

Type of the Paper (Article, Review, Communication, etc.)

Meta-analysis of transcriptome-wide association studies across 13 brain tissues identified novel clusters of genes associated with nicotine addiction

Zhenyao Ye ^{1,2,†}, Chen Mo ^{1,2,†}, Hongjie Ke ^{3,†}, Qi Yan ⁴, Chixiang Chen ², Peter Kochunov ¹, L.Elliot Hong ¹, Braxton D. Mitchell ⁵, Shuo Chen ^{1,2,*}, and Tianzhou Ma ^{3,*}

¹ Maryland Psychiatric Research Center, Department of Psychiatry, School of Medicine, University of Maryland, Baltimore, Maryland, United States of America;

² Division of Biostatistics and Bioinformatics, Department of Epidemiology and Public Health, School of Medicine, University of Maryland, Baltimore, Maryland, United States of America;

³ Department of Epidemiology and Biostatistics, School of Public Health, University of Maryland, College Park, Maryland, United States of America;

⁴ Department of Obstetrics & Gynecology, Irving Medical Center, Columbia University, New York, NY, United States of America;

⁵ Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland, United States of America;

* Correspondence: SC: shuochen@som.umaryland.edu; TM: tma0929@umd.edu

† Equal contribution

Abstract: Genome-wide association studies (GWAS) have identified and reproduced thousands of diseases associated loci but many of them are not directly interpretable due to the strong linkage disequilibrium among variants. Transcriptome-wide association studies (TWAS) incorporated expression quantitative trait loci (eQTL) cohorts as reference panel to detect associations with the phenotype at the gene level and were gaining popularity in recent years. For nicotine addiction, several important susceptible genetic variants were identified by GWAS, but TWAS that detected genes associated with nicotine addiction and unveiled the underlying molecular mechanism were still lacking. In this study, we used eQTL data from the Genotype-Tissue Expression (GTEx) consortium as reference panel to conduct tissue specific TWAS on cigarettes per day (CPD) over 13 brain tissues in two large cohorts: UK Biobank (UKBB; N=142,202) and the GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN; N=143,210), and then meta-analyzed the results across tissues while considering the heterogeneity across tissues. We identified three major clusters of genes with different meta-patterns across tissues consistent in both cohorts, including homogenous genes associated with CPD in all brain tissues, partially homogeneous genes associated with CPD in cortex, cerebellum and hippocampus tissues, and lastly the tissue-specific genes associated with CPD in only few specific brain tissues. Downstream enrichment analyses on each gene cluster identified unique biological pathways associated with CPD and provided important biological insights into the regulatory mechanism of nicotine dependence in the brain.

Keywords: genome-wide association study; transcriptome-wide association study; meta-analysis; expression quantitative trait loci; nicotine addiction.

1. Introduction

The past decade has witnessed the explosion in Genome-wide association studies (GWAS) research, that identified thousands of robust reproducible genetic risk variants associated with complex diseases and traits [1, 2]. These findings have contributed to a better understanding of disease biology and the relative roles of genes vs. environment in disease risk [3, 4]. However, the loci identified by GWAS are not directly interpretable due

to the strong linkage disequilibrium (LD) that obscure the causal variants and GWAS data alone can hardly determine the causal genes and the underlying regulatory mechanism [5]. To fill this gap, transcriptome-wide association studies (TWAS) are developed to utilize expression quantitative trait loci (eQTL) cohorts (e.g. Genotype-Tissue Expression (GTEx) [6]) which include both genotype and gene expression data as reference panel to infer association with a trait at the gene level [7]. In short, TWAS involve training a predictive model of expression from genotype in the reference panel, then using the trained model to predict the expression in the GWAS data, which are used to find the genes associated with the trait [7, 8]. Various statistical methods and computational tools for implementing TWAS have been developed to date [9, 10]. Since gene expression and eQTL regulation are tissue-dependent, TWAS are usually conducted in tissue-specific manner. For example, PrediXcan [8] is the first ever TWAS tool that leverages the SNP-gene associations identified in a single tissue to infer gene-trait associations. S-PrediXcan [11] is an extension of PrediXcan that takes GWAS summary statistics as the input. Considering the similarity in transcription regulation across tissues, MultiXcan [12] and UTMOST [13] fit models to integrate the information of SNP-gene associations across multiple tissues to infer the gene-trait associations.

To date, many genetic researches have revealed an important role of genetic factors on nicotine dependence [14, 15]. For example, GWAS have identified susceptible genetic variants located in nicotinic acetylcholine receptors (nAChRs) [16], metabolic enzyme encoded gene *CYP2A6* [17] and lung-specific genes *TENM2* [18] associated with nicotine addiction. But how these genetic compositions contribute to human nicotine dependence behaviors and the underlying molecular regulatory mechanism in the brain remained largely unknown. Palmer et al., [19] conducted a cross-species TWAS analysis of tobacco consumption through integrating human GWAS data from UK Biobank and mRNA expression references from brains of multiple animal species, and identified 10 homologous genes associated with cigarette per day in different animal models to illustrate the genetic mechanisms of human tobacco consumption. However, the heterogeneity among eQTL datasets and tissues-dependent nature of transcription regulation have impeded the ability of TWAS to provide further insights into the genetic basis of diseases [20].

