

Dissecting the Epigenomic Differences between Smoking and Nicotine Dependence in a Veteran Cohort

Sheila Tiemi Nagamatsu^{1,2}, Robert H Pietrzak^{1,2,3}, Ke Xu^{1,2,3}, John H Krystal^{1,2,3}, Joel Gelernter^{1,2,3},
Janitza Liz Montalvo-Ortiz^{1,2,3}

Author Affiliations:

1 Division of Human Genetics, Department of Psychiatry, Yale University School of Medicine, New Haven, CT;

2 VA CT Healthcare Center, West Haven, CT;

3 U.S. Department of Veterans Affairs National Center of Posttraumatic Stress Disorder, Clinical Neurosciences Division, West Haven, CT.

Corresponding Author:

Janitza L. Montalvo-Ortiz, Ph.D.

Errera Community Care Center-Orange Annex, 200 Edison Road, Office #1339, Orange, CT 06477; Phone: (203)508-7772; email: janitza.montalvo-ortiz@yale.edu

Short title: Epigenomic markers of Smoking and Nicotine Dependence

Keywords: Smoking, Nicotine dependence, epigenomics, DNA methylation, co-methylation, Veterans

ABBREVIATIONS:

SS = smoking status

ND = nicotine dependence

EA = European American

DNAm = DNA methylation

GWS = genome-wide significant

DMS = differentially methylated sites

EWAS = epigenome-wide association studies

NHRVS = National Health and Resilience in Veterans Study

VA = Veterans Affairs

FTND = Fagerström Test for Nicotine Dependence

PPI = protein-protein interaction analysis

OER = Observed/expected rate

NUMBER OF WORDS, ABSTRACT: 250

NUMBER OF WORDS, TEXT: 3875

NUMBER OF TABLES:

Table 1. Demographic and clinical characteristics.

Table 2. Top 10 SS-related DMS.

Table 3. Top 10 ND-related DMS.

NUMBER OF FIGURES:

Figure 1. Manhattan plot for EWAS analysis.

Figure 2. Functional annotation of SS-related sites.

Figure 3. SS- and ND-related DMS.

Figure 4. Co-methylation analysis to SS and ND.

Figure 5. Functional annotation to SS- and ND-related modules.

Figure 6. Violin plot.

Figure 7. Nicotine addiction sites in steelblue module.

NUMBER OF SUPPLEMENTARY ITEMS:

Supplementary Figure S1. Distribution of Fagerström test score.

Supplementary Figure S2. Density plots before and after normalization.

Supplementary Figure S3. QQ plot to EWAS analysis.

Supplementary Figure S4. Violin plot to the top 10 ND- and SS-related DMS.

Supplementary Figure S5. Module membership and gene significance.

ABSTRACT

Background: Smoking is a serious public health issue linked to more than 8 million deaths per year worldwide. It also may lead to nicotine dependence (ND). Smoking can induce long-lasting epigenetic changes. Although epigenetic alterations related to tobacco smoke have been largely studied, few works have investigated ND and its interaction with smoking status (SS). **Objective:** We investigated the peripheral epigenomic profile of SS and ND in a U.S. male veteran cohort. **Methods:** DNA from saliva was collected from 1,135 European American (EA) male U.S. military veterans. DNAm was assessed using the Illumina Infinium Human MethylationEPIC BeadChip array. SS was evaluated as: current smokers ($n=137$; 12.1%) and non-current smokers (never and former smokers; $n=998$; 87.9%). ND was assessed using the Fagerström Test for Nicotine Dependence (FTND). EWAS and co-methylation analyses were conducted for SS and ND. **Results:** A total of 450 and 22 genome-wide significant differentially methylated sites (DMS) were associated with SS and ND, respectively (fifteen overlapped sites). We identified 97 DMS (43 genes) in SS-EWAS previously reported in the literature, including *AHRR*, and *F2RL3* genes (p -value range: 1.95×10^{-83} to 4.5×10^{-33}). ND novel DMS mapped to *NEUROG1*, *ANPEP*, and *SLC29A1*. Co-methylation analysis identified 386 modules (11 SS-related and 19 ND-related). SS-related modules showed enrichment for alcoholism, chemokine signaling pathway, and neurogenesis; while ND-related modules were enriched for cellular adhesion, and nicotine addiction. **Conclusions:** This study confirms previous findings and identifies novel and -potentially specific - epigenetic signatures for SS and ND in a sample of EA male veterans.

INTRODUCTION

Smoking is a serious public health issue associated with increased risk of numerous diseases such as cancer, stroke, and chronic obstructive pulmonary disease. Tobacco smoking is responsible for more than 8 million deaths per year, with highest prevalence in males aged between 45 to 54 years on a global scale (1). In the U.S., 13.7% of adults are current smokers (2), reaching even higher rates among U.S. veterans than non-veterans, about 29.2% according to the 2010-2015 National Survey on Drug Use and Healthy data (3).

Nicotine is the main component of tobacco smoking associated with addiction, but other constituents of tobacco smoke also contribute to the addiction process (4). Nicotine and other constituents of tobacco smoke act on brain reward circuits underlying both recreational and compulsive use (5). Regular tobacco use can produce nicotine dependence (ND), identified clinically by withdrawal symptoms upon abrupt discontinuation. Several distinguishable stages of the smoking addiction process are highly heritable, including smoking persistence ($h=46-59\%$) and ND ($h=\text{up to } 75\%$) (6). Genome-wide association studies (GWAS) have identified genetic risk variants for both smoking behavior and ND (5, 7, 8). However, as is often the case for complex traits, these variants only explain a small proportion of the variance in this disorder.

The interplay of genetic and environmental mechanisms is often expressed through epigenetic processes. DNA methylation (DNAm), one of the most studied epigenetic mechanisms, has been previously shown to be associated with smoking behavior. Epigenome-wide association studies (EWAS) approaches have identified epigenomic biomarkers associated with smoking in multiple tissues: blood (9-15), newborn blood (16), lung (17, 18), adipose tissue (19), buccal cells (20), and saliva (21)). While a large number of CpG sites have been associated with tobacco smoking status (SS), few studies to date have examined epigenomic changes associated with ND. One of the most important sets of findings from large-scale

GWAS of substance use disorders (SUDs) is that dependence traits and quantity/frequency traits often have fundamentally different genetic substrates (22-24). Quantity/frequency measures have been extensively studied for nicotine; dependence traits have been less commonly evaluated. Thus, one of the goals of this study, in separating our dependent and nondependent nicotine use, was to investigate whether similar use/dependence differences exist in the context of epigenetics. In this study, we conducted EWAS and co-methylation analyses for SS and ND in a veteran cohort (n=1,135) to elucidate the epigenomic signatures (shared and unique) of these two traits.

METHODS AND MATERIALS

Study population

Our cohort included 1,135 male veterans who self-reported European ancestry from the National Health and Resilience in Veterans Study (NHRVS) (25, 26). The sociodemographic and clinical characteristics of the study sample are presented in **Table 1**. Participants were recruited from the Knowledge Networks research panel, which comprises more than 50,000 households developed and maintained by the Ipsos survey research firm. The study was approved by the Human Subjects Subcommittee of the Veterans Affairs (VA) Connecticut Healthcare System and VA Office of Research & Development and all participants provided informed consent.

Smoking characteristics and FTND assessment

SS was evaluated as current smokers and non-current smokers (including both never and former smokers). ND was defined using the Fagerström Test for Nicotine Dependence (FTND) assessment (27). The study sample showed a 12.1% prevalence of current smokers (N=137; of which 48.9% are nicotine dependent) and lifetime FTND score distribution ranging from 0 to 10 (mean = 2.58 and standard deviation = 2.83; N=238 with FTND ≥ 5) (**Supplementary Figure S1**).

DNA extraction, sample preparation, and array processing

Genomic DNA (500ng) from saliva samples was extracted using Oragene kits (DNA Genotek, Ottawa, Ontario, Canada). The DNA was treated with bisulfite reagents using the EZ-96 DNA methylation kit (Zymo Research, Orange, CA, USA) following the standard protocol. DNAm was assessed with the Illumina Infinium Human MethylationEPIC BeadChip (Illumina, San Diego, CA, USA), which detected >850,000 loci across the genome. The DNAm assay was conducted at the Yale Center for Genome Analysis (YCGA) and GenomeStudio software (Illumina) was used to generate β values for all CpG sites. The β values were calculated from the total methylated signal (M) and the unmethylated signal (U) whereby $\beta = M/(M+U)$, which vary from 0.0 to 1.0.

Quality control and data normalization

Quality control analysis was conducted in R (version 3.4.1) using the ‘minfi’ package (Bioconductor 1.8.9) (28). Probes filtered out included those with detection p-value > 0.001, probes located in known single nucleotide polymorphisms (SNPs) at SBE/CpGs sites, probes mapped in multiple places in the genome, and probes in sex chromosomes. To correct for batch effects, the ComBat function was applied using the ‘sva’ package (29) and normalization was conducted using internal probes to control for technical variation. To evaluate the distribution of beta values, density plots were generated before and after normalization (**Supplementary Figure S2**).

Cell proportion estimation analysis was conducted using an adjusted version of the Houseman method (30). The analysis enabled an estimate of each cell type proportion (e.g. CD14, CD34, and buccal cells) in the heterogeneous peripheral saliva samples, which were included in the model. Principal component analysis was also performed to adjust for population stratification using the Barfield method (31), which uses sets of CpGs within 50kb of SNPs with minor allele frequency (MAF) > 0.1 described in the 1000 Genomes Project. The first 10 principal components were included in the model. After quality control, a total of 756,573 CpG sites (71.93%) were used for EWAS and co-methylation analyses.

