Integrative Genomics Analysis Identifies *ACVR1B* as a Candidate Causal Gene of Emphysema Distribution

Adel Boueiz^{1,2}, Betty Pham¹, Robert Chase¹, Andrew Lamb¹, Sool Lee¹, Zun Zar Chi Naing¹, Michael H. Cho^{1,2}, Margaret M. Parker¹, Phuwanat Sakornsakolpat¹, Craig P. Hersh^{1,2}, James D. Crapo³, Andrew B. Stergachis⁴, Ruth Tal-Singer⁵, Dawn L. DeMeo^{1,2}, Edwin K. Silverman^{1,2}, Xiaobo Zhou^{1,2}, and Peter J. Castaldi^{1,6}; on behalf of Genetic Epidemiology of COPD (COPDGene) Investigators*

¹Channing Division of Network Medicine, ²Division of Pulmonary and Critical Care Medicine, ⁴Division of Genetics, and ⁶Division of General Medicine and Primary Care, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; ³Pulmonary Medicine, National Jewish Health, Denver, Colorado; and ⁵GSK, King of Prussia, Pennsylvania

ORCID IDs: 0000-0002-4308-0974 (P.S.); 0000-0002-1342-4334 (C.P.H.).

Abstract

Genome-wide association studies (GWAS) have identified multiple associations with emphysema apicobasal distribution (EABD), but the biological functions of these variants are unknown. To characterize the functions of EABD-associated variants, we integrated GWAS results with 1) expression quantitative trait loci (eQTL) from the Genotype Tissue Expression (GTEx) project and subjects in the COPDGene (Genetic Epidemiology of COPD) study and 2) cell type epigenomic marks from the Roadmap Epigenomics project. On the basis of these analyses, we selected a variant near *ACVR1B* (activin A receptor type 1B) for functional validation. SNPs from 168 loci with *P* values less than 5×10^{-5} in the largest GWAS meta-analysis of EABD were analyzed. Eighty-four loci overlapped eQTL, with 12 of these loci showing greater than 80% likelihood of harboring a single, shared GWAS and eQTL causal variant. Seventeen cell types were enriched for overlap between EABD loci and Roadmap Epigenomics marks (permutation $P < 0.05$), with the strongest enrichment observed in CD4⁺, CD8⁺, and regulatory

T cells. We selected a putative causal variant, rs7962469, associated with *ACVR1B* expression in lung tissue for additional functional investigation, and reporter assays confirmed allele-specific regulatory activity for this variant in human bronchial epithelial and Jurkat immune cell lines. *ACVR1B* expression levels exhibit a nominally significant association with emphysema distribution. EABD-associated loci are preferentially enriched in regulatory elements of multiple cell types, most notably T-cell subsets. Multiple EABD loci colocalize to regulatory elements that are active across multiple tissues and cell types, and functional analyses confirm the presence of an EABD-associated functional variant that regulates *ACVR1B* expression, indicating that transforming growth factor- β signaling plays a role in the EABD phenotype.

Clinical trial registered with www.clinicaltrials.gov (NCT00608764).

Keywords: integrative genomics; emphysema distribution; chronic obstructive pulmonary disease; *ACVR1B* gene; transforming growth factor- β signaling

(Received in original form March 27, 2018; accepted in final form October 15, 2018)

*A complete list of COPDGene Investigators may be found before the beginning of the REFERENCES.

Supported by National Heart, Lung, and Blood Institute (NHLBI) grants K08HL141601-01, U01HL089897, R01HL089897, R01HL089856, R01HL124233, R01HL126596, R01HL113264, P01HL05339, and P01HL114501. The COPDGene study (COPD Genetic Epidemiology study; NCT00608764) is also supported by the COPD Foundation through contributions made to an industry advisory board comprised of AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Novartis, Pfizer, Siemens, and Sunovion. The Norway GenKOLS (Genetics of Chronic Obstructive Lung Disease; GSK code RES11080) and ECLIPSE (Evaluation of Chronic Obstructive Pulmonary Disease to Longitudinally Identify Predictive Surrogate Endpoints; NCT00292552; GSK code SCO104960) studies were funded by GlaxoSmithKline. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NHLBI or the National Institutes of Health.

Author Contributions: P.J.C.: had full access to all of the data in the study, takes responsibility for the integrity of the data and the accuracy of the data analysis, and had authority over manuscript preparation and the decision to submit the manuscript for publication; study concept and design: A.B. and P.J.C.; acquisition, analysis, or interpretation of data: all authors; drafting of the manuscript: A.B. and P.J.C.; critical revision of the manuscript for important intellectual content: all authors; statistical analysis: A.B. and P.J.C.; obtained funding: P.J.C., J.D.C., and E.K.S.; and study supervision: all authors. All authors gave final approval of the version to be published and have agreed to be accountable for all aspects of the work.

Correspondence and requests for reprints should be addressed to Peter J. Castaldi, M.D., M.Sc., Channing Division of Network Medicine, Brigham and Women's Hospital, 181 Longwood Avenue, Boston, MA 02115. E-mail: peter.castaldi@channing.harvard.edu.

This article has a data supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Am J Respir Cell Mol Biol Vol 60, Iss 4, pp 388–398, Apr 2019

Copyright © 2019 by the American Thoracic Society

Originally Published in Press as DOI: 10.1165/rcmb.2018-0110OC on October 18, 2018

Internet address: www.atsjournals.org

The success of genome-wide association studies (GWAS) in discovering novel disease-causing loci presents new challenges with respect to prioritizing disease-associated variants for further functional investigation. Many complex disease-associated loci alter gene expression (1, 2), but for diseases such as chronic obstructive pulmonary disease (COPD), in which the number of candidate cell types is large, identifying the proper cell type in which to conduct functional work is challenging. In the present study, we used data from two large compendia of gene regulatory annotations, the Genotype Tissue Expression (GTEx) and Roadmap Epigenomics projects, to pursue a comprehensive *in silico* approach to prioritizing GWAS-associated regulatory loci for functional follow-up in relevant tissues and cell types.

COPD is a clinical syndrome with multiple clinical manifestations, and this phenotypic heterogeneity has prognostic and therapeutic implications (3–5). Interestingly, among patients with emphysema, the patterns of lung destruction are often asymmetric (6). This asymmetry is clinically relevant, and it impacts the severity of airflow limitation, disease progression, and response to lung volume reduction procedures (7–11). A recent GWAS of vertical asymmetry of emphysema patterns (i.e., emphysema apicobasal distribution [EABD]) identified five genome-wide significant loci in smokers without alpha-1 antitrypsin deficiency. In addition to the genome-wide significant associations, additional analyses indicated that EABD has a substantial polygenic component, implying that there are many true EABD-causing associations that fall below the genome-wide significance level (12).