Meta-analysis is a set of powerful statistical tools that combines multiple related studies for various biological purposes and has gained popularity in both GWAS and omics research in recent years [21, 22]. Traditional meta-analysis methods such as Fisher's and Stouffer's methods combine p-values from multiple studies without further exploring the association patterns across studies [23, 24]. New meta-analysis methods have been proposed to account for the heterogeneity across studies and categorize biomarkers (e.g., genes) by their cross-study patterns while combining the studies [25-27]. In this study, we performed tissue-specific TWAS of nicotine addiction (measured by cigarettes per day (CPD)) for 13 brain tissues based on the GWAS data from UK Biobank (UKBB) [28] and GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN) [29], using eQTL cohorts from GTEx (version 8) [6] as the reference panel. We then conducted meta-analysis of the TWAS results while considering the heterogeneity across tissues and clustered the nicotine addiction associated genes by their cross-tissue patterns. Such a procedure was shown to be more powerful than multi-tissue TWAS tool (e.g., S-MultiXcan) and detected novel clusters of genes with different meta-patterns across brain tissues. Downstream enrichment analysis on the different clusters of genes identified important nicotine addiction related pathways in different brain tissues and provided more insights into the molecular regulatory mechanisms underlying nicotine dependence inside the brain.

2. Materials and Methods

2.1 Study cohorts

In this study, we performed TWAS analysis on two large cohorts that include both genotype and nicotine addiction phenotype data:

1) UK Biobank (UKBB): a large prospective study that recruited ~500,000 participants aged between 40-69 years in 2006-2010 in 22 assessment centers throughout the UK and collected abundant phenotypic and genomic data [28]. We focused on N=142,202 individuals with white ethnicity backgrounds (British, Irish, and any other white background) that had both genotype and nicotine dependence related smoking phenotypic data available.

2) GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN): a meta-analysis of up to 35 GWAS cohorts of European ancestry including around 1.2 million individuals (depending on traits) [29]. Smoking related phenotypes in the GSCAN were self-reported responses gathered by multiple teleconferences [29]. We used the GSCAN data excluding UKBB and 23andMe ("Minus23andMeUKBB" [refer to their website]; N=143,210) as a validation cohort. For GSCAN, only meta-analyzed GWAS summary data were available.

2.2 Nicotine dependence related smoking phenotype

Cigarettes per day (CPD) is one of the well-known traits related to nicotine addiction and widely used in many published studies [30, 31]. We used CPD as the phenotype of interest in our study. For UKBB cohort, CPD was defined as the average number of cigarettes smoked per day by participants who were either current or past smokers using phenotype codes 2887 (number of cigarettes previously smoked daily), 3456 (number of cigarettes currently smoked daily), and 6183 (number of cigarettes previously smoked daily (current cigar/pipe smokers)). The CPD values of participants who smoked less than one cigarette per day were recoded to 0; and CPD values of those who smoked more than 60 cigarettes per day were recoded to 60. CPD was denoted as CigDay in GSCAN cohort. The detailed data processing procedure of the CigDay can be found in Liu et al., [29].

2.3 Reference panel

The reference panel of eQTL cohort used to perform TWAS analysis was obtained from the Genotype-Tissue Expression (GTEx) project (version 8) [6]. It included both genotype data of 838 donors of mainly European ancestry and gene expression data of these donors in 13 brain tissues including amygdala, anterior cingulate cortex (BA24), caudate (basal ganglia), cerebellum, cerebellar hemisphere, cortex, frontal cortex (BA9), hippocampus, hypothalamus, nucleus accumbens (basal ganglia), putamen (basal ganglia), spinal cord (cervical c-1) and substantia nigra. The single-tissue predicted weights and single-/across-tissue LD reference files from GTEx used in S-PrediXcan [11] and S-MultiXcan [12] were provided by PredictDB (<https://hakyimlab.org/post/2020/01/07/predictdb-transcriptome-prediction-model-repository/>) [32, 33] for use in our study.

2.4 TWAS analysis

In this study, we first conducted tissue-specific TWAS (TS-TWAS) of CPD for each of the 13 brain tissues by combining GWAS data with the eQTL reference panel and then performed meta-analysis to combine the TS-TWAS results and categorize the CPD associated genes by their meta-patterns across the tissues (Figure 1). Below, we described the two steps of our analysis in details.

2.4.1 Tissue-specific TWAS

In the first step, we conducted TS-TWAS for each of the 13 brain tissues using S-PrediXcan [11]. For UKBB, we first performed GWAS on CPD of 142,202 participants (Mean Age = 57.57 (7.83); 48.12% are Female) using PLINK (version 1.9, www.cog-genomics.org/plink/1.9/) [34] under an additive genetic model. We performed quality control and removed variants with minor allele frequency below 0.01, Hardy-Weinberg equilibrium P-value below 0.001 and missing genotype rate at 5%, and to excluded individuals with more than 2% missing genotypes. The analysis was adjusted by the following variables: sex, age, body mass index (BMI), genotyping chip type and top ten principal components of population admixture generated from PLINK (version 2.0, www.cog-genomics.org/plink/2.0/) [34]. For GSCAN, GWAS summary on CPD was directly obtained from the University of Minnesota library [29]. We integrated GWAS summary statistics of

both cohorts with the pre-trained prediction models over 13 brain tissues to obtain TS-TWAS results using S-PrediXcan [11], an extension of PrediXcan [8] that used only summary level GWAS statistics to estimate Z-score of association between gene expression and trait. The tissue-specific Z-score for the g -th gene in s -th tissue can be estimated as follows:

$$z_{gs} \approx \sum_{l \in Model_{gs}} w_{lgs} \frac{\hat{\sigma}_l}{\hat{\sigma}_{gs}} \frac{\hat{\beta}_l}{se(\hat{\beta}_l)}$$

where $Model_{gs}$ is the pre-trained prediction model from GTEx reference panel consisting of SNPs used to predict the gene expression for g -th gene in s -th tissue, w_{lgs} is the predicted weight of l -th SNP on the g -th gene in s -th tissue in the pre-trained prediction model, directly obtained from PredictDB. $\hat{\beta}_l$ is the GWAS estimate for l -th SNP; $se(\hat{\beta}_l)$ is GWAS standard error of $\hat{\beta}_l$; $\hat{\sigma}_l$ is the variance of l -th SNP and $\hat{\sigma}_{gs}$ is the variance of the predicted expression for g -th gene in s -th tissue. The SNP variance term $(\frac{\hat{\sigma}_l}{\hat{\sigma}_{gs}})$ calculated from 1000 Genomes data were also obtained from PredictDB. We computed the p-value for the g -th gene in s -th tissue as $p_{gs} = 2(1 - \Phi(z_{gs}))$, where $\Phi(\cdot)$ is the cumulative density function of standard normal distribution.

