Abstract
Previous expression quantitative trait loci (eQTL) studies have performed genetic association studies for gene expression, but most of these studies examined lymphoblastoid cell lines from non-diseased individuals. We examined the genetics of gene expression in a relevant disease tissue from chronic obstructive pulmonary disease (COPD) patients to identify functional effects of known susceptibility genes and to find novel disease genes. By combining gene expression profiling on induced sputum samples from 130 COPD cases from the ECLIPSE Study with genomewide single nucleotide polymorphism (SNP) data, we found 4315 significant cis-eQTL SNP-probe set associations (3309 unique SNPs). The 3309 SNPs were tested for association with COPD in a genomewide association study (GWAS) dataset, which included 2940 COPD cases and 1380 controls. Adjusting for 3309 tests (p < 1.5e-5), the two SNPs which were significantly associated with COPD were located in two separate genes in a known COPD locus on chromosome 15: CHRNA5 and IREB2. Detailed analysis of chromosome 15 demonstrated additional eQTLs for IREB2 mapping to that gene. eQTL SNPs for CHRNA5 mapped to multiple linkage disequilibrium (LD) bins. The eQTLs for IREB2 and CHRNA5 were not in LD. Seventy-four additional eQTL SNPs were associated with COPD at p < 0.10. These were genotyped in two COPD populations, finding replicated associations with a SNP in PSORS1C1, in the HLA-C region on chromosome 6. Integrative analysis of GWAS and gene expression data from relevant tissue from diseased subjects has located potential functional variants in two known COPD genes and has identified a novel COPD susceptibility locus.
Introduction

Gene expression levels in humans are highly heritable [1,2]. Multiple published studies have examined the associations between single nucleotide polymorphism (SNP) variation and microarray gene expression measurements to identify expression Quantitative Trait Loci (eQTLs), single nucleotide polymorphisms (SNPs) that influence gene expression [3–5]. However, most of the published studies have examined gene expression in lymphoblastoid cell lines (LCL) from unphenotyped individuals [3,4], though a recent paper has described eQTLs in peripheral blood CD4+ lymphocytes of patients with asthma [6]. Integrative genomic analyses can provide functional information regarding significant SNPs found through genomewide association studies (GWAS) or identify the key genes within a locus identified through GWAS. For example, genome-wide expression profiling in LCL from children with asthma [5] was used to localize ORMDL3 (ORM1-like 3 (S. cerevisiae) [MIM 610075]) as the likely gene for childhood asthma in the multi-gene chromosome 17q21 locus found through GWAS [7]. However, this study did not determine whether the eQTLs identified were relevant in primary human tissues in asthma. Integrative genomics studies can also be used to implicate novel genes for complex traits, such as the association between MMP20 (matrix metalloproteinase 20 [MIM 604629]) and age related decline in kidney function [8].

Chronic obstructive pulmonary disease (COPD [MIM 606963]), which includes emphysema and chronic bronchitis, is a complex disease with genetic and environmental influences [9]. COPD is a major source of morbidity and mortality in the U.S. and worldwide [10]. Previous GWAS have identified three susceptibility loci for COPD, including HHIP (hedgehog interacting protein [MIM 606178]), FAM13A (family with sequence similarity 13, member A [MIM 613299]), and a multi-gene locus on chromosome 15q25 containing candidate genes CHRNA5 (cholinergic receptor, nicotinic, alpha 5 [MIM 118505]), CHRNA3 (MIM 118503), and IREB2 (iron-responsive element binding protein 2 [MIM 147582]) [11–13]. Cough and phlegm production is common among COPD patients, and sputum samples may provide a non-invasive window into pathobiologic processes in the lungs of COPD patients. Therefore, we integrated GWAS data with microarray gene expression profiles from induced sputum samples from well-characterized COPD subjects participating in the Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE) Study [14]. We addressed two hypotheses: (1) eQTL analysis will improve understanding of previously known COPD susceptibility loci, such as chromosome 15q25; and (2) eQTL SNPs can be used to identify novel COPD susceptibility genes. Limiting the search to functional eQTL SNPs can reduce the multiple testing burden found in traditional GWAS. Although eQTL studies have now been performed in several human tissues besides blood, our study represents one of the first integrative genomics analyses performed in affected patients in order to gain insights into a common disease.

Methods

Ethics Statement

Study subjects provided written informed consent, and all studies were approved by the Institutional Review Boards at Partners Healthcare and all participating centers.

