Race-Specific Differential DNA Methylation Marks and Gene Comethylation Modules of COPD

Share Your Story

Title: Race-Specific Differential DNA Methylation Marks and Gene Comethylation Modules of COPD

Authors: Robert Busch, MD[1], Weiliang Qiu, PhD[1], Jessica Lasky-Su, PhD[1], Jarrett Morrow, PhD[1], Gerard Criner, MD[2], and Dawn DeMeo, MD MPH[1]

Affiliations: [1]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA/USA; [2]Temple Lung Center, Temple University Health System, Philadelphia, PA/USA

Email addresses:

Robert Busch   rerbu@channing.harvard.edu

Weiliang Qiu   stwxq@channing.harvard.edu

Jessica Lasky-Su   jessica.a.su@gmail.com

Jarrett Morrow   remdj@channing.harvard.edu

Gerard Criner   Gerard.Criner@tuhs.temple.edu

Dawn DeMeo   redld@channing.harvard.edu

Corresponding Author:

Robert Busch, MD

181 Longwood Ave

Room 456

Boston, MA, USA, 02111

rerbu@channing.harvard.edu

Abstract:

Rationale: Chronic obstructive pulmonary disease (COPD) is the third-leading cause of death worldwide. Identifying COPD-associated DNA methylation marks in African-Americans may contribute to our understanding of racial disparities in COPD susceptibility. We determined differentially methylated genes and co-methylation network modules associated with COPD in African-Americans from the Pennsylvania Study of Chronic Obstructive Pulmonary Exacerbations (PA-SCOPE) cohort.

Methods: We assessed DNA methylation from peripheral blood samples in 362 African-American smokers in the PA-SCOPE cohort using the Illumina Infinium HumanMethylation27 BeadChip Array. Final analysis included 19302 CpG probes annotated to the nearest gene transcript after quality control. We tested methylation associations with COPD case-control status using mixed linear models. Weighted gene comethylation networks were constructed using WGCNA and network modules were analyzed for association with COPD.

Results: There were five differentially methylated CpG probes significantly associated with COPD among African-Americans at an FDR less than 5%, and seven additional probes that approached significance at an FDR less than 10%. The top ranked gene association was *MAML1*, which has been shown to affect *NOTCH*-dependent angiogenesis in murine lung. Network modeling yielded the "yellow" and "blue" comethylation modules which were significantly associated with COPD (p-value $4x10^{-10}$ and $4x10^{-9}$, respectively). The yellow module was enriched for gene sets related to inflammatory pathways known to be relevant to COPD. The blue module contained the top ranked genes in the concurrent differential methylation analysis (*FXYD1/LGI4*, gene significance p-value $1.2x10^{-26}$; *MAML1*, p-value $2.0x10^{-26}$; *CD72*, p-value $2.1x10^{-25}$; and *LPO*, p-value $7.2x10^{-25}$), and was significantly associated with lung development processes in Gene Ontology gene-set enrichment analysis.

Conclusions: We identified 12 differentially methylated CpG sites associated with COPD that mapped to biologically plausible genes. Network module comethylation patterns have identified candidate genes that may be contributing to racial differences in COPD susceptibility and severity. COPD-associated comethylation modules contained genes

previously associated with lung disease and inflammation and recapitulated known COPD-associated genes. The genes implicated by differential methylation and WGCNA analysis may provide mechanistic targets for explaining disparities in COPD susceptibility and outcomes between African-Americans and non-Hispanic Whites.

Trial Registration: NCT00774176, Registry: ClinicalTrials.gov, URL: www.clinicaltrials.gov, Date Registered: 04 January 2008 (retrospectively registered).

Background

Chronic obstructive pulmonary disease (COPD) is an incurable lung disease characterized by progressive airflow obstruction involving emphysematous destruction of lung parenchyma and mucus hypersecretion with chronic bronchitis. Over 12 million Americans are affected by COPD, which is the third leading cause of death in the US[1], and projected to become the third leading cause of death worldwide[2]. Recent data suggest that the prevalence of emphysema, chronic bronchitis, and COPD hospitalizations are increasing among African-Americans (AA) [3-5], and that AA may develop COPD at a younger age than non-Hispanic Whites (NHW)[5]. In addition, AA males have one of the highest prevalence rates of smoking (25.5%) among racial groups in the United States[6], leading to a predictable growing burden of lung disease in this group. AA individuals present with similar severity of airflow obstruction as NHW, despite fewer pack-years of smoking[5].   Once they have developed COPD, AA have lower quality of life scores[7]. Despite these alarming trends, COPD has been understudied in African-Americans.

Race is an important contributor to genetic[8] and epigenetic variability, and recent studies have identified epigenetic association signals that differ between racial groups[9]. Similarly, the results of differential methylation association studies of complex traits in single racial-ancestry cohorts may miss epigenetic risk factors in another racial-ancestry cohort, and may not be generalizable to other racial cohorts at all[10, 11]. Recent methylation studies have shown a subset of methylation signals particular to AA smokers[9], but to our knowledge race-specific investigations of epigenetic associations with COPD have not been previously performed. Understanding the epigenetic associations of smoking and COPD in AA current and former smokers may provide insights into race-specific features relevant to COPD-related disparities in AA.

DNA methylation patterns are determined at multiple time points in the life of an individual[12], including *in utero* during imprinting, tissue-specific methylation during development, and changes in the methylation of genes in response to major environmental exposures. Differential methylation impacts gene regulation, which may lead to clinically relevant changes in disease-related phenotypes. Modules of genes with correlated comethylation profiles may identify groups of genes under similar regulation that are associated with COPD risk. Prior research has identified differential

methylation signals related to tobacco smoke exposure that may influence risk for development of COPD[13-18]. The majority of these studies have focused on NHW subjects as the largest proportion of their cohorts. Our investigation focused on the identification of differential methylation sites associated with COPD as well as COPD-associated comethylation modules in an AA cohort. Our hypothesis was that race-specific patterns of DNA methylation would identify differentially methylated genes or comethylation networks that may contribute to racial disparities in COPD. A better understanding of the epigenetic factors associated with the features of COPD in AA smokers may provide insights into new diagnostic options, drive the discovery and targeting of therapeutics, and improve primary prevention strategies in this susceptible population.

Results

After quality control, the PA-SCOPE AA dataset included methylation data on 19302 probes measured in 93 subjects with COPD defined by GOLD spirometry criteria (GOLD I-IV), as well as 269 smoking controls. A quality control schematic is provided in Figure 1. Baseline statistics for the two groups showed expected differences in metrics used to define COPD severity including forced expiratory volume in 1 second as percent predicted (FEV1), forced vital capacity as percent predicted (FVC), and the ratio of FEV1 to FVC (FEV1/FVC), while pack-year history of smoking (PYH) was similar (Table 1).

Differential DNA Methylation Analysis: We used linear mixed models to identify COPD-associated CpG sites, using the method of Benjamini and Hochberg to control type I error. Five differentially methylated CpG sites were associated with COPD (Table 2) at FDR< 5%. All of these five CpG sites exhibited hypomethylation in association with the presence of COPD. The mean difference in percent methylation between cases and controls among the top five associated sites ranged from 5.3 to 9.6 %. The top differentially methylated CpG site was cg16361890 (unadjusted p-value 8.188×10^{-8}, percent methylation change -7.8%), which mapped to the *MAML1* gene. Gene annotation for the remaining four differentially methylated CpG sites associated with COPD included *RBFOX2*, *CD72*, *GRASP*, and *SH3TC1*.

Seven additional differentially methylated CpG sites were associated with COPD at FDR of less than 10%, (Table 2). These seven additional CpG sites (annotated to *FOLR3*, *ELOVL3*, *FAXDC2*, *FXYD1*/LGI4, *GABRR1*, *IFNGR2*, and *LPO*) also

5

exhibited relative hypomethylation in association with COPD. The mean difference in methylation ranged from 7.5 to 11.9% (see Supplemental Figures). These differential methylation results are presented in a volcano plot (Figure 2) and a Manhattan Plot (Figure 3).

Comparison to NHW Dataset: We examined the differentially methylated CpG sites from the PA-SCOPE AA cohort in the International COPD Genetics Network (ICGN) NHW cohort methylation data[19]. We qualitatively compared the magnitude of change in differential methylation between AA and NHW using the difference between the test-statistic for differential methylation at each CpG site (PA-SCOPE minus ICGN) from the mixed linear model output. We posited that a large difference in test statistic of differential methylation at a given CpG site between the PA-SCOPE and ICGN studies may indicate a difference in differential methylation between AA and NHW. This difference metric was found to have a normal distribution (mean= -0.316, sd= 1.515) across the ~19000 probes remaining after data cleaning. Seven of the twelve differentially methylated genes identified at FDR less than 10% in the primary PA-SCOPE analysis were found to have test-statistic-difference values in the lower 2.5 percentile tail of this distribution, suggesting potential race-specific differential methylation at these sites(see Supplemental Figures).

Weighted Gene Comethylation Network Analysis:

Weighted Gene Comethylation Network Analysis (WGCNA) was used to create a scale-free comethylation network[20] (see Supplemental Figures). The resultant network contained ten modules (see Figure 4), of which two were significantly correlated with COPD case-control status (labeled "blue" and "yellow" modules). The blue module contained 5009 probes (p-value 4×10^{-9} for module association with COPD status), and the yellow module contained 1698 probes (p-value 4×10^{-10} for module association with COPD status). The association of the blue and yellow module eigengenes remained statistically significant (p-value 1.6×10^{-4} for blue module, 5.5×10^{-5} for yellow module) in a logistic regression model of COPD affection status after controlling for age, gender, and pack-years of smoking history.

We investigated the two COPD-associated network modules for genes previously related to COPD. This investigation included genes from the differential methylation analysis results and genes found to be associated with COPD and lung

6

function measurements in genome-wide association studies (GWAS) (see Table 3). CpG sites marking 11 of the 12 top differential methylation sites found in our differential methylation analysis (*GRASP*, *FXYD1/LGI4*, *MAML1*, *FOLR3*, *CD72*, *LPO*, *GABRR1*, *SH3TC1*, *RBFOX2*, *IFNGR2*, *ELOVL3*) were part of the blue module. CpG sites marking the COPD- and lung function-associated genes *NOTCH4*, *SERPINA1*, *FAM13A*, *TNS1*, *PPT2*, *CHRNA5*, *PPAP2B*, *CHRNA3*, *RARB*, *CHRNA4*, *ARMC2*, *CCDC38*, *MECOM*, *ADAMTS19*, *HHIP*, and *ZKSCAN3/ZSCAN31* were also all found within the blue module. We limited both the blue and yellow modules to those genes with a stringent module membership (kME value) cutoff of 0.85 for further analysis[20], yielding a gene set of 317 members and 151 members, respectively.

Gene set enrichment analyses of the limited modules were performed using GO Biological Processes ontology, KEGG, and Reactome pathways. The limited blue module of 317 genes was enriched for developmental gene sets, with statistically significant enrichment found in GO pathways such as anatomical structure development (GO:0048856; adjusted p-value 3.6×10^{-10}), branching morphogenesis of an epithelial tube (GO:0048754; adjusted p-value 1.5×10^{-4}), lung morphogenesis (GO:0060425; adjusted p-value 0.01), lung lobe morphogenesis (GO:0060425; adjusted p-value 0.0246), lung development (GO:0030324; adjusted p-value), and lung lobe development (GO:0060428; p-value 1.38×10^{-3}). The limited yellow module of 151 genes was enriched for immune/inflammatory gene sets, such as response to other organism (GO:0051707; adjusted p-value 8.3×10^{-4}), immune response (GO:0006955; adjusted p-value 7.3×10^{-3}), defense response to bacterium (GO:0071219; adjusted p-value 4.2×10^{-3}), and chemotaxis of natural killer cells, neutrophils, and eosinophils (GO:0035747, 0030593, 0048245, respectively; adjusted p-values 3.5×10^{-2} to 7.2×10^{-3}).