Statistical Analysis

To identify DMS, we conducted an EWAS analysis using the ‘cpg.assoc’ function from the ‘minfi’ Bioconductor R package (28). Covariates included: age, estimates of cell type proportions (CD14, CD34, Buccal cells), and the first 10 principal components (PCs). In the ND EWAS analysis, SS was included as a covariate. Bonferroni correction was used to adjust for multiple testing.

SS EWAS model:

$$\beta \sim SS + Age + cell\ type\ proportion\ (CD14, CD34, Buccal\ cells) + PC1 - 10$$

ND EWAS model:

$$\beta \sim FTND + Age + cell\ type\ proportion\ (CD14, CD34, Buccal\ cells) + PC1 - 10 + SS$$

Gene Ontology analysis

The ‘gometh’ function was performed using the Bioconductor R package “missMethyl” (32), which accounts for the varying number of CpG sites per gene by assigning a prior probability for each gene based on gene length. This method applies a modified hypergeometric test for over-representation of a gene set. For this analysis, we selected CpG sites that showed the suggestive significance of $p < 1.0 \times 10^{-5}$. To correct for multiple testing, false discovery rate (FDR) was set at 0.05.

Interactome analysis

The protein-protein interaction (PPI) analysis was conducted using the STRING database (Version 11.0) (33). STRING combines the probability of different sources including genomic prediction, high-throughput experiments, co-expression analysis, text mining in available papers, and databases. It generates a score for each interaction to indicate the confidence in that interaction. SS- and ND-related genome-wide significant (GWS) DMS were used as input with a medium confidence of 0.4.

Co-methylation analysis

Normalized beta values were used to perform the co-methylation analysis, which was conducted using the “WGCNA” R package (Version 1.69) (34). Methylation level of CpG sites was clustered in modules represented by eigengene, defined as the first principal component of the module. We identified modules associated with SS and ND with $\text{correlation} \geq |0.07|$ and $p\text{-value} \geq 0.05$. For correlated modules, we evaluated the gene significance and module membership. Finally, we identified the modules enriched for DMS. GO enrichment analysis was conducted on all modules associated with SS or ND. Modules containing only one DMS were removed and the observed/expected rate (OER) was calculated. Functional annotation analysis was performed for the correlated modules and modules with $\text{OER} > 1$.

RESULTS

SS- and ND-related differential methylated sites

Manhattan plots for SS and ND are depicted in **Figure 1**. Quantile-quantile plots of each analysis are shown in **Supplementary Figure S3**, with lambda values of 1.41 and 1.11 for SS and ND, respectively. After Bonferroni correction, 450 DMS were identified for SS (**Supplementary Table 1**), and 22 DMS for ND (**Supplementary Table 2**).

GWS DMS associated with SS included 97 CpGs previously described in the literature as smoking-related, including the commonly identified *AHRR* (cg05575921, cg11554391, cg12806681, cg21161138, cg24090911, cg24688690, cg25648203, cg26703534), *F2RL3* (cg03636183), sites near *ALPPL2* gene (cg21566642, cg01940273, cg03329539), *RARA* (cg17739917, cg19572487), *CNTNAP2* (cg21322436, cg25949550), *CYP1B1* (cg20408276), *GFII* (cg09935388, cg12876356), *IER3* (cg14753356, cg15342087, cg24859433), *IGHJ6* (cg03686998, cg15310518, cg27271698), *PRSS23* (cg00475490, cg11660018, cg14391737). GWS DMS previously identified in the literature are listed in **Supplementary Table 1**. Further, we also identified a novel EWS DMS located near the *CYP1B1* gene (cg19753864), a gene previously described in epigenome studies as associated with smoking.

The ND-associated GWS DMS (**Table 3**) included 14 CpGs previously identified in the smoking EWAS literature mapping to *ALPPL2* (cg21566642, cg01940273), *AHRR* (cg05575921), *F2RL3* (cg03636183, cg21911711), *PRSS23* (cg14391737, cg00475490), *PPCDC* (cg18110140), *AC068134.6* (cg16841366), and intergenic region (cg01940273, cg14753356, cg00045592, cg07251887, cg06644428). In addition, we found eight novel GWS DMS associated with ND in *ANPEP* (cg06344992, cg23432008, cg02008229), within *CYP1B1* (cg19753864), *SLC29A1* (cg03634967, cg04175292), *NEUROG1* (cg04330449), and intergenic region (cg06581475).

GO (**Supplementary Table 3**) and KEGG (**Supplementary Table 4**) analyses were performed for SS- and ND-related GWS DMS. SS-associated GWS DMS were enriched to GO terms related to secondary metabolic processes, response to stimulus and xenobiotic metabolic processes (**Figure 2A**), and KEGG pathways (**Figure 2B**) related to chemokine signaling pathways, drug metabolism, and chemical carcinogenesis. For ND-related GWS DMS, no enriched terms were identified with FDR < 0.05.

Comparison between SS- and ND-associated DMS

SS- and ND-associated GWS DMS were examined for overlap to identify the common DMS between the two traits and those specific for each trait (**Figure 3A and 3B**). We identified 435 SS-specific DMS, 15 DMS shared between SS and ND, and 7 ND-specific DMS.

From the 435 SS-specific DMS identified, 152 were hypermethylated and 283 hypomethylated. DMS were identified mapped to genes previously associated with smoking, as *CSRNP1*, *ZMIZ1*, *CPAMD8*, *GFII*, *MTSS1*, *NFE2L2*, and *AKR1B1*.

All overlapping DMS between SS and ND showed the same direction of effect (**Supplementary Figure S4**), from which five were consistent with the smoking literature (cg21566642, cg05575921, cg01940273, cg14753356, cg19572487); there were no previous reports for the other DMS. The association was

statistically stronger in ND than in SS for the majority of ND-associated sites, except cg14391737 (SS: 3.25E-10, ND: 6.70E-21; *PRSS23*), cg00475490 (SS: 1.17E-12, ND: 1.93E-13; *PRSS23*), and cg06581475 (SS: 1.98E-08, ND: 8.25E-09).

From the 7 ND-specific DMS identified, 6 were hypermethylated in ND and 1 was hypomethylated (cg06644428; previously associated as smoking-related in the literature). Novel DMS were located in *ANPEP* (cg06344992, cg23432008 and cg02008229), *SLC29A1* gene (cg03634967 and cg04175292), *NEUROG1* gene (cg04330449). PPI analysis (**Figure 3C**) showed interactions between the gene products identified in SS and ND EWAS as SLC29A1 (ND) and SLC22A4 (SS).

Co-methylated modules related to SS and ND

A total of 386 modules were identified in the co-methylation analysis, of which 26 had more than 1000 CpGs. The eigengenes calculated to represent each module were tested for correlation with SS and ND. Low correlation was observed between SS and ND, reaching a maximum of 0.09 and a minimum of -0.14. We identified 11 modules associated with SS (**Figure 4A**) and 19 modules associated with ND (**Figure 4B**) with $\text{cor} \geq |0.07|$ and $p\text{-value} < 0.05$. No overlap was observed between SS and ND. OER are shown for modules containing DMS associated with SS (**Figure 4C**) and ND (**Figure 4D**).

Functional enrichment analysis (**Figure 5**) identified seven significantly enriched modules. For SS, enriched modules included yellow (positively correlated), brown (positively correlated and OER>1), blue (negatively correlated and OER >1), and pink (OER > 1). For ND, enriched modules included steelblue (negatively correlated), turquoise (positively correlated), pink (OER>1), brown (OER>1), and grey (OER>1). Yellow and blue modules were only correlated with SS and were enriched for Ras signaling pathway, cell adhesion, and ion transmembrane transport. Steelblue, turquoise, and grey modules were only correlated with ND and showed enrichment for nicotine addiction, synaptic signaling, and cellular

component organization. Brown and pink modules were detected in both SS and ND and enriched for cell communication, neuroactive ligand-receptor interaction, neurogenesis, and nervous system development.

DISCUSSION

Few works have investigated epigenetics marks associated with ND and evaluate its differences with SS. In this work, we studied DNA derived from saliva, i.e. peripheral tissue, in a cohort of male EA U.S. military veterans. We identified SS-associated GWS DMS in saliva trait previously found in blood, lung tissue, and adipose tissue. Our results identified 435 SS- and 7 ND-specific GWS DMS while also identifying those common between the two traits (15 DMS). Six sites associated with ND were novel, as were 356 associated with SS. The ND analysis identified three novel DMS in *NEUROGI*, *ANPEP*, and *SLC29A1*. Co-methylation analysis identified methylomic networks based on the intercorrelation between CpG sites. The SS analysis showed correlated modules enriched for neurodegenerative disorders such as Alzheimer and Parkinson's disease, as well as alcoholism. ND co-methylation analysis identified correlated modules involved in nicotine addiction.

From the 450 GWS DMS identified in the SS EWAS, at least 97 were previously described in the literature as being associated with smoking literature. Our top hit identified in the SS model (cg05575921, p -value=4.54E-83) is within the *AHRR* gene, a well-known locus that acts in the aryl hydrocarbon receptor signaling cascade. This CpG was hypomethylated in current-smokers compared to non-smokers, as reported in other peripheral tissues including blood (9-12), lung (17, 18), adipose tissue (19), and buccal cells (20). Further, 13 additional GWS DMS in *AHRR* were detected (minimum p -value = 5.73E-8), of which seven were previously described. The methylation pattern in the cg05575921 site is a known biomarker of smoking behavior, showing a reversible pattern after smoking cessation (9). *AHRR* variant (rs1051730) was also described as a likely causal genetic marker attributable to heavy smoking behavior and the hypomethylation in cg05575921 has been related to high smoking-related morbidity and with high all-cause mortality (9).

