudy) (Accessed 3.16.22).
- Thorgeirsson, T.E., Geller, F., Sulem, P., Rafnar, T., Wiste, A., Magnusson, K.P., Manolescu, A., Thorleifsson, G., Stefansson, H., Ingason, A., Stacey, S.N., Bergthorsson, J.T., Thorlacius, S., Gudmundsson, J., Jonsson, T., Jakobsdottir, M., Saemundsdottir, J., Olafsdottir, O., Gudmundsson, L.J., Bjornsdottir, G., Kristjansson, K., Skuladottir, H., Isaksson, H.J., Gudbjartsson, T., Jones, G.T., Mueller, T., Gottsäter, A., Flex, A., Aben, K.K.H., de Vegt, F., Mulders, P.F.A., Isla, D., Vidal, M.J., Asin, L., Saez, B., Murillo, L., Blondal, T., Kolbeinsson, H., Stefansson, J. G., Hansdottir, I., Runarsdottir, V., Pola, R., Lindblad, B., van Rij, A.M., Dieplinger, B., Haltmayer, M., Mayordomo, J.I., Kiemeney, L.A., Matthiasson, S.E., Oskarsson, H., Tyrfingsson, T., Gudbjartsson, D.F., Gulcher, J.R., Jonsson, S., Thorsteinsdottir, U., Kong, A., Stefansson, K., 2008. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 452, 638–642. https://doi.org/10.1038/nature06846.
- Tian, C., Gregersen, P.K., Seldin, M.F., 2008. Accounting for ancestry: population substructure and genome-wide association studies. Hum. Mol. Genet. 17, 143–150. https://doi.org/10.1093/hmg/ddn268.
- Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A.N., Knowles, D.A., Golan, D., Ermel, R., Ruusalepp, A., Quertermous, T., Hao, K., Björkegren, J.L.M., Im, H.K., Pasaniuc, B., Rivas, M.A., Kundaje, A., 2019. Opportunities and challenges for transcriptome-wide association studies, 2019 514 Nat. Genet. 51, 592–599. https://doi.org/10.1038/s41588-019-0385-z.
- Wang, H., Yang, J., Schneider, J.A., De Jager, P.L., Bennett, D.A., Zhang, H.Y., 2020. Genome-wide interaction analysis of pathological hallmarks in Alzheimer's disease. Neurobiol. Aging 93, 61–68. https://doi.org/10.1016/J. NEUROBIOLAGING.2020.04.025.
- Wang, X., Whelan, E., Liu, Z., Liu, C.F., Smith, W.W., 2021. Controversy of TMEM230 associated with Parkinson's disease. Neuroscience 453, 280–286. https://doi.org/ 10.1016/J.NEUROSCIENCE.2020.11.004.
- Wang, Y.Y., Liu, Y., Ni, X.Y., Bai, Z.H., Chen, Q.Y., Zhang, Y.E., Gao, F.G., 2014. Nicotine promotes cell proliferation and induces resistance to cisplatin by α7 nicotinic acetylcholine receptor-mediated activation in Raw264.7 and El4 cells. Oncol. Rep. 31, 1480–1488. https://doi.org/10.3892/OR.2013.2962/HTML.
- Wen, L., Jiang, K., Yuan, W., Cui, W., Li, M.D., 2016. Contribution of variants in CHRNA5/A3/B4 gene cluster on chromosome 15 to tobacco smoking: from genetic association to mechanism. Mol. Neurobiol. 53, 472–484. https://doi.org/10.1007/ s12035-014-8997-x.
- Widén, E., Junna, N., Ruotsalainen, S., Surakka, I., Mars, N., Ripatti, P., Partanen, J.J., Aro, J., Mustonen, P., Tuomi, T., Palotie, A., Salomaa, V., Kaprio, J., Partanen, J., Hotakainen, K., Pöllänen, P., Ripatti, S., 2022. How communicating polygenic and clinical risk for atherosclerotic cardiovascular disease impacts health behavior: an observational follow-up study. Circ. Genom. Precis. Med. https://doi.org/10.1161/ CIRCGEN.121.003459.
- Willer, C.J., Li, Y., Abecasis, G.R., 2010. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinform. Appl. Note 26, 2190–2191. https://doi. org/10.1093/bioinformatics/btq340.
- Yang, J., Lee, S.H., Goddard, M.E., Visscher, P.M., 2011. GCTA: A tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88, 76–82. https://doi.org/10.1016/j. ajhg.2010.11.011.
- Zhu, Z., Chen, B., Na, R., Fang, W., Zhang, W., Zhou, Q., Zhou, S., Lei, H., Huang, A., Chen, T., Ni, D., Gu, Y., Liu, J., Rao, Y., Fang, F., 2020. A genome-wide association study reveals a substantial genetic basis underlying the Ebbinghaus illusion, 2020 663 J. Hum. Genet. 66, 261–271. https://doi.org/10.1038/s10038-020-00827-4.