GWAS and Systems Biology Analysis of Depressive Symptoms among Smokers from the COPDGene Cohort

Jonathan T. Heinzman, BA[1], Karin F. Hoth, PhD[1], Michael H. Cho, MD[2,3], Phuwanat Sakornsakolpat, MD[2], Elizabeth A. Regan, MD[4,5], Barry J. Make, MD[4], Gregory L. Kinney, PhD[5], Frederick S. Wamboldt, MD[4,6], Kristen E. Holm, PhD[4], Nicholas Bormann, BA[1], Julian Robles, BA[1], Victor Kim, MD[7], Anand S. Iyer, MD[8,9,10], Edwin K. Silverman, MD[2,3], James D. Crapo, MD[4], Shizhong Han, PhD[1,11], James B. Potash, MD[1,11], and Gen Shinozaki, MD[1] on behalf of the COPDGene Investigators

[1]Department of Psychiatry, University of Iowa Hospitals and Clinics, Iowa City, Iowa, USA

[2]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA

[3]Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA

[4]Department of Medicine, National Jewish Health, Denver, CO, USA

[5]Department of Epidemiology, Colorado School of Public Health, University of Colorado, Aurora, CO, USA

[6]Department of Psychiatry, University of Colorado School of Medicine at the Anschutz Medical Campus, Aurora, CO, USA

[7]Department of Thoracic Medicine and Surgery, Temple University School of Medicine, Philadelphia, PA, USA

[8]Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA

Corresponding Author: Gen Shinozaki, MD, University of Iowa Hospitals and Clinics, Medical Laboratories, 25 South Grand Avenue, c/o B002FE, Iowa City, IA 52242, Phone: (319) 384-4932, Fax: (319) 353-4568, gen-shinozaki@uiowa.edu.

[9]Lung Health Center, University of Alabama at Birmingham, Birmingham, AL, USA

[10]Health Services, Outcomes, and Effectiveness Research Training Program, University of Alabama at Birmingham, Birmingham, AL, USA

[11]Department of Psychiatry, Johns Hopkins University, Baltimore, MD, USA

Abstract

Background: Large sample GWAS is needed to identify genetic factors associated with depression. This study used genome-wide genotypic and phenotypic data from the COPDGene study to identify genetic risk factors for depression.

Methods: Data were from 9,716 COPDGene subjects with 10 pack-year history. Depression was defined as antidepressant use and/or a HADS depression subscale score 8. Non-Hispanic White (6,576) and African-American (3,140) subsets were analyzed. A GWAS pipeline identified SNPs associated with depression in each group. Network analysis software analyzed gene interactions through common biological pathways, genetic interactions, and tissue-specific gene expression.

Results: The mean age was 59.4 years (SD 9.0) with 46.5% female subjects. Depression was in 24.7% of the NHW group (1,622) and 12.5% of the AA group (391). No SNPs had genome-wide significance. One of the top SNPs, rs12036147 ($p=1.28 \times 10^{-6}$), is near CHRM3. Another SNP was near MDGA2 (rs17118176, $p=3.52 \times 10^{-6}$). Top genes formed networks for synaptic transmission with statistically significant level of more co-expression in brain than other tissues, particularly in the basal ganglia ($p=1.00 \times 10^{-4}$).

Limitations: Limitations included a depression definition based on antidepressant use and a limited HADS score subgroup, which could increase false negatives in depressed patients not on antidepressants. Antidepressants used for smoking cessation in non-depressed patients could lead to false positives.

Conclusions: Systems biology analysis identified statistically significant pathways whereby multiple genes influence depression. The gene set pathway analysis and COPDGene data can help investigate depression in future studies.

Keywords

Introduction

Major depressive disorder (MDD) is the most common psychiatric disorder in the United States, with an estimated prevalence of 17% (Kessler et al., 2005). Depression is likely a heterogenous disorder with multiple synergetic effects from many genetic variants; however, a single genetic susceptibility factor with a large effect size has not been found. Though twin studies such as the one led by Kendler et al. (2006) support a genetic predisposition influencing depression, until recently, depression genome-wide association studies (GWAS)

have not shown genome-wide significance in almost all studies reported (Flint and Kendler, 2014).

The difficulty in detecting genome-wide significance was likely due to the small effect sizes of specific genetic variants and relatively small sample sizes until more recent larger meta-analyses. One of the largest single GWAS of depression from a cohort of 5,303 Han Chinese women and 5,337 controls reported two genome-wide significant loci (CONVERGE Consortium, 2015). A recent meta-analysis of 75,607 subjects of European descent and 231,747 controls identified 15 loci (Hyde et al., 2016). Most recently, the Psychiatric Genome Consortium MDD group reported 44 genome-wide significant loci using 130,664 cases and 330,570 controls (Wray and Sullivan, 2017). To detect such small effect sizes at the genome-wide significance level, it is vital to include large samples; however, collecting both genotypic and phenotypic data in a single study is still very difficult. To overcome such challenges, the present study used a rare opportunity to analyze GWAS and phenotypic data from over 9,000 participants collected through the single largest genetic study among smokers with and without COPD—the Genetic Epidemiology of COPD (COPDGene) study (Regan et al., 2010).

The current analysis focused on identifying genetic risk factors associated with depressive symptoms among smokers (10 pack-year smoking history). Of note, although the COPDGene cohort is enriched with COPD patients, not all subjects necessarily had COPD. Thus, a GWAS of the COPDGene sample in the current study investigated depressive symptoms among current and former smokers with and without COPD.

