Journal of Affective Disorders 243 (2019) 16–22

Contents lists available at ScienceDirect

Journal of Affective Disorders

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

Research paper

GWAS and systems biology analysis of depressive symptoms among smokers from the COPDGene cohort

Jonathan T. Heinzman[a], Karin F. Hoth[a], Michael H. Cho[b,c], Phuwanat Sakornsakolpat[b], Elizabeth A. Regan[d,e], Barry J. Make[d], Gregory L. Kinney[e], Frederick S. Wamboldt[d,f], Kristen E. Holm[d], Nicholas Bormann[a], Julian Robles[a], Victor Kim[g], Anand S. Iyer[h,i,j], Edwin K. Silverman[b,c], James D. Crapo[d], Shizhong Han[a,k], James B. Potash[a,k], Gen Shinozaki[a,*], on behalf of the COPDGene Investigators

[a] Department of Psychiatry, University of Iowa Hospitals and Clinics, Iowa City, IA, USA
[b] Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA
[c] Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA
[d] Department of Medicine, National Jewish Health, Denver, CO, USA
[e] Department of Epidemiology, Colorado School of Public Health, University of Colorado, Aurora, CO, USA
[f] Department of Psychiatry, University of Colorado School of Medicine at the Anschutz Medical Campus, Aurora, CO, USA
[g] Department of Thoracic Medicine and Surgery, Temple University School of Medicine, Philadelphia, PA, USA
[h] Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA
[i] Lung Health Center, University of Alabama at Birmingham, Birmingham, AL, USA
[j] Health Services, Outcomes, and Effectiveness Research Training Program, University of Alabama at Birmingham, Birmingham, AL, USA
[k] Department of Psychiatry, Johns Hopkins University, Baltimore, MD, USA

ARTICLE INFO

Keywords:
Major depressive disorder
Systems biology
Chronic obstructive pulmonary disease
Smokers
Genome-wide association study

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 9716 COPDGene subjects with ≥10 pack-year history. Depression was defined as antidepressant use and/or a HADS depression subscale score ≥8. Non-Hispanic White (6576) and African-American (3140) 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 (1622) 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 a 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.

1. 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

* Corresponding author.
E-mail address: gen-shinozaki@uiowa.edu (G. Shinozaki).

https://doi.org/10.1016/j.jad.2018.09.003
Received 4 April 2018; Received in revised form 24 July 2018; Accepted 6 September 2018
Available online 07 September 2018
0165-0327/

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 5303 Han Chinese women and 5337 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 9000 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% to 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.

2. Methods

2.1. 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 8000 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 are 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 2000 participants from their Phase 2 visit.

2.2. 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 (~2000) 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 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.

2.3. 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 1000 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,b), ConsensusPathDB (Kamburov et al.2009, 2011), and GLITTER (Liu et al., 2016) to identify biological pathways associated