Discussion

Within the PASCOPE AA cohort, we identified 5 differentially methylated CpG sites significantly associated with COPD using an FDR of 5%, and 7 additional associations that approached significance using an FDR of 10%. We used WGCNA to identify comethylation modules associated with COPD that were enriched for genes related to lung development and immune response and contained biologically relevant genes associated with COPD and lung function. Differentially methylated CpG sites associated with COPD mapped to genes that were biologically plausible candidates for COPD pathogenesis. Notable functions among these genes included *NOTCH4*-dependent lung angiogenesis, alveolar

macrophage response pathways, and airway defense mechanisms targeting bacteria, as detailed below. Our results

included genes and CpGs implicated in previous studies of obstructive lung disease and exacerbations, including *GRASP*

and *IFNGR2* (previous genetic associations with asthma) as well as *FYXD1* (differential methylation associated with

response to systemic steroids and to COPD[15, 21]). Only 1 of these 12 differentially methylated CpG sites (cg27461196,

mapped to *LGI4*/*FXYD1*) was statistically significantly associated with COPD at an FDR of 5% in an independent, larger

NHW COPD methylation dataset cleaned and processed in a comparable way. In addition, in a qualitative comparison of

the difference in differential methylation of CpG sites between AA and NHW, many of our results were preferentially

differentially methylated in AA.

The majority of our significantly associated (FDR <10%) differential methylation CpG sites are located in genes that are

biologically plausible genes for lung disease that may affect the pulmonary, immune, and vascular biology of COPD.

Many of these genes are expressed in either lung or whole blood based on GTEx data[22].

Two of these genes were associated with COPD in prior studies. Folate Receptor Gamma (FOLR3) was found to be 15- to

20-fold upregulated during stable COPD and acute exacerbations of COPD in previous studies[23], although the

mechanistic and functional implications of this upregulation are unclear. Differential methylation of Phospholemman

(FXYD1) was shown to be associated with COPD in the ICGN cohort by Qiu et al[15], and this gene was also previously

found to be differentially methylated in response to systemic steroid use in COPD[21]. Both of these genes are notably

related to acute exacerbations of COPD as well as a preferred treatment modality (systemic steroids) for acute

exacerbations. Given that the PASCOPE study was ascertained from subjects who were hospitalized for COPD, the

methylation pattern of these genes may be related to acute exacerbations of COPD.

Five CpG sites were annotated to genes related to pulmonary and airway physiology.   Lactoperoxidase (LPO) is secreted

by submucosal glands in human bronchi and plays a role in human airway host defense against bacteria[24]. Gamma-

aminobutyric acid Receptor1 (GABBR1) has been shown to affect alveolar fluid homeostasis in alveolar epithelial type II

cells[25].   Upregulated gene expression of Very Long Chain Fatty Acid Elongase3 (ELOVL3) has been proposed to

contribute to dysregulated lipid droplet formation in pulmonary surfactant in response to particulate exposure[26].

Rare missense mutations of GRP1-Associated Scaffold Protein (GRASP) were previously associated with asthma in a

Latino cohort[27].   The function of SH3 Domain and Tetratricopeptide Repeats 1 (SH3TC1) has not been adequately

described in the lung; however, it is implicated in networks related to bronchial airway epithelial cells and cigarette

smoking [28].

An additional three CpG sites were annotated to genes related to immune response and steroid synthesis. Cluster of

Differentiation 72 (CD72) is a CD5 co-ligand involved in hypersensitivity reactions and sarcoidosis, highly expressed in

pulmonary alveolar macrophages[29]. Fatty Acid Hydroxylase Domain Containing2 (FAXDC2) is implicated in "steroid

biosynthesis" through KEGG pathways[30]. Interferon Gamma Receptor 2 (IFNGR2) plays a role in activation of

macrophages and regulation of Th1 response to intracellular pathogens, with genetic variants previously associated with

atopic asthma[31] and pulmonary tuberculosis.

The final two CpG sites were related to cardiovascular processes. Mastermind-Like1 protein (MAML1) effects

angiogenesis during organ development through NOTCH-dependent signaling in murine lung[32]. RNA-binding Protein

Fox-1 Homolog 2 (RBFOX2) is a splicing regulator implicated in differentiation of myofibroblasts to skeletal muscle, and

diminished expression previously associated with pressure-overload-mediated progression of dilated

cardiomyopathy/heart failure[33]; potential impact on airway smooth muscle has not been described.

We present data showing that many of our top COPD-associated CpG sites lie in the lower tail of a histogram of the

difference in test-statistic between CpG sites in the African-American PA-SCOPE dataset and the non-Hispanic White

ICGN dataset (see Supplemental Data) . This finding may represent qualitative evidence that these sites are

preferentially differentially methylated in African-Americans and not in non-Hispanic Whites, although this conclusion

must be seen as hypothesis-generating only without a separate properly controlled and matched study design that

would be free of confounding by technical artifacts related to batch. The potential race-specificity of these differentially

methylated CpG sites could be explained by several scenarios.   The most mechanistically attractive possibility is that

these CpG sites represent differential methylation events in response to gene-environment interactions. The second mechanistic possibility is that these CpG sites represent blood methylation quantitative trait loci (mQTL) that are influenced by the genetic architecture specific to the population substructure[34] of African-Americans. In both of the preceding scenarios, the differential methylation could in turn impact damage and airflow obstruction through changes in gene expression and protein production, which could present unique targets for intervention. Finally, the differential methylation may simply be a marker of a confounder between methylation state and COPD, tagging a prior exposure that directly contributed to both disease and CpG methylation through distinct mechanisms.

WGCNA identifies modules of co-methylated genes starting from the level of thousands of CpG probes and correlates these modules to phenotypic variables. The network creation and module-building processes of WGCNA are informed purely by gene methylation levels, without consideration of case-control status for COPD. Individual genes within the module can then be related to the module eigengenes by measures of module membership and gene significance to the module. This technique identifies driver genes for the module that may help identify biologically meaningful pathways. In our dataset the yellow and blue modules showed significant association with COPD. Yellow module measures of gene significance were predominantly positive (indicating relative hypermethylation in the presence of COPD) while the blue module contained primarily negative measures of gene significance (indicating relative hypomethylation in the presence of COPD).

Further investigation of the blue module showed a network with biological significance for obstructive lung disease. The module was statistically enriched for pathways related to lung development, and also contained multiple genes previously associated with COPD and lung function. *SERPINA1* is the gene responsible for alpha-1-antitrypsin deficiency[35], a known genetic cause of COPD, and this gene was found to be highly significant in the blue module. The blue module also contained multiple genes previously associated with COPD or lung function measurements through GWAS. Many of the CpG sites found in the differential methylation analysis were also found in the blue module with high measures of module membership (indicating importance of the gene to the module) and large measures of gene

significance to COPD. The recapitulation of these CpG sites in the same module as previously known COPD- and lung-function-related genes adds validation to our differential methylation results.

The yellow module, by comparison, contained genes enriched for immune response pathways. Chronic inflammation in response to airway damage from cigarette smoking as well as external pathogens are recognized as integral parts of the pathogenesis of COPD and exacerbations [36-38]. Enrichment for the chemotaxis of effector cells that are known to play a role in COPD pathogenesis (neutrophils[39], eosinophils[40, 41], and natural killer cells[42]) were found using yellow module genes with high module membership values. The PA-SCOPE population was ascertained using subjects with disease exacerbations, so this population may have been enriched for signals associated with acute inflammation and immune response[43, 44].

The PA-SCOPE dataset was a retrospective case-control study and so no direct causation can be implied from results, only associations of CpG sites with disease. DNA methylation in response to smoking is a dynamic process, and effects may be dependent upon recent or remote smoking practices[16, 45]. The methylation dataset for PA-SCOPE did not contain data related to current smoking or time since quitting smoking, and could not assess the effects that these might have on our differential methylation results. Longitudinal data was not available in PA-SCOPE, so further conclusions integrating clinical stability, clinical progression, or other lung function trajectories[46] associated with CpG sites cannot be made using these data. Without paired gene expression data, it is unclear what effect these differentially methylated sites have on expression of the associated gene products. While both PA-SCOPE and ICGN were studies of COPD subjects and smoking controls, differences in ascertainment of the datasets may influence the conclusions. Similarly, the comparison of test-statistic differences between ICGN and PA-SCOPE could be influenced by factors other than racial disparities in differential methylation related to COPD. Batch effects between the PA-SCOPE and ICGN assays, differences in ascertainment and study design, and baseline differences in the two populations other than racial make-up could account for the difference in test statistics among these populations, so we present these data points as qualitative and hypothesis-generating. The Illumina Infinium HumanMethylation27 BeadChip Array interrogates only a subset of CpG sites in the human epigenome, and additional unmeasured sites may be differentially methylated in

association with COPD. Future studies including large populations of both AA and NHW would be needed to further validate both the differential methylation results as well as the race-specificity of the results. The recapitulation of many of our differentially methylated genes in network modules strongly associated with COPD provides some biological validation of the importance of these sites to COPD using a different methodology. While further research is needed to investigate the biological consequences of differential methylation of these genes in experimental models, they represent promising candidate genes for mechanistic investigations of the racial disparity in COPD susceptibility and severity.

Conclusions

In conclusion, we performed differential methylation analysis in African-American subjects and identified 12 CpG sites statistically significantly associated with COPD at an FDR less than 10%, of which seven appear more relevant for the AA population compared to a similar NHW population. We also performed weighted gene comethylation network analysis and identified two comethylation modules associated with COPD, one of which included multiple genes related to obstructive lung disease and COPD. This module was enriched for lung-specific gene sets and our results add to insights into molecular mechanisms that may partially explain lung disease disparities in African-Americans. These race-specific DNA methylation marks could reflect differences in exposures among AA and NHW that may influence disease mechanisms that affect COPD susceptibility. Molecular mechanisms for racial disparities in COPD-related outcomes have not been adequately described from a genetic and mechanistic standpoint, and this investigation is a step towards better describing and finding treatments for the biological mechanisms underlying these disparities.

Methods

Subjects and Data Collection: Researchers at Temple University and the Pennsylvania Department of Public Health designed the Pennsylvania Study of Chronic Obstructive Pulmonary Exacerbations (PA-SCOPE, ClinicalTrials.gov Identifier: NCT00774176) as a collaborative observational study to identify demographic and genetic factors that

contributed to COPD exacerbations among AA smokers with COPD in urban and rural Pennsylvania. Subject recruitment and data gathering occurred between June 2004 and May 2008. PA-SCOPE data types for each of 371 AA subjects in the methylation cohort consisted of questionnaire data, spirometry and pulmonary function tests, and DNA from peripheral blood samples. Subjects were male and female AA smokers (>20 pack-years) with and without COPD (defined by FEV1/FVC $\leq$ 70% or FEV1 $\leq$ 70% predicted) ages 40-80 years. Race was determined by self-report. We excluded subjects with asthma or chronic respiratory diseases other than COPD. Participants provided written consent to participate in this study, and the study was approved by the institutional review boards at all participating institutions (Partners IRB: 2005P000453/BWH).

We assessed DNA samples from 371 COPD cases and smoking controls for genome wide differential methylation using the Illumina (San Diego, CA) Infinium HumanMethylation27 BeadChip (Illumina27K). The Illumina27K array assays 27,758 CpG dinucleotides[47] for quantitative measurements of DNA methylation, covering over 14,000 genes. We performed the assay using the manufacturer's suggested protocol, including standard controls for bisulfite conversion, amplification, hybridization, and extension. We report the percent methylation values ranging from 0 to 100% (corresponding to beta value of 0 to 1), calculated as the ratio of the fluorescent intensity of the methylated bead type (meth) to the combined locus intensity of methylated and unmethylated bead types (meth + unmeth) plus an offset (beta = meth/(meth+unmeth + 100)). The log2-ratio of the methylated to unmethylated intensities (M-value) was used for association testing[48]. The absolute difference in percent methylation between cases and controls was used to quantify the effect size.