Depression is highly relevant to those with COPD, due to the high prevalence of depression as reported by van Ede et al. (1999), van Manen et al. (2002), and Kunik et al. (2005), and the association of depression with increased mortality (Almagro et al., 2002; de Voogd et al., 2009; Fan et al., 2007; Gudmundsson et al., 2012; Ng et al., 2007; Papaioannou et al., 2013; Stage et al., 2005; Yohannes et al., 2005). Depression is also associated with smoking, although the directionality of the relationship is still debated (Fluharty et al., 2017). Studies have shown that the prevalence of depression or depressive symptoms in COPD patients ranges from 30–60% (Kunik et al., 2005; van Ede et al., 1999; van Manen et al., 2002). Wide variation in the prevalence estimates is likely due to different definitions of depression and depressive symptoms. Furthermore, mortality among COPD patients with depression is higher than those without depression. Although there is variation in the odds ratio estimates, ranging from 0.30 to 3.60, the majority of the odds ratios suggest increased mortality in COPD patients with depression compared to those without depression (Almagro et al., 2002; de Voogd et al., 2009; Fan et al., 2007; Gudmundsson et al., 2012; Ng et al., 2007; Papaioannou et al., 2013; Stage et al., 2005; Yohannes et al., 2005). Additionally, the only study the authors are aware of that investigated the genetics of depression and COPD focused on a single gene (Ishii et al., 2011). Thus, the genetic risk factors of depression among smokers and COPD patients with a more unbiased GWAS approach are a critical issue to investigate.

Using data from a well-characterized population of smokers with and without COPD, we sought to identify genetic risk factors for depressive symptoms among smokers and to

expand upon the existing GWAS studies of depression that have been published in other populations. Because it is still challenging to achieve genome-wide significance even with the relatively large sample size of the present study from the COPDGene cohort, network analyses can supplement our understanding of the genetics of depression and fill gaps left by GWAS through exploration of interactions between genes.

Methods

Participants

The COPDGene study collected both genome-wide genetic data and phenotypic information on over 10,000 subjects (Regan et al., 2010). COPDGene Phase 1 began in 2007 and included over 10,000 subjects with information about medication use, including antidepressants, and smoking history, including past and current smoking. COPDGene Phase 2 collected 5-year follow-up data for approximately 8,000 subjects returning from Phase 1. The primary inclusion criteria were self-identified ethnicity as Non-Hispanic White (NHW) or African-American (AA) between 45 and 80 years old at study enrollment, with at least a 10 pack-year smoking history. Some examples of exclusion criteria were a history of other lung disease except asthma, previous surgical excision of at least one lung lobe, and a first- or second-degree relative enrolled in the study. The complete exclusion criteria is listed in the COPDGene Study (Regan et al., 2010). Depressive symptoms were measured in Phase 2 using the Hospital Anxiety and Depression Scale depression subscale score (HADS-D), and all participants reported their use of antidepressant medications in Phase 1 and Phase 2 (Zigmong and Snaith, 1983). HADS-D data were available from approximately 2,000 participants from their Phase 2 visit.

Definition of depression phenotype

COPDGene subjects were categorized into two groups, one with evidence of the depression phenotype and another without evidence of the phenotype. We defined the phenotype of depression as subjects reporting antidepressant use at the time of the Phase 1 COPDGene study visit using a medication record available for the entire cohort, or a self-reported HADS-Depression Subscale score   8 at the Phase 2 visit for a sub group (~2,000) of study participants. This was due to data availability at the time of analysis. We defined antidepressant use as a binary variable based on a previous analysis by Iyer et al. (submitted) as shown in Supplemental Table S1. The HADS-D subscale score was used instead of the total HADS score to focus on only depressive symptoms and not anxiety symptoms, as the HADS total score also includes anxiety-specific questions. We then separated subjects into Non-Hispanic White (NHW) and African-American (AA) datasets and analyzed them separately to avoid confounders due to ethnic stratification.

Genome Wide Association Study (GWAS) and Systems Biology Analysis

The initial analysis of each dataset was a GWAS, and the details of the quality control can be found at the COPDGene website (http://www.copdgene.org/study-design); details of this process and imputation have been previously described. Briefly, genotyping quality control was performed following previously described guidelines (Cho et al., 2014). DNA samples from COPDGene subjects were genotyped on the Illumina HumanOmniExpress array.

GenomeStudio quality control, including manual review of cluster plots, was performed. Genetic ancestry was adjusted by principal components to identify racial mismatches and population outliers. Markers with low minor allele frequency (<1%) were additionally excluded for the primary analysis. Imputation to 1,000 Genomes Phase One v3 EUR used MaCH and minimax. We performed logistic regression adjusting for ancestry principal components using an additive genetic model as implemented in plink 1.9.

The gene annotation was to the closest gene from the SNPs. When we had multiple SNPs annotated from the same gene, we chose the SNPs with the smallest p-value to rank the order of the top hit genes.