2.4.2 Meta-analysis of TS-TWAS over 13 brain tissues and downstream analysis

AW-Fisher' method [25] is a meta-analysis method extending the conventional Fisher's method that combines p-values from multiple studies while taking the study to study heterogeneity into account. In this paper, we treated different brain tissues as studies and applied AW-Fisher's method to meta-analyze the TS-TWAS results from all $S=13$ brain tissues. The null hypothesis in meta-analysis is commonly considered as

$$H_0: \theta_{g1} = \dots = \theta_{gs} = 0,$$

where θ_{gs} is the gene effect of g -th gene in the s -th tissue. For alternative hypothesis, we aimed to detect genes associated with CPD in at least one tissue, i.e., $H_a: \theta_{gs} \neq 0$ for some $1 \leq s \leq S$. For AW-Fisher's method,

$$U_g(\omega_g) = - \sum_{s=1}^S \omega_{gs} \log(p_{gs}),$$

where p_{gs} is the p-value of the g -th gene in the s -th tissue from TS-TWAS, ω_{gs} is the 0-1 binary weight assigned to the s -th tissue and $\omega_g = (\omega_{g1}, \dots, \omega_{gs})$. For a specific ω_g , the p-value of the observed weighted statistic $p_U(u_g(\omega_g))$ under the null hypothesis can be obtained via permutation. The AW-Fisher's statistic was defined as the minimal p-value among all possible weights. For inference, there is no closed-form distribution for AW-Fisher's statistics under the null, so permutation tests and importance sampling are used to obtain the p-values p_g^{AW} and control the false discovery rate (FDR). More details can be found in the original AW-Fisher paper [25].

After meta-analysis, we focused on genes passing FDR threshold of 0.05 in both cohorts (i.e., take the intersection) and performed downstream analysis. We categorized the genes by their meta-patterns across brain tissues using hierarchical clustering with Ward linkage on $-\log_{10}(p_{gs})$. For each category of genes, we further performed pathway enrichment analysis using three popular pathways datasets: Gene Ontology (GO) [35], Kyoto Encyclopedia of Genes and Genomes (KEGG) [36] and Reactome [37]. Top enriched pathways (e.g., Fisher's exact test p-value < 0.05) helped us understand the unique functions for each category of genes associated with nicotine addiction but with different cross-tissue patterns. We also applied the S-MultiXcan method [12] for multi-tissue TWAS analysis across 13 brain tissues as a comparison.

3. Results

We first performed TS-TWAS and then meta-analyzed the TS-TWAS results over 13 brain tissues by AW-Fisher’s method. The meta-analysis of TS-TWAS identified 48 genes significantly associated with CPD at FDR < 0.05 common in both UKBB and GSCAN cohorts (Table 1). Comparing to S-MultiXcan and TS-TWAS, meta-analysis was overall more powerful in identifying more nicotine addiction associated genes (Table 1; Figure 2 highlighted in red, Figure S1), especially among genes with heterogeneous association patterns across tissues (Figure S1, Supplementary file 1). These included multiple nicotine addiction associated genes reported in previous studies [17, 18, 31]. We focused on the 48 genes at FDR < 0.05 for biomarker categorization and downstream analysis.

Gene categorization by meta-patterns identified three clusters of genes common to both cohorts (Figure 3): (i) homogeneous genes, which were associated with CPD in all 13 brain tissues; (ii) partially homogeneous genes, which were associated with CPD in a majority of tissues but not significant in the rest; (iii) tissue-specific or heterogeneous genes, which had very unique association patterns in different tissues, reflecting a high degree of heterogeneity across tissues. The two cohorts had a large proportion of genes matched in each cluster (38 out of 48 genes in total; Table 1). The first cluster included 20 genes homogeneously associated with CPD in all brain tissues, including well-known smoking-related genes *CHRNA5*, *NCKIPSD* and *SIRT6*. The second cluster consisted of 8 genes including *PSMA4* and *RPRD2*, which are highly expressed and associated with CPD in fore-brain regions such as frontal cortex BA9 and anterior cingulate BA24 and hindbrain regions including cerebellum and cerebellar hemisphere. The third cluster included 10 heterogeneous genes only associated with CPD in very few specific brain tissues.

We then performed pathway analysis to each cluster of identified genes. Cluster 1 of homogeneous genes (e.g. *CHRNA5*) were mainly enriched in pathways related to presynaptic and postsynaptic nicotinic acetylcholine receptors (Figure 4, see the full pathway analysis results in Supplementary file 2), that play versatile roles in neuronal apoptosis [38] and neurotransmission (e.g., Ca^{2+} signaling [39] and dopamine [40]). Cluster 2 of partially homogeneous genes (e.g. *PSMA4*) were enriched in pathways related to proteasomal activity (e.g., KEGG Proteasome), intercellular bivalent cations Mg^{2+} (e.g., GO:MF magnesium ion binding) and chromosome segregation (e.g., GO:BP chromosome segregation) that can be highly impacted by cigarette smoking to inhibit proteasomal activity, cause mental disorders disease and induce segregation anomalies separately reported in the previous studies [41–43]. Cluster 3 of heterogeneous genes (e.g., *CHRNA3*) were enriched in pathways GO:MF acetylcholine-activated cation-selective channel activity and GO:CC acetylcholine-gated channel complex. Genes with different meta-patterns across tissues were functionally specific that are worth further investigation in future studies.