We also identified two GWS DMSs in the *F2RL3* gene (cg21911711, p-value = 4.48E-17; cg03636183, p-value = 5.61E-43). From a prior EWAS of smoking in blood, its methylation pattern may reflect tobacco exposure status, varying from less methylation in current smokers to higher levels in never smokers (35). A causal relationship has been suggested between *F2RL3* methylation pattern, smoking, and cardiovascular diseases related to smoking (35). An association of *F2RL3* hypomethylation with smoking has been also reported with lung cancer incidence and mortality (36), so it is also considered a strong predictor of mortality, including in smoking-related diseases (37).

We detected three DMS in the *PRSS23* gene (cg14391737, cg00475490 and cg11660018), which encodes a member of the trypsin family of serine proteases. Methylation changes in *PRSS23* gene was recently associated with cadmium and smoking exposure in a EWAS study of the Strong Heart Study (38). Further, it was reported the *PRSS23* is essential for the cardiac valvulogenesis in zebrafish embryos (39), and is associated with tumor progression in cancer (40, 41). Besides, the repression of *PRSS23* gene by the miR-532 showed a protective effect in cultured cardiac endothelial cells (42). Another notable site is cg21566642, located near the *ALPPL2* gene. *ALPPL2* is a member of the alkaline phosphatase family and has high homology with the placental alkaline phosphatase (*ALPP*) (43). Its expression was detected in different types of mesotheliomas, being classified as a tumor-specific antigen (43).

SS EWAS DMSs show enrichment of genes involved in metabolic pathways and biological functions related to toxic metabolic processes, xenobiotics metabolic processes, and response to xenobiotic stimulus. This is consistent with the physiological necessity to clear the products of combustion inhaled by the smoker. In a study examining metabolite composition of smokers, xenobiotics have been identified as associated with smoking behavior (44). Further, we also identified enrichment of genes involved in the drug metabolism pathway and drug metabolism by cytochrome P450, a membrane-bound in cellular organelles (mitochondria and endoplasmic reticulum) involved in drug, steroid, and carcinogen metabolism (45).

In the ND EWAS, at least 14 GWS DMS (from 22) were previously described in the smoking literature, including *AHRR*, *F2RL3*, and a CpG near *ALPPL2* (top DMS in ND analysis). From all 22, we observed 15 DMS overlapped with SS, of which twelve showed a smaller effect size ($(\text{Mean}_{\beta\text{SS}} - \text{Mean}_{\beta\text{non-smoker}})/\text{Sd}$) in the analysis of SS relative to ND. For three DMS with a larger effect size in ND than in SS, two were within the *PRSS23* gene. A violin plot (**Figure 6**) of cg21566642(*ALPPL2*), cg14391737 (*PRSS23*), cg05575921 (*AHRR*), and cg04330449(*NEUROG1*) showed a similar direction between current-smokers and ND after excluding smokers and individuals with nicotine dependence.

ND-specific GWS DMS map to *NEUROG1*, *ANPEP*, and *SLC29A1*. Of the ND-specific DMS, only one was previously reported as SS-related: cg06644428 (10, 19). From the ND-specific GWS DMSs, though six DMSs were not previously reported in EWAS of smoking, three genes were already described in smoking literature. The *NEUROG1* gene encodes a transcriptional regulator that may promote neuronal differentiation (46). Differential methylation in this gene was previously associated with maternal smoking during pregnancy in the offspring cord blood (47). Though the study did not differentiate between SS and ND and they detected a DMS in a different CpG site, the direction effect was the same as in our study. However, no information about nicotine dependence prevalence in the mothers were provided or tested in that study. Our study suggests that differential methylation at *NEUROG1* may be specific to ND and therefore may be a potential biomarker for ND. An additional ND-specific DMS was located in the *ANPEP* gene. *ANPEP* is a diabetes risk gene (48), which was recently identified as SS-associated in a EWAS examining the effects of cadmium and smoking in the Strong Heart Study cohort in blood tissue (38). *SLC29A1* mediates uptake of nucleosides and is not correlated to smoking in EWAS.

The network containing the annotated genes in SS and ND-related EWAS showed a correlation between all genes, including mostly textmining and co-expression information, further supporting the functionality/biological relevance of our findings. More research is needed to replicate these ND findings in independent cohorts to confirm these associations. No enrichment was found to ND-associated GWS DMS.

In the SS co-methylation analysis, we identified two positively correlated modules, yellow and brown. The yellow module showed enrichment for ion transmembrane transport, regulation of Rho protein signal transduction, and cytoskeleton organization. In a study using epithelial 3D culture, a significant reduction in ion channel function in cells treated with tobacco product preparation was observed, while no effect was observed using nicotine alone (49). The brown module showed enrichment related to cell communication and cell adhesion. This biological process has been previously identified in a smoking transcriptome analysis (50). This indicates that methylation changes observed in our work may impact gene expression.

We also found one negatively correlated module (blue), which showed enriched terms related to cell junctions' organization, extracellular matrix organization, and actin filament-based process. Previous studies have found that cigarette smoking changes the secretion of proteins related to extracellular matrix organization in lung epithelial cells responded, indicating smoking-induced tissue remodeling (51). These three modules included DMS, however, just brown and blue modules showed OER > 1. Our findings demonstrate the effects of smoking on multiple biological processes via epigenetic changes.

In the ND co-methylation analysis, the top correlated modules included the steelblue (negatively correlated) and the turquoise (positively correlated). The steelblue module is enriched for chemical synaptic transmission, cell-cell signaling, and nicotine addiction. The genes related to nicotine addiction in those modules are depicted in **Figure 7** and include *GABRA1* (cg12340947), *GABRA2* (cg09430391), *GABRA4* (cg24154839), *GABRB2* (cg11042722), and *GRIN1* (cg10538151). *GABRA1*, *GABRA2*, *GABRA4*, *GABRB2* encode parts of the GABA receptor. GWAS studies have found that variations in *GABRA2* gene influences the risk and susceptibility to alcohol dependence (52, 53). Further, genetic variants in *GABRA2* gene have been associated with ND (54) and alcohol use (55) in the candidate gene era; they await genomewide-level replication. Although mutations in GABA receptors have not been directly associated with smoking, studies suggest a role in nicotine addiction. Animal studies have shown that *GABA_B* receptors moderate the effects of reward induced by nicotine (56). In bronchial epithelial cells

of rhesus monkeys, nicotine induces an upregulation of GABA signaling (57). The turquoise module showed enriched terms related to cellular organization and focal adhesion. Previous studies showed that cell adhesion metabolism has been implicated in the maintenance of connectivity in the nervous system, in which neurexins are involved (58). Multiple CpGs within Neurexin-1 β (*NRXN1*) were identified in the turquoise module. *NRXN1* is involved in nicotinic acetylcholine receptors' activity and neurotransmitter release in the presynaptic and postsynaptic sites (59). We also identified one module for ND with OER > 1, which indicates an enrichment of ND-associated DMS (grey). The grey module showed enrichment for chemical synaptic transmission and neuroactive ligand-receptor interaction. It has been shown that nicotine exposure during cortical development induces alterations in histone methylation responsible for regulating the expression of genes involved in maintenance of glutaminergic synapses (60). Our findings suggest a link between the neuronal effect of ND and methylation patterns in saliva. Two modules indicate enrichment for DMS from both SS and ND analysis, the brown (previously discussed) and pink module. The pink module showed enrichment of pathways related to regulation of cell differentiation, neurogenesis, and nervous system development. A recent study showed that exposure to intrauterine smoking might alter DNAm patterns associated with mature neuronal content (61).

The main limitation of the study is the inclusion of former smokers in the non-current smokers group, which may reduce the contrast between the methylation pattern of smokers and non-current smokers. Further, while efforts were made to carefully ascertain ND by adding SS as a covariate in the analysis, this approach may not entirely capture the effects of SS in our ND analysis, as allowed the identification of genes related with toxicity in ND approach. Therefore, research is needed to further confirm potential ND-specific associations, analyze the impact of nicotine vs. non-nicotine constituents on toxicity and addiction, and replicate these findings in independent cohorts. While this helps to improve statistical power of a comparatively small sample, ongoing efforts include more samples, the addition of women as well as other ancestries, such as African, to evaluate whether these associations can also be observed

across sexes and populations. Future work will focus on examining these associations in human postmortem brain tissue and animal models to further dissect the differences between SS and ND.

In conclusion, we detected 450 SS-related DMS and 22 ND-related DMS, which 357 and six was novel findings in smoking literature for SS and ND analyses, respectively. We replicated previous epigenetic findings of studies evaluating in blood, lung tissue, and adipose tissue, which indicates the potential of saliva to study epigenetic biomarkers of smoking based on the easy access of this tissue. Further, co-methylation analysis revealed different functional modules between SS and ND modules, for example, ion transmembrane transport for smoking and nicotine addiction for ND. These findings add to the growing literature of epigenetic changes of smoking by evaluating these using two different approaches, examining saliva, and studying a veteran cohort, while also revealing epigenetic changes potentially specific to ND. As has been observed for genetics of several SUDs, epigenetics of nicotine use traits shows apparent distinction between use and dependence. Our EWAS findings identified GWS DMS for both SS and ND. Future studies are needed to confirm and evaluate the specificity of these findings in larger independent cohorts.

ACKNOWLEDGMENTS

This work is supported by the Department of Veterans Affairs via 1IK2CX002095-01A1 (JLMO), NIDA R21DA050160 (JLMO), **P50AA012870** (JHK), DA12890 (JG), VA MERIT I01CX001849 (JG), and the U.S. Department of Veterans Affairs National Center for Posttraumatic Stress Disorder.