ECLIPSE Study

ECLIPSE was a three year observational study conducted at 46 centers in 12 countries [14]. ECLIPSE recruited 2083 COPD subjects ages 40–75 with a smoking history of at least 10 pack-years (cigarettes smoked per day multiplied by years smoked, divided by 20 to convert to packs), 332 control smokers with at least 10 pack-years smoking history and normal lung function, and 237 non-smoking controls [15]. COPD was defined by GOLD stage 2 or greater (FEV1/FVC<0.7 with FEV1<80% predicted) [10]. Genome-wide SNP genotyping was performed on all ECLIPSE subjects using the Illumina HumanHap550 BeadChip. GWAS analysis included 1736 cases COPD cases and 175 controls [11]. Sputum induction was performed on a subset of COPD cases at 14 sites, using a standard protocol [16]. RNA was extracted from sputum cell pellets using TRIzol and amplified with the Nugen Ovation RNA Amplification kit. Gene expression profiling was performed on RNA extracted from sputum samples of 145 COPD cases (all ex-smokers) using the Affymetrix Human U133 Plus+2 array [17]. MIAME-compliant array data are available in the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo), accession GSE22148. Only Caucasian subjects were included in this analysis.

Other GWAS Populations

Subjects from two additional COPD case-control studies were merged with the ECLIPSE subjects in the combined GWAS analysis and the GWAS meta-analysis [11]. COPD cases and control smokers were Caucasians recruited in Bergen, Norway [18,19]. Cases were defined by GOLD stage 2 or greater COPD; smoking controls had normal lung function. Both cases and controls had smoking history of at least 2.5 pack-years. GWAS included 838 cases and 791 controls, genotyped using the Illumina HumanHap550 BeadChip [12].

The National Emphysema Treatment Trial (NETT) cases have FEV1<54.5% predicted and emphysema on chest CT scan [20,21]. Thus, NETT cases have COPD severity of GOLD Stage 3 or greater. All NETT Genetics Ancillary Study subjects are former smokers; only white subjects are included in this analysis. The Normative Aging Study (NAS) is a cohort study of initially healthy men followed by the Boston VA [22]. To define a control group for comparison to NETT cases, we selected Caucasian subjects meeting the following criteria: FEV1>80% predicted, FEV1/FVC>90% predicted, at least 10 pack-years of smoking, and an adequate DNA sample [23]. Genomewide SNP genotyping has been performed in the NETT-NAS study (366 cases, 414 controls) using the Illumina 610-Quad BeadChip [11].

Replication Populations

The International COPD Genetics Network (ICGN) was a family-based study of COPD at ten centers in North America and Europe [18,24]. Proband were ages 45-65 with post-bronchodilator FEV1<60% predicted, FEV1/FVC<90% predicted, a smoking history of at least 5 pack-years, and at least one sibling with ≥5 pack-year smoking history. Genotyping was performed on Caucasian subjects only (Table 1).

The Genetic Epidemiology of COPD Study (COPDGene) enrolled COPD cases and control smokers at 21 clinical centers throughout the United States [25]. Subjects are 45–80 years old and have a smoking history of at least 10 pack-years. This analysis included the first 994 non-Hispanic white case and control subjects enrolled in COPDGene (Table 1). In these samples, a set of 77 ancestry informative markers has been previously genotyped and did not show evidence of population stratification [11].

SNP genotyping in ICGN and COPDGene SNPs was done using the iPLEX Gold assay on the Sequenom (San Diego, CA) MassARRAY system [26] or the TaqMan 5’ exonuclease assay (Applied Biosystems, Foster City, CA) [27].
Table 1. Characteristics of Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE) study subjects in the integrative genomics analysis as well as subjects from the International COPD Genetics Network (ICGN) and the Genetic Epidemiology of COPD study (COPDGene) included in follow-up analyses.