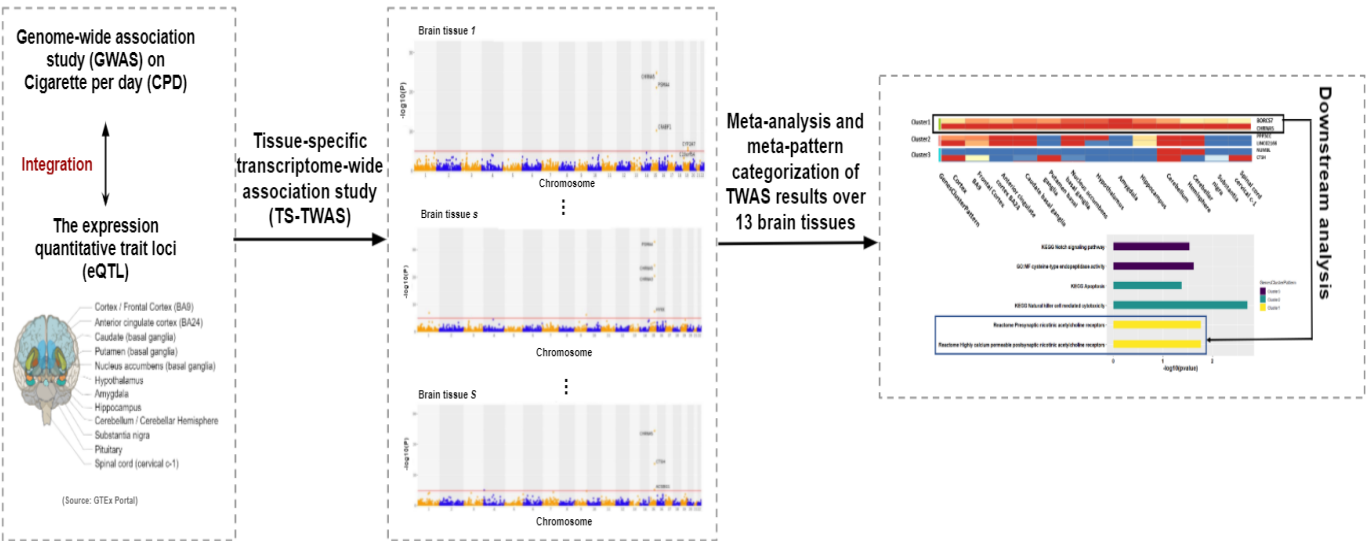


Figure 1. Study scheme. We integrated GWAS summary statistics with eQTL reference panel from GTEx to conduct tissue-specific TWAS analysis (TS-TWAS) for each of the 13 brain tissues using S-PrediXcan. We then performed meta-analysis of the TS-TWAS results across tissues using AW-Fisher’s method and clustered the genes by their meta-patterns across tissues. We additionally performed downstream analysis (e.g., pathway enrichment analysis) to each category of genes with unique meta-pattern.

Table 1. Summary of number of CPD-associated genes detected by meta-analysis and S-MultiXcan and in each category of unique meta-pattern in both UKBB and GSCAN cohorts and their intersection.

Cohort		UKBB	GSCAN	Intersection
S-MultiXcan (FDR < 0.05)		60	13	8
Meta-analysis by AW-Fisher's method (FDR < 0.05)		245	217	48
Meta-pattern categorization (48 genes in intersection at FDR<0.05)	Cluster 1 (homogeneous genes)	24	22	20
	Cluster 2 (partially homogeneous genes)	12	12	8
	Cluster 3 (tissue-specific or heterogeneous genes)	12	14	10