DISCLOSURES

JHK Financial Disclosure: Consultant (Note: – The Individual Consultant Agreements listed below are less than \$5,000 per year): Aptinyx, Inc.; Atai Life Sciences; AstraZeneca Pharmaceuticals; Biogen, Idec,; MA; Biomedisyn Corporation; Bionomics, Limited (Australia); Boehringer Ingelheim International; Cadent Therapeutics, Inc.; Clexio Bioscience, Ltd.; COMPASS Pathways, Limited, United

Kingdom; Concert Pharmaceuticals, Inc.; Epiodyne, Inc.; EpiVario, Inc.; Greenwich Biosciences, Inc.; Heptares Therapeutics, Limited (UK); Janssen Research & Development; Jazz Pharmaceuticals, Inc.; Otsuka America Pharmaceutical, Inc.; Perception Neuroscience Holdings, Inc.; Spring Care, Inc.; Sunovion Pharmaceuticals, Inc.; Takeda Industries; Taisho Pharmaceutical Co., Ltd. Board of Directors: Freedom Biosciences, Inc.. Scientific Advisory Board: Biohaven Pharmaceuticals; BioXcel Therapeutics, Inc. (Clinical Advisory Board); Cadent Therapeutics, Inc. (Clinical Advisory Board); Cerevel; Therapeutics, LLC; EpiVario, Inc.; Eisai, Inc.; Jazz Pharmaceuticals, Inc.; Lohocla Research Corporation; Novartis Pharmaceuticals Corporation; PsychoGenics, Inc.; RBNC Therapeutics, Inc.; Tempero Bio, Inc.; Terran Biosciences, Inc.. Stock: Biohaven Pharmaceuticals; Sage Pharmaceuticals; Spring Care, Inc.. Stock Options: Biohaven Pharmaceuticals Medical Sciences; EpiVario, Inc.; RBNC Therapeutics, Inc.; Terran Biosciences, Inc.; Tempero Bio, Inc.

JHK Financial Disclosure: Income Greater than \$10,000. Editorial Board: Editor - Biological Psychiatry.

Patents and Inventions: 1. Seibyl JP, Krystal JH, Charney DS. Dopamine and noradrenergic reuptake inhibitors in treatment of schizophrenia. US Patent #:5,447,948. September 5, 1995; 2. Vladimir, Coric, Krystal, John H, Sanacora, Gerard – Glutamate Modulating Agents in the Treatment of Mental Disorders. US Patent No. 8,778,979 B2 Patent Issue Date: July 15, 2014. US Patent Application No. 15/695,164: Filing Date: 09/05/2017; 3. Charney D, Krystal JH, Manji H, Matthew S, Zarate C., - Intranasal Administration of Ketamine to Treat Depression United States Patent Number: 9592207, Issue date: 3/14/2017. Licensed to Janssen Research & Development; 4. Zarate, C, Charney, DS, Manji, HK, Mathew, Sanjay J, Krystal, JH, Yale University “Methods for Treating Suicidal Ideation”, Patent Application No. 15/379,013 filed on December 14, 2016 by Yale University Office of Cooperative Research; 5. Arias A, Petrakis I, Krystal JH. – Composition and methods to treat addiction. Provisional Use Patent Application no.61/973/961. April 2, 2014. Filed by Yale University Office of Cooperative Research.; 6. Chekroud, A., Gueorguieva, R., & Krystal, JH. “Treatment Selection for Major Depressive Disorder” [filing date 3rd June 2016, USPTO docket number Y0087.70116US00]. Provisional patent submission by Yale University; 7. Gihyun, Yoon, Petrakis I, Krystal JH – Compounds, Compositions and

Methods for Treating or Preventing Depression and Other Diseases. U. S. Provisional Patent Application No. 62/444,552, filed on January 10, 2017 by Yale University Office of Cooperative Research OCR 7088 US01; 8. Abdallah, C, Krystal, JH, Duman, R, Sanacora, G. Combination Therapy for Treating or Preventing Depression or Other Mood Diseases. U.S. Provisional Patent Application No. 62/719,935 filed on August 20, 2018 by Yale University Office of Cooperative Research OCR 7451 US01; 9. John Krystal, Godfrey Pearlson, Stephanie O'Malley, Marc Potenza, Fabrizio Gasparini, Baltazar Gomez-Mancilla, Vincent Malaterre. Mavoglurant in treating gambling and gaming disorders. U.S. Provisional Patent Application No. 63/125,181 filed on December 14, 2020 by Yale University Office of Cooperative Research OCR 8065 US00; NON Federal Research Support: AstraZeneca Pharmaceuticals provides the drug, Saracatinib, for research related to NIAAA grant "Center for Translational Neuroscience of Alcoholism [CTNA-4]; Novartis provides the drug, Mavoglurant, for research related to NIAAA grant "Center for Translational Neuroscience of Alcoholism [CTNA-4]; Cerevel provides the drug PF-06412562 for A Translational and Neurocomputational Evaluation of a D1R Partial Agonist for Schizophrenia (1 U01 MH121766-01).

No other co-author reports conflict of interest.

REFERENCES

1. WHO (2018): Global report on trends in prevalence of tobacco smoking 2000–2025. Second edition ed.
2. CDC (2019): Current cigarette smoking among adults in the United States. Center for Disease Control and Prevention.
3. Odani S, Agaku IT, Graffunder CM, Tynan MA, Armour BS (2018): Tobacco Product Use Among Military Veterans - United States, 2010-2015. *MMWR Morb Mortal Wkly Rep.* 67:7-12.

4. Cosgrove KP, McKay R, Esterlis I, Kloczynski T, Perkins E, Bois F, et al. (2014): Tobacco smoking interferes with GABAA receptor neuroadaptations during prolonged alcohol withdrawal. *Proc Natl Acad Sci U S A*. 111:18031-18036.
5. Hancock DB, Markunas CA, Bierut LJ, Johnson EO (2018): Human Genetics of Addiction: New Insights and Future Directions. *Curr Psychiatry Rep*. 20:8.
6. Markunas CA, Semick SA, Quach BC, Tao R, Deep-Soboslay A, Carnes MU, et al. (2021): Genome-wide DNA methylation differences in nucleus accumbens of smokers vs. nonsmokers. *Neuropsychopharmacology*. 46:554-560.
7. Quach BC, Bray MJ, Gaddis NC, Liu M, Palviainen T, Minica CC, et al. (2020): Expanding the genetic architecture of nicotine dependence and its shared genetics with multiple traits. *Nat Commun*. 11:5562.
8. Chen J, Loukola A, Gillespie NA, Peterson R, Jia P, Riley B, et al. (2020): Genome-Wide Meta-Analyses of FTND and TTFC Phenotypes. *Nicotine Tob Res*. 22:900-909.
9. Bojesen SE, Timpson N, Relton C, Davey Smith G, Nordestgaard BG (2017): AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. *Thorax*. 72:646-653.
10. Sayols-Baixeras S, Lluís-Ganella C, Subirana I, Salas LA, Vilahur N, Corella D, et al. (2015): Identification of a new locus and validation of previously reported loci showing differential methylation associated with smoking. The REGICOR study. *Epigenetics*. 10:1156-1165.
11. Su D, Wang X, Campbell MR, Porter DK, Pittman GS, Bennett BD, et al. (2016): Distinct Epigenetic Effects of Tobacco Smoking in Whole Blood and among Leukocyte Subtypes. *PLoS One*. 11:e0166486.
12. Imboden M, Wielscher M, Rezwan FI, Amaral AFS, Schaffner E, Jeong A, et al. (2019): Epigenome-wide association study of lung function level and its change. *Eur Respir J*. 54.
13. Wiklund P, Karhunen V, Richmond RC, Parmar P, Rodriguez A, De Silva M, et al. (2019): DNA methylation links prenatal smoking exposure to later life health outcomes in offspring. *Clin Epigenetics*. 11:97.

14. Richard MA, Huan T, Ligthart S, Gondalia R, Jhun MA, Brody JA, et al. (2017): DNA Methylation Analysis Identifies Loci for Blood Pressure Regulation. *Am J Hum Genet.* 101:888-902.
15. Miao L, Yin RX, Zhang QH, Hu XJ, Huang F, Chen WX, et al. (2019): Integrated DNA methylation and gene expression analysis in the pathogenesis of coronary artery disease. *Aging (Albany NY).* 11:1486-1500.
16. Sikdar S, Joehanes R, Joubert BR, Xu CJ, Vives-Usano M, Rezwan FI, et al. (2019): Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking. *Epigenomics.* 11:1487-1500.
17. Stueve TR, Li WQ, Shi J, Marconett CN, Zhang T, Yang C, et al. (2017): Epigenome-wide analysis of DNA methylation in lung tissue shows concordance with blood studies and identifies tobacco smoke-inducible enhancers. *Hum Mol Genet.* 26:3014-3027.
18. de Vries M, van der Plaat DA, Nedeljkovic I, Verkaik-Schakel RN, Kooistra W, Amin N, et al. (2018): From blood to lung tissue: effect of cigarette smoke on DNA methylation and lung function. *Respir Res.* 19:212.
19. Tsai PC, Glastonbury CA, Eliot MN, Bollepalli S, Yet I, Castillo-Fernandez JE, et al. (2018): Smoking induces coordinated DNA methylation and gene expression changes in adipose tissue with consequences for metabolic health. *Clin Epigenetics.* 10:126.
20. Jessen WJ, Borgerding MF, Prasad GL (2018): Global methylation profiles in buccal cells of long-term smokers and moist snuff consumers. *Biomarkers.* 23:625-639.
21. Barcelona V, Huang Y, Brown K, Liu J, Zhao W, Yu M, et al. (2019): Novel DNA methylation sites associated with cigarette smoking among African Americans. *Epigenetics.* 14:383-391.
22. Gelernter J, Polimanti R (2021): Genetics of substance use disorders in the era of big data. *Nat Rev Genet.*
23. Johnson EC, Demontis D, Thorgeirsson TE, Walters RK, Polimanti R, Hatoum AS, et al. (2020): A large-scale genome-wide association study meta-analysis of cannabis use disorder. *Lancet Psychiatry.* 7:1032-1045.