We hypothesized that a subset of EABD loci affect phenotype through the regulation of gene expression. We further hypothesized that integration of EABD GWAS results with extensive compendia of multitissue eQTL and multicell epigenetic marks would 1) prioritize candidate functional variants for further experimental characterization and 2) identify specific tissues and cell types through which these variants are likely to exert their phenotypic effects. To test these hypotheses, we integrated EABD GWAS results with eQTL from subjects with COPD and GTEx and cell-based

epigenetic marks from the Roadmap Epigenomics project. We used colocalization and permutation-based tools to quantify the enrichment of EABD GWAS regions in eQTL and epigenomic marks. We also performed functional evaluation and demonstrated allele-specific enhancer activity for an emphysema distribution-associated variant that alters the lung expression of *ACVR1B* (activin A receptor type 1B), an activin receptor in the transforming growth factor (TGF)- β signaling pathway. Some of these results were previously reported in the form of an abstract (13).

Methods

Study Populations

We analyzed 11,532 non-alpha-1 antitrypsin-deficient current and former smokers with complete genotype and computed tomographic densitometry data from four cohorts: the COPDGene NHW (Genetic Epidemiology of COPD study non-Hispanic whites), COPDGene African Americans, the GenKOLS (Genetics of Chronic Obstructive Lung Disease), and the ECLIPSE (Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints) studies. Detailed descriptions, including study populations, genotyping quality control, and genotyping imputation, were published previously (12–14).

Computed Tomographic Measurements

Quantitative assessment of emphysema was performed using three-dimensional Slicer density mask analysis (www.chestimagingplatform.org) to determine the percentage of lung voxels with attenuation less than -950 Hounsfield units at maximal inspiration (15). From these measurements, two correlated but complementary measures of emphysema distribution were constructed: 1) the difference between upper-third and lower-third emphysema (diff950) and 2) the ratio of upper-third to lower-third emphysema (ratio950) (12). A rank-based inverse normal transformation was applied to both phenotypes to reduce the impact of outliers and deviations from normality (12). In the present study, given that ratio950 had a higher heritability and was associated with more genome-wide significant signals than

diff950 (12), we performed fine mapping of ratio950-associated variants.

Peripheral Blood Gene Expression

As previously described (16), total RNA was extracted from 2.5 ml of whole blood per subject from the COPDGene study using the PAXgene Blood miRNA kit (QIAGEN). Samples with nondegraded RNA (RNA integrity number, >6) and concentration greater than or equal to $25 \mu\text{g}/\mu\text{l}$ were sequenced. Globin reduction and sequencing library generation were performed with TruSeq Stranded Total RNA with the Ribo-Zero Globin Kit (Illumina). Seventy-five-base pair paired-end sequencing was performed on HiSeq 2000 sequencers (Illumina), generating an average of 20 million reads per sample. Read pairs were aligned to the human genome (Genome Reference Consortium Human Build 38) using the STAR (Spliced Transcripts Alignment to a Reference) aligner (version 2.4.0h). Subsequently, gene counts were generated using RSubreads with the Ensembl version 81 annotation. Data quality control was performed using the RNA-SeQC and FastQC packages. Samples were excluded from subsequent analysis if the number of mapped reads was less than or equal to 10 million, mapping rate was below 80% to the reference genome, more than 10% of R1 reads were in the sense orientation, Pearson correlation was below 0.9 with samples in the same library construction batch, or reported sex and X-inactive specific transcript and Y-chromosome expression was inconsistent. Genotype concordance was determined by comparing DNA genotypes with those derived by RNA sequencing (RNA-seq).

Cis-eQTL Analysis

Three hundred eighty-five NHW subjects from the COPDGene study with available eQTL data were analyzed. Transcript-level expression count data were normalized for library size using the trimmed mean of M-values method and then subjected to inverse normal transformation (17). eQTL associations were tested for biallelic autosomal SNPs with minor allele frequency greater than 0.05 and mapping to a dbSNP 142 Reference SNP number. Cis-eQTL analysis was performed for all SNPs within 1 Mb of the target gene using Matrix eQTL with a linear model adjusting for age, sex, library preparation batch, three

principal components of genetic ancestry, and 35 probabilistic estimation of expression residuals factors of gene expression (18). A total of 5,815,008 SNPs were tested for association with 27,277 transcripts. The threshold for significance was a false discovery rate (FDR) of 10%, using the FDR procedure implemented in Matrix eQTL (19).

eQTL Emphysema Distribution GWAS Colocalization Analysis

We integrated 168 lead emphysema distribution GWAS SNPs ($P < 5 \times 10^{-5}$ in the largest GWAS meta-analysis of EABD [12]) with multitissue eQTL data from the GTEx project version 6 (20) (<https://gtexportal.org/home/datasets>; date of download, November 18, 2015) and whole-blood eQTL from 385 COPDGene NHW subjects. GWAS-eQTL integration was performed as previously described (21). Briefly, we selected all the SNPs within 1 Mb of each of the 168 lead emphysema distribution GWAS SNPs. We then integrated these SNPs with significant eQTL at FDR of 10% from the 44 tissues in GTEx and blood samples in COPDGene. For each independent association, Bayesian colocalization tests were performed for all SNPs within a 250-kb window of the lead GWAS variant at that locus to quantify the probability that the GWAS and eQTL associations were due to a single, shared causal variant (22). This probability corresponds to the posterior probability PP_4 number described in the original publication. The workflow of this GWAS-eQTL integration analysis is summarized in Figure 1.

Enrichment in Tissue-/Cell Type-Specific Chromatin States

Genetic variants associated with complex diseases have been shown to overlap regulatory enhancer elements (1, 2). On the basis of this finding, we quantified the overlap of emphysema distribution GWAS SNPs ($P < 5 \times 10^{-5}$) and their associated high-linkage disequilibrium SNPs ($r^2 > 0.8$ in the 1000 Genomes Project European samples) with regulatory elements identified in the Roadmap Epigenomics project (release 9) (23). For the 127 Roadmap Epigenomics cell types, we downloaded ChromImpute imputed annotations (24), and we analyzed ChromImpute DNase I-hypersensitive (DHS) peaks and ChromImpute enhancer

marks (defined as chromatin states 13 through 18) for all 127 cell types with available data. We also examined DHS hotspots (i.e., broader regions of DNase I hypersensitivity encompassing DNase peaks) that were available for 39 Roadmap Epigenomics cell types. The hotspot identification algorithm has been described previously (25). We also analyzed DNase I digital genomic footprinting (DGF) data from 42 uniformly processed cell types in the Roadmap Epigenomics project. DNase I footprints are coverage troughs in deeply sequenced DNase I hypersensitivity data that represent narrow genomic regions shielded from DNase I digestion because of transcription factors, and thus they represent likely transcription factor binding sites.