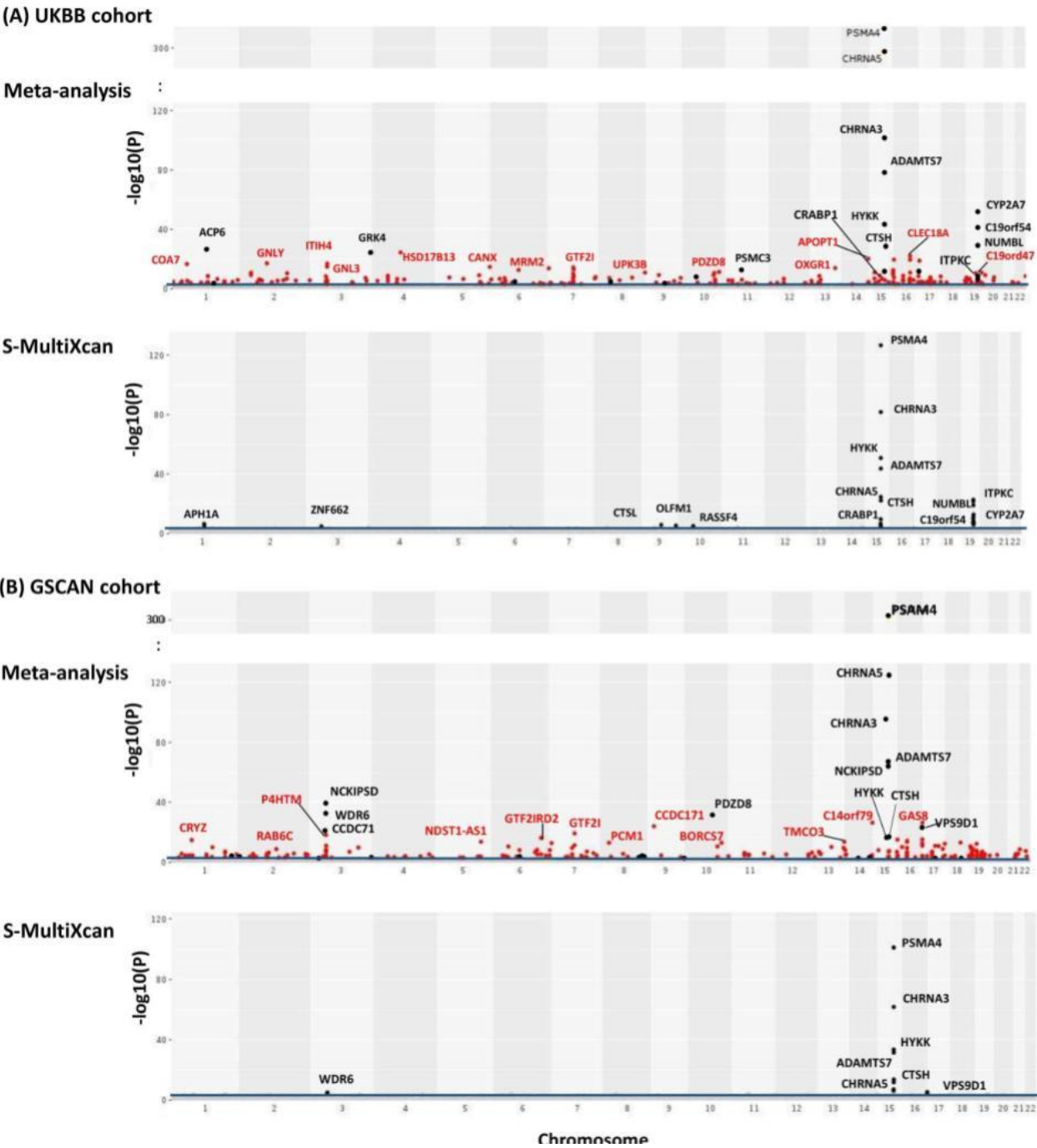


Figure 2. Manhattan plots of meta-analysis of TS-TWAS results across all 13 brain tissues for both UKBB (A) and GSCAN (B). Y-axis is the $-\log_{10}(p_g^{AW})$ from AW-Fisher. Results from S-MultiXcan are used for comparison. The blue line indicates an FDR cutoff of 0.05. Genes passing Bonferroni cutoff (i.e., $p < 0.05/\#genes$) were labeled and genes detected by meta-analysis but not by S-MultiXcan were red highlighted.

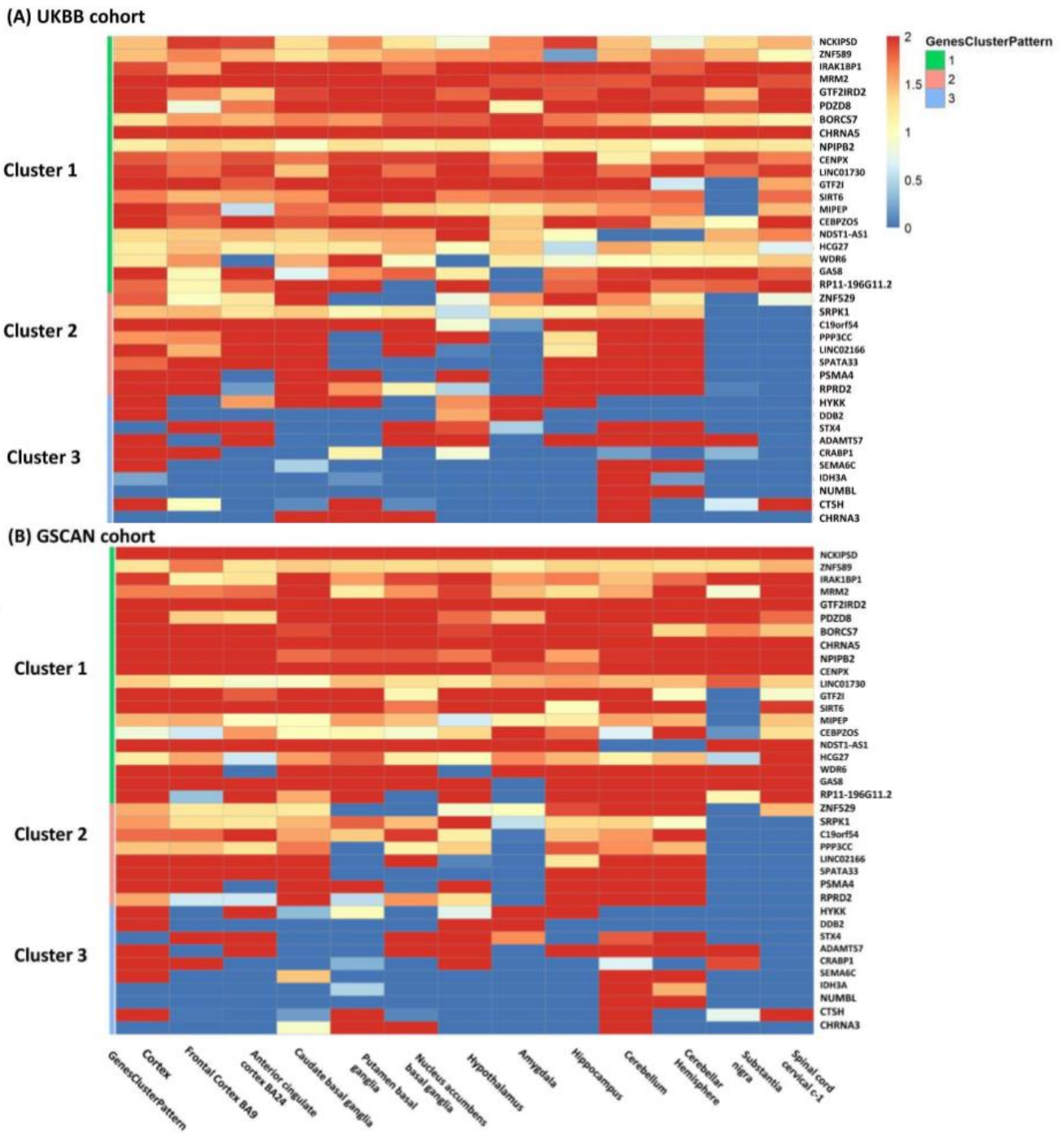


Figure 3. The heatmap included the 38 genes (Cluster 1: 20; Cluster 2: 8; Cluster 3: 10) with the same clustering patterns and passing meta-analysis $FDR < 0.05$ threshold in both cohorts and was colored by $-\log_{10}(p_{gs})$ of TS-TWAS in each brain tissue (on columns) from both cohorts (Panel (A) for UKBB and Panel (B) for GSCAN). On the rows, the genes were clustered into three categories common to two cohorts: cluster 1 was homogeneous genes, cluster 2 was partially homogeneous genes and cluster 3 was tissue-specific or heterogeneous genes.