24. Zhou H, Sealock JM, Sanchez-Roige S, Clarke TK, Levey DF, Cheng Z, et al. (2020): Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. *Nat Neurosci.* 23:809-818.
25. Fuehrlein BS, Kachadourian LK, DeVlyder EK, Trevisan LA, Potenza MN, Krystal JH, et al. (2018): Trajectories of alcohol consumption in U.S. military veterans: Results from the National Health and Resilience in Veterans Study. *Am J Addict.*
26. Stefanovics EA, Potenza MN, Pietrzak RH (2018): The physical and mental health burden of obesity in U.S. veterans: Results from the National Health and Resilience in Veterans Study. *J Psychiatr Res.* 103:112-119.
27. Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO (1991): The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br J Addict.* 86:1119-1127.
28. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. (2014): Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* 30:1363-1369.
29. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD (2012): The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics.* 28:882-883.
30. Houseman EA, Kile ML, Christiani DC, Ince TA, Kelsey KT, Marsit CJ (2016): Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. *BMC Bioinformatics.* 17:259.
31. Barfield RT, Almli LM, Kilaru V, Smith AK, Mercer KB, Duncan R, et al. (2014): Accounting for population stratification in DNA methylation studies. *Genet Epidemiol.* 38:231-241.
32. Phipson B, Maksimovic J, Oshlack A (2016): missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics.* 32:286-288.
33. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. (2019): STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47:D607-D613.

34. Langfelder P, Horvath S (2008): WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 9:559.
35. Zhang Y, Yang R, Burwinkel B, Breitling LP, Brenner H (2014): F2RL3 methylation as a biomarker of current and lifetime smoking exposures. *Environ Health Perspect*. 122:131-137.
36. Zhang Y, Schottker B, Ordonez-Mena J, Holleczeck B, Yang R, Burwinkel B, et al. (2015): F2RL3 methylation, lung cancer incidence and mortality. *Int J Cancer*. 137:1739-1748.
37. Zhang Y, Yang R, Burwinkel B, Breitling LP, Holleczeck B, Schottker B, et al. (2014): F2RL3 methylation in blood DNA is a strong predictor of mortality. *Int J Epidemiol*. 43:1215-1225.
38. Domingo-Relloso A, Riffo-Campos AL, Haack K, Rentero-Garrido P, Ladd-Acosta C, Fallin DM, et al. (2020): Cadmium, Smoking, and Human Blood DNA Methylation Profiles in Adults from the Strong Heart Study. *Environ Health Perspect*. 128:67005.
39. Chen IH, Wang HH, Hsieh YS, Huang WC, Yeh HI, Chuang YJ (2013): PRSS23 is essential for the Snail-dependent endothelial-to-mesenchymal transition during valvulogenesis in zebrafish. *Cardiovasc Res*. 97:443-453.
40. Han B, Yang Y, Chen J, He X, Lv N, Yan R (2019): PRSS23 knockdown inhibits gastric tumorigenesis through EIF2 signaling. *Pharmacol Res*. 142:50-57.
41. Chan HS, Chang SJ, Wang TY, Ko HJ, Lin YC, Lin KT, et al. (2012): Serine protease PRSS23 is upregulated by estrogen receptor alpha and associated with proliferation of breast cancer cells. *PLoS One*. 7:e30397.
42. Bayoumi AS, Teoh JP, Aonuma T, Yuan Z, Ruan X, Tang Y, et al. (2017): MicroRNA-532 protects the heart in acute myocardial infarction, and represses prss23, a positive regulator of endothelial-to-mesenchymal transition. *Cardiovasc Res*. 113:1603-1614.
43. Su Y, Zhang X, Bidlingmaier S, Behrens CR, Lee NK, Liu B (2020): ALPPL2 Is a Highly Specific and Targetable Tumor Cell Surface Antigen. *Cancer Res*. 80:4552-4564.
44. Gu F, Derkach A, Freedman ND, Landi MT, Albanes D, Weinstein SJ, et al. (2016): Cigarette smoking behaviour and blood metabolomics. *Int J Epidemiol*. 45:1421-1432.

45. Guengerich FP (2019): Cytochrome P450 research and The Journal of Biological Chemistry. *J Biol Chem.* 294:1671-1680.
46. Song Z, Jadali A, Fritzsche B, Kwan KY (2017): NEUROG1 Regulates CDK2 to Promote Proliferation in Otic Progenitors. *Stem Cell Reports.* 9:1516-1529.
47. Kupers LK, Xu X, Jankipersadsing SA, Vaez A, la Bastide-van Gemert S, Scholtens S, et al. (2015): DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring. *Int J Epidemiol.* 44:1224-1237.
48. Ligthart S, Steenaard RV, Peters MJ, van Meurs JB, Sijbrands EJ, Uitterlinden AG, et al. (2016): Tobacco smoking is associated with DNA methylation of diabetes susceptibility genes. *Diabetologia.* 59:998-1006.
49. Rayner RE, Makena P, Prasad GL, Cormet-Boyaka E (2019): Cigarette and ENDS preparations differentially regulate ion channels and mucociliary clearance in primary normal human bronchial 3D cultures. *Am J Physiol Lung Cell Mol Physiol.* 317:L295-L302.
50. Vink JM, Jansen R, Brooks A, Willemsen G, van Grootheest G, de Geus E, et al. (2017): Differential gene expression patterns between smokers and non-smokers: cause or consequence? *Addict Biol.* 22:550-560.
51. Mossina A, Lukas C, Merl-Pham J, Uhl FE, Mutze K, Schamberger A, et al. (2017): Cigarette smoke alters the secretome of lung epithelial cells. *Proteomics.* 17.
52. Edenberg HJ, Dick DM, Xuei X, Tian H, Almasy L, Bauer LO, et al. (2004): Variations in GABRA2, encoding the alpha 2 subunit of the GABA(A) receptor, are associated with alcohol dependence and with brain oscillations. *Am J Hum Genet.* 74:705-714.
53. Lappalainen J, Krupitsky E, Remizov M, Pchelina S, Taraskina A, Zvartau E, et al. (2005): Association between alcoholism and gamma-amino butyric acid alpha2 receptor subtype in a Russian population. *Alcohol Clin Exp Res.* 29:493-498.

54. Agrawal A, Pergadia ML, Balasubramanian S, Saccone SF, Hinrichs AL, Saccone NL, et al. (2009): Further evidence for an association between the gamma-aminobutyric acid receptor A, subunit 4 genes on chromosome 4 and Fagerstrom Test for Nicotine Dependence. *Addiction*. 104:471-477.
55. Yang BZ, Arias AJ, Feinn R, Krystal JH, Gelernter J, Petrakis IL (2017): GRIK1 and GABRA2 Variants Have Distinct Effects on the Dose-Related Subjective Response to Intravenous Alcohol in Healthy Social Drinkers. *Alcohol Clin Exp Res*. 41:2025-2032.
56. Varani AP, Pedron VT, Aon AJ, Hocht C, Acosta GB, Bettler B, et al. (2018): Nicotine-induced molecular alterations are modulated by GABAB receptor activity. *Addict Biol*. 23:230-246.
57. Fu XW, Wood K, Spindel ER (2011): Prenatal nicotine exposure increases GABA signaling and mucin expression in airway epithelium. *Am J Respir Cell Mol Biol*. 44:222-229.
58. Comoletti D, Miller MT, Jeffries CM, Wilson J, Demeler B, Taylor P, et al. (2010): The macromolecular architecture of extracellular domain of alphaNRXN1: domain organization, flexibility, and insights into trans-synaptic disposition. *Structure*. 18:1044-1053.
59. Cheng SB, Amici SA, Ren XQ, McKay SB, Treuil MW, Lindstrom JM, et al. (2009): Presynaptic targeting of alpha4beta 2 nicotinic acetylcholine receptors is regulated by neurexin-1beta. *J Biol Chem*. 284:23251-23259.
60. Jung Y, Hsieh LS, Lee AM, Zhou Z, Coman D, Heath CJ, et al. (2016): An epigenetic mechanism mediates developmental nicotine effects on neuronal structure and behavior. *Nat Neurosci*. 19:905-914.
61. Chatterton Z, Hartley BJ, Seok MH, Mendelev N, Chen S, Milekic M, et al. (2017): In utero exposure to maternal smoking is associated with DNA methylation alterations and reduced neuronal content in the developing fetal brain. *Epigenetics Chromatin*. 10:4.

TABLES AND FIGURE LEGENDS

Table 1. Demographic and clinical characteristics.

Table 2. Top 10 SS-related DMS. The effect (E) is the difference between beta values from SS and non-SS samples ($\text{Effect} = (\text{Mean}\beta_{\text{smoking}} - \text{Mean}\beta_{\text{non-smoker}})/\text{Sd}$).

Table 3. Top10 ND-related DMS. The effect (E) is the difference between beta values from ND and non-ND samples ($\text{Effect} = (\text{Mean}\beta_{\text{ND}} - \text{Mean}\beta_{\text{non-ND}})/\text{Sd}$).

Figure 1. Manhattan plot for EWAS analysis. A) SS model; B) ND model.

Figure 2. Functional annotation of smoking-related sites. A) GO enrichment; and B) KEGG enrichment to the CpGs selected after Bonferroni correction.

Figure 3. Smoking- and ND-related DMS. A) Venn diagram of SS- and ND-related CpGs; B) Venn diagram of the annotated genes to the SS- and ND-related DMS. C) protein-protein interaction SS-related DMS and ND-related DMS (medium-confidence rate). The green nodes represent smoking, the pink represents ND and blue represents the overlapping.

Figure 4. Co-methylation analysis to SS and ND. A) Correlated modules to SS ($\text{cor} \geq |0.07|$); B) Correlated modules to ND ($\text{cor} \geq |0.07|$); C) Enrichment rate of SS-related CpGs; D) Enrichment rate of ND-related CpGs. Abbreviations: rylb = royalblue; brwn = brown; mdng = midnightblue; prpl = purple; mgnt = magenta; trqs = turquoise; ylw = yellow; blk = black.

