Online Data Supplement

Genetic regulation of gene expression in the lung identifies CST3 and CD22 as potential causal genes for airflow obstruction

Maxime Lamontagne*[1], Wim Timens*[2], Ke Hao*[3], Yohan Bossé[1,4], Michel Laviolette[1], Katrina Steiling[5], Joshua D Campbell[5], Christian Couture[1], Massimo Conti[1], Karen Sherwood[6], James C. Hogg[6,7], Corry-Anke Brandsma[2], Maarten van den Berge[8], Andrew Sandford[6,9], Stephen Lam[10], Marc E Lenburg[5], Avrum Spira[5], Peter D Paré[6,9], David Nickle[11], Don D. Sin[6,9], Dirkje S. Postma[†8]

Methods

Subject selection

The methods for subject selection and phenotyping as well as for interrogation of gene expression and genotype were recently described[1]. The lung tissue used for discovery of eQTLs was from 1,111 human subjects who underwent lung surgery at three academic sites, Laval University, University of British Columbia (UBC), and University of Groningen, henceforth referred to as Laval, UBC, and Groningen, respectively. For the present study, the principal aim was to examine smoking-related airway obstruction. Thus, we excluded subjects whose lung function may have been influenced by lung diseases other than COPD. Exclusion criteria were: 1) missing data for lung function, i.e. for both Forced Expiratory Volume in 1 second as a percentage of its predicted value (FEV_1 % predicted) and that divided by Forced Vital Capacity (FEV_1/FVC) and 2) patients with asthma, extensive pneumonia, cystic fibrosis, bronchiectasis, pulmonary fibrosis, pulmonary hypertension, primary bullous emphysema, mesothelioma, diffuse alveolar damage and alpha-1-antitrypsin deficiency. We did not exclude individuals with lung cancer, focal pneumonia or atelectasis, but in these cases lung tissue samples were taken as far away as possible from involved areas. Although lung function in theory could be compromised in these cases, the vast majority of subjects included in the biobanks from the three participating sites had

relatively small, non-obstructing lung tumors which were unlikely to significantly interfere with the subjects' lung function. Of the 1,111 subjects in whom there were data on genotype and gene expression, 848 with sufficient phenotypic information and without a lung disease (other than COPD and lung cancer) were included in the analysis. The demographic and clinical features of the subjects in the three cohorts are described in **Table 1**.

COPD phenotypes

For the primary analyses, three phenotypes were used: FEV_1 % predicted and FEV_1/FVC as continuous variables, and COPD defined dichotomously based on an FEV_1/FVC < 0.7 cutoff. We used post-bronchodilator spirometry when available; otherwise, pre-bronchodilator values were used.

Genome-wide association study

Genome-wide association was performed on the three phenotypes using linear or logistic regression models. Single-marker association tests were run within each cohort adjusting for age, gender and smoking status. Furthermore, we conducted a fixed-effects meta-analysis combining the three cohorts using inverse standard

error weighting. We performed genomic control (GC) correction for individual cohorts and for the meta-analysis. To avoid over-correction, we computed the genomic inflation factor (□) on the 90% of SNPs with the largest p-values. The □ estimates were small (≤1.03) for the individual cohort analyses and the meta-analysis.

Causality Models

We evaluated three competing causality models to describe the relationship between lung eQTL-SNPs, RNA expression and COPD phenotypes (**Figure 1**). We tested three linear equations in describing the molecular relationships,

$$T_i = \alpha_1 + \beta_1 L_i + \varepsilon_{1i} \quad \text{(eq 1)}$$

$$G_i = \alpha_2 + \beta_2 T_i + \beta_3 L_i + \varepsilon_{2i} \quad \text{(eq 2)}$$

$$T_i = \alpha_3 + \beta_4 G_i + \beta_5 L_i + \varepsilon_{3i} \quad \text{(eq 3)}$$