DHS ChromImpute peaks were downloaded from <https://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidatedImputed/narrowPeak/> on July 13, 2016. DHS hotspots were downloaded from <https://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidated/broadPeak/DNase/> on February 20, 2015. Enhancer marks were downloaded from <https://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/imputed12marks/jointModel/final/> on December 23, 2015. DGF data were downloaded from <https://egg2.wustl.edu/roadmap/data/byDataType/dgfootprints/> on July 13, 2016.

To determine the extent of GWAS-epigenomic annotation overlap, we identified independent emphysema distribution GWAS signals at $P < 5 \times 10^{-5}$ within 1-Mb windows. We then used Genomic Annotation Shifter (GoShifter) to calculate the enrichment for these variants in Roadmap Epigenomics annotations (DHS, enhancer marks, DGF) (26). This method uses a local permutation strategy to account for the local density of a given epigenomic mark. One thousand permutations were performed using linkage disequilibrium information from the 1000 Genomes Project European population with an r^2 threshold of 0.8.

Probabilistic Identification of Causal SNPs: rs7962469

The online probabilistic identification of causal SNPs (PICS) algorithm is a fine-mapping algorithm which calculates the probability that an individual SNP is a causal variant given the haplotype structure

and observed pattern of association at the locus (<https://pubs.broadinstitute.org/pubs/finemapping/pics.php>) (27). We used this algorithm to generate the 95% credible SNP set for the GWAS association identified in the region near the *ACVR1B* (activin A receptor type 1B), *MEI1* (meiotic double-stranded break formation protein 1), *DCBLD1* (discoidin, CUB and LCCL domain containing 1), and *IGHV3* (immunoglobulin heavy variable 3) genes.

Luciferase Reporter Assay: rs7962469

Two approximately 500-bp genomic segments including rs7962469 were obtained from human bronchial epithelial (16HBE) cells heterozygous at rs7962469 and cloned into the XhoI and BglII sites of a pGL4.23[luc2/minP] vector. Each luciferase construct was cotransfected with thymidine kinase Renilla, a luciferase control reporter, in 16HBE and Jurkat immune cell lines at approximately 60–70% confluency by using Lipofectamine 3000 Reagent (Invitrogen), following the manufacturer's protocol. Four separate transfections were performed in triplicate at a concentration of 300 ng per well and the thymidine kinase Renilla was used at 15 ng per well. Empty luciferase vector, pGL4.23[luc2/minP], was also transfected in triplicate as a control. All plasmids used were confirmed by sequencing. Promoter activity was quantified 48 hours after transfection using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's protocol. Luminescence signals were captured in a Wallac VICTOR³ 1420 plate reader (PerkinElmer) and normalized by the Renilla luciferase readings for each well. Allele-specific transcription effects were quantified by using the Wilcoxon rank-sum test to compare 1) normalized luciferase values for rs7962469-G versus rs7962469-A constructs and 2) luciferase values for each allele with the pGL4.23 empty vector values. *P* values less than 0.05 were considered significant.

ACVR1B Differential Gene Expression Analysis

We performed *ACVR1B* differential gene expression in 1,045 subjects in the COPDGene study with complete RNA-seq and phenotype data. *ACVR1B* transcripts (Ensembl Gene ID: ENSG00000135503) with one count per million or more mapped reads in 10 or more subjects were analyzed for differential gene expression

using *limma/voom* implemented in the R package *limma* (28). Age, sex, race, current smoking, pack-years, batch, cell counts, and surrogate variables were used as covariates in the analysis. The phenotype analyzed was *ratio950* emphysema distribution. Statistical significance for differentially expressed *ACVR1B* was defined on the basis of a *P* value less than 0.05.

Allele-Specific Expression in *ACVR1B*

To assess allele-specific expression (ASE) in *ACVR1B*, we combined whole-genome sequencing and whole-blood RNA-seq data from 1,100 individuals in the COPDGene study. Allelic counts at heterozygous sites at the *ACVR1B* locus were quantified using the Genome Analysis Toolkit ASEReadCounter. Analyzed sites were restricted to those with at least 15 reads in the heterozygote individual. We performed binomial tests across heterozygous sites with the null value equal to the mean reference ratio of 0.5.

Web Access to Study Results

Searchable tables of the formal metrics, results, and LocusZoom plots for each genomic region are available to the public as companion sites for this paper (https://cdnm.shinyapps.io/eabd_eqtlcolocalization/ [GWAS-eQTL] and https://cdnm.shinyapps.io/eabd_gwas_roadmap_goshifter/ [GWAS-epigenomic annotations]).

Results

Subject Characteristics

A previously published GWAS was performed using data from 6,215 NHW and 2,955 African American subjects from COPDGene, 1,538 subjects from ECLIPSE, and 824 subjects from GenKOLS with complete phenotype and genotype data (12). The characteristics of these 11,532 subjects are shown in Table E1 in the data supplement, and the characteristics of the smokers enriched for COPD with available RNA-seq eQTL data are shown in Table E2.

Multitissue GWAS-eQTL Colocalization

We identified 168 genomic regions associated with EABD at $P < 5 \times 10^{-5}$, which we refer to as “EABD-associated loci.” To identify promising loci for additional functional investigations, we used previously established tools to

integrate evidence from the EABD GWAS with two large compendia of gene regulatory functional information relating to eQTL studies and epigenomic marks.

We integrated the 168 EABD-associated loci with eQTL results from 44 tissues from GTEx and a novel set of blood eQTL generated from 385 smokers in the COPDGene study (Figure 1, Table E3). The lead GWAS SNP from 84 (50%) EABD-associated loci overlapped a significant *cis*-eQTL SNP in at least one of the studied tissues. We refer to these 84 loci as “EABD GWAS-eQTL loci.” Forty-eight (57%) of these lead SNPs were a significant eQTL in more than one tissue (Figures 2A and 2B), and there was a nonsignificant trend toward loci with lower GWAS *P* values

having significant eQTL in a greater number of tissues (Spearman correlation between number of overlap tissues and log $P = 0.06$).