Figure 5. Functional annotation to SS- and ND-related modules. A) GO and KEGG enrichment of steelblue module; B) GO and KEGG enrichment of brown module; C) GO and KEGG enrichment of pink module; D) GO and KEGG enrichment of grey module; E) GO and KEGG enrichment of blue module; F) GO enrichment of yellow module; G) GO enrichment of turquoise module.

Figure 6. Violin plot. The graph shows the beta values in the y-axis and four groups on the x-axis: 1) Both grouped samples that are smokers and nicotine dependents; 2) ND includes only nicotine-dependent; 3) non-smokers included the control group for both ND and smoking; and 4) Smoker included only samples that are current smokers.

Figure 7. Nicotine addiction sites in steelblue module. The networking shows the proteins encoded by the annotated genes in the steelblue module. In red are the proteins associated with the nicotine addiction pathway.

Table 1.

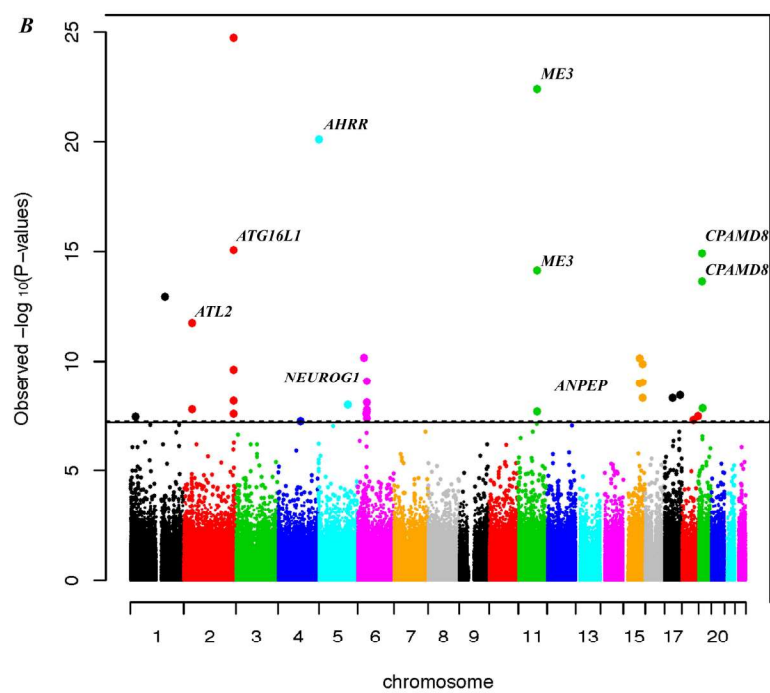
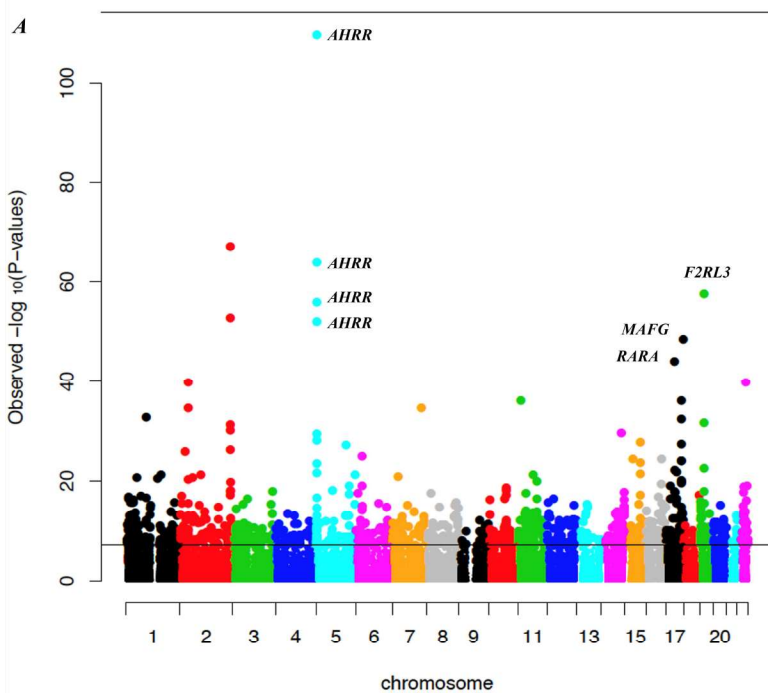
		Total
		N = 1,135
Age		
	All sample	64.3 ± 12.2
	Current smokers	58.7 ± 11.5
	Non-smokers	65.0 ± 12.1
Race		
	European Americans	1135 (100%)
Smoking status		
	Current smokers	137 (12.1%)
	Non-smokers	998 (87.9%)
Nicotine Dependence		
	ND cases	236
Alcohol dependence		
	Current smokers	38
	Non-smokers	147
Drug dependence		
	Current smokers	17
	Non-smokers	52

Table 2.

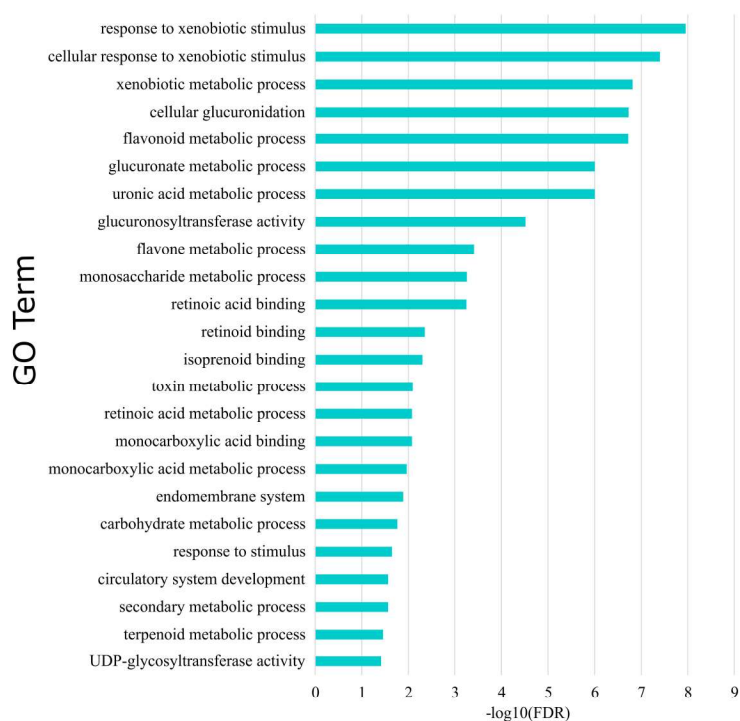
CpG Name	Chr	Position	Gene Annotation	P-value	Effect	Literature
						References
cg05575921	chr5	373378	AHRR	4.54E-83	-2.0189	B, SB, T, Wk, D
cg21566642	chr2	233284661	ALPPL2	3.87E-50	-1.7522	S, SB, T, D
cg25648203	chr5	395444	AHRR	8.91E-48	-0.3967	S, SB, T, D
cg03636183	chr19	17000585	F2LR3	8.91E-48	-1.1037	S, SB, D
cg21161138	chr5	399360	AHRR	9.05E-42	-1.5547	S, SB, Wk, D
cg01940273	chr2	233284934	ALPPL2	2.14E-39	-1.8733	S, SB, T, D
cg26703534	chr5	377358	AHRR	6.93E-39	-0.3132	S, D
cg07824483	chr17	79882042	MAFG	3.79E-36	-0.0212	

Table 3.

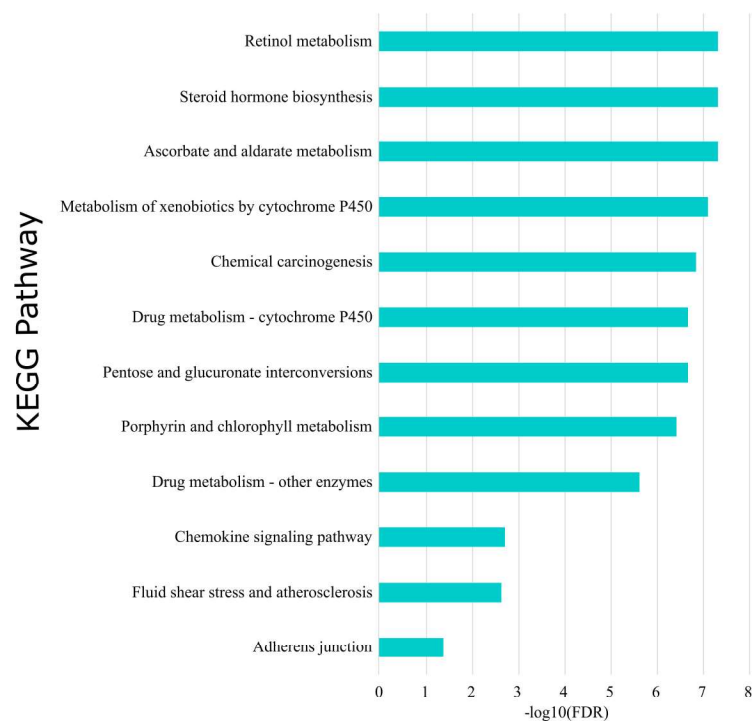
CpG Name	Chr	Position	Gene Annotation	P-value	Effect	Literature
						References
cg21566642	chr2	233284661	ALPPL2	5.71E-23	-0.8035	S, SB, T, D
cg14391737	chr11	86513429	PRSS23	6.70E-21	-0.7414	D
cg05575921	chr5	373378	AHRR	8.56E-19	-0.8692	B, SB, T, Wk, D
cg01940273	chr2	233284934	ALPPL2	2.87E-14	-0.5798	S, SB, T, D
cg21911711	chr19	16998668	F2RL3	3.90E-14	-0.5741	D
cg00475490	chr11	86517110	PRSS23	1.93E-13	-0.6170	D
cg03636183	chr19	17000585	F2RL3	5.23E-13	-0.6784	S, SB, D
cg00045592	chr1	160714299		2.30E-12	-0.4137	D