|                  | ECLIPSE | ICGN         | COPDGene |
|------------------|---------|--------------|----------|
|                  | Cases   | Probands     | Relatives| Cases   | Controls  |
| N                | 131     | 983          | 1876     | 496     | 498       |
| Male sex         | 87 (66.4%) | 580 (59.0%) | 970 (51.7%) | 244 (49.2%) | 251 (50.4%) |
| Age              | 64.9 (±5.5) | 58.4 (±5.4) | 57.9 (±9.5) | 64.7 (±8.1) | 60.3 (±8.6) |
| Pack-years of smoking | 46.8 (±28.3) | 52.7 (±29.5) | 39.2 (±25.3) | 54.8 (±26.8) | 38.8 (±21.0) |
| Current smoker   | 0       | 320 (32.6%) | 956 (51.0%) | 149 (30.0%) | 168 (33.7%) |
| Post-bronchodilator FEV₁, % predicted | 49.3 (±15.3) | 35.5 (±13.1) | 83.3 (±26.0) | 48.8 (±18.4) | 98.0 (±11.3) |
| Post-bronchodilator FEV₁/FVC      | 0.43 (±0.12) | 0.37 (±0.12) | 0.64 (±0.14) | 0.48 (±0.13) | 0.78 (±0.05) |

Values are presented as mean (±SD) or N (%).

**Results**

Sputum eQTL Analysis

Characteristics of the 131 ECLIPSE COPD subjects in the eQTL analysis are shown in Table 1. On average, COPD subjects had a heavy smoking history and severely impaired lung function, similar to the full set of ECLIPSE GWAS cases [11]. The data analysis is outlined in Figure 1. Combining the gene expression data with genomewide SNP data and limiting analysis to potential cis-acting SNPs (within 50 kb of TSS) yielded 562,787 SNP-probe set association tests. Of these, 4313 SNP-probe set associations were significant at FDR-adjusted p<0.05 (corresponding to unadjusted p<3.8e-4), representing 3309 unique SNPs and 1399 unique probe sets, covering 1066 genes (Table S1).

The top eQTL was for SNP rs104664 within the gene *FAM118A* (family with sequence similarity 118, member A). This SNP was found to be highly associated with COPD loci.

We also tested for eQTL SNPs influencing the expression of genes in previously identified COPD loci. On chromosome 13q25, we defined a region starting 50 kb centromeric from *IREB2* extending 50 kb telomeric from *CHRN4* (approx. 300 kb total) and tested all genotyped SNPs within this region for association with expression levels of probe sets for six genes: *IREB2, AGPDH1, FSHA4, CHRNA5, CHRNA3, and CHRN4*. For the other two COPD loci, we expanded the cis-eQTL analysis to all SNPs with 200 kb of the TSS of the genes *HHIP* and *FAM118A*.

**Sputum eQTLs associated with COPD**

We queried the 3309 significant cis-eQTL SNPs in the combined GWAS dataset including ECLIPSE, Norway, and NETT-NAS subjects [11]. Using a strict Bonferroni correction, there were two
cis-eQTL SNPs significantly associated with COPD at p<0.05/3309 = 1.5e-5 (Table 2). These two SNPs on chromosome 15q25 are located in CHRNA5 and IREB2, genes with known COPD associations.[12,36] At a nominal threshold of p<0.01, there were 64 cis-eQTL SNPs associated with COPD (Table S2). These 64 SNPs were significantly associated with COPD in the combined GWAS analysis of the ECLIPSE, Bergen Norway, and National Emphysema Treatment Trial (NETT)-Normative Aging Study (NAS) [11]. COPD = chronic obstructive pulmonary disease. eQTL = expression quantitative trait locus. FDR = false discovery rate. ICGN = International COPD Genetics Network. SNP = single nucleotide polymorphism. doi:10.1371/journal.pone.0024395.g001

Replication Studies

Characteristics of the ICGN and COPDGene subjects in the replication analysis are reported in Table 1. The two SNPs in Table 2 were analyzed in previous reports [12,36] and were not retested. Of the remaining 74 SNPs, 69 were successfully genotyped in ICGN and COPDGene. Screening in the larger ICGN study found 8 SNPs with p<0.1 (Table 3). Of these, only one had p<0.1 in COPDGene. SNP rs1265098 was significantly associated with COPD in ICGN and had a trend for significance in COPDGene. The effect direction for rs1265098 was consistent in ICGN, COPDGene, and the combined GWAS; the minor allele was associated with increased COPD risk in all three studies. SNP rs1265098 maps to the gene PSORS1C1 [psoriasis susceptibility-1 candidate 1 [MIM 613525]] on chromosome 6, yet is associated with transcript levels of the neighboring gene PSORS1C3 (p = 0.2e-05, FDR-adjusted p = 0.016) (Figure 2).