Because there are many eQTL across the genome, some overlap between GWAS and eQTL may be due to chance, but overlaps due to chance are unlikely to show a strong concordance in eQTL and GWAS association statistics at a given locus. To distinguish chance from causal overlaps, we performed a Bayesian test for colocalization at each EABD GWAS-eQTL overlap region. Of the 84 candidate loci, 29 had a reasonable likelihood ($>50\%$) of harboring a shared causal variant, and 12 loci had a high likelihood ($>80\%$) of harboring a shared causal variant, including regions

Figure 1. Workflow and summary of the results of the colocalization analysis of emphysema distribution-associated genetic variants with expression quantitative trait loci (eQTL) and epigenomic data. *ACVR1B* = activin A receptor type 1B; *DCBLD1* = discoidin, CUB and LCCL domain containing 1; GWAS = genome-wide association studies; HBE = human bronchial epithelial; *IGHV3* = immunoglobulin heavy variable 3; *MEI1* = meiotic double-stranded break formation protein 1; PICS = probabilistic identification of causal SNPs.

Figure 2. Overlap of emphysema apicobasal distribution (EABD) risk variants in multiple tissues and cell types. (A) The majority of EABD loci overlap eQTL regions that are active in more than one tissue, and (B) there is trend toward stronger EABD GWAS association and a greater number of active eQTL tissues at a given locus. (C) The majority of EABD loci also lie within regions of DNase I peaks and (D) enhancer elements that are active in more than one cell type.

affecting the expression of the *ACVR1B*, *MEI1*, *DCBLD1*, and *IGHV3* genes in lung tissues.

Table 1 shows the significant colocalizations of EABD GWAS loci with eQTL in three tissues likely to be relevant to emphysema distribution, namely lung, blood, and lymphoblastoid cell lines. The complete set of colocalization results can be viewed interactively at https://cdnm.shinyapps.io/eabd_eqtlcolocalization/.

EABD GWAS Loci Overlap Broadly Active Regulatory Elements

Active regulatory elements can be identified by characteristic epigenomic marks, and the Encyclopedia of DNA Elements (ENCODE) and Roadmap Epigenomics projects have generated genome-wide regulatory maps of these marks in purified cell types or cell lines. These epigenomic marks provide cell-specific regulatory information that is

complementary to eQTL tissue studies. To determine whether EABD-associated loci are enriched in cell type-specific regulatory elements, we performed enrichment analysis for EABD loci and four epigenomic marks (DHS peaks and hotspots, enhancer regions, and DNase I footprints) from 127 cell types generated by the ENCODE and Roadmap Epigenomics projects (29, 30). The four marks covered, on average, 0.44–2.87% of the genome per cell type (Table E4). Ninety-nine EABD loci (58.9%) overlapped at least one of the four studied epigenomic annotations in at least one cell type. The number of overlapping loci by epigenomic marks ranged from 45 loci (DNase I footprints) to 86 loci (DHS hotspots) (Table E5).

To determine the cell types showing the strongest enrichment for EABD-associated loci in regulatory regions while controlling for the overall regulatory activity of each cell

type, we used the local permutation approach implemented in the GoShifter program (26). As illustrated in Figure 3, a total of 17 different cell types exhibited evidence of enrichment of EABD loci (permutation $P < 0.05$) in at least one set of epigenomic marks, with the most significant enrichment observed for $CD4^+$, $CD8^+$, and regulatory T cells. These results indicate that, on average, EABD-associated loci are particularly overrepresented in immune cell types, but other cell types, such as fetal tissues and fibroblasts, also showed evidence of enrichment. Finally, the ENCODE and Roadmap Epigenomics projects do not include regulatory maps for some important lung-related cell types, such as bronchial epithelial cells. Table 2 shows the number of cell types with overlaps between emphysema distribution GWAS-eQTL loci in lung, blood, and lymphoblastoid cell lines and DHS sites, DNase hotspots, enhancer regions, and DNase I footprints. The complete set of results can be viewed at https://cdnm.shinyapps.io/eabd_gwas_roadmap_goshifter/.

When EABD-associated loci overlapped regulatory elements, these elements tended to be active in multiple cell types. The proportions of overlap loci active in more than one cell type were 91%, 91%, 87%, and 73% for DHS peaks, hotspots, enhancer regions, and footprints, respectively. In some annotations, most notably DHS peaks, there was a biphasic pattern of cell-specific activity in which most loci were either active in a small number of cell types or constitutively active in all cell types (Figures 2C and 2D).

Functional Validation of the rs7962469 Variant Near *ACVR1B* and Evidence for TGF- β Signaling Pathways in EABD

On the basis of the integrative analyses described above, we focused our subsequent experiments on four loci that colocalized with lung eQTL and showed evidence of regulating *ACVR1B*, *MEI1*, *DCBLD1*, and *IGHV3* expression. All these loci overlapped broadly active regulatory elements in multiple cell types. For a GWAS-identified locus, it can be challenging to identify the disease-causing functional variant from statistical evidence alone, because SNPs in close physical proximity can show high correlation owing to linkage disequilibrium. For these four loci, we used the PICS algorithm to

Table 1. Significant Colocalization of Emphysema Distribution-associated Genome-Wide Association Study Variants with Expression Quantitative Trait Loci in Lung, Blood, and Lymphoblastoid Cell Lines from the Genotype Tissue Expression Project and the COPDGene Study, Ordered by the Lead SNP Genome-Wide Association Study *P* Values

Lead GWAS SNP	Chromosome	Position	HUGO Gene Annotation	GWAS <i>P</i> Value	eQTL Tissue	eQTL <i>Q</i> Value	Colocalization Probability
rs5758407	22	42076956	<i>MEI1</i>	5.90×10^{-6}	Transformed fibroblasts (GTEx)	4.47×10^{-18}	0.93
					Lung (GTEx)	1.59×10^{-11}	0.91
					Whole blood (GTEx)	4.45×10^{-18}	0.83
					EBV-transformed lymphocytes (GTEx)	2.27×10^{-6}	0.82
rs7962469	12	52348259	<i>ACVR1B</i>	1.70×10^{-5}	Lung (GTEx)	0.0005	0.91
rs4468504	14	107122186	<i>IGHV3-66</i>	2.70×10^{-5}	Lung (GTEx)	6.15×10^{-18}	0.91
			<i>IGHV3-66</i>		Whole blood (GTEx)	2.43×10^{-21}	0.91
			<i>IGHV3-64</i>		Lung (GTEx)	0.0008	0.89
			<i>IGHV4-61</i>		Whole blood (GTEx)	1.52×10^{-9}	0.87
			<i>IGHV3-66</i>		Lung (GTEx)	6.70×10^{-8}	0.86
			<i>IGHV3-64</i>		Whole blood (GTEx)	2.41×10^{-6}	0.83
		106666170	<i>IGHV1-69-2</i>		Whole blood (COPDGene NHW)	3.39×10^{-18}	0.91
			<i>IGHV3-64</i>		Whole blood (COPDGene NHW)	4.35×10^{-5}	0.83
rs210611	6	117836184	<i>DCBLD1</i>	3.90×10^{-5}	Lung (GTEx)	0.049	0.82

Definition of abbreviations: *ACVR1B* = activin A receptor type 1B; COPDGene study = Genetic Epidemiology of COPD study; *DCBLD1* = discoidin, CUB and LCCL domain containing 1; EBV = Epstein-Barr virus; eQTL = expression quantitative trait loci; GTEx = Genotype Tissue Expression project; GWAS = genome-wide association study; HUGO = Human Genome Organization; *IGHV* = immunoglobulin heavy variable; *MEI1* = meiotic double-stranded break formation protein 1; NHW = non-Hispanic white individuals.