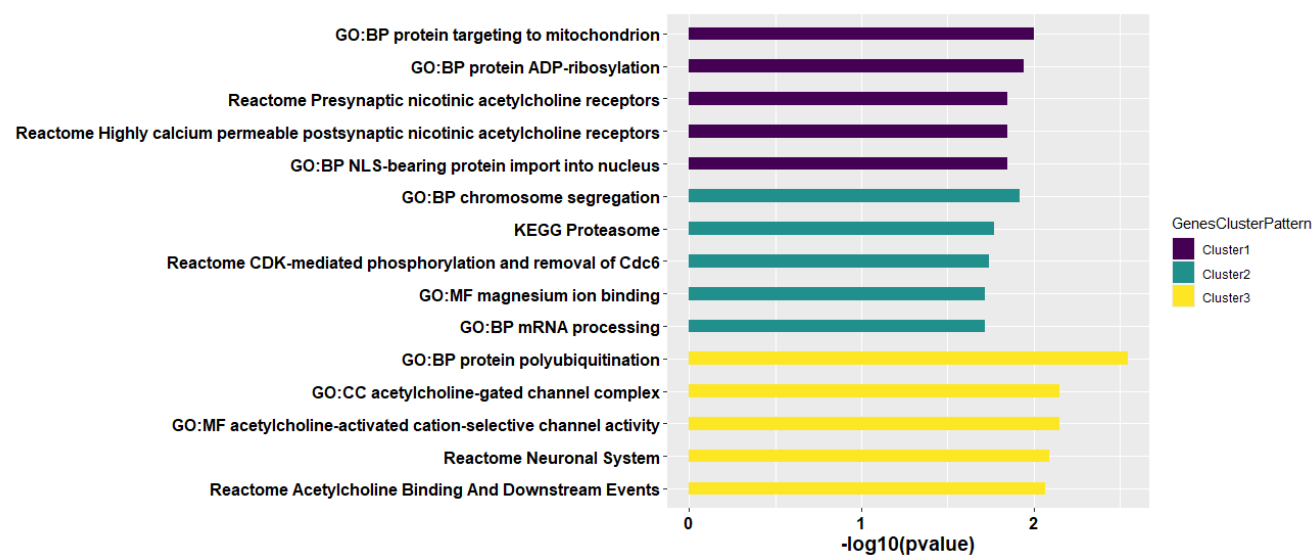


Figure 4. Top five pathways enriched by each cluster of genes identified sorted by both p-value and pathway size. The p-value is from the Fisher’s exact test.

4. Discussion

In this study, we used eQTL reference panel from GTEx to conduct meta-analysis of TS-TWAS on nicotine addiction over 13 brain tissues in two large cohorts, UKBB and GSCAN. The meta-analysis was shown more powerful than multi-tissue TWAS method implemented in S-MultiXcan by detecting more nicotine addiction associated genes while accounting for the heterogeneity across multiple brain tissues. In addition to detecting more associated genes, gene categorization by meta-patterns identified three novel clusters of genes common to both cohorts, including 20 genes homogeneously associated in all brain tissues, 8 genes partially homogeneously associated mainly in cortex and cerebrum, and 10 genes with tissue-specific association. Several well-known nicotine addictions associated genes including *CHRNA5*, *PSMA4* and *CHRNA3*, were identified and their cross brain tissue association patterns were revealed. To the best of our knowledge, our study was the first comprehensive meta-analysis of TWAS on nicotine addiction across 13 major brain tissues, investigated and validated in two large-scale epidemiological cohorts UKBB and GSCAN.

The first cluster of genes were enriched in pathways related to presynaptic and postsynaptic nicotinic acetylcholine receptors, as marked by the gene *CHRNA5*. Previous GWAS have identified multiple reproducible variants in the *CHRNA5* [31, 44], which were attributed to functions in both enhancement and aversion of nicotine intake [45]. We further showed in our study *CHRNA5* was highly expressed in all brain tissues and its association with CPD was consistent throughout the brain. The second cluster was marked by the proteasome gene *PSMA4*, which was identified the association with CPD mainly in the frontal lobe (e.g., frontal cortex BA9 and anterior cingulate BA24), cerebellum/cerebellar hemisphere and hippocampus tissues. The enriched biological pathways were driven by *PSMA4*, playing a central role in decreasing neuronal proteasome activity [46]. Another nicotinic acetylcholine receptor gene *CHRNA3* was observed in the third cluster with tissue-specific association with CPD only in putamen basal ganglia and nucleus accumbens basal ganglia. The elevated dopamine activity in the basal ganglia region of cigarette smoker have been identified for the enriched pathways driven by *CHRNA3* in the previous studies [47, 48]. These findings showed the strength of our comprehensive meta-analysis of TWAS on CPD that identified novel clusters of genes with unique meta-patterns across tissues, inferring different biological function.