Systems biology analysis was then performed using online software—GeneMANIA (Warde-Farley et al., 2010), DAVID (Huang et al., 2009a; Huang et al., 2009b), ConsensusPathDB (Kamburov et al., 2011; Kamburov et al., 2009), and GLITTER (Liu et al., 2016) to identify biological pathways associated with the depressed phenotype. GeneMANIA (http://www.genemania.org) uses a given input list and extends the list with functionally similar genes using data from GEO, BioGRID, Pathway Commons, and I2D (Warde-Farley et al., 2010). DAVID (http://www.david.ncifcrf.gov) uses a given input list and agglomerates millions of gene and protein identifiers based on genome databases such as NCBI, PIR, and UniProt into secondary gene clusters based on gene ontology functionality (Huang et al., 2009a; Huang et al., 2009b). ConsensusPathDB (http://www.cpdb.molgen.mpg.de) uses a given input list and computes functional interaction maps based on genomic data from 12 databases, including Reactome, KEGG, Bio GRID and more (Kamburov et al., 2011; Kamburov et al., 2009). GLITTER (http://www.han-lab.org/GLITTER) is designed to examine the functional relatedness of disease susceptibility genes according to tissue-specific gene expression profiles. It can also shed light on the specific tissues in which susceptibility genes might exert their functions. The gene expression profiles are from the RNA-Seq data (v.6) from the Genotype-Tissue Expression Project (GTEx) for 49 different tissues (Liu et al., 2016). Different numbers of top hit genes were used with each program during analysis. This is because of the different recommendations given by each program for producing high-quality results as reported on their respective websites.

Results

Analyzed study subjects

We had access to data on 9,970 subjects for the current analysis with genotypic information for the GWAS. Of the 9,970 subjects genotyped, 254 subjects provided limited consent for using their blood samples and were excluded. Thus, a total of 9,716 subjects remained for the final analysis, with 5,198 (53.5%) male and 4,518 (46.5%) female. The mean age was 59.4 (SD = 9.0) and 6,576 (67.7%) of the subjects were NHW, while 3,140 (32.3%) were AA. Overall, 1,622 (24.7%) of the NHW subjects and 391 (12.5%) of the AA subjects were defined as the depressed phenotype. While all subjects were smokers, they were further classified by COPD status based on the GOLD Stages classification system (Vestbo et al., 2013). The breakdown of the classification of the subjects into different GOLD stages is shown in Table 1. The COPD group trended towards being older, more likely male, and with a higher pack-year smoking history, when compared to non-COPD patients. It is worth

noting that while all subjects were smokers, the prevalence of COPD in the sample was 49%. This prevalence estimate is higher than in previous literature, which suggests a prevalence of around 20%, although there is a wide variation in prevalence estimates (Vestbo et al., 2013).

GWAS

The GWAS for NHW subjects was conducted and top hits are listed in order of descending p-value (Table 2). The genes listed in the table are the genes nearest to the corresponding SNPs. The results from the GWAS revealed top hits in the $p = \sim10^{-7}$ range—not reaching genome wide statistical significance at the $\sim 10^{-8}$ cutoff based on correction for GWAS multiple testing. The results for the African-American analysis did not reach statistical significance either, likely due to the even smaller sample size (3,140 subjects). Because the sample was smaller and neither GWAS nor subsequent network analysis demonstrated a strong signal, all African-American dataset results are shown in the supplementary tables. The results for the GWAS can be found in Supplemental Table S2.

GeneMANIA

The GeneMANIA analysis using the top 35 NHW gene hits graphically depicted the interactions between the top hit genes from the GWAS as shown in Figure 1.

In this figure, the lines connecting the circles represent a combination of co-expression, co-localization, and genetic interactions. It indicates that these genes are regulating each other in concert and may be working together. The GeneMANIA figure showed potential tight interconnections between the top hit genes, both directly and indirectly through other genes. Of note, some top hit genes are not shown in the figure, as the genes were either not in the GeneMANIA reference gene databases (FAM222A-AS1, A2MP1, and RNU6–83) or had no interconnections within the identified network (MYEOV2). The top hit genes from the AA dataset did not show strong connections among top hit genes. The AA results can be found in Supplemental Figure S1.

DAVID and ConsensusPathDB

Next, to identify pathways by which top genes may interact and function together, gene ontology analysis software such as DAVID and ConsensusPathDB were utilized. The following tables show the results from these analyses. While the programs identified many pathways and gene ontologies, the following tables show the noteworthy ones specifically.

The DAVID analysis identified various nominally significant gene ontologies, such as central nervous system development, synaptic signaling, neuron projection development, and neuron projection extension (Table 3). There was also some signal for neurotransmitter transport and acetylcholine receptor activity. However, none of these gene ontologies were statistically significant when corrected for multiple comparisons.

Table 3 also shows the top gene ontologies from the ConsensusPathDB analysis. ConsensusPathDB provided similar ontologies to that of DAVID; however, it should be noted that transcription corepressor activity survived at a statistically significant level after

correction for multiple testing. Other strong signals were from neurotransmitter secretion and synaptic transmission. The full results for the NHW analysis are shown in Supplemental Tables S3 and S4. The full results from the AA analysis are shown in Supplemental Tables S5 and S6.

GLITTER

GLITTER software was used to determine the relative co-expression patterns of the top hit genes in different tissues in the human body. The results from GLITTER are shown in Table 4 and indicate that the top hit genes from this analysis are preferably co-expressed in human brain tissues relative to other non-brain tissues in the human body. Brain-related tissues are bolded in Table 4.