A

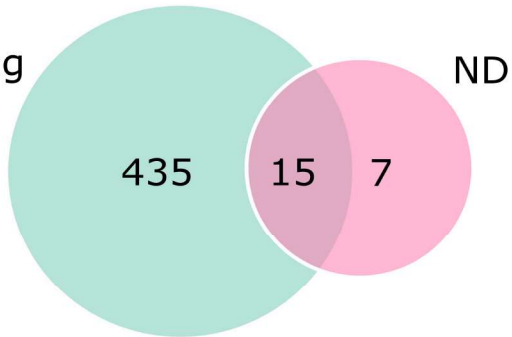


B



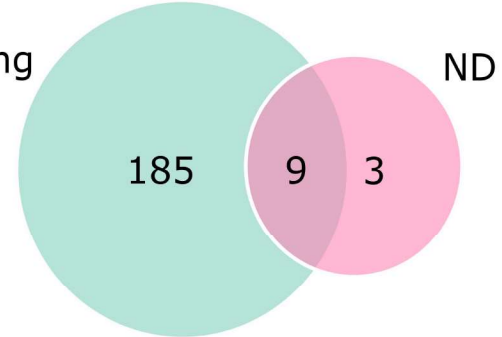
A

Smoking

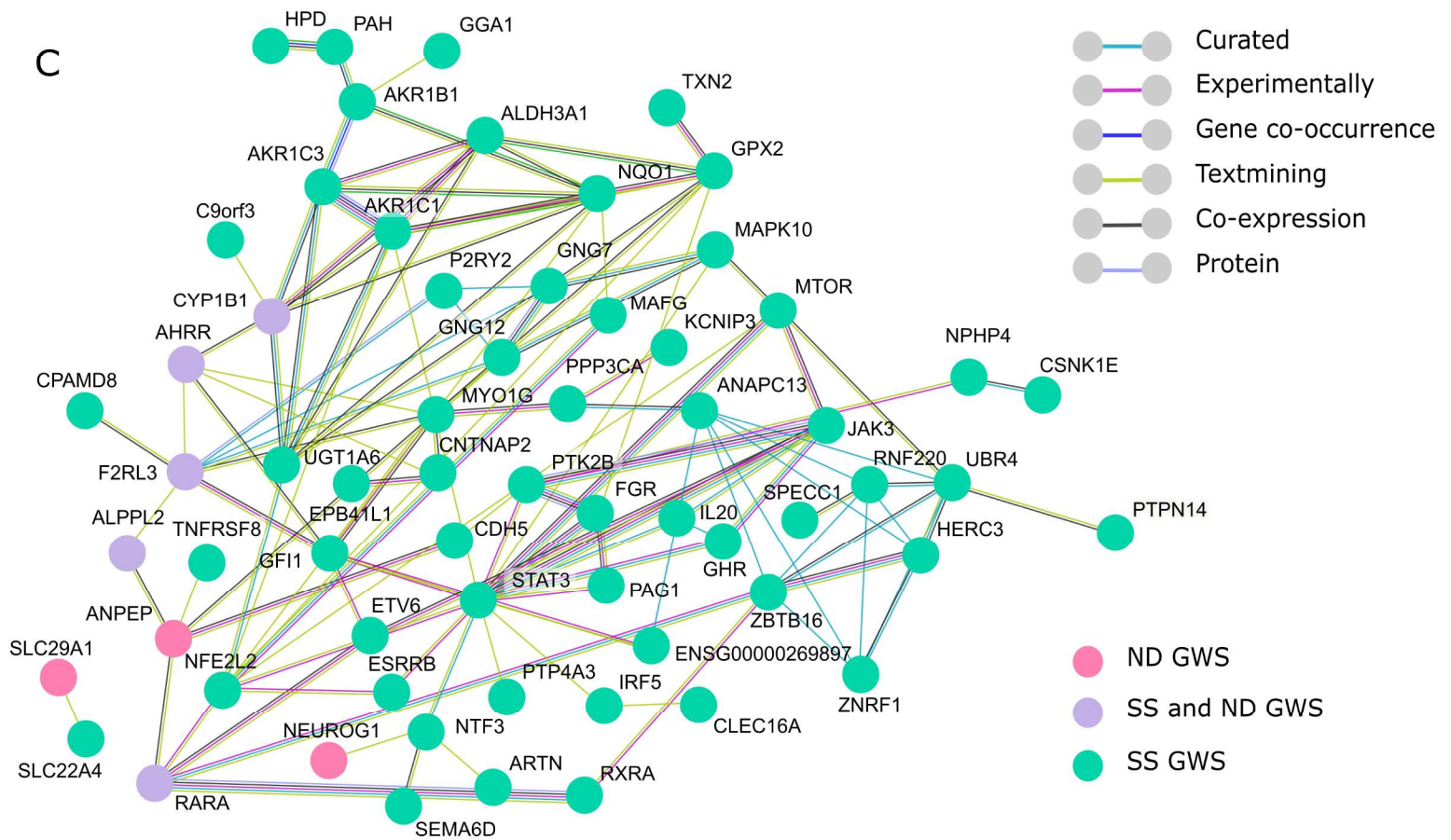


B

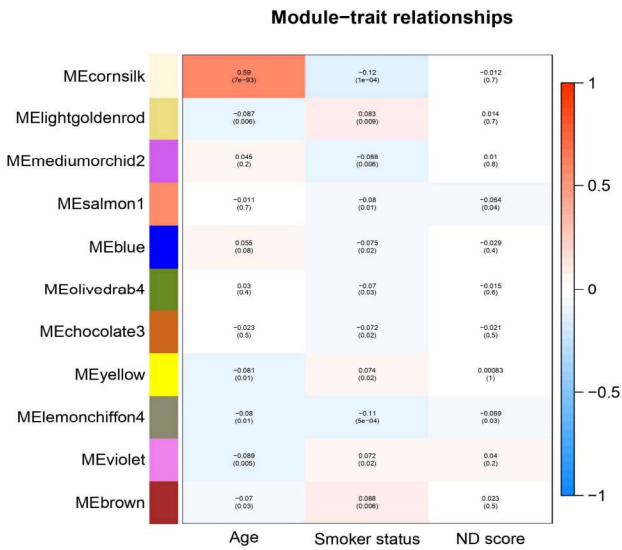
Smoking



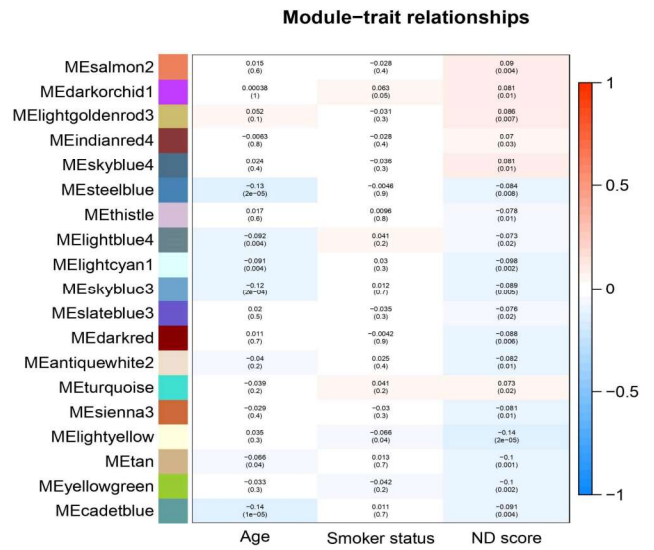
C



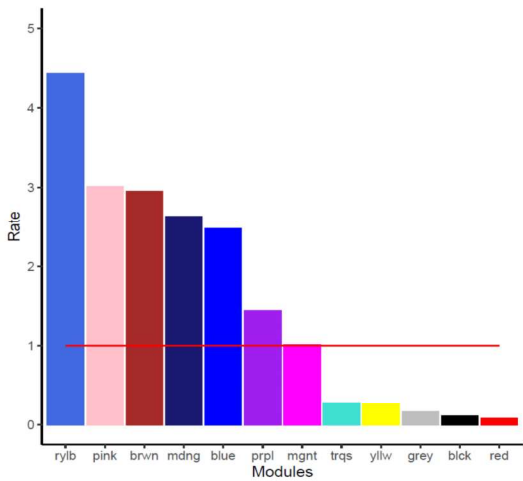
A



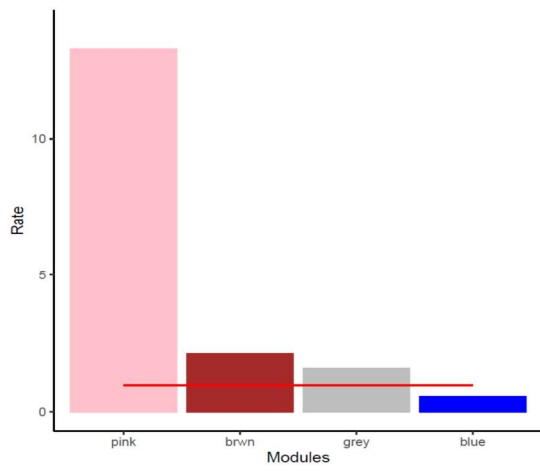
B