Colocalization probability is the posterior probability that the same causal variant is associated with emphysema distribution GWAS and eQTL signals. Lead SNP is the SNP with the lowest GWAS *P* value for association in a given GWAS-identified region. Listed are all instances for which the Bayesian colocalization results indicate that there is high posterior probability ($PP_4, \geq 0.8$) that the same causal variant is associated with emphysema distribution GWAS and eQTL signals in lung, blood, and lymphoblastoid cell lines from the GTEx project and the COPDGene study.

estimate the probability that the lead GWAS variant (i.e., the SNP with the lowest *P* value in a given association region) is the disease-causing variant. The PICS causal probabilities for rs7962469 near *ACVR1B*, rs5758407 near *MEI1*, rs210611 near *DCBLD1*, and rs4468504 near *IGHV3* were estimated to be 65.7%, 3.7%, 100%, and 4.3%, respectively. Because rs7962469 has a high PICS probability and is clearly a lead variant in both the GWAS and eQTL associations (Figure 4A), we selected this variant for further functional characterization.

The rs7962469 variant is present in a regulatory element that is active in over 60 cell types. Because it is a significant eQTL in lung tissue and has strong enrichment in CD4⁺, CD8⁺, and regulatory T cells, we selected the 16HBE human bronchial epithelial cell line and a human immune cell line (Jurkat human leukemic T-cell lymphoblast) in which to test the allele-specific regulatory activity of rs7962469. We observed that the G variant of rs7962469 has significantly decreased expression relative to the A variant (Figures

5 and E1) and that the direction of effect from the reporter assay was consistent with the eQTL result. The rs7962469 G variant has previously been shown to be associated with COPD susceptibility (odds ratio, 1.13; SE, 1.03; GWAS *P* = 0.002) (31) and with upper lobe emphysema predominance (effect size, 0.06; SE, 0.01; GWAS *P* = 1.7×10^{-5}) (12).

To determine whether the allelic enhancer effect observed in Jurkat cells was corroborated by GTEx results in immune cells, we queried the association between rs7962469 and *ACVR1B* expression in Epstein-Barr virus-transformed lymphocytes in GTEx, and we observed a nominal association in the same direction observed in lung tissue (*P* = 0.017) (Figure E2).

We also performed differential gene expression in 1,045 subjects in the COPDGene study with complete RNA-seq and phenotype data, and we observed that, after adjustment for demographic factors and cell counts, *ACVR1B* gene expression levels were significantly associated with emphysema distribution (log fold change,

0.017; average expression, 4.53; *P* = 0.005). We also performed ASE analysis to determine whether *ACVR1B* showed evidence of ASE in these subjects, but there was no evidence of ASE. This may be attributable to the fact that the ASE analysis was performed in whole blood with a mixture of different cell types.

ACVR1B mediates TGF- β signaling through interaction with SMAD proteins, including SMAD3. Interestingly, one of the top colocalizing EABD associations is a GWAS-eQTL colocalization for *SMAD3* in esophageal mucosa from GTEx (Figure 4B). SNPs in this locus also overlie active enhancer elements in multiple lung-specific cell types, including IMR-90 fetal lung fibroblasts, adult lung fibroblasts, and the A549 lung epithelial cell line.

Discussion

Using broad regulatory compendia of eQTL from 45 tissues and epigenomic marks from 127 cell types, we performed an integrated genetic-epigenomic study to further our

Figure 3. Plot of the cell type enrichment $-\log_{10} P$ value from the Genomic Annotation Shifter (GoShifter) analyses of the overlaps between emphysema distribution-associated genetic variants with DNase I-hypersensitive peaks, DNase I hotspots, enhancer marks, and digital DNase I footprints. The cell types with enrichment P values less than 0.05 are colored red in A and shown in more detail in B. iPS = induced pluripotent stem cell; PMA-I = phorbol 12-myristate 13-acetate plus ionomycin.

functional understanding of common variants associated with EABD, a clinically important manifestation of COPD. The following were the primary findings from this analysis:

1. EABD-associated loci that overlap eQTL and epigenomic marks tend to affect
2. Analysis of epigenomic marks identifies multiple cell types enriched for overlap of EABD loci in regulatory elements, most notably T-cell subsets.
3. Colocalization analysis gives strong evidence of a shared causal variant in 12 EABD loci.
4. Reporter assays confirmed allele-specific regulatory activity for the EABD-associated variant, rs7962469, near

Table 2. Number of Cell Types with Overlaps between Emphysema Distribution-associated Genome-Wide Association Study Variants with Expression Quantitative Trait Loci in Lung, Blood, and Lymphoblastoid Cell Lines and Each or Any One of DNase I-Hypersensitive Sites, DNase Hotspots, Enhancer Regions, and DNase I Footprints

Lead GWAS SNP	DHS	Hotspots	Enhancer Regions	DNase I Footprints	Any Annotation
rs5758407	125	39	126	42	169
rs7962469	62	11	0	0	63
rs4468504	27	9	28	13	43
rs210611	1	0	2	0	3

Definition of abbreviations: DHS = DNase I hypersensitive site; eQTL = expression quantitative trait loci; GWAS = genome-wide association study. Emphysema distribution-associated GWAS eQTL loci in lung, blood, and lymphoblastoid cell lines are loci with SNPs that have emphysema distribution-associated GWAS P values less than 5×10^{-5} and *cis*-eQTL colocalization probability less than or equal to 0.8 in lung, blood, and lymphoblastoid cell lines. Loci are defined by the regions bounded by common SNPs in linkage disequilibrium with the lead SNP at $r^2 > 0.8$. Lead SNPs are SNPs with the lowest GWAS P values for association in a given GWAS-identified region.

ACVR1B with the G allele associated with decreased reporter gene expression, increased COPD susceptibility, and upper lobe emphysema predominance.