Of note, the top hit, Brain—Caudate (basal ganglia), was statistically significant after multiple corrections. Table 4 also illustrates that the brain-related tissues are concentrated towards the top of the table. Results for the African-American analysis can be found in Supplemental Table S7.

Discussion

This is the first GWAS that focused on depressive symptoms among a large cohort of current and former smokers. The analysis used data from the COPDGene study and has several strengths. First, a single study enrolling over 9,000 subjects with a uniform protocol allowed us to evaluate the genetic disposition of depressive symptoms among a large sample of smokers. The NHW sample included over 6,500 subjects, and the AA sample included over 3,000 subjects. Although the top hits from the GWAS did not reach genome-wide significance, systems biology analysis using GeneMANIA, DAVID, ConsensusPathDB, and GLITTER for the NHW sample demonstrated that the top hit genes function together to form pathways relevant to brain function. Systems biology results for the AA dataset did not demonstrate such findings. The lack of replication is possibly due to the limited identification of the depressed phenotype as evidenced by the much lower prevalence of depression cases among AA (12.5%) compared to NHW (24.7%).

When looking at the top hit genes found in this GWAS, many have been reported previously in the psychiatric literature. This includes the Cholinergic Receptor Muscarinic 3 (*CHRM3*, rs12036147; $p = 1.28 \times 10^{-6}$), the MAM Domain Containing Glycosylphosphatidylinositol Anchor 2 (*MDGA2*, rs17118176; $p = 3.52 \times 10^{-6}$), the Solute Carrier Family 1 Member 3 (*SLC1A3*, rs13171357; $p = 5.27 \times 10^{-6}$), the Calcium Voltage-Gated Channel Auxiliary Subunit alpha2delta3 (*CACNA2D3*, rs17744447; $p = 7.86 \times 10^{-6}$), and the Proline Rich Membrane Anchor 1 (*PRIMA1*, rs11160142; $p = 2.38 \times 10^{-5}$). The *CHRM3* gene is a cholinergic receptor and has been implicated in schizophrenia and may be a potential treatment target in COPD patients as speculated in a previous study (Wain et al., 2017; Wang et al., 2016). The *MDGA2* gene is a membrane anchor that was a top hit in a neuroticism GWAS of 1,200 subjects (van den Oord et al., 2008). Although the SNP found in the neuroticism study and the SNPs found in this COPDGene cohort are not in strong linkage disequilibrium, it is intriguing to find signals in the same gene associated with neuroticism, because neuroticism is a personality trait highly relevant to depression (Farmer et al., 2002).

The *SLC1A3* gene codes for glutamate transport and has been implicated in ADHD (Turic et al., 2005), while the *CACNA2D3* gene codes for a calcium channel complex subunit that has been implicated in autism (Heyes et al., 2015). Another calcium channel gene, the *CACNA1C* gene has been associated with bipolar disorder in multiple studies (Fiorentino et al., 2014; Starnawska et al., 2016), highlighting the potential importance of the *CACNA2D3* gene to mood. The *PRIMA1* gene assists with acetylcholinesterase and was a top hit in a previous genome-wide DNA methylation association study that included 39 postmortem frontal cortex samples (Sabunciyan et al., 2012). Pathway analysis by DAVID and ConsensusPathDB showed nominally significant pathways, including cholinergic activity, that were not significant after multiple testing correction. This was likely due to the small sample size relative to GWAS standards and also phenotypic definition challenges not based on diagnostic criteria for major depressive disorder. The GLITTER analysis showed that the caudate of the basal ganglia was a brain region where the top hit genes are expressed together. The caudate nucleus is innervated by dopaminergic neurons, suggesting the role of neurotransmission, especially that of dopamine, in the pathophysiology of depressive symptoms among smokers. Although not included in the top hits listed in the table, the serotonin receptor gene was also identified, suggesting potential involvement of serotonin in depression pathophysiology.

It is worth emphasizing that all subjects in the study were smokers. The results from this analysis may contribute to a better understanding of the genetic link of the risk for depressive symptoms in the context of nicotine addiction. It is possible that the genes contributing to the risk of depressive symptoms among smokers could be different from the genetic risk factors for depression among the general population. These genes may share expression patterns and interact with each other at multiple levels within the brain. This would be a potential area for future study.

Based on the promising signals relevant to depressive symptoms and brain function, if a standardized psychiatric evaluation is conducted for the COPDGene cohort, it would be possible to improve the phenotypic definition of depression and produce stronger GWAS signals for genome-wide significance. Moreover, Exome Chip and Exome Sequence data will be available for further enrichment analysis against other depression GWAS analyses and could validate the signals obtained here.

Limitations

This study has several limitations. First, the definition of the depression phenotype was based on antidepressant use information available from the entire cohort, with HADS-D scores available for only a sub-group of ~2,000 participants. Using the best available definition based on antidepressant use and HADS-D together likely increased the rate of false negatives (i.e., subjects with currently elevated depressive symptoms who were "missed" and categorized as non-depressed if they were not taking antidepressant medications). Another limitation of using antidepressant use to code for the depression phenotype is that antidepressants (particularly bupropion) can be used for smoking cessation among non-depressed individuals and that tricyclic antidepressants can be used for sleep and/or chronic pain. Also, antidepressant use history may include subjects with anxiety and

not depression, who are being treated with those medications. Thus, our categorization of the depressed phenotype may also have resulted in false positives. Our analysis did not find significant GWAS associations, and the subsequent analysis assigned the closest gene to the top variant and did not account for biases in gene size or distribution.