TWAS are getting popular over recent years as a promising complement to GWAS by incorporating the functional annotation information and analyzing association with the trait at the gene level. Despite the foreseen success, most TWAS methods to date are

tissue specific and ignore the similarity in transcription regulation across tissues, and usually have limited effective sample size thus underpowered [13]. Our study performed the meta-analysis of TS-TWAS across 13 brain tissues on nicotine addiction and categorized the identified genes by their meta-patterns across tissues. Such a meta-analytical framework can also be widely applicable to analyze the other traits through targeting at other tissues and eQTL reference panels (e.g., blood [49] and lung [50]). One of the main challenges of TWAS approach is hard to prioritize causal genes due to co-regulation [7]. A Mendelian randomization framework incorporated into the TWAS for identification of putative causal inference need to be conducted to carve out this issue [10, 51, 52]. Further studies such as the application of fine-mapping methods (e.g., FOCUS [53]) will be needed to confirm our meta-TWAS results and to distinguish the causal genes for nicotine addiction, which can improve our understanding of the genetic basis of brain related disorders.

Supplementary Materials: Figure S1: Manhattan plots of the tissue-specific (TS-TWAS) results from UKBB cohort. Y-axis is the $-\log_{10}(p_{gs})$ from S-PrediXcan. An FDR cutoff of 0.05 was shown as a blue line. Genes passing this FDR cutoff were labeled in black color. Supplementary file 1: A comparison of brain tissue-specific genes between the associations with nicotine dependence (e.g., CPD) from both UKBB cohort and GSCAN cohort. Supplementary file 2: Full output of pathway enrichment analysis for each our detected meta-pattern category.

Author Contributions: ZY, CM and HK performed the analysis and wrote the manuscript. TM and SC supervised the project and took the lead in editing the manuscript. QY, CC, PK, LEH and BDM contributed to manuscript writing and polishing. All authors provided critical feedback and helped to shape the research, analysis, and manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Institute on Drug Abuse (NIDA) of National Institute of Health under the award number 1DP1DA048968-01 to Z.Y., M.C., S.C. and T.M., by the University of Maryland MPower Brain Health and Human Performance seed grant to P.K., S.C. and T.M.

Data Availability Statement: The raw genetic and phenotypic data used in the current study are available from the UK Biobank (UKB), which can be accessed via <https://www.ukbiobank.ac.uk/>.

Acknowledgments: The authors would like to thank Dr. Mengzhen Liu for providing us meta-analyzed GWAS summary statistics used for the analyses described in this manuscript, which were obtained from (<https://conservancy.umn.edu/handle/11299/201564>). The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The picture included in Figure 1 in this manuscript were obtained from <https://gtexportal.org/home/samplingSitePage> The GTEx Portal on 11/16/2021.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mills, M.C. and C. Rahal, *A scientometric review of genome-wide association studies*. Communications biology, 2019. **2**(1): p. 1-11.
2. Tam, V., et al., *Benefits and limitations of genome-wide association studies*. Nature Reviews Genetics, 2019. **20**(8): p. 467-484.
3. Visscher, P.M., et al., *10 years of GWAS discovery: biology, function, and translation*. The American Journal of Human Genetics, 2017. **101**(1): p. 5-22.
4. Visscher, P.M., et al., *Five years of GWAS discovery*. The American Journal of Human Genetics, 2012. **90**(1): p. 7-24.
5. Altshuler, D., M.J. Daly, and E.S. Lander, *Genetic mapping in human disease*. science, 2008. **322**(5903): p. 881-888.
6. Lonsdale, J., et al., *The genotype-tissue expression (GTEx) project*. Nature genetics, 2013. **45**(6): p. 580-585.
7. Wainberg, M., et al., *Opportunities and challenges for transcriptome-wide association studies*. Nature genetics, 2019. **51**(4): p. 592-599.
8. Gamazon, E.R., et al., *A gene-based association method for mapping traits using reference transcriptome data*. Nature genetics, 2015. **47**(9): p. 1091-1098.