Multiple lines of evidence provide strong support for the allelic effect of the emphysema-associated variant rs7962469 on expression of *ACVR1B*. Seventy-three separate Roadmap Epigenomics/ENCODE experiments indicate that the rs7962469 variant lies in a region of open chromatin. Our integrative genomics analysis prioritized rs7962469 as the most likely functional variant in this region. eQTL data from GTEx link this variant to *ACVR1B* gene expression in lung and Epstein-Barr virus-transformed lymphocytes, and we demonstrated allele-specific enhancer activity in 16HBE and Jurkat cell lines. These observations, together with the observation of an EABD colocalizing signal with *SMAD3* expression, suggest that genetic susceptibility to emphysema is mediated in part through TGF- β signaling.

ACVR1B, also known as *ALK-4*, acts as a transducer of activin-like ligands that are growth and differentiation factors belonging to the TGF- β superfamily of signaling proteins. Although *ACVR1B* has not previously been associated with emphysema distribution, prior genetic studies have demonstrated an association of gene polymorphisms of the TGF- β superfamily with COPD (32, 33), and TGF β 2 (transforming growth factor- β 2) expression levels were reduced in a set of Lung Tissue Research Consortium COPD

lung tissue samples compared with controls (31). A network analysis incorporating COPD GWAS and protein-protein interaction data included *ACVR1B* in a 10-gene consensus network module associated with COPD case-control status (34), and a separate lung eQTL network analysis also highlighted *ACVR1B* as a potential COPD candidate gene (35). Our demonstration of a COPD-associated functional variant associated with *ACVR1B* expression in the lung adds to multiple lines of prior evidence implicating TGF- β signaling in COPD, but more work is warranted to elucidate the functional mechanisms by which *ACVR1B* contributes to the development of COPD and emphysema.

It is well established that GWAS-identified regulatory loci often target genes that are not the closest genes to the causal variant, as measured by linear distance along the chromosome (36). With recently developed integrative genomics methods and a growing amount of publicly available epigenomic data, integrating GWAS with known functional annotations can be an effective strategy for identifying and prioritizing a small number of disease-associated variants for low-throughput functional validation assays. Using this approach, we identified a subthreshold EABD association that appears to modulate *ACVR1B* expression (and presumably TGF- β signaling) across a wide range of cell types, and we successfully narrowed the candidate variants in this region to a single variant with allelic gene regulatory activity in bronchial epithelial and Jurkat cell lines.

Our bioinformatic screening approach identified many promising regions (i.e., 12 loci with $>80\%$ colocalization probability across all eQTL tissues), more than can be routinely screened with low-throughput functional methods. In addition, the most effective follow-up functional strategy varies on the basis of the pattern of association present in a given region. The association near *ACVR1B* is particularly well suited for low-throughput enhancer validation because it has the same single well-defined lead variant for both the GWAS and eQTL associations. Many other loci, such as *MEI1*, with high colocalization scores did not have clearly identified causal variants per the PICS analysis, and functional investigations of these loci would require high-throughput functional screening approaches. The functional validation of rs7962469 is proof of concept of the utility of our bioinformatic prioritization approach, but more comprehensive efforts to validate this approach for all prioritized loci could be achieved in the future by using massively parallel reporter assays.

Although some disease-associated loci act in a cell type-specific manner, other disease loci have been shown to reside in regulatory elements that are broadly active across many cell types (37). Epigenomic enrichment analyses prioritized T-cell subsets as having the strongest enrichment for EABD-associated loci, in line with extensive prior work demonstrating that the adaptive immune response plays an important role in the pathophysiology of COPD and emphysema (32–34). However, it is important to note that many cell types attained nominal significance in our enrichment analyses, supporting a model in which multiple cell types contribute to the EABD phenotype. In addition, some EABD variants such as rs7962469 are broadly active across cell types. The association between rs7962469 genotype and *ACVR1B* expression in lung tissue may indicate an allelic effect in bronchial epithelial cells, infiltrating immune cells, or any other prominent lung cell type.

The strengths of this study are the breadth of functional data used for integrative functional analysis, application of Bayesian and permutation-based methods to assess the significance of the observed overlaps accounting for the genomic abundance of candidate eQTL and epigenomic annotations, and the novel

Figure 4. GWAS locus and eQTL plots for (A) the rs7962469 variant near *ACVR1B* in lung tissue and (B) for rs10152544 near *SMAD3* in esophageal mucosa. cMMb = centimorgan per megabase.

demonstration of allele-specific enhancer activity for a candidate causal variant near *ACVR1B*. These results implicate TGF- β signaling in the pathogenesis of emphysema distribution, pointing to the importance of

further work to understand this disease mechanism.

This study also has important limitations. First, we limited our analysis to *cis*-eQTL, excluding other classes of gene

regulatory variants, including *trans*-eQTL, isoform ratio QTL, and variants implicated by ASE analyses. Future studies will be strengthened by the inclusion of these emerging features of eQTL studies (38, 39).

Figure 5. Reporter assays of the rs7962469 variant near the *ACVR1B* gene in 16HBE cells showing significantly increased expression of the A variant relative to the G variant ($P = 0.008$) and the empty luciferase vector (pGL4.23) ($P = 0.00002$). The P value comparing the G variant with pGL4.23 is 0.02. $**P < 0.05$. Data are from four independent transfections, each performed in triplicate. TK = thymidine kinase.

Second, the colocalization and enrichment methods that we used have limitations. The colocalization method does not account for multiple independent signals, and the GoShifter approach may be biased against regions that are dense in epigenomic annotations (22, 26, 40). There were also many EABD-associated regions for which no regulatory overlaps were identified. These instances of nonoverlap could be due to false-positive GWAS associations at the reduced stringency levels used for this analysis, limited power of the included eQTL analyses (32), limited assessment of the dynamic nature of epigenomic marks in included cell-type data, and lack of representation of relevant tissues and cell types for emphysema in GTEx and the Roadmap Epigenomics project, respectively. Finally, although we provide compelling evidence for rs7962469 as a COPD-associated functional variant, additional functional work is required to further elucidate the transcriptional mechanisms altered by genetic variation at this locus.

Conclusions

This study provides proof of concept that integrative analyses leveraging public compendia of eQTL and epigenomic maps can effectively

prioritize disease-associated loci and specific variants for further functional investigation. Enrichment analyses implicated a wide range of cells and tissues, emphasizing the importance of having comprehensive compendia of regulatory annotation with respect to tissues, cell types, diseases, and environmental exposures. On the basis of these integrative analyses, we prioritized and functionally validated a COPD- and emphysema-associated variant involved in TGF- β signaling. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The authors acknowledge Tuuli Lappalainen and Stephane Castel for helpful discussions regarding allele-specific expression.