We observed a higher prevalence of the depression phenotype in the NHW sample than in the AA sample. Previous research has found a lower rate of symptom endorsement for the AA population in general (Connor et al., 2010). It is also possible that depression in African-Americans is under-recognized by medical providers, leading to lower rates of antidepressant use. Thus, it is possible that the AA sample had more false negatives than the NHW sample. Concern for a falsely low prevalence of the depression phenotype as well as the smaller sample size (~3,000) for the AA sample likely contributed to the absence of signal relevant to psychiatric conditions in the AA sample. Focusing on understanding genetic risk for psychiatric symptoms in potentially underserved populations is an important area for future work.

Conclusions

The COPDGene cohort provides important data in which to study depressive symptoms due to the large sample size of the cohort with genetic data and the high prevalence of depression in smokers. GWAS performed on this cohort revealed top hit genes in the $p = \sim 10^{-7}$ range, which did not survive multiple corrections for genome-wide significance. Given that depression is likely caused by many genes and the cumulative impact of many small effect sizes, the top hit genes were analyzed from a systems biology approach. GeneMANIA indicated that the top hit genes were tightly interconnected and that some top hit genes have been implicated in psychiatric disorders. Analysis using DAVID and ConsensusPathDB identified brain-related pathways through which the genes are interconnected through central nervous system development, synaptic signaling, neuron projection development, and cholinergic function. GLITTER analysis showed that the top hit genes are primarily co-expressed in the human brain as compared to other non-brain tissues, specifically in the caudate of the basal ganglia at a statistically significant level.

In conclusion, the COPDGene dataset shows signals for genes previously cited in the psychiatric literature and suggests such genes function through central nervous system development, neuronal projection, and dopaminergic activity. Future studies with the dataset may significantly contribute to a better understanding of the relationship between genetics and depression among smokers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institute of Health. We would also like to acknowledge all the COPDGene Investigators.

References

Almagro P, Calbo E, Ochoa de Echaguen A, Barreiro B, Quintana S, et al. (2002). Mortality after hospitalization for COPD. Chest. 121(5), 1441–1448. [PubMed: 12006426]

Cho MH, McDonald ML, Zhou X, Mattheisen M, Castaldi PJ, et al. (2014). Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. Lancet Respir Med. 2(3), 214–225. [PubMed: 24621683]

Connor K, Copeland V, Grote N, Koeske G, Rosen D, et al. (2010). Mental health treatment seeking among older adults with depression: the impact of stigma and race. Am J Geriatr Psychiatry. 18(6), 531–543. [PubMed: 20220602]

Consortium CONVERGE. (2015). Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature. 523(7562), 588–591. [PubMed: 26176920]

de Voogd J, Wempe J, Koeter G, Postema K, van Sonderen E, et al. (2009). Depressive Symptoms as Predictors of Mortality in Patients With COPD. Chest. 135(3), 619–625. [PubMed: 19029432]

Fan VS, Ramsey SD, Giardino ND, Make BJ, Emery CF, et al. (2007). Sex, depression, and risk of hospitalization and mortality in chronic obstructive pulmonary disease. Arch Intern Med. 167(21), 2345–2353. [PubMed: 18039994]

Farmer A, Redman K, Harris T, Mahmood A, Sadler S, et al. (2002). Neuroticism, extraversion, life events and depression. The Cardiff Depression Study. Br J Psychiatry. 181, 118–122. [PubMed: 12151281]

Fiorentino A, O'Brien NL, Locke DP, McQuillin A, Jarram A, et al. (2014). Analysis of ANK3 and CACNA1C variants identified in bipolar disorder whole genome sequence data. Bipolar Disord. 16(6), 583–591. [PubMed: 24716743]

Flint J, and Kendler K (2014). The genetics of major depression. Neuron. 81(3), 484–503. [PubMed: 24507187]

Fluharty M, Taylor AE, Grabski M, and Munafo MR (2017). The Association of Cigarette Smoking With Depression and Anxiety: A Systematic Review. Nicotine Tob Res. 19(1), 3–13. [PubMed: 27199385]

Gudmundsson G, Ulrik CS, Gislason T, Lindberg E, Brondum E, et al. (2012). Long-term survival in patients hospitalized for chronic obstructive pulmonary disease: a prospective observational study in the Nordic countries. Int J Chron Obstruct Pulmon Dis. 7, 571–576. [PubMed: 23055707]

Heyes S, Pratt WS, Rees E, Dahimene S, Ferron L, et al. (2015). Genetic disruption of voltage-gated calcium channels in psychiatric and neurological disorders. Prog Neurobiol. 134, 36–54. [PubMed: 26386135]

Huang D, Sherman B, and Lempicki R (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37(1), 1–13. [PubMed: 19033363]

Huang D, Sherman B, and Lempicki R (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 4(1), 44–57. [PubMed: 19131956]

Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, et al. (2016). Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat Genet. 48(9), 1031–1036. [PubMed: 27479909]

Ishii T, Wakabayashi R, Kurosaki H, Gemma A, and Kida K (2011). Association of serotonin transporter gene variation with smoking, chronic obstructive pulmonary disease, and its depressive symptoms. J Hum Genet. 56(1), 41–46. [PubMed: 20981038]

Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H, et al. (2011). ConsensusPathDB: toward a more complete picture of cell biology. Nucleic Acids Res. 39(Database issue), D712–717. [PubMed: 21071422]

Kamburov A, Wierling C, Lehrach H, and Herwig R (2009). ConsensusPathDB-a database for integrating human functional interaction networks. Nucleic Acids Res. 37, D623–D628. [PubMed: 18940869]

Kendler KS, Gatz M, Gardner CO, and Pedersen NL (2006). A Swedish national twin study of lifetime major depression. Am J Psychiatry. 163(1), 109–114. [PubMed: 16390897]

Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, et al. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry. 62(6), 593–602. [PubMed: 15939837]

Kunik ME, Roundy K, Veazey C, Souchek J, Richardson P, et al. (2005). Surprisingly high prevalence of anxiety and depression in chronic breathing disorders. Chest. 127(4), 1205–1211. [PubMed: 15821196]

Liu X, Yu P, Cheng C, Potash JB, and Han S (2016). GLITTER: a web-based application for gene link inspection through tissue-specific coexpression. Sci Rep. 6, 33460. [PubMed: 27623690]

Ng TP, Niti M, Tan WC, Cao Z, Ong KC, et al. (2007). Depressive symptoms and chronic obstructive pulmonary disease: effect on mortality, hospital readmission, symptom burden, functional status, and quality of life. Arch Intern Med. 167(1), 60–67. [PubMed: 17210879]

Papaioannou AI, Bartziokas K, Tsikrika S, Karakontaki F, Kastanakis E, et al. (2013). The impact of depressive symptoms on recovery and outcome of hospitalised COPD exacerbations. Eur Respir J. 41(4), 815–823. [PubMed: 22878874]

Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, et al. (2010). Genetic epidemiology of COPD (COPDGene) study design. COPD. 7(1), 32–43. [PubMed: 20214461]

Sabunciyan S, Aryee MJ, Irizarry RA, Rongione M, Webster MJ, et al. (2012). Genome-wide DNA methylation scan in major depressive disorder. PLoS One. 7(4), e34451. [PubMed: 22511943]

Stage KB, Middelboe T, and Pisinger C (2005). Depression and chronic obstructive pulmonary disease (COPD). Impact on survival. Acta Psychiatr Scand. 111(4), 320–323. [PubMed: 15740469]

Starnawska A, Demontis D, Pen A, Hedemand A, Nielsen AL, et al. (2016). CACNA1C hypermethylation is associated with bipolar disorder. Transl Psychiatry. 7(6), e831.

Turic D, Langley K, Williams H, Norton N, Williams NM, et al. (2005). A family based study implicates solute carrier family 1-member 3 (SLC1A3) gene in attention-deficit/hyperactivity disorder. Biol Psychiatry. 57(11), 1461–1466. [PubMed: 15950021]

van den Oord EJ, Kuo PH, Hartmann AM, Webb BT, Moller HJ, et al. (2008). Genomewide association analysis followed by a replication study implicates a novel candidate gene for neuroticism. Arch Gen Psychiatry. 65(9), 1062–1071. [PubMed: 18762592]

van Ede L, Yzermans CJ, and Brouwer HJ (1999). Prevalence of depression in patients with chronic obstructive pulmonary disease: a systematic review. Thorax. 54(8), 688–692. [PubMed: 10413720]

van Manen JG, Bindels PJ, Dekker FW, IJzermans C, van der Zee JS, et al. (2002). Risk of depression in patients with chronic obstructive pulmonary disease and its determinants. Thorax. 57(5), 412–416. [PubMed: 11978917]

Vestbo J, Hurd S, Agusti A, Jones P, Vogelmeier C, et al. (2013). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. American Journal of Respiratory and Critical Care Medicine. 187(4), 347–365. [PubMed: 22878278]

Wain LV, Shrine N, Artigas MS, Erzurumluoglu AM, Noyvert B, et al. (2017). Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. Nat Genet. 49(3), 416–425. [PubMed: 28166213]

Wang Q, Cheng W, Li M, Ren H, Hu X, et al. (2016). The CHRM3 gene is implicated in abnormal thalamo-orbital frontal cortex functional connectivity in first-episode treatment-naive patients with schizophrenia. Psychol Med. 46(7), 1523–1534. [PubMed: 26959877]

Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, et al. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 38(Web Server issue), W214–220. [PubMed: 20576703]

Wray NR, and Sullivan PF (2017). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. bioRxiv.

Yohannes A, Baldwin R, and Connolly M (2005). Predictors of 1-year mortality in patients discharged from hospital following acute exacerbation of chronic obstructive pulmonary disease. Age Ageing. 34(5), 491–496. [PubMed: 16107452]

Zigmong A, and Snaith R (1983). The Hospital Anxiety and Depression Scale. Acta Psychiatrica Scandinavica. 67(6), 361–370. [PubMed: 6880820]

Highlights:

- GWAS analysis of the COPDGene study showed several top hit SNPs.

- Genes nearest top hit SNPs are associated with depression in the literature.

- Such genes showed potential association with neurotransmitter networks.

- The genes are expressed together in the brain, especially in the basal ganglia.