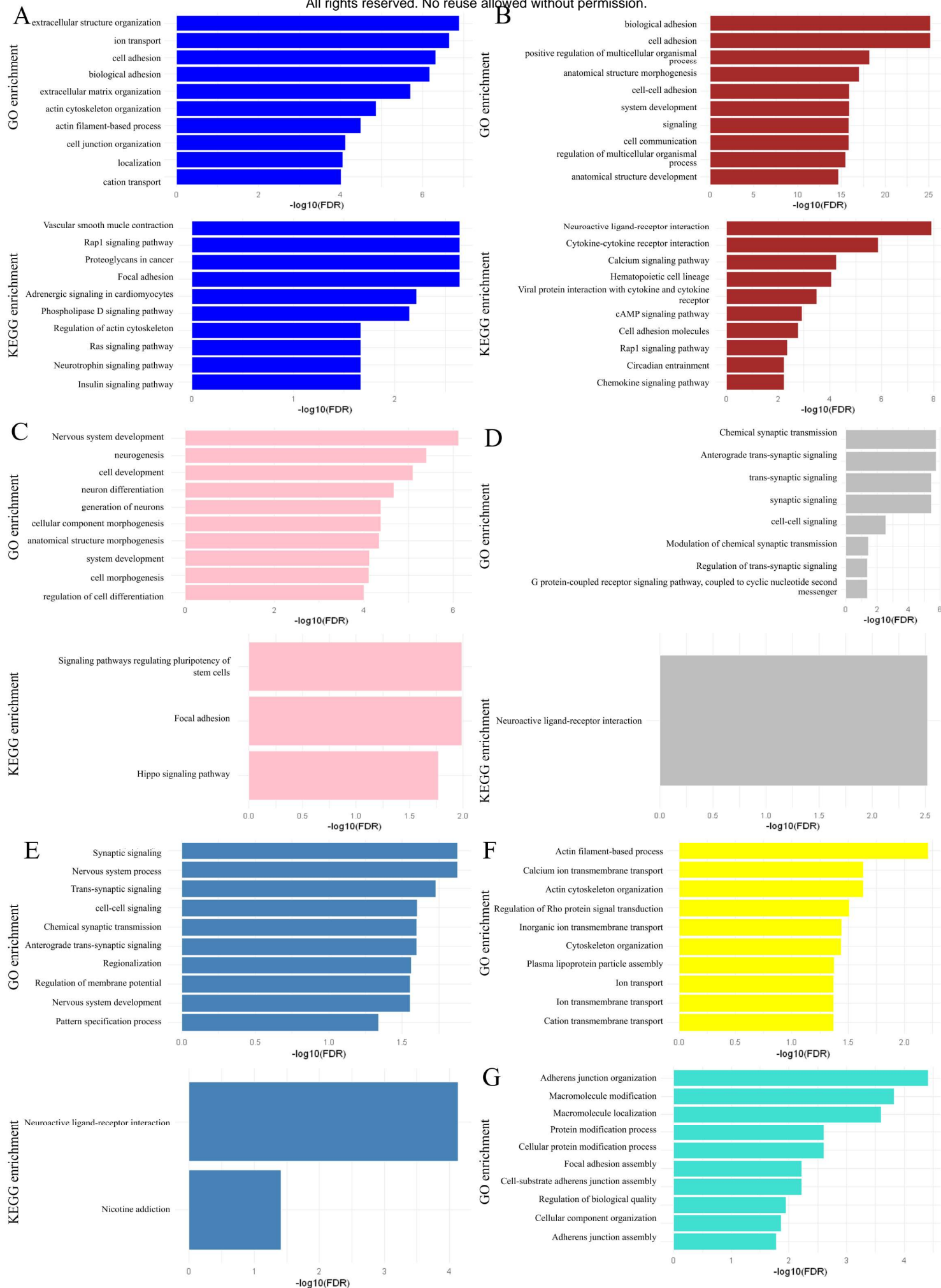


C

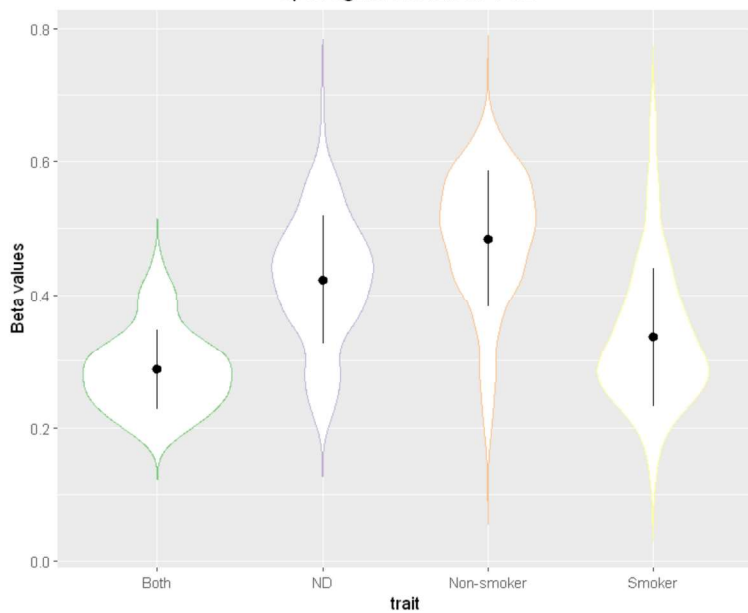


D

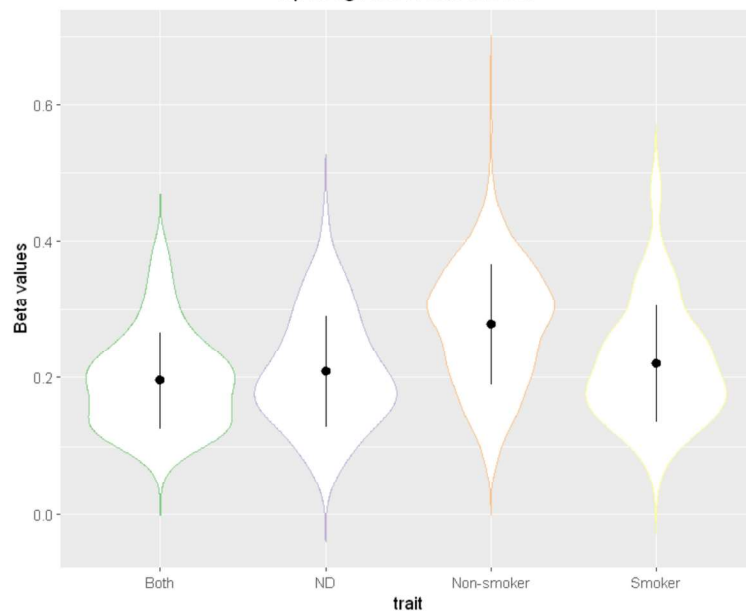




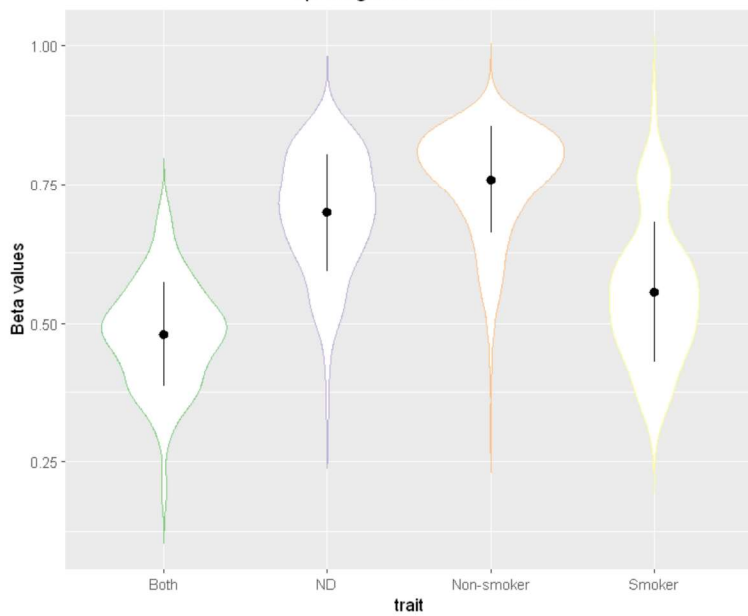
CpG: cg21566642/ALPPL2



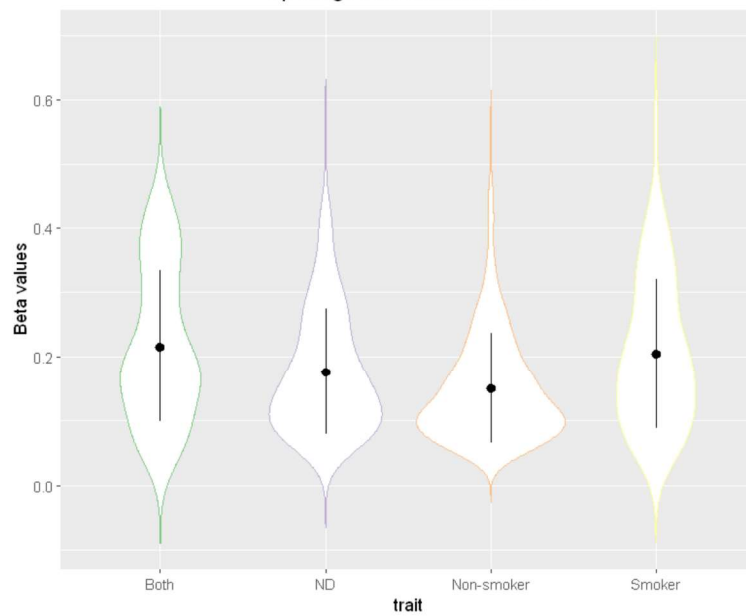
CpG: cg14391737/PRSS23



CpG: cg05575921/AHRR



CpG: cg04330449/NEUROG1



Fagerstrom score distribution