COPD Gene investigators, by Core Units:

Administrative Center: James D. Crapo, M.D. (principal investigator); Edwin K. Silverman, M.D., Ph.D. (principal investigator); Barry J. Make, M.D.; Elizabeth A. Regan, M.D., Ph.D.

Genetic Analysis Center: Terri Beaty, Ph.D.; Ferdouse Begum, Ph.D.; Peter J. Castaldi, M.D., M.Sc.; Michael Cho, M.D.; Dawn L. DeMeo, M.D., M.P.H.; Adel Boueiz, M.D., M.M.Sc.; Marilyn G. Foreman, M.D., M.S.; Eitan Halper-Stromberg; Lystra P. Hayden, M.D., M.M.Sc.; Craig P. Hersch, M.D., M.P.H.; Jacqueline Hetmanski, M.S., M.P.H.; Brian D. Hobbs, M.D.; John E. Hokanson, M.P.H., Ph.D.; Nan Laird, Ph.D.; Christoph Lange, Ph.D.; Sharon M. Lutz, Ph.D.; Merry-Lynn McDonald, Ph.D.; Margaret M. Parker, Ph.D.; Dandi Qiao, Ph.D.; Elizabeth A. Regan, M.D., Ph.D.; Edwin K. Silverman, M.D., Ph.D.; Emily S. Wan, M.D.; Sungho Won, Ph.D.; Phuwanat Sakornsakolpat, M.D.; Dmitry Prokopenko, Ph.D.

Imaging Center: Mustafa Al Qaisi, M.D.; Harvey O. Coxson, Ph.D.; Teresa Gray; MeiLan K. Han, M.D., M.S.; Eric A. Hoffman, Ph.D.; Stephen Humphries, Ph.D.; Francine L. Jacobson, M.D., M.P.H.; Philip F. Judy, Ph.D.; Ella A. Kazerooni, M.D.; Alex Kluiber; David A. Lynch, M.B.; John D. Newell, Jr., M.D.; Elizabeth A. Regan, M.D., Ph.D.; James C. Ross, Ph.D.; Raul San Jose Estepar, Ph.D.; Joyce Schroeder, M.D.; Jered Sieren; Douglas Stinson; Berend C. Stoel, Ph.D.; Juerg Tschirren, Ph.D.; Edwin Van Beek, M.D., Ph.D.; Bram van Ginneken, Ph.D.; Eva van Rikxoort, Ph.D.; George Washko, M.D.; Carla G. Wilson, M.S.

PFT QA Center, Salt Lake City, Utah: Robert Jensen, Ph.D.

Data Coordinating Center and Biostatistics, National Jewish Health, Denver, Colorado: Douglas Everett, Ph.D.; Jim Crooks, Ph.D.; Camille Moore, Ph.D.; Matt Strand, Ph.D.; Carla G. Wilson, M.S.

Epidemiology Core, University of Colorado Anschutz Medical Campus, Aurora, Colorado: John E. Hokanson, M.P.H., Ph.D.; John Hughes, Ph.D.; Gregory Kinney, M.P.H., Ph.D.; Sharon M. Lutz, Ph.D.; Katherine Pratte, M.S.P.H.; Kendra A. Young, Ph.D.

Mortality Adjudication Core: Surya Bhatt, M.D.; Jessica Bon, M.D.; MeiLan K. Han, M.D., M.S.; Barry Make, M.D.; Carlos Martinez, M.D., M.S.; Susan Murray, Sc.D.; Elizabeth Regan, M.D.; Xavier Soler, M.D.; Carla G. Wilson, M.S.

Biomarker Core: Russell P. Bowler, M.D., Ph.D.; Katerina Kechris, Ph.D.; Farnoush Banaei-Kashani, Ph.D.

ECLIPSE Investigators:

Bulgaria: Y. Ivanov, Plevan; K. Kostov, Sofia.

Canada: J. Bourbeau, Montreal, Quebec; M. Fitzgerald, Vancouver, British Columbia; P. Hernandez, Halifax, Nova Scotia; K. Killian, Hamilton, Ontario; R. Levy, Vancouver, British Columbia; F. Maltais, Montreal, Quebec; D. O'Donnell, Kingston, Ontario.

Czech Republic: J. Krepelka, Prague.

Denmark: J. Vestbo, Hvidovre.

The Netherlands: E. Wouters, Horn-Maastricht.

New Zealand: D. Quinn, Wellington.

Norway: P. Bakke, Bergen.

Slovenia: M. Kosnik, Golnik.

Spain: A. Agusti, J. Sauleada, Palma de Mallorca.

Ukraine: Y. Feschenko, V. Gavrisyuk, L. Yashina, Kiev; N. Monogorova, Donetsk.

United Kingdom: P. Calverley, Liverpool; D. Lomas, Cambridge; W. MacNee, Edinburgh; D. Singh, Manchester; J. Wedzicha, London.

United States: A. Anzueto, San Antonio, Texas; S. Braman, Providence, Rhode Island; R. Casaburi, Torrance, California; B. Celli, Boston, Massachusetts; G. Giessel, Richmond, Virginia; M. Gotfried, Phoenix, Arizona; G. Greenwald, Rancho Mirage, California; N. Hanania, Houston, Texas; D. Mahler, Lebanon, New Hampshire; B. Make, Denver, Colorado; S. Rennard, Omaha, Nebraska; C. Rochester, New Haven, Connecticut; P. Scanlon, Rochester, Minnesota; D. Schuller, Omaha, Nebraska; F. Sciurba, Pittsburgh, Pennsylvania; A. Sharafkhaneh, Houston, Texas; T. Siler, St. Charles, Missouri; E. Silverman, Boston, Massachusetts; A. Wanner, Miami, Florida; R. Wise, Baltimore, Maryland; R. ZuWallack, Hartford, Connecticut.

ECLIPSE Steering Committee: H. Coxson (Canada), C. Crim (GlaxoSmithKline, United States), L. Edwards (GlaxoSmithKline, United States), D. Lomas (United Kingdom), W. MacNee (United Kingdom), E. Silverman (United States), R. Tal-Singer (cochair, GlaxoSmithKline, United States), J. Vestbo (cochair, Denmark), J. Yates (GlaxoSmithKline, United States).

ECLIPSE Scientific Committee: A. Agusti (Spain), P. Calverley (United Kingdom), B. Celli (United States), C. Crim (GlaxoSmithKline, United States), B. Miller (GlaxoSmithKline, United States), W. MacNee (chair, United Kingdom), S. Rennard (United States), R. Tal-Singer (GlaxoSmithKline, United States), E. Wouters (the Netherlands), J. Yates (GlaxoSmithKline, United States).