A SNP at a specific locus is denoted by L, gene expression for a specific transcript by G, and a measured clinical endpoint by T (i.e. FEV_1 % predicted, FEV_1/FVC or COPD yes/no). Dependencies are likely to exist between certain pairs of covariates in the preceding three models. We inferred a causal relationship using conditional correlation[2]. We required a p-value for $\beta_1 < 1 \times 10^{-3}$ before examining a particular (L, G, T) variable triplet. If a p-value for $\beta_3 < 0.05$ and p-value for $\beta_5 > 0.05$, the causal model was selected. If the converse was true, a reactive model was

selected. If both p-values were < 0.05, an independence model was selected. If both p-values were > 0.05 then no call was made. Model 1 indicates a potential causal relationship and was the main interest of this work.

Analyses were performed in the three cohorts separately and then combined into a meta-analysis. We then conducted sample bootstrapping and repeated the causality test for N=1000 realizations. The reliability score of the molecular relationship is the fraction of bootstrap realizations that supports the call observed.

Manual curation of significant and reliable causal models

To explore the biology and possible role of genes in the causal model, experts in the pathobiology of COPD at each of the participating sites manually reviewed the literature for functionality and known genetic associations. Several strategies were used including PubMed searches using the gene's name and COPD or emphysema, interrogation of NCBI PubMed Gene site, GeneRIFs (Gene Reference Into Functions database), PheGenI for eQTL, and phenotype association data.

Pathway analyses on causality results

Pathway analyses were performed on causality results in order to search for enrichments of specific gene pathways. Genes that fitted Model 1 (causal) with

reliability scores > 0.8 were analyzed using the Ingenuity Pathways Analysis (IPA, Ingenuity® Systems, www.ingenuity com). Causal genes were mapped to corresponding gene objects in IPA using official gene symbols and overlaid onto canonical pathways contained in the Ingenuity Pathways Knowledge Base. The latter analysis identified pathways from the IPA library of canonical pathways that were most enriched with causal genes. The significance of the association with canonical pathways was determined using a right-tailed Fisher's exact test. This test compared the number of causal genes versus total genes in a canonical pathway beyond that expected by chance alone. A total of 181 canonical pathways were tested. A nominal p-value < 0.05 was considered significant. Similarly the same list of causality genes was examined for enrichment in gene ontology and functional categories using MetaCore™ (version 6.12, build 42289, GeneGo, Inc.) and mapped to both the Gene Ontology (GO) project and proprietary ontologies in the MetaCore™ knowledge database. The significance of functional enrichment of genes was determined by using a False Discovery Rate (FDR) cutoff of 0.05. Network analysis was also carried out using canonical networks in the Metacore™ knowledge database. Direct interaction networks based on our gene lists were built using seed nodes and their direct interactions were assessed using curated, known interactions. Pathway analysis was also carried out using Partek Genomic Suite 6.6 software (Partek GS, version 6.12.0907, www.partek.com). Causality genes were

mapped to official gene symbols in Partek GS and parsed to Partek Pathway software. Partek Pathway interrogates the REACTOME (www.reactome.org) and KEGG (www.genome.jp/kegg/) databases and identifies canonical pathways that are enriched with causality genes. Enrichment score p-values were derived using Fisher's Exact test. A p-value < 0.05 was considered significant.

Replication studies

Bronchial airway epithelial dataset. Bronchial airway brushings were obtained during bronchoscopy from active and former smokers who were enrolled in a lung cancer screening program at the British Columbia Cancer Research Agency[3]. Institutional Review Board approval was obtained at participating institutions, and all subjects provided written informed consent. RNA isolated from bronchial brushings of 238 lung cancer-free active and former smokers with and without COPD was profiled using Affymetrix Human Gene 1.0 ST Arrays (Affymetrix Inc., Santa Clara, CA, USA). Microarray data were deposited in the Gene Expression Omnibus (GSE37147). Gene expression estimates were derived and normalized as previously described[3]. Gene expression levels associated with FEV_1 % predicted, FEV_1/FVC and the presence of COPD were determined using linear modeling as previously described[3].