9. Xie, Y., et al., *Transcriptome wide association studies: general framework and methods*. Quantitative Biology: p. 0.
10. Zhu, H. and X. Zhou, *Transcriptome-wide association studies: a view from Mendelian randomization*. Quantitative Biology, 2020: p. 1-15.
11. Barbeira, A.N., et al., *Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics*. Nature communications, 2018. **9**(1): p. 1-20.
12. Barbeira, A.N., et al., *Integrating predicted transcriptome from multiple tissues improves association detection*. PLoS genetics, 2019. **15**(1): p. e1007889.
13. Hu, Y., et al., *A statistical framework for cross-tissue transcriptome-wide association analysis*. Nature genetics, 2019. **51**(3): p. 568-576.
14. Kendler, K.S., et al., *A population-based twin study in women of smoking initiation and nicotine dependence*. Psychological medicine, 1999. **29**(2): p. 299-308.
15. Vink, J.M., G. Willemsen, and D.I. Boomsma, *Heritability of smoking initiation and nicotine dependence*. Behavior genetics, 2005. **35**(4): p. 397-406.
16. Lassi, G., et al., *The CHRNA5–A3–B4 gene cluster and smoking: from discovery to therapeutics*. Trends in neurosciences, 2016. **39**(12): p. 851-861.
17. Buchwald, J., et al., *Genome-wide association meta-analysis of nicotine metabolism and cigarette consumption measures in smokers of European descent*. Molecular psychiatry, 2021. **26**(6): p. 2212-2223.
18. Quach, B.C., et al., *Expanding the genetic architecture of nicotine dependence and its shared genetics with multiple traits*. Nature communications, 2020. **11**(1): p. 1-13.
19. Palmer, R.H., et al., *Multi-omic and multi-species meta-analyses of nicotine consumption*. Translational psychiatry, 2021. **11**(1): p. 1-10.
20. Nariai, N., et al., *Efficient prioritization of multiple causal eQTL variants via sparse polygenic modeling*. Genetics, 2017. **207**(4): p. 1301-1312.
21. Tseng, G.C., D. Ghosh, and E. Feingold, *Comprehensive literature review and statistical considerations for microarray meta-analysis*. Nucleic acids research, 2012. **40**(9): p. 3785-3799.
22. Begum, F., et al., *Comprehensive literature review and statistical considerations for GWAS meta-analysis*. Nucleic acids research, 2012. **40**(9): p. 3777-3784.
23. Fisher, R.A., *Statistical methods for research workers*, in *Breakthroughs in statistics*. 1992, Springer. p. 66-70.
24. Stouffer, S.A., et al., *The american soldier: Adjustment during army life.(studies in social psychology in world war ii), vol. 1*. 1949.
25. Li, J. and G.C. Tseng, *An adaptively weighted statistic for detecting differential gene expression when combining multiple transcriptomic studies*. The Annals of Applied Statistics, 2011. **5**(2A): p. 994-1019.
26. Ma, T., F. Liang, and G. Tseng, *Biomarker detection and categorization in ribonucleic acid sequencing meta-analysis using bayesian hierarchical models*. Journal of the Royal Statistical Society. Series C, Applied statistics, 2017. **66**(4): p. 847.
27. Ye, Z., et al., *Biomarker categorization in transcriptomic meta-analysis by concordant patterns with application to Pan-cancer studies*. Frontiers in genetics, 2021. **12**.
28. Sudlow, C., et al., *UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age*. Plos med, 2015. **12**(3): p. e1001779.
29. Liu, M., et al., *Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use*. Nature genetics, 2019. **51**(2): p. 237-244.
30. Xu, K., et al., *Genome-wide association study of smoking trajectory and meta-analysis of smoking status in 842,000 individuals*. Nature Communications, 2020. **11**(1): p. 1-11.
31. Erzurumluoglu, A.M., et al., *Meta-analysis of up to 622,409 individuals identifies 40 novel smoking behaviour associated genetic loci*. Molecular psychiatry, 2019: p. 1-18.

32. Gamazon, E.R., et al., *A gene-based association method for mapping traits using reference transcriptome data*. Nat Genet, 2015. **47**(9): p. 1091-8.
33. Barbeira, A.N., et al., *Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics*. Nat Commun, 2018. **9**(1): p. 1825.
34. Chang, C.C., et al., *Second-generation PLINK: rising to the challenge of larger and richer datasets*. Gigascience, 2015. **4**(1): p. s13742-015-0047-8.
35. Ashburner, M., et al., *Gene ontology: tool for the unification of biology*. Nature genetics, 2000. **25**(1): p. 25-29.
36. Kanehisa, M., et al., *KEGG: new perspectives on genomes, pathways, diseases and drugs*. Nucleic acids research, 2017. **45**(D1): p. D353-D361.
37. Fabregat, A., et al., *The reactome pathway knowledgebase*. Nucleic acids research, 2016. **44**(D1): p. D481-D487.
38. Hassa, P.O., et al., *Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going?* Microbiology and Molecular Biology Reviews, 2006. **70**(3): p. 789-829.
39. Shen, J.-x. and J.L. Yakel, *Nicotinic acetylcholine receptor-mediated calcium signaling in the nervous system*. Acta Pharmacologica Sinica, 2009. **30**(6): p. 673-680.
40. Exley, R. and S. Cragg, *Presynaptic nicotinic receptors: a dynamic and diverse cholinergic filter of striatal dopamine neurotransmission*. British journal of pharmacology, 2008. **153**(S1): p. S283-S297.
41. Pereira, C.S., et al., *Smoking-induced chromosomal segregation anomalies identified by FISH analysis of sperm*. Molecular cytogenetics, 2014. **7**(1): p. 1-8.
42. Rezvani, K., et al., *Nicotine regulates multiple synaptic proteins by inhibiting proteasomal activity*. Journal of Neuroscience, 2007. **27**(39): p. 10508-10519.
43. Nechifor, M., *Magnesium and zinc involvement in tobacco addiction*. Journal of Addiction Research and Therapy. S, 2012. **2**: p. 1-5.
44. Liu, M., et al., *Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use*. Nat Genet, 2019. **51**(2): p. 237-244.
45. Wills, L. and P.J. Kenny, *Addiction - related neuroadaptations following chronic nicotine exposure*. Journal of Neurochemistry, 2021. **157**(5): p. 1652-1673.
46. Tai, H.-C., *Characterization of the brain proteasome and its interacting proteins and their regulation by neuronal activity*. 2010: California Institute of Technology.
47. Salokangas, R.K., et al., *High levels of dopamine activity in the basal ganglia of cigarette smokers*. American Journal of Psychiatry, 2000. **157**(4): p. 632-634.
48. Lee, S.-H., et al., *Combined genetic influence of the nicotinic receptor gene cluster CHR5/A3/B4 on nicotine dependence*. BMC genomics, 2018. **19**(1): p. 1-11.
49. Li, B., et al., *Tissue specificity-aware TWAS (TSA-TWAS) framework identifies novel associations with metabolic, immunologic, and virologic traits in HIV-positive adults*. PLoS genetics, 2021. **17**(4): p. e1009464.
50. Bossé, Y., et al., *Transcriptome -wide association study reveals candidate causal genes for lung cancer*. International journal of cancer, 2020. **146**(7): p. 1862-1878.
51. Xue, H., W. Pan, and A.S.D.N. Initiative, *Some statistical consideration in transcriptome -wide association studies*. Genetic epidemiology, 2020. **44**(3): p. 221-232.
52. Zhu, Z., et al., *Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets*. Nature genetics, 2016. **48**(5): p. 481-487.
53. Mancuso, N., et al., *Probabilistic fine-mapping of transcriptome-wide association studies*. Nature genetics, 2019. **51**(4): p. 675-682.