- The COPDGene study provides future opportunities to study depression among smokers.
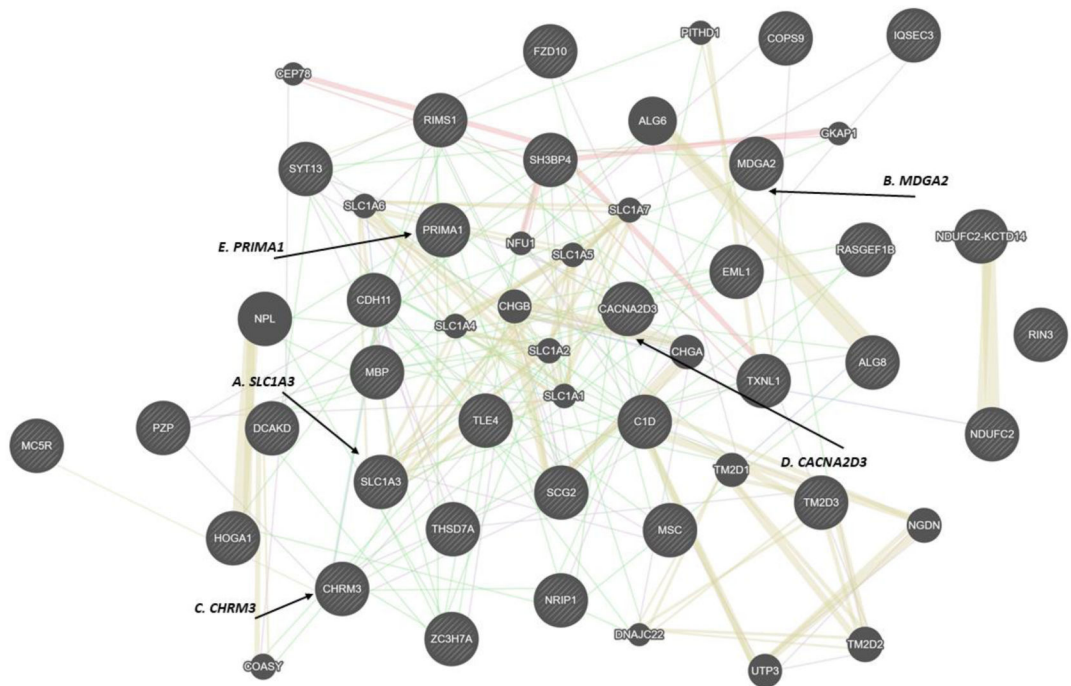
Figure 1.

Outcome from GeneMANIA based on NHW top 35 gene hit results. The genes from the top hit GWAS are shown as larger circles, while genes from the GeneMANIA extension are smaller. A) SLC1A3; p= 5.27×10^{-6} B) MDGA2; p= 3.52×10^{-6} C) CHRM3; p= 1.28×10^{-6} D) CACNA2D3; p= 7.86×10^{-6} E) PRIMA1; p= 2.38×10^{-5}.

Table 1.

Summary statistics of the 6,576 NHW subjects analyzed in this study.GOLD −1 = Preserved ratio, impaired spirometry, (FEV1/FVC   0.7, FEV1% predicted < 80%); GOLD 0 = Normal spirometry; GOLD 1 = FEV1/FVC < 70%, FEV1   80% predicted; GOLD 2 = FEV1/FVC < 70%, 50% < FEV1% predicted < 80%; GOLD 3 = FEV1/FVC < 70%, 30% < FEV1% predicted < 50%; GOLD 4 = FEV1/FVC < 70%, FEV1   30% predicted.

	Non-COPD			COPD				
GOLD Stage	**−1**	**0**	*Total Non-COPD*	**1**	**2**	**3**	**4**	*Total COPD*
# of Subjects	692	2,509	3,201	595	1,407	892	481	3,375
Ave. Age	60.1	60.0	59.7	63.1	64.2	65.2	64.8	64.4
# of Males	303	1,240	1,543	349	755	513	281	1,898
Sex (% Male)	44%	49%	48%	59%	54%	58%	58%	56%
Ave. Pack-Years	46.5	38.0	39.8	46.3	54.0	58.0	60.4	54.5

Table 2.

Top 35 Non-Hispanic White GWAS Results. SNPs are the top hit SNPs from the GWAS analysis. Nearest gene is the annotated nearest gene to the SNP. Many SNPs were closest to the same gene. The table omits the duplicate genes to show unique annotated top genes. A1 is effect allele. A2 is other allele. MAF is the frequency of A1.