Regional lung tissue dataset. Whole lungs were explanted from patients with severe COPD (n=6) and from donors (n=2). Each lung was sampled in eight consecutive regions from the apex to the base of the lung. This study was approved by the institutional review board and written informed consent was obtained from each patient prior to surgery or from the next of kin of the persons who the donated lung. The gene expression profiles were obtained in 64 samples (8 patients x 8 regions) using the Human Exon 1.0 ST array. The gene expression dataset was deposited in the Gene Expression Omnibus (GSE27597). Details of sample collection and processing as well as analytical methods to normalize and obtain gene expression values have been previously described[4]. The degree of emphysema in each sample was quantified by measuring the mean linear intercept (Lm) on micro-computed tomography (CT) scans for tissues. Lm represents a morphological measurement of alveolar destruction and is a surrogate for emphysema severity. In this study, genes were associated with regional emphysema severity within the same lung using linear models as described before[4].

Table S2. Models of biological relevance identified by manual curation

SNPs	Gene (probe set)	Gene role & references
FEV1 % predicted		
rs769178	*NCR3* (100125842_TGI_at)	SNPs near *NCR3* gene were associated with lung function[5].
rs6048956	*CST3* (100307577_TGI_at)	Protease-antiprotease balance[6].
rs6515375	*CST3* (100125967_TGI_at)	Protease-antiprotease balance[6].
rs2270859	*CSTA* (100148334_TGI_at)	Expression of CSTA is modulated by genotype, smoking, COPD and lung cancer[7].
rs4550905	*PPARGC1A* (100131093_TGI_at)	The PPAR□-PGC-1α pathway regulates antioxidant genes and protects against COPD in rats[8].
rs3803761	*FLCN* (100135396_TGI_at)	Mutations in *FLCN* cause Birt-Hogg-Dubé syndrome, a monogenic disorder characterized by spontaneous pneumothorax. Genetic variants in *FLCN* were not associated with severe, early-onset COPD[9].
rs1543438	*BCL2L1* (100158784_TGI_at)	Cigarette smoke extract induces the expression of BCL2L1in human dendritic cells and augments survival of these cells in COPD patients[10]
rs9880397	*CADM2* (100162763_TGI_at)	SNPs in *CADM2* were associated with lung function and asthma (unpublished).
rs2466183	TNFRSF10B (100153254_TGI_at)	Involved in T cell and eosinophil regulation in bronchial smooth muscle cell death in asthma[11]. Increased expression of the protein in the lung of subjects with

		emphysema[12].
FEV_1/FVC		
rs17754977	*GSTO2* (100132911_TGI_at)	GSTO enzymes protect against oxidative stress. Involved in the biotransformation of arsenic found in cigarette smoke[13]. SNPs in *GSTO2* were associated with FEV_1, FVC[14] and COPD[15].
rs9987135	*DEPDC6* (100154484_TGI_at)	DEPDC6 is a negative regulator of mTORC1 and mTORC2 signaling pathways. Decreased TOR activity has been found to slow aging in yeast, worms, flies, and mice[16]. The mTOR inhibitor, rapamycin, increases the lifespan in mice and reduces bronchial hyperresponsiveness and airway remodelling[17].
rs10411704	*CD22* (100154732_TGI_at)	CD22 is present on the surface of (mature) B cells and prevents over activation of the immune system and development of autoimmune system.
rs12179536	*MUC22* (100304000_TGI_at)	Polymorphisms in *MUC22* were associated with panbronchiolitis[18]. MUC22 is expressed in the lung.
rs2287765	*SPINK5* (100305138_TGI_at)	*SPINK5* is a candidate gene for asthma and allergy[19]. Upregulation of SPINK5 in epithelial cell line increases inflammatory responses[20].

REFERENCES

1. Hao K, Bosse Y, Nickle DC, et al. Lung eQTLs to Help Reveal the Molecular Underpinnings of Asthma. *PLoS Genet* 2012;8:e1003029.