Quality Control: The R programming language[49] and the BioConductor[50] suite of software (packages included methylumi[51], GenomicRanges[52], and wateRmelon[53]) were used for data annotation, probe quality control and pruning , and subject-level quality control. The Illumina27K probes were annotated to their nearest gene using hg19 coordinates and investigated for potential sources of probe bias. To eliminate bias due to SNPs underlying probe regions, probes containing a CpG within 5 base pairs upstream or downstream of a known genomic SNP were eliminated. Probes underlying genomic repeat regions were eliminated. Probes interrogating the sex chromosomes

13

were also eliminated. Subject level quality control eliminated one sample having 1% of sites with a detection p-value greater than 0.05. CpG sites with a beadcount of less than 3 in greater than 5% of samples, and sites having greater than 1% of samples with a detection p-value greater than 0.05 were also removed; 8 subjects' data were incomplete and were removed for the final analysis.

Cell Type Deconvolution: Differential methylation signals arising primarily from cell type composition of whole blood can bias methylation analysis. To control for this bias, we performed cell type deconvolution using the method and software provided by Houseman et al[54].

Differential DNA Methylation Analysis: Differential methylation analysis was performed on 362 samples after quality control using the limma package[55]. For the COPD analysis, associations between differentially methylated probes and COPD case-control status were modeled in a logistic mixed model controlling for age, sex, pack-years of smoking, batch number, and cell type deconvolution (covariates accounting for natural killer cells, CD8 T-cells, CD4 T-cells, B-cells, and monocytes). Family-wise type I error was controlled using the method of Benjamini-Hochberg to achieve a genome-wide false discovery rate (FDR) threshold of significance of less than 5%. Additionally, we examined the results that approached statistical significance at less than FDR 10% level for assessment of additional biologically-plausible targets. The absolute difference in the mean percent methylation value was used to quantify the magnitude and direction of effect for differentially methylated CpG sites; positive values of delta beta correspond to relative hypermethylation among the cases.

Comparison to NHW Dataset: An analysis of differential DNA methylation sites associated with COPD was performed previously by our group using NHW subjects in the International COPD Genetic Network (ICGN)[15]. Whole blood samples were assayed on the Illumina27k array. This dataset included 692 COPD NHW cases by GOLD spirometry criteria, as well as 437 NHW controls. We passed these data through annotation, quality control, probe control, cell type deconvolution, and differential DNA methylation analysis steps that were identical to those used for the PA-SCOPE dataset. We hypothesized that race-specific disparities in differential methylation of genes may reflect biological

disparities in disease mechanisms or disease-relevant exposures. We used this ICGN NHW dataset to compare

differential methylation values to those found in the PA-SCOPE AA cohort. We compared the differential methylation

characteristics of CpG sites between AAs and NHWs qualitatively in order to identify sites with strong differential

methylation in AA without corresponding differential methylation in NHW. In order to identify those sites at which the

difference in differential methylation was qualitatively greatest between AAs and NHWs, we calculated the difference in

test statistic (retaining direction of effect) for each CpG site between the PA-SCOPE AA analysis and the ICGN NHW

analysis.

Weighted Gene Comethylation Network Analysis: Weighted gene comethylation networks were constructed using the

quality controlled PA-SCOPE dataset as input to the WGCNA R-package by Langfelder and Horvath[56], and network

modules were analyzed for association with COPD in a signed correlation network. Scale-free properties were achieved

with a soft thresholding value of 12, resulting in ten modules. Analysis of the eigengenes of significantly COPD-

associated modules (measured as the correlation of the gene's methylation profile with the module eigenvector) was

performed to evaluate driver genes for the module (genes whose methylation is most highly correlated to the eigengene

methylation). Gene significance is quantified statistically by the Student's t-test statistic for differential methylation

between COPD cases and controls, and larger values of gene significance indicate more significant p-values for the

gene's significance to COPD. Results were inspected for genes in the differential methylation top results, as well as for

known obstructive lung disease associations from GWAS studies of lung function[57-59] and COPD[60-65]. For each

significantly associated eigengene, logistic regression models of COPD case-control status were constructed using the

module eigengenes and clinical covariates of age, gender, and pack-years of smoking history to evaluate the robustness

of the eigengene association to COPD. COPD-associated WGCNA modules were limited to genes with a stringent

module membership (kME value) cutoff of >0.85, as per previously published methods[20]. Gene set enrichment

analysis on these limited modules was performed using ConsensusPathDB[66] to compare the COPD-associated module

genes to evaluate for enrichment in the Gene Ontology (GO) Biological Processes ontology, Kyoto Encyclopedia of Genes

and Genomes (KEGG), and Reactome pathways using hypergeometric testing controlled for multiple testing using the

false discovery rate method applied to the number of included genes.

Declarations

List of Abbreviations:

AA - African-American

COPD - Chronic obstructive pulmonary disease

CpG - Cytosine phosphate-bond guanine dinucleotide

FEV1 - Forced expiratory volume in one second

FVC - Forced vital capacity

GO - Gene Ontology

ICGN - International COPD Genetics Network

KEGG - Kyoto Encyclopedia of Genes and Genomes

mQTL - Methylation quantitative trait loci

NHW - non-Hispanic White

PA-SCOPE - Pennsylvania Study of Chronic Obstructive Pulmonary Exacerbations

PYH - Pack-year history of smoking

WGCNA - Weighted gene comethylation network analysis

Ethics Approval and Consent to Participate:

Participants provided written consent to participate in this study, and the study was approved by the institutional review

boards at all participating institutions (Partners IRB: 2005P000453/BWH).

Consent for Publication:

Not Applicable

Availability of Data and Material:

The PA-SCOPE dataset is not currently available on dbGaP or related public data institutions.

Authors' Contributions:

All of the authors listed have contributed sufficiently to the project to be included as authors by ICMJE guidelines, and all those who are qualified to be authors are listed in the author byline. We have included statements on funding sources and acknowledgements of the study teams for each included study. The authors have approved the manuscript and its submission, and agreed to be accountable for the content of the work.

Mondo Medico di I.F.I.M. srl, Multidisciplinary and Rehabilitation Outpatient Clinic, Borgomanero (NO), Italy; Robert D. Levy, University of British Columbia, Vancouver, Canada; Peter D. Paré, University of British Columbia, Vancouver, Canada; Stephen Rennard, Section of Pulmonary & Critical Care, University of Nebraska Medical Center, Omaha, NE, USA; Jørgen Vestbo, University of Manchester, UK.

Prior Publication Declaration:

This manuscript is original, has not been published previously, and is not under consideration at any other journal. By BMC editorial policies, limited findings from this paper have been published as abstracts/posters at the ATS and ASHG academic meetings:

Busch R, Qiu W, Criner G, DeMeo DL. Race-Specific Differential DNA Methylation Marks of COPD; Abstract #444/F. Presented as a Poster at the 65th Annual Meeting of The American Society of Human Genetics Annual Meeting, October 9, 2015, Baltimore, MD.

Busch R, Qiu W, Lasky-Su J, Morrow J, Criner G, DeMeo DL. DNA Co-Methylation Modules Associate with COPD in African Americans; Abstract #A2925. Accepted for Poster Discussion Session at The American Thoracic Society International Conference, May 16, 2016, San Francisco, CA.

References:

1. Kochanek KD, Murphy SL, Xu J: **Deaths: Final Data for 2011**. *Natl Vital Stat Rep* 2015, **63**(3):1-120.
2. Vestbo J, Hurd SS, Agusti AG, Jones PW, Vogelmeier C, Anzueto A, Barnes PJ, Fabbri LM, Martinez FJ, Nishimura M *et al*: **Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary**. *Am J Respir Crit Care Med* 2013, **187**(4):347-365.
3. **Trends in COPD (Chronic Bronchitis and Emphysema): Morbidity and Mortality** [http://www.lung.org/assets/documents/research/copd-trend-report.pdf]
4. Dransfield MT, Bailey WC: **COPD: racial disparities in susceptibility, treatment, and outcomes**. *Clin Chest Med* 2006, **27**(3):463-471, vii.
5. Foreman MG, Zhang L, Murphy J, Hansel NN, Make B, Hokanson JE, Washko G, Regan EA, Crapo JD, Silverman EK *et al*: **Early-onset chronic obstructive pulmonary disease is associated with female sex, maternal factors, and African American race in the COPDGene Study**. *Am J Respir Crit Care Med* 2011, **184**(4):414-420.
6. **State of Lung Disease in Diverse Communities: 2010** [http://action.lung.org/site/DocServer/state-of-lung-disease-in-diverse-communities-2010.pdf]
7. Han MK, Curran-Everett D, Dransfield MT, Criner GJ, Zhang L, Murphy JR, Hansel NN, DeMeo DL, Hanania NA, Regan EA *et al*: **Racial differences in quality of life in patients with COPD**. *Chest* 2011, **140**(5):1169-1176.
8. Manichaikul A, Hoffman EA, Smolonska J, Gao W, Cho MH, Baumhauer H, Budoff M, Austin JH, Washko GR, Carr JJ *et al*: **Genome-wide study of percent emphysema on computed tomography in the general population. The Multi-Ethnic Study of Atherosclerosis Lung/SNP Health Association Resource Study**. *Am J Respir Crit Care Med* 2014, **189**(4):408-418.
9. Sun YV, Smith AK, Conneely KN, Chang Q, Li W, Lazarus A, Smith JA, Almli LM, Binder EB, Klengel T *et al*: **Epigenomic association analysis identifies smoking-related DNA methylation sites in African Americans**. *Hum Genet* 2013, **132**(9):1027-1037.
10. Fraser HB, Lam LL, Neumann SM, Kobor MS: **Population-specificity of human DNA methylation**. *Genome Biol* 2012, **13**(2):R8.
11. Barfield RT, Almli LM, Kilaru V, Smith AK, Mercer KB, Duncan R, Klengel T, Mehta D, Binder EB, Epstein MP *et al*: **Accounting for population stratification in DNA methylation studies**. *Genet Epidemiol* 2014, **38**(3):231-241.
12. Bjornsson HT, Fallin MD, Feinberg AP: **An integrated epigenetic and genetic approach to common human disease**. *Trends Genet* 2004, **20**(8):350-358.
13. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H: **Tobacco-smoking-related differential DNA methylation: 27K discovery and replication**. *Am J Hum Genet* 2011, **88**(4):450-457.
14. Sakao S, Tatsumi K: **The importance of epigenetics in the development of chronic obstructive pulmonary disease**. *Respirology* 2011, **16**(7):1056-1063.
15. Qiu W, Baccarelli A, Carey VJ, Boutaoui N, Bacherman H, Klanderman B, Rennard S, Agusti A, Anderson W, Lomas DA *et al*: **Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function**. *Am J Respir Crit Care Med* 2012, **185**(4):373-381.
16. Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, Agusti A, Anderson W, Lomas DA, Demeo DL: **Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome**. *Hum Mol Genet* 2012, **21**(13):3073-3082.
17. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A *et al*: **Tobacco smoking leads to extensive genome-wide changes in DNA methylation**. *PLoS One* 2013, **8**(5):e63812.
18. Saccone NL, Wang JC, Breslau N, Johnson EO, Hatsukami D, Saccone SF, Grucza RA, Sun L, Duan W, Budde J *et al*: **The CHRNA5-CHRNA3-CHRNB4 nicotinic receptor subunit gene cluster affects risk for nicotine dependence in African-Americans and in European-Americans**. *Cancer Res* 2009, **69**(17):6848-6856.
19. Zhu G, Warren L, Aponte J, Gulsvik A, Bakke P, Anderson WH, Lomas DA, Silverman EK, Pillai SG: **The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two large populations**. *Am J Respir Crit Care Med* 2007, **176**(2):167-173.
20. Rickabaugh TM, Baxter RM, Sehl M, Sinsheimer JS, Hultin PM, Hultin LE, Quach A, Martinez-Maza O, Horvath S, Vilain E *et al*: **Acceleration of age-associated methylation patterns in HIV-1-infected adults**. *PLoS One* 2015, **10**(3):e0119201.

21. Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, Agusti A, Anderson WH, Lomas DA, DeMeo DL: **Systemic steroid exposure is associated with differential methylation in chronic obstructive pulmonary disease**. *Am J Respir Crit Care Med* 2012, **186**(12):1248-1255.