GenKOLS Investigators:

Per Bakke and Amund Gulsvik

References

- Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* 2010;6:e1000888.
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012;337:1190–1195.
- Marsh SE, Travers J, Weatherall M, Williams MV, Aldington S, Shirlcliffe PM, et al. Proportional classifications of COPD phenotypes. *Thorax* 2008;63:761–767.
- Celli BR. Roger S. Mitchell lecture. Chronic obstructive pulmonary disease phenotypes and their clinical relevance. *Proc Am Thorac Soc* 2006;3:461–465.
- de Torres JP, Bastarrika G, Zagaceta J, Sáiz-Mendiguren R, Alcaide AB, Seijo LM, et al. Emphysema presence, severity, and distribution has little impact on the clinical presentation of a cohort of patients with mild to moderate COPD. *Chest* 2011;139:36–42.
- Han MK, Bartholmai B, Liu LX, Murray S, Curtis JL, Sciurba FC, et al. Clinical significance of radiologic characterizations in COPD. *COPD* 2009;6:459–467.
- Fishman A, Martinez F, Naunheim K, Piantadosi S, Wise R, Ries A, et al.; National Emphysema Treatment Trial Research Group. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med* 2003;348:2059–2073.
- Venuta F, Anile M, Diso D, Carillo C, De Giacomo T, D'Andrilli A, et al. Long-term follow-up after bronchoscopic lung volume reduction in patients with emphysema. *Eur Respir J* 2012;39:1084–1089.
- Deslée G, Mal H, Dutau H, Bourdin A, Vergnon JM, Pison C, et al.; REVOLENS Study Group. Lung volume reduction coil treatment vs usual care in patients with severe emphysema: the REVOLENS randomized clinical trial. *JAMA* 2016;315:175–184.
- Sciurba FC, Chandra D, Bon J. Bronchoscopic lung volume reduction in COPD: lessons in implementing clinically based precision medicine. *JAMA* 2016;315:139–141.
- Martinez FJ, Foster G, Curtis JL, Criner G, Weinmann G, Fishman A, et al.; NETT Research Group. Predictors of mortality in patients with emphysema and severe airflow obstruction. *Am J Respir Crit Care Med* 2006;173:1326–1334.
- Boueiz A, Lutz SM, Cho MH, Hersh CP, Bowler RP, Washko GR, et al.; COPDgene and ECLIPSE Investigators. Genome-wide association study of the genetic determinants of emphysema distribution. *Am J Respir Crit Care Med* 2017;195:757–771.
- El Boueiz A, Chase R, Lamb A, Naing ZZ, Parker MM, Hersh CP, et al. Integrative analysis identifies candidate causal genes of emphysema distribution in non-alpha 1-antitrypsin deficient smokers [abstract]. *Am J Respir Crit Care Med* 2017;195:A7350.
- Cho MH, Castaldi PJ, Hersh CP, Hobbs BD, Barr RG, Tal-Singer R, et al.; NETT Genetics, ECLIPSE, and COPDgene Investigators. A genome-wide association study of emphysema and airway quantitative imaging phenotypes. *Am J Respir Crit Care Med* 2015;192:559–569.
- Coxson HO, Rogers RM, Whittall KP, D'yachkova Y, Paré PD, Sciurba FC, et al. A quantification of the lung surface area in emphysema using computed tomography. *Am J Respir Crit Care Med* 1999;159:851–856.
- Parker MM, Chase RP, Lamb A, Reyes A, Saferali A, Yun JH, et al. RNA sequencing identifies novel non-coding RNA and exon-specific effects associated with cigarette smoking. *BMC Med Genomics* 2017;10:58.
- Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 2010;11:R25.
- Stegle O, Parts L, Durbin R, Winn J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput Biol* 2010;6:e1000770.
- Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 2012;28:1353–1358.
- GTEX Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013;45:580–585.
- Castaldi PJ, Cho MH, Zhou X, Qiu W, McGeachie M, Celli B, et al. Genetic control of gene expression at novel and established chronic obstructive pulmonary disease loci. *Hum Mol Genet* 2015;24:1200–1210.
- Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014;10:e1004383.
- Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al.; Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518:317–330.
- Ernst J, Kellis M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat Biotechnol* 2015;33:364–376.
- John S, Sabo PJ, Thurman RE, Sung MH, Biddie SC, Johnson TA, et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet* 2011;43:264–268.
- Trynka G, Westra HJ, Slowikowski K, Hu X, Xu H, Stranger BE, et al. Disentangling the effects of colocalizing genomic annotations to functionally prioritize non-coding variants within complex-trait loci. *Am J Hum Genet* 2015;97:139–152.
- Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 2015;518:337–343.
- Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 2014;15:R29.
- Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. *Nature* 2012;489:75–82.
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.
- Cho MH, McDonald ML, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, et al.; NETT Genetics, ICGN, ECLIPSE and COPDgene Investigators. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med* 2014;2:214–225.
- Königshoff M, Kneidinger N, Eickelberg O. TGF- β signaling in COPD: deciphering genetic and cellular susceptibilities for future therapeutic regimen. *Swiss Med Wkly* 2009;139:554–563.
- Warburton D, Shi W, Xu B. TGF- β -Smad3 signaling in emphysema and pulmonary fibrosis: an epigenetic aberration of normal development? *Am J Physiol Lung Cell Mol Physiol* 2013;304:L83–L85.
- McDonald ML, Mattheisen M, Cho MH, Liu YY, Harshfield B, Hersh CP, et al.; GenKOLS, COPDgene and ECLIPSE study investigators. Beyond GWAS in COPD: probing the landscape between gene-set associations, genome-wide associations and protein-protein interaction networks. *Hum Hered* 2014;78:131–139.
- Morrow JD, Cho MH, Platig J, Zhou X, DeMeo DL, Qiu W, et al. Ensemble genomic analysis in human lung tissue identifies novel genes for chronic obstructive pulmonary disease. *Hum Genomics* 2018;12:1.
- Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS discovery: biology, function, and translation. *Am J Hum Genet* 2017;101:5–22.
- Obeidat M, Nie Y, Fishbane N, Li X, Bossé Y, Joubert P, et al. Integrative genomics of emphysema-associated genes reveals potential disease biomarkers. *Am J Respir Cell Mol Biol* 2017;57:411–418.
- Joehanes R, Zhang X, Huan T, Yao C, Ying SX, Nguyen QT, et al. Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol* 2017;18:16.
- Sun W, Hu Y. eQTL mapping using RNA-Seq data. *Stat Biosci* 2013;5:198–219.
- Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell* 2017;169:1177–1186.