SNP	Nearest Gene	A1	A2	MAF	P-Value
rs6487668	PZP	C	A	0.2181	7.06×10^{-7}
rs35426224	EML1	G	A	0.9255	1.07×10^{-6}
rs734152	FAM222A-AS1	G	C	0.8804	1.13×10^{-6}
rs12036147	CHRM3	A	T	0.9116	1.28×10^{-6}
rs79309730	SH3BP4	C	T	0.9427	1.70×10^{-6}
rs7298108	A2MP1	C	T	0.8217	2.01×10^{-6}
rs17118176	MDGA2	A	G	0.9373	3.52×10^{-6}
rs6485603	SYT13	G	A	0.5506	4.76×10^{-6}
rs143756124	MC5R	T	C	0.9320	4.90×10^{-6}
rs6743806	SCG2	C	G	0.7326	5.08×10^{-6}
rs13171357	SLC1A3	G	A	0.7185	5.27×10^{-6}
rs7194099	ZC3H7A	A	G	0.5561	5.48×10^{-6}
rs180931486	MBP	T	C	0.9334	6.21×10^{-6}
rs536504	RASGEF1B	C	G	0.6995	6.79×10^{-6}
rs17744447	CACNA2D3	A	G	0.3458	7.86×10^{-6}
rs567697	RNU6–83	T	C	0.8451	8.09×10^{-6}
rs76088896	NPJP1	T	C	0.8925	8.66×10^{-6}
rs13257626	MSC	G	T	0.9769	9.06×10^{-6}
rs570456	ALG8	C	T	0.7392	9.23×10^{-6}
rs117574193	CDH11	G	A	0.9750	9.63×10^{-6}
rs10773937	IQSEC3	G	C	0.4928	9.86×10^{-6}
rs58200410	C1D	C	A	0.9545	1.02×10^{-5}
rs7786342	THSD7A	C	A	0.6497	1.08×10^{-5}
rs12450972	DCAKD	C	T	0.5978	1.21×10^{-5}
rs609952	NDUFC2	C	T	0.8459	1.43×10^{-5}
rs75277622	RIN3	C	T	0.9361	1.63×10^{-5}
rs112680866	TM2D3	A	T	0.9749	1.67×10^{-5}
rs143236826	TLE4	A	C	0.9342	1.67×10^{-5}
rs61863283	HOGA1	G	A	0.7112	1.75×10^{-5}
rs13389043	MYEOV2	T	G	0.8353	1.79×10^{-5}
rs56277520	FZD10	G	A	0.5515	1.82×10^{-5}
rs7767783	RIMS1	T	C	0.6306	1.86×10^{-5}
rs140245790	TXNL1	A	G	0.9792	1.96×10^{-5}
rs655445	NDUFC2-KCTD14	G	A	0.8484	1.97×10^{-5}

SNP	Nearest Gene	Al	A2	MAF	P-Value
rs11160142	PRTMA1	A	G	0.6436	2.38×10^{-5}

Table 3.

Pertinent gene ontologies from DAVID and ConsensusPathDB. While gene ontologies in DAVID, such as CNS development, synaptic signaling, neuron projection development, and neuron projection extension showed nominally significant signals, they did not survive correction for multiple testing. In the top 30 gene analysis in ConsensusPathDB, transcription corepressor activity remained significant even after correction for multiple testing. The gene ontology abbreviation BP is Biological Process, MF is Molecular Function, and CC is Cellular Component.

	Number of Input Genes	Gene Ontology Term	Gene Ontology	P-Value	Q-Value
DAVID	20	Synaptic signaling	BP	0.02	0.99
		Central nervous system development	BP	0.05	0.98
	50	Synaptic signaling	BP	0.01	0.99
		Neurotransmitter transport	BP	0.07	0.99
		Synaptic membrane	CC	0.03	0.95
		Presynaptic active zone	CC	0.07	0.94
		Acetylcholine receptor activity	MF	0.07	0.82
	100	Neuron projection development	BP	0.01	0.77
		Central nervous system development	BP	0.02	0.82
		Neuron projection extension	BP	0.02	0.81
		Regulation of axon extension	BP	0.06	0.89
		Regulation of neurogenesis	BP	0.07	0.89
Consensus PathDB	20	Synaptic transmission	BP	0.004	0.13
		Central nervous system development	BP	0.008	0.13
		Neurotransmitter secretion	BP	0.009	0.13
	30	Transcription corepressor activity	MF	0.0002	0.002

Table 4.

GLITTER results from the top 100 gene hits analysis sorted by p-value. Brain-related tissues are bolded.

Tissue	P-value
Brain-Caudate (basal ganglia)	0.00010
Brain-Putamen (basal ganglia)	0.0057
Brain-Substantia nigra	0.019
Stomach	0.026
Brain-Nucleus accumbens (basal ganglia)	0.037
Brain-Cortex	0.052
Brain-Spinal cord (cervical C1)	0.059
Adipose-Subcutaneous	0.070
Brain-Hypothalamus	0.079
Brain-Hippocampus	0.13
Prostate	0.14
Brain-Amygdala	0.15
Vagina	0.17
Colon-Transverse	0.19
Brain-Cerebellum	0.24
Brain-Frontal Cortex(BA9)	0.24
Brain-Anterior cingulate cortex(BA24)	0.28
Ovary	0.28
Pituitary	0.33
Skin-Not Sun Exposed(Suprapubic)	0.35
Artery-Coronary	0.38
Small Intestine-Terminal Ileum	0.39
Breast-Mammary Tissue	0.47
Colon-Sigmoid	0.48
Uterus	0.50
Brain-Cerebellar Hemisphere	0.51
Heart-Left Ventricle	0.65
Kidney-Cortex	0.66
Muscle-Skeletal	0.72
Minor Salivary Gland	0.72
Esophagus-Mucosa	0.73
Adipose-Visceral (Omentum)	0.73
Testis	0.80
Pancreas	0.82
Esophagus-Muscularis	0.82
Heart-Atrial Appendage	0.83
Esophagus-Gastroesophageal Junction	0.84
Lung	0.84
Nerve-Tibial	0.85

Tissue	P-value
Cells-Transformed fibroblasts	0.88
Skin-Sun Exposed (Lower leg)	0.88
Whole Blood	0.88
PPI	0.89
Liver	0.98