22. **Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans**. *Science* 2015, **348**(6235):648-660.

23. Wu X, Sun X, Chen C, Bai C, Wang X: **Dynamic gene expressions of peripheral blood mononuclear cells in patients with acute exacerbation of chronic obstructive pulmonary disease: a preliminary study**. *Crit Care* 2014, **18**(6):508.

24. Wijkstrom-Frei C, El-Chemaly S, Ali-Rachedi R, Gerson C, Cobas MA, Forteza R, Salathe M, Conner GE: **Lactoperoxidase and human airway host defense**. *Am J Respir Cell Mol Biol* 2003, **29**(2):206-212.

25. Jin N, Kolliputi N, Gou D, Weng T, Liu L: **A novel function of ionotropic gamma-aminobutyric acid receptors involving alveolar fluid homeostasis**. *J Biol Chem* 2006, **281**(47):36012-36020.

26. Rowan-Carroll A, Halappanavar S, Williams A, Somers CM, Yauk CL: **Mice exposed in situ to urban air pollution exhibit pulmonary alterations in gene expression in the lipid droplet synthesis pathways**. *Environ Mol Mutagen* 2013, **54**(4):240-249.

27. Igartua C, Myers RA, Mathias RA, Pino-Yanes M, Eng C, Graves PE, Levin AM, Del-Rio-Navarro BE, Jackson DJ, Livne OE *et al*: **Ethnic-specific associations of rare and low-frequency DNA sequence variants with asthma**. *Nat Commun* 2015, **6**:5965.

28. **TCNG: The Cancer Network Galaxy** [http://tcng.hgc.jp/index.html]

29. Baughman RP, Culver DA: **Sarcoidosis**. *Clin Chest Med* 2015, **36**(4):xv.

30. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes**. *Nucleic Acids Res* 2000, **28**(1):27-30.

31. Kumar A, Das S, Agrawal A, Mukhopadhyay I, Ghosh B: **Genetic association of key Th1/Th2 pathway candidate genes, IRF2, IL6, IFNGR2, STAT4 and IL4RA, with atopic asthma in the Indian population**. *J Hum Genet* 2015, **60**(8):443-448.

32. Li X, Zhang X, Leathers R, Makino A, Huang C, Parsa P, Macias J, Yuan JX, Jamieson SW, Thistlethwaite PA: **Notch3 signaling promotes the development of pulmonary arterial hypertension**. *Nat Med* 2009, **15**(11):1289-1297.

33. Wei C, Qiu J, Zhou Y, Xue Y, Hu J, Ouyang K, Banerjee I, Zhang C, Chen B, Li H *et al*: **Repression of the Central Splicing Regulator RBFox2 Is Functionally Linked to Pressure Overload-Induced Heart Failure**. *Cell Rep* 2015.

34. Smith AK, Kilaru V, Kocak M, Almli LM, Mercer KB, Ressler KJ, Tylavsky FA, Conneely KN: **Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type**. *BMC Genomics* 2014, **15**:145.

35. Carrell RW, Jeppsson JO, Laurell CB, Brennan SO, Owen MC, Vaughan L, Boswell DR: **Structure and variation of human alpha 1-antitrypsin**. *Nature* 1982, **298**(5872):329-334.

36. Bhat TA, Panzica L, Kalathil SG, Thanavala Y: **Immune Dysfunction in Patients with Chronic Obstructive Pulmonary Disease**. *Ann Am Thorac Soc* 2015, **12 Suppl 2**:S169-175.

37. Retamales I, Elliott WM, Meshi B, Coxson HO, Pare PD, Sciurba FC, Rogers RM, Hayashi S, Hogg JC: **Amplification of inflammation in emphysema and its association with latent adenoviral infection**. *Am J Respir Crit Care Med* 2001, **164**(3):469-473.

38. Lommatzsch M, Bratke K, Knappe T, Bier A, Dreschler K, Kuepper M, Stoll P, Julius P, Virchow JC: **Acute effects of tobacco smoke on human airway dendritic cells in vivo**. *Eur Respir J* 2010, **35**(5):1130-1136.

39. Hoenderdos K, Condliffe A: **The neutrophil in chronic obstructive pulmonary disease**. *Am J Respir Cell Mol Biol* 2013, **48**(5):531-539.

40. Brightling CE, Monteiro W, Ward R, Parker D, Morgan MD, Wardlaw AJ, Pavord ID: **Sputum eosinophilia and short-term response to prednisolone in chronic obstructive pulmonary disease: a randomised controlled trial**. *Lancet* 2000, **356**(9240):1480-1485.

41. Brightling CE, McKenna S, Hargadon B, Birring S, Green R, Siva R, Berry M, Parker D, Monteiro W, Pavord ID *et al*: **Sputum eosinophilia and the short term response to inhaled mometasone in chronic obstructive pulmonary disease**. *Thorax* 2005, **60**(3):193-198.

42. Culley FJ: **Natural killer cells in infection and inflammation of the lung**. *Immunology* 2009, **128**(2):151-163.

43.   Wedzicha JA: **Mechanisms of Chronic Obstructive Pulmonary Disease Exacerbations**. *Ann Am Thorac Soc* 2015, **12 Suppl 2**:S157-159.
44.   Papi A, Bellettato CM, Braccioni F, Romagnoli M, Casolari P, Caramori G, Fabbri LM, Johnston SL: **Infections and airway inflammation in chronic obstructive pulmonary disease severe exacerbations**. *Am J Respir Crit Care Med* 2006, **173**(10):1114-1121.
45.   Guida F, Sandanger TM, Castagne R, Campanella G, Polidoro S, Palli D, Krogh V, Tumino R, Sacerdote C, Panico S *et al*: **Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation**. *Hum Mol Genet* 2015, **24**(8):2349-2359.
46.   Lange P, Celli B, Agusti A, Boje Jensen G, Divo M, Faner R, Guerra S, Marott JL, Martinez FD, Martinez-Camblor P *et al*: **Lung-Function Trajectories Leading to Chronic Obstructive Pulmonary Disease**. *N Engl J Med* 2015, **373**(2):111-122.
47.   Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL: **Genome-wide DNA methylation profiling using Infinium(R) assay**. *Epigenomics* 2009, **1**(1):177-200.
48.   Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM: **Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis**. *BMC Bioinformatics* 2010, **11**:587.
49.   Team RC: **R: A Language and Environment for Statistical Computing**. In*.* Vienna, Austria; 2015.
50.   Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T *et al*: **Orchestrating high-throughput genomic analysis with Bioconductor**. *Nat Methods* 2015, **12**(2):115-121.
51.   Pan Du SD, Sven Bilke, Tim Triche, Jr., Moiz Bootwalla: **methylumi: Handle Illumina methylation data**. In*.*, 2.16 edn; 2014.
52.   Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ: **Software for computing and annotating genomic ranges**. *PLoS Comput Biol* 2013, **9**(8):e1003118.
53.   Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC: **A data-driven approach to preprocessing Illumina 450K methylation array data**. *BMC Genomics* 2013, **14**:293.
54.   Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT: **DNA methylation arrays as surrogate measures of cell mixture distribution**. *BMC Bioinformatics* 2012, **13**:86.
55.   Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies**. *Nucleic Acids Res* 2015, **43**(7):e47.
56.   Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis**. *BMC Bioinformatics* 2008, **9**:559.
57.   Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, Zhao JH, Ramasamy A, Zhai G, Vitart V *et al*: **Genome-wide association study identifies five loci associated with lung function**. *Nat Genet* 2010, **42**(1):36-44.
58.   Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, Zhai G, Zhao JH, Smith AV, Huffman JE *et al*: **Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function**. *Nat Genet* 2011, **43**(11):1082-1090.
59.   Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, Franceschini N, van Durme YM, Chen TH, Barr RG *et al*: **Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function**. *Nat Genet* 2010, **42**(1):45-52.
60.   Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Feng S, Hersh CP, Bakke P, Gulsvik A *et al*: **A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci**. *PLoS Genet* 2009, **5**(3):e1000421.
61.   Celedon JC, Lange C, Raby BA, Litonjua AA, Palmer LJ, DeMeo DL, Reilly JJ, Kwiatkowski DJ, Chapman HA, Laird N *et al*: **The transforming growth factor-beta1 (TGFB1) gene is associated with chronic obstructive pulmonary disease (COPD)**. *Hum Mol Genet* 2004, **13**(15):1649-1656.
62.   DeMeo DL, Mariani T, Bhattacharya S, Srisuma S, Lange C, Litonjua A, Bueno R, Pillai SG, Lomas DA, Sparrow D *et al*: **Integration of genomic and genetic approaches implicates IREB2 as a COPD susceptibility gene**. *Am J Hum Genet* 2009, **85**(4):493-502.
63.   Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, DeMeo DL, Hunninghake GM, Litonjua AA, Sparrow D *et al*: **Variants in FAM13A are associated with chronic obstructive pulmonary disease**. *Nat Genet* 2010, **42**(3):200-202.

64. Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, Demeo DL, Himes BE, Sylvia JS, Klanderman BJ, Ziniti JP *et al*: **A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13**. *Hum Mol Genet* 2012, **21**(4):947-957.

65. Cho MH, McDonald ML, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, Demeo DL, Sylvia JS, Ziniti J, Laird NM *et al*: **Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis**. *Lancet Respir Med* 2014, **2**(3):214-225.

66. Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H, Herwig R: **ConsensusPathDB: toward a more complete picture of cell biology**. *Nucleic Acids Res* 2011, **39**(Database issue):D712-717.
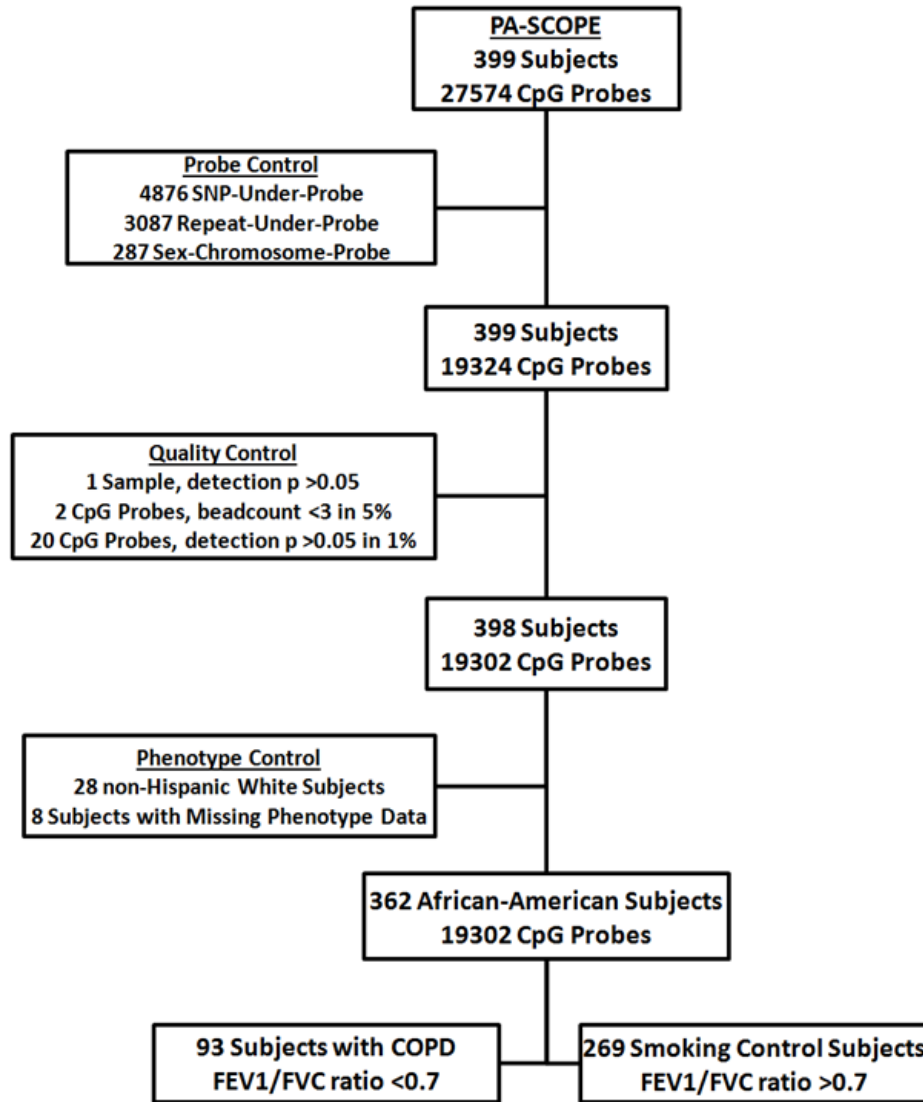
Figure 1: Subject- and Probe-Level Quality Control Chart. Quality control of the PA-SCOPE methylation dataset included Probe-level controls and Subject-level quality controls (see Methods for details). Final analysis included 93 COPD cases and 269 smoking controls.
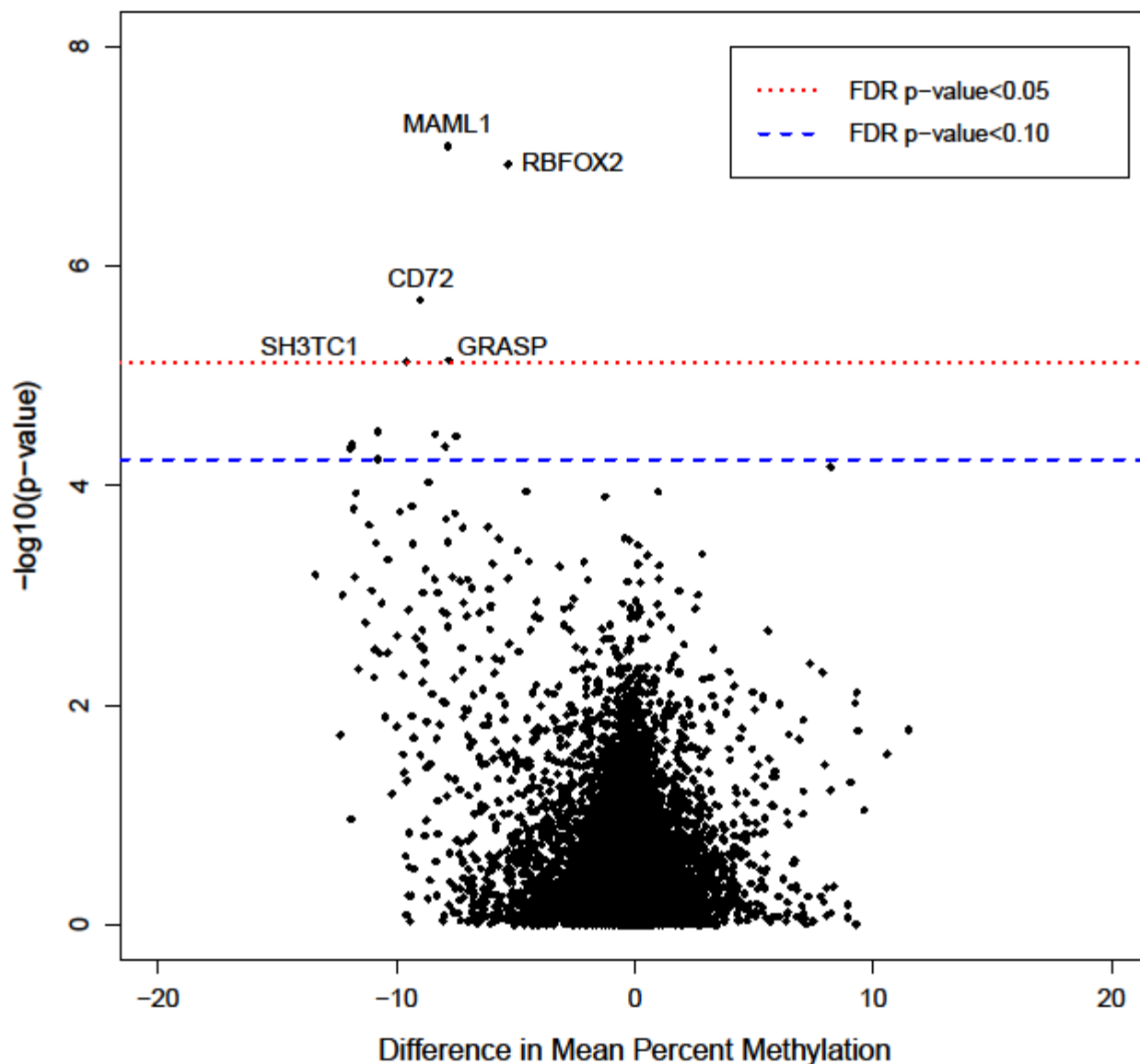
Figure 2: Differentially Methylated CpG Probes Associated with COPD. Differential methylation analysis revealed 12 CpG sites in 12 genes significantly associated with COPD with an FDR-corrected p-value less than 0.10. Difference in mean percent methylation represents the difference in mean methylation between COPD cases and smoking controls. The y-axis represents the negative log of the association p-value from linear mixed models adjusted for age, pack years of smoking, assay batch, and cell type. The name of the nearest gene is included with each of the top five CpG results.
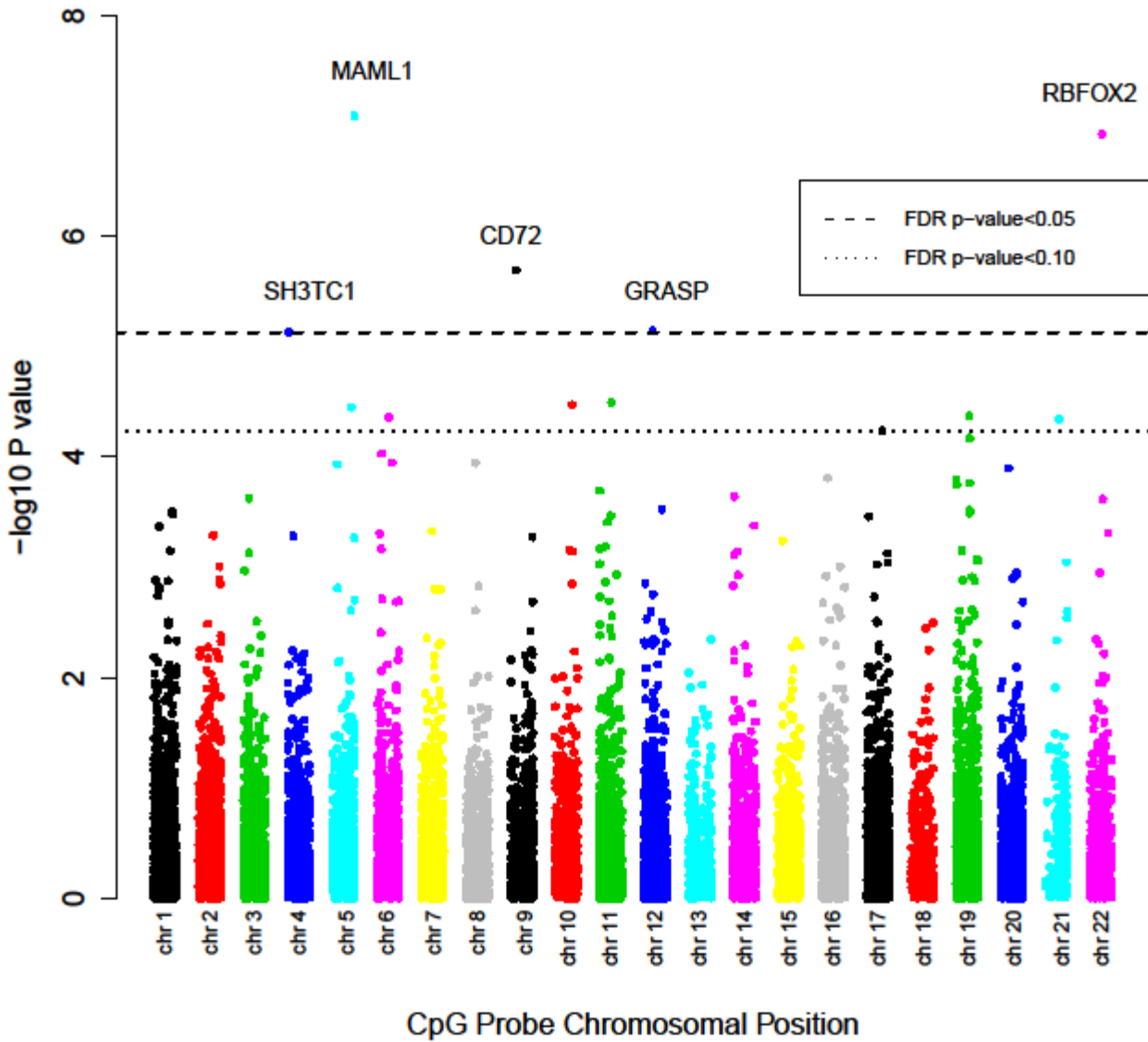
Figure 3: Manhattan Plot of Differential Methylation Analysis Results. Differential methylation analysis results presented by chromosomal location (x-axis). The y-axis represents the negative log of the association p-value from linear mixed models adjusted for age, pack years of smoking, assay batch, and cell type. The name of the nearest gene is included with each of the top five CpG results**.**
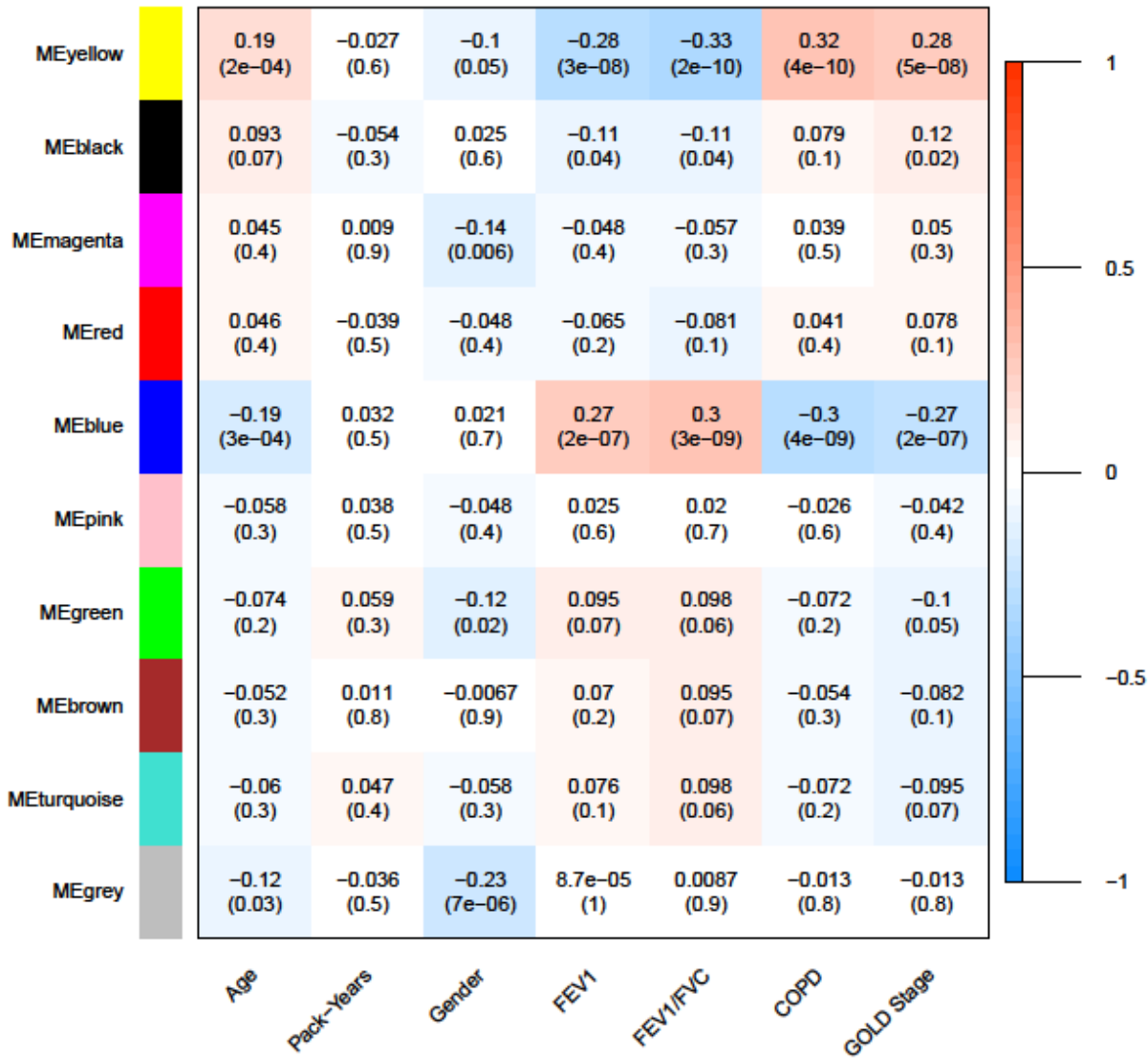
Module-trait Heatmap