2. Schadt EE, Lamb J, Yang X, et al. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet* 2005;37:710-7.

3. Steiling K, van den Berge M, Hijazi K, et al. A dynamic bronchial airway gene expression signature of chronic obstructive pulmonary disease and lung function impairment. *Am J Respir Crit Care Med* 2013;187:933-42.

4. Campbell JD, McDonough JE, Zeskind JE, et al. A gene expression signature of emphysema-related lung destruction and its reversal by the tripeptide GHK. *Genome medicine* 2012;4:67.

5. Soler Artigas M, Loth DW, Wain LV, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011;43:1082-90.

6. Abboud RT, Vimalanathan S. Pathogenesis of COPD. Part I. The role of protease-antiprotease imbalance in emphysema. *Int J Tuberc Lung Dis* 2008;12:361-7.

7. Butler MW, Fukui T, Salit J, et al. Modulation of cystatin A expression in human airway epithelium related to genotype, smoking, COPD, and lung cancer. *Cancer Res* 2011;71:2572-81.

8. Li J, Dai A, Hu R, et al. Positive correlation between PPARgamma/PGC-1alpha and gamma-GCS in lungs of rats and patients with chronic obstructive pulmonary disease. *Acta Biochim Biophys Sin (Shanghai)* 2010;42:603-14.

9. Cho MH, Klanderman BJ, Litonjua AA, et al. Folliculin mutations are not associated with severe COPD. *BMC Med Genet* 2008;9:120.

10. Vassallo R, Walters PR, Lamont J, et al. Cigarette smoke promotes dendritic cell accumulation in COPD; a Lung Tissue Research Consortium study. *Respir Res* 2010;11:45.

11. Solarewicz-Madejek K, Basinski TM, Crameri R, et al. T cells and eosinophils in bronchial smooth muscle cell death in asthma. *Clin Exp Allergy* 2009;39:845-55.

12. Morissette MC, Vachon-Beaudoin G, Parent J, et al. Increased p53 level, Bax/Bcl-x(L) ratio, and TRAIL receptor expression in human emphysema. *Am J Respir Crit Care Med* 2008;178:240-7.

13. Lesseur C, Gilbert-Diamond D, Andrew AS, et al. A case-control study of polymorphisms in xenobiotic and arsenic metabolism genes and arsenic-related bladder cancer in New Hampshire. *Toxicol Lett* 2012;210:100-6.

14. Wilk JB, Walter RE, Laramie JM, et al. Framingham Heart Study genome-wide association: results for pulmonary function measures. *BMC Med Genet* 2007;8 Suppl 1:S8.

15. Yanbaeva DG, Wouters EF, Dentener MA, et al. Association of glutathione-S-transferase omega haplotypes with susceptibility to chronic obstructive pulmonary disease. *Free Radic Res* 2009;43:738-43.

16. Evans DS, Kapahi P, Hsueh WC, et al. TOR signaling never gets old: aging, longevity and TORC1 activity. *Ageing Res Rev* 2011;10:225-37.

17. Kramer EL, Hardie WD, Mushaben EM, et al. Rapamycin decreases airway remodeling and hyperreactivity in a transgenic model of noninflammatory lung disease. *J Appl Physiol* 2011;111:1760-7.

18. Hijikata M, Matsushita I, Tanaka G, et al. Molecular cloning of two novel mucin-like genes in the disease-susceptibility locus for diffuse panbronchiolitis. *Hum Genet* 2011;129:117-28.

19. Kabesch M, Carr D, Weiland SK, et al. Association between polymorphisms in serine protease inhibitor, kazal type 5 and asthma phenotypes in a large German population sample. *Clin Exp Allergy* 2004;34:340-5.

20. Birben E, Sackesen C, Turgutoglu N, et al. The role of SPINK5 in asthma related physiological events in the airway epithelium. *Respir Med* 2012;106:349-55.