	Age	Pack-Years	Gender	FEV1	FEV1/FVC	COPD	GOLD Stage
MEyellow	0.19 (2e−04)	−0.027 (0.6)	−0.1 (0.05)	−0.28 (3e−08)	−0.33 (2e−10)	0.32 (4e−10)	0.28 (5e−08)
MEblack	0.093 (0.07)	−0.054 (0.3)	0.025 (0.6)	−0.11 (0.04)	−0.11 (0.04)	0.079 (0.1)	0.12 (0.02)
MEmagenta	0.045 (0.4)	0.009 (0.9)	−0.14 (0.006)	−0.048 (0.4)	−0.057 (0.3)	0.039 (0.5)	0.05 (0.3)
MEred	0.046 (0.4)	−0.039 (0.5)	−0.048 (0.4)	−0.065 (0.2)	−0.081 (0.1)	0.041 (0.4)	0.078 (0.1)
MEblue	−0.19 (3e−04)	0.032 (0.5)	0.021 (0.7)	0.27 (2e−07)	0.3 (3e−09)	−0.3 (4e−09)	−0.27 (2e−07)
MEpink	−0.058 (0.3)	0.038 (0.5)	−0.048 (0.4)	0.025 (0.6)	0.02 (0.7)	−0.026 (0.6)	−0.042 (0.4)
MEgreen	−0.074 (0.2)	0.059 (0.3)	−0.12 (0.02)	0.095 (0.07)	0.098 (0.06)	−0.072 (0.2)	−0.1 (0.05)
MEbrown	−0.052 (0.3)	0.011 (0.8)	−0.0067 (0.9)	0.07 (0.2)	0.095 (0.07)	−0.054 (0.3)	−0.082 (0.1)
MEturquoise	−0.06 (0.3)	0.047 (0.4)	−0.058 (0.3)	0.076 (0.1)	0.098 (0.06)	−0.072 (0.2)	−0.095 (0.07)
MEgrey	−0.12 (0.03)	−0.036 (0.5)	−0.23 (7e−06)	8.7e−05 (1)	0.0087 (0.9)	−0.013 (0.8)	−0.013 (0.8)

Figure 4: WCGNA Module Trait Relationship Heatmap: Heatmap showing comethylation module correlation with phenotypic trait and associated p-value for these correlations within PA-SCOPE. Positive or negative correlation magnitude with COPD is presented with p-value for the correlation with COPD in parenthesis. The yellow and blue modules were both significantly associated with COPD affection status, labeled "COPD".

	Cases		Controls	
N=	93		269	
Age (years)	61.4	± 7.7	48.9	± 6.4
FEV1/FVC Ratio	0.434	± 0.123	0.799	± 0.053
FEV1 percent predicted	36.8	± 15.4	89.9	± 12.5
FVC percent predicted	67.5	± 22.4	90.7	± 12.3
Gender (% male)	50	(53.8)	195	(72.5)
Pack-Years of Smoking	42.2	± 22.1	39.2	± 20.5

Table 1. Baseline Statistics Among African-Americans in PA-SCOPE. Data is presented as count with proportion in parentheses or mean with standard deviation. FEV1 is the forced expiratory volume in one second, and FVC is the forced vital capacity.

Chromosome	CpG Probe	Nearest Gene	Mean Change in Percent Methylation	p-value	FDR Adjusted p-value
5	cg16361890	*MAML1*	-7.8	8.19E-08	0.001
22	cg00615377	*RBFOX2*	-5.3	1.20E-07	0.001
9	cg12971694	*CD72*	-9.0	2.05E-06	0.013
12	cg22566906	*GRASP*	-7.8	7.27E-06	0.029
4	cg02635407	*SH3TC1*	-9.6	7.47E-06	0.029
11	cg25634666	*FOLR3*	-10.8	3.23E-05	0.080
10	cg18390025	*ELOVL3*	-8.4	3.40E-05	0.080
5	cg10257049	*FAXDC2*	-7.5	3.58E-05	0.080
19	cg27461196	*FXYD1/LGI4*	-11.9	4.27E-05	0.080
6	cg00333528	*GABRR1*	-8.0	4.43E-05	0.080
21	cg17356733	*IFNGR2*	-11.9	4.56E-05	0.080
17	cg24489015	*LPO*	-10.8	5.82E-05	0.094

Table 2. Differentially Methylated CpG Probes Associated with COPD Among African-Americans in PA-SCOPE. Data are presented in decreasing order of significance. Mean change in percent methylation represents the mean difference in percent methylation between cases and controls; p-values were adjusted for multiple comparisons using the method of Benjamini and Hochberg.

CpG site	Nearest Gene	Gene Significance to COPD	Gene Significance p-value	Module Membership to Blue Module	Module Membership p-value
cg22566906	GRASP	-0.52	7.06E-27	0.76	8.62E-71
cg27461196	FXYD1/LGI4	-0.52	1.19E-26	0.76	7.10E-70
cg16361890	MAML1	-0.52	1.98E-26	0.83	1.28E-93
cg14700707	NOTCH4	-0.52	2.61E-26	0.72	1.52E-59
cg25634666	FOLR3	-0.51	1.82E-25	0.78	3.50E-76
cg12971694	CD72	-0.51	2.11E-25	0.84	7.77E-99
cg24489015	LPO	-0.50	7.21E-25	0.79	2.57E-78
cg00333528	GABRR1	-0.50	7.29E-25	0.83	5.25E-94
cg02635407	SH3TC1	-0.50	7.93E-25	0.85	1.02E-102
cg00615377	RBFOX2	-0.49	1.25E-23	0.87	1.20E-111
cg11394785	MAML1	-0.48	5.73E-23	0.78	2.39E-75
cg17356733	IFNGR2	-0.48	1.79E-22	0.83	6.93E-95
cg18390025	ELOVL3	-0.48	1.83E-22	0.84	4.19E-97
cg02181506	SERPINA1	-0.45	8.76E-20	0.74	7.93E-65
cg04536922	FAM13A	-0.45	1.62E-19	0.87	1.49E-114
cg00431050	ELOVL3	-0.39	1.57E-14	0.69	3.83E-52
cg18328334	TNS1	-0.37	2.87E-13	0.54	4.75E-29
cg04845063	FAM13A	-0.36	1.38E-12	0.80	1.97E-84
cg12585943	PPT2	-0.29	2.72E-08	0.59	5.39E-35
cg22563815	CHRNA5	-0.29	2.84E-08	0.52	1.72E-26
cg16505550	PPAP2B	-0.23	7.46E-06	0.38	5.96E-14
cg04603031	CHRNA3	-0.22	2.28E-05	0.66	7.85E-47
cg27486427	RARB	-0.22	3.08E-05	0.90	2.67E-130
cg00318573	CHRNA4	-0.19	1.77E-04	0.58	9.87E-35
cg03874127	ARMC2	-0.18	6.61E-04	0.57	2.36E-33
cg25229172	CCDC38	-0.17	8.09E-04	0.43	8.49E-18
cg19848683	MECOM	-0.14	6.09E-03	0.75	5.35E-67
cg13701109	ADAMTS19	-0.14	7.38E-03	0.71	1.57E-57
cg06334284	PPAP2B	-0.13	1.11E-02	0.82	7.50E-92
cg14228238	MECOM	-0.12	2.55E-02	0.60	1.41E-37
cg14580567	HHIP	-0.04	4.24E-01	0.61	3.32E-39
cg21750589	ZKSCAN3/ZSCAN31	-0.04	4.95E-01	0.26	4.78E-07

Table 3: COPD-related Genes from WGCNA blue comethylation module. Summary of "gene signficance to phenotype" values and "blue module membership" values of genes previously associated with COPD and lung function through GWAS, as well as differentially methylated sites discovered in the differential methylation analysis.

**Race-Specific Differential DNA Methylation Marks and Gene Comethylation Modules of COPD**

Authors: Robert Busch, MD[1], Weiliang Qiu, PhD[1], Jessica Lasky-Su, PhD[1], Jarrett Morrow, PhD[1], Gerard Criner, MD[2], and Dawn DeMeo, MD MPH[1]

Affiliations: [1]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA/USA; [2]Temple Lung Center, Temple University Health System, Philadelphia, PA/USA

Supplemental Figures:
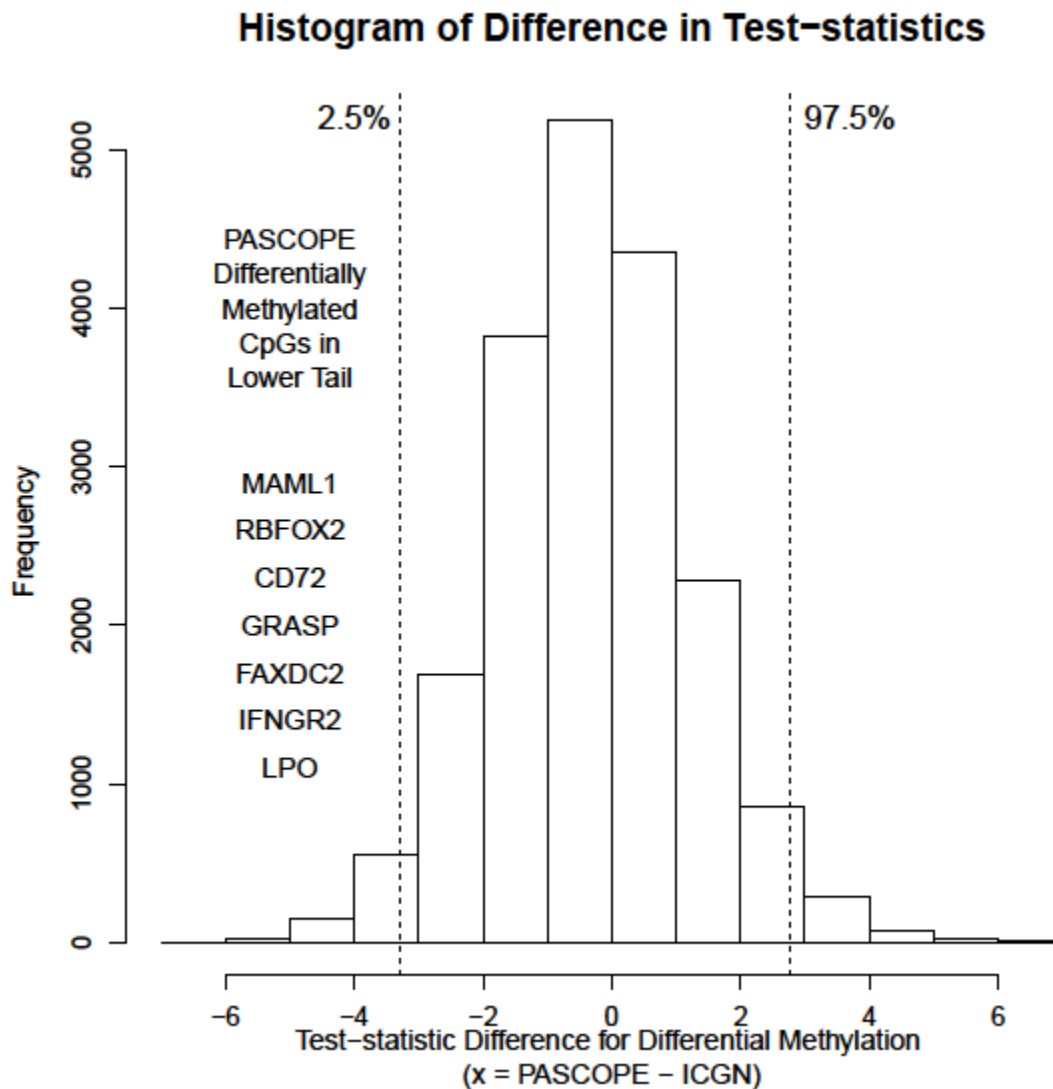
Supplemental Figure 1: Comparison to NHW Dataset. For each CpG site, test statistics were compared between the PA-SCOPE and ICGN datasets. The difference in test statistic was normally distributed over the 19302 probes analyzed. Notably, many of the CpG sites found to be differentially methylated in association with COPD in PA-SCOPE were in the lower tail of this distribution.
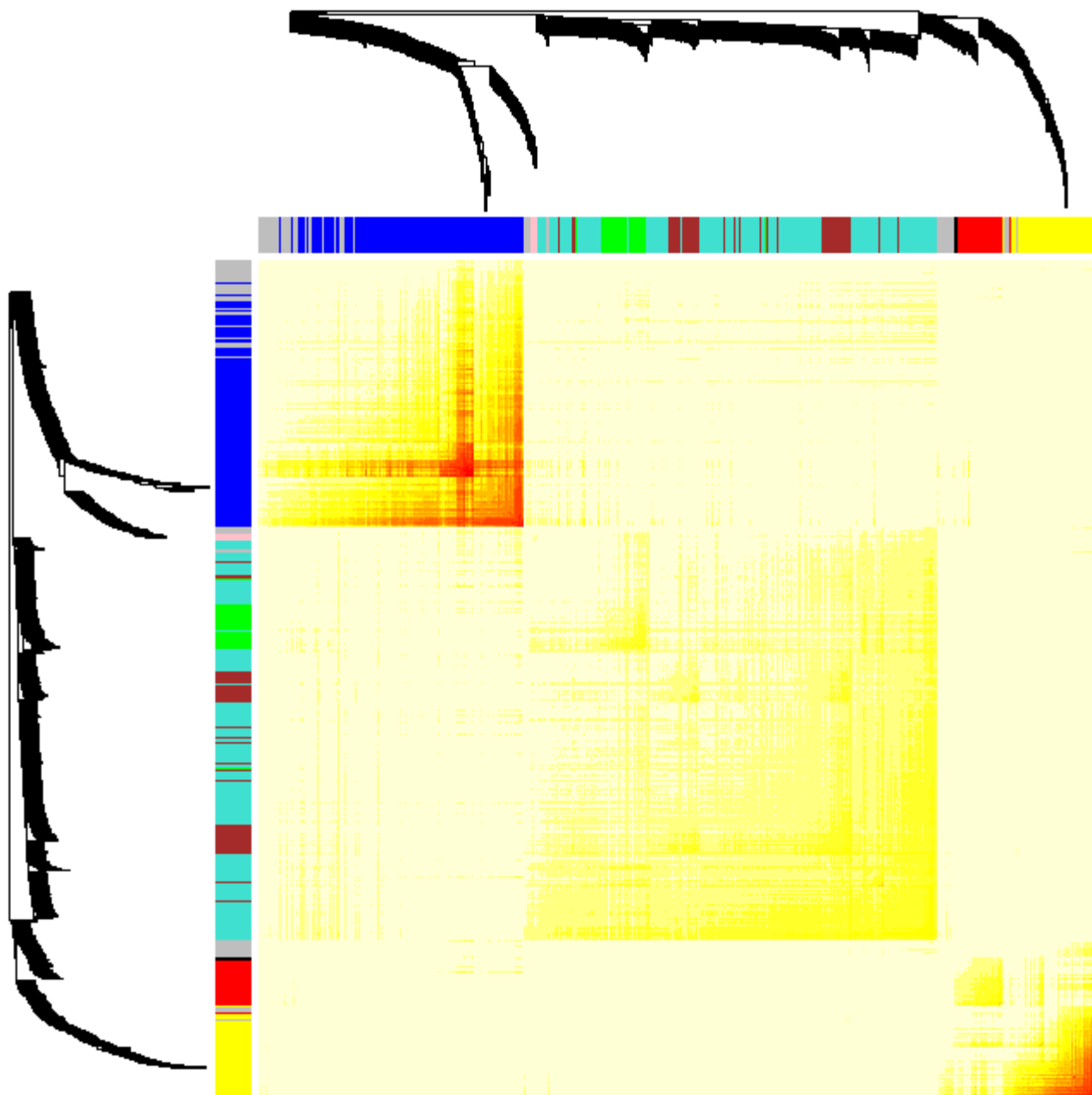
Supplemental Figure 2: Network Topology Heatmap. Heatmap representation of the topological overlap-based dissimilarity matrix calculated from PA-SCOPE weighted gene comethylation network analysis. Hierarchical clustering was used to construct clustering dendrograms to separate CpG sites into modules based on comethylation correlation. Module assignment colors based on clustering are represented along the x- and y-axis.

Supplemental Figure 3: Differential Methylation Boxplots. Boxplots showing the mean beta (absolute methylation difference ranging from 0 to 1, corresponding to 0 to 100% methylated) for COPD cases and controls for each of the twelve statistically significantly differentially methylated CpG sites reported in the Results section. These boxplots do not account for covariate adjustment.
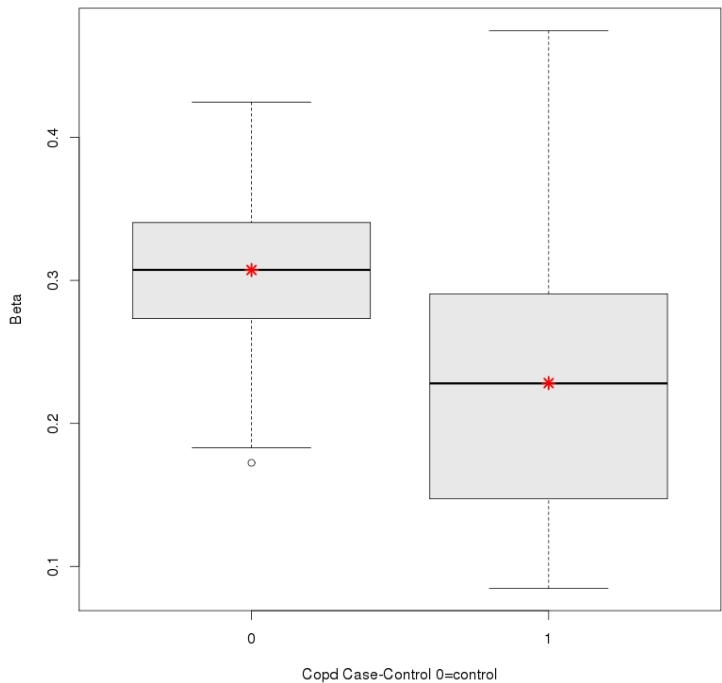
31
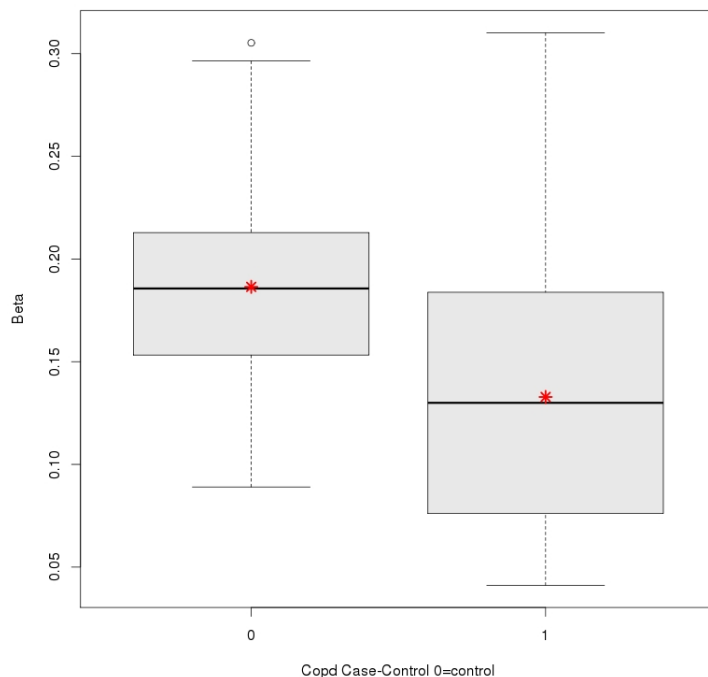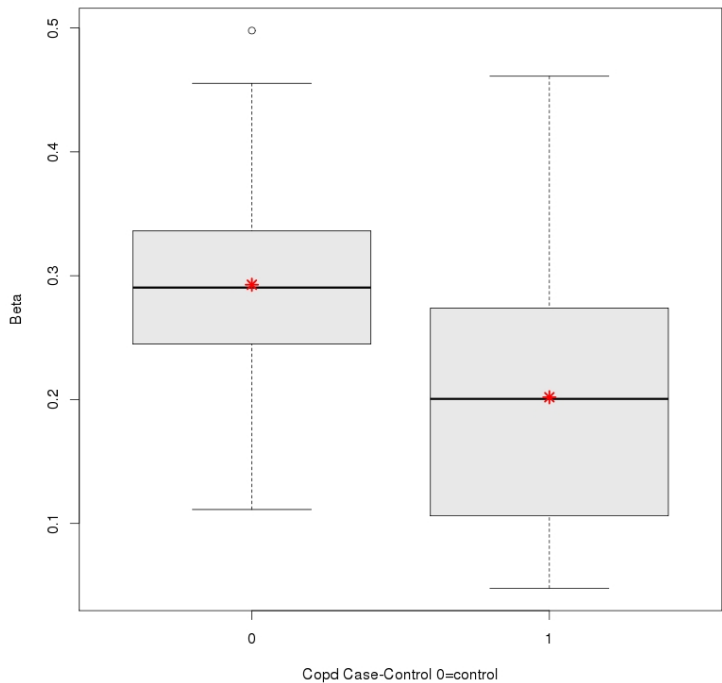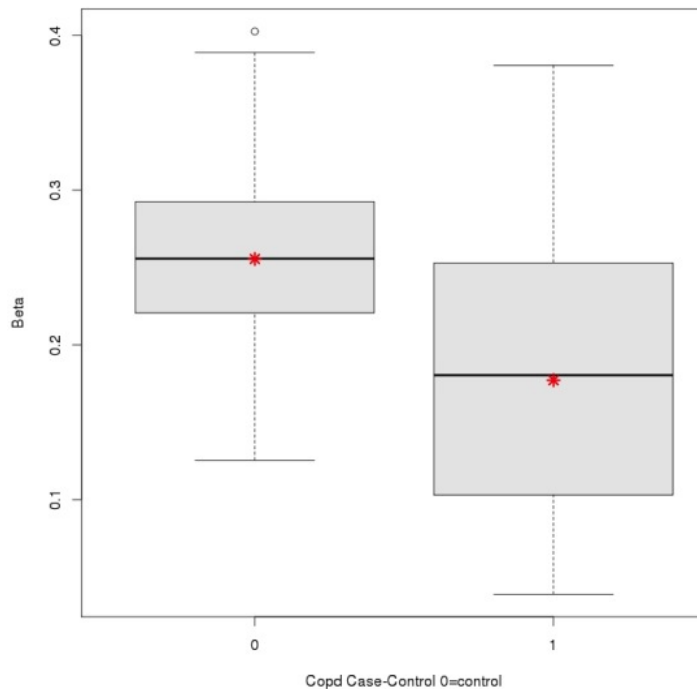
Boxplot of Betas for cg02635407, chr4, SH3TC1
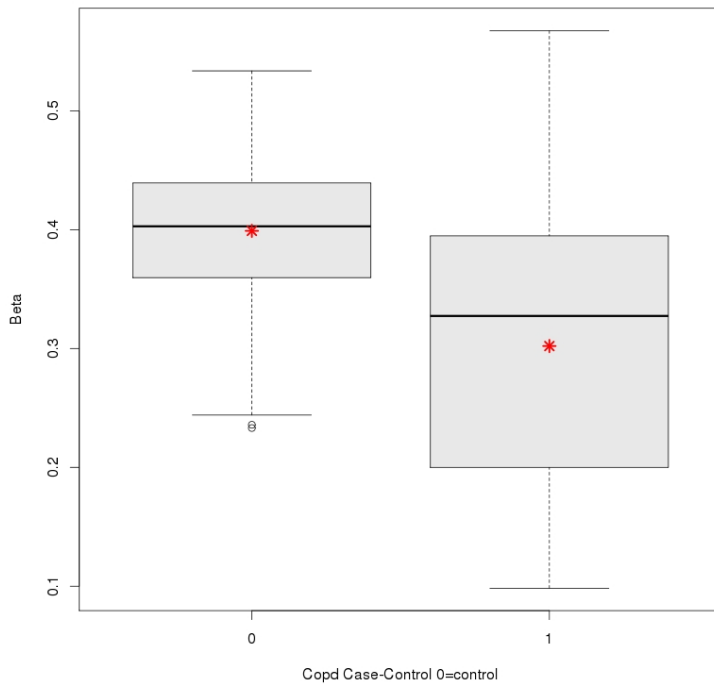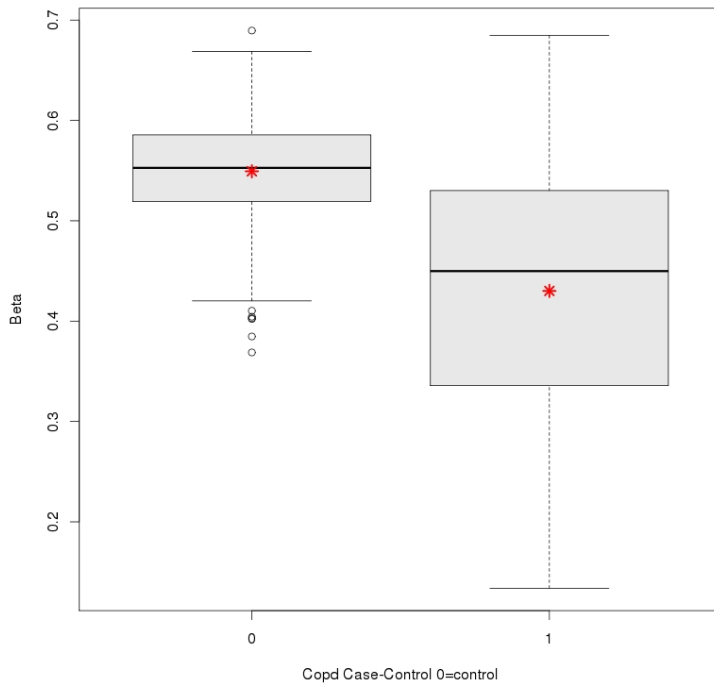
Boxplot of Betas for cg25634666, chr11, FOLR3
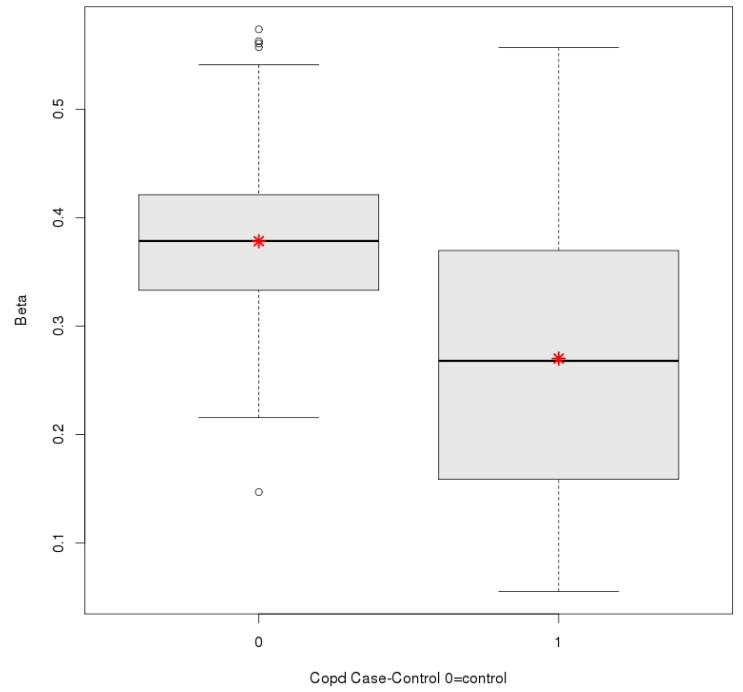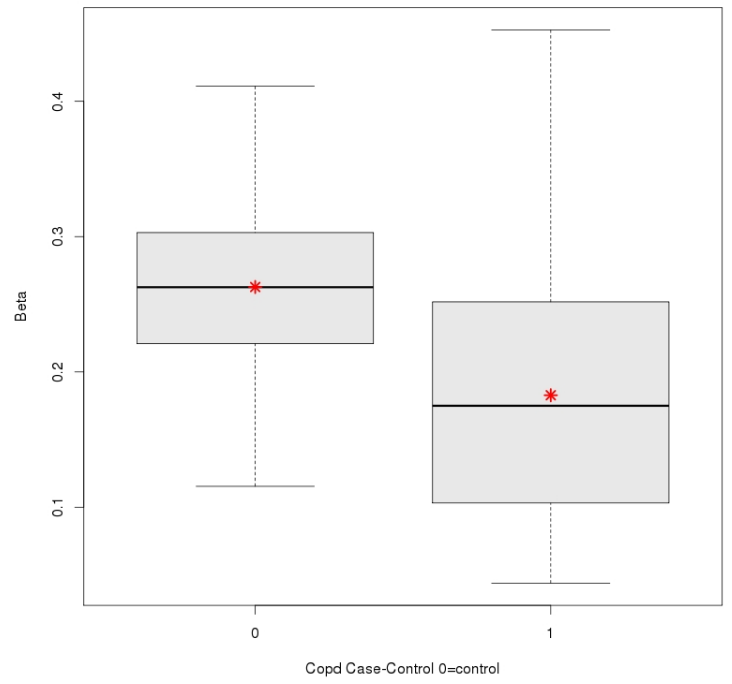
Boxplot of Betas for cg18390025, chr10, ELOVL3
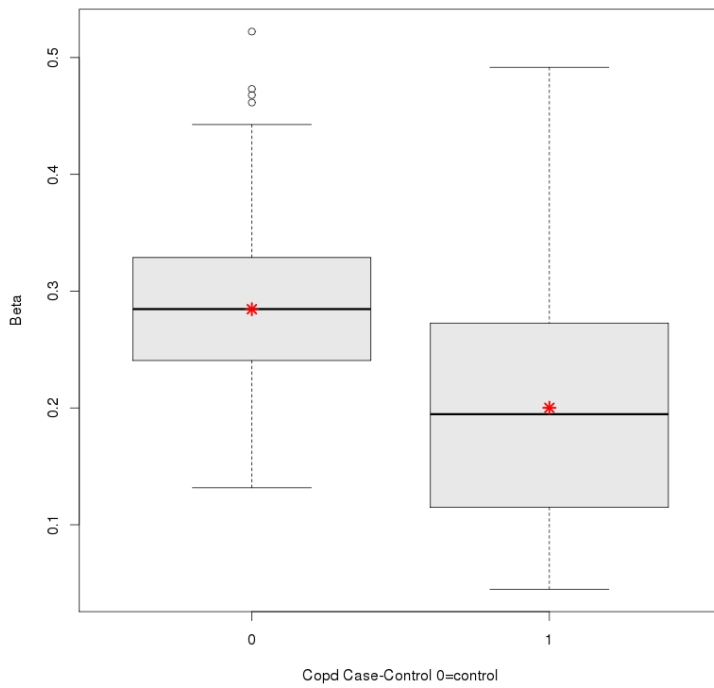
Boxplot of Betas for cg10257049, chr5, FAXDC2
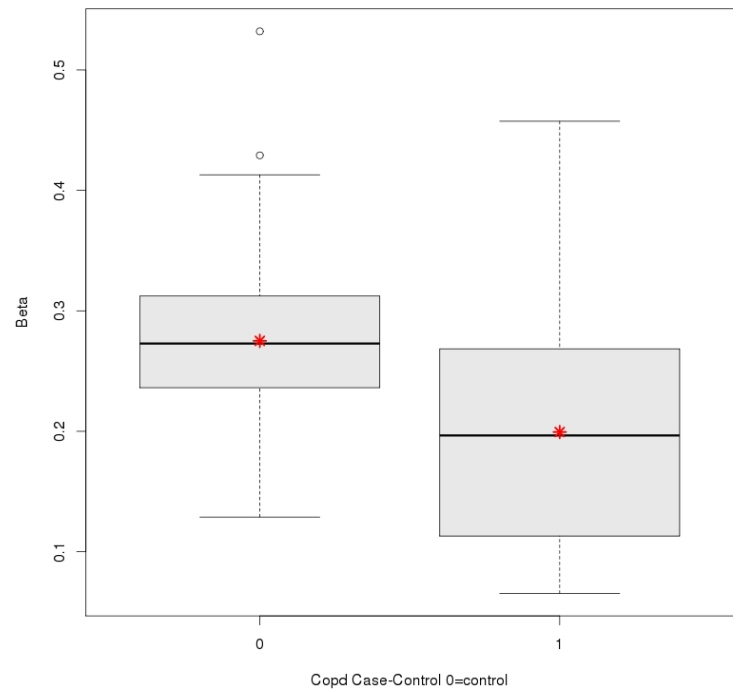
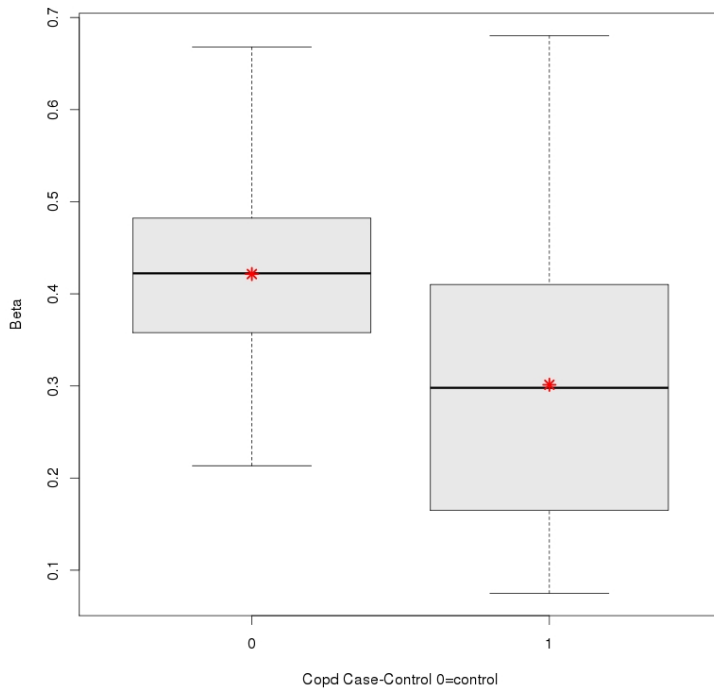**Boxplot of Betas for cg27461196, chr19, FXYD1**

Boxplot of Betas for cg00333528, chr6, GABRR1

Boxplot of Betas for cg17356733, chr21, IFNGR2

Boxplot of Betas for cg24489015, chr17, LPO