Sputum eQTLs in COPD Candidate Loci

Previous GWAS have identified three loci associated with COPD susceptibility: HHIIP on chromosome 4q31 [12,13], FAM13A on chromosome 4q22 [11], and a region on chromosome 15q25 encompassing candidate genes CHRNA5, CHRNA3 and IREB2, among others [12,36]. On chromosome 15q25, cis-eQTL associations for IREB2 mapped to that gene (Figure 3a). Genetic regulation of CHRNA5 was more complex. Previous studies have demonstrated cis-acting effects of multiple SNPs on CHRNA5 expression. Saccone et al. defined 4 LD bins surrounding CHRNA5 with varying associations with cigarette smoking, lung cancer, and COPD [37]. Bins 1–3 were represented in our dataset, tagged by SNPs rs1051730, rs9360692, and rs6495306, respectively (Table 4). SNPs in bins 1 and 3 were associated with CHRNA5 expression in sputum (Figure 2), as has been demonstrated in brain [30] and lung tissue [39]. SNPs in bin 2 were not eQTLs for CHRNA5. We added additional SNPs to these bins, based on strong LD with tag SNPs in the larger ECLIPSE GWAS dataset. We also identified 3 sets of SNPs (3a, b, c in Table 4) with cis-eQTL associations for CHRNA5 and moderate LD with SNPs in bin 3 (r^2 0.57–0.76). SNPs in bins 1 and 2, but not bin 3, showed evidence of association with COPD in the combined GWAS dataset, though they were not genomewide significant (bin 1: rs1051730, p = 2.8e-6; bin 2: rs9360692, p = 5.6e-5; bin 3: rs6495306, p = 0.2). These results suggest that the COPD-associated SNP rs1051730 (bin 1) may influence phenotype by its effect on gene expression, while COPD-associated SNPs in bin 2 (tagged by rs9360692) may exert their effect through other mechanisms. SNPs in bin 3, although eQTLs, were not associated with COPD risk.

SNPs in IREB2 were both cis-eQTLs for that gene (Table 4, Figure 2) and were associated with COPD in the combined GWAS (rs13180, p = 5.0e-7). Even though some of the significant eQTL SNPs for CHRNA5 mapped to IREB2 (bin 3a), SNPs in all 3 bins were not in LD with the IREB2 eQTL SNPs (Figure 3b). No SNPs were significantly associated with ACPHD1, PSMA4, or CHRNBN4 gene expression. For the other two COPD GWAS loci, HHIIP and FAM13A, we found no significant cis-eQTL SNPs.
within 50 kb, so we expanded the assessment of cis-eQTLs to all SNPs within 200 kb of the TSS of each gene. There were no significant cis-eQTLs within 50 kb, so we expanded the assessment of cis-eQTLs to all SNPs within 200 kb of either HHIP or FAM13A.

### Discussion

In a cohort of well-characterized COPD subjects, we integrated genomewide SNP and gene expression data derived from induced sputum, a biologically-relevant tissue in COPD, to identify a set of eQTL SNPs affecting gene expression levels. The SNPs were then tested for association with the clinical phenotype of COPD; gene expression was not tested for association with disease status in this set of COPD cases only. Using the eQTL results, we implicated two distinct COPD susceptibility genes in a previously identified region of chromosome 15q25. Additionally, we provide evidence for a potential novel COPD susceptibility locus in the HLA region on chromosome 6.

The initial GWAS in COPD found significant associations on chromosome 15q25, with SNPs in the genes CHRNA5 and CHRNA3, encoding two subunits of the nicotinic acetylcholine receptor [12]. This region has also been associated with lung cancer, peripheral arterial disease, and smoking behavior [37,40–43], so it is not clear whether these genes have a direct effect on COPD susceptibility, or their effects are at least partially influenced through cigarette smoking, the major environmental risk factor for COPD [44,45]. In terms of genetic regulation of expression of the chromosome 15q25 genes, we found similar eQTL associations with CHRNA5 expression in induced sputum as has been found in brain [38] and lung tissue [39]. We found additional sputum eQTL SNPs for CHRNA5 in moderate LD with previously defined eQTLs. The previous papers on brain and lung tissue gene expression did not report testing IREB2, a gene previously associated with COPD [11,36]. The specific IREB2 SNPs associated in GWAS (rs13180) [11] and in a candidate gene analysis of differentially expressed genes (rs20560699) [36] were in only moderate LD ($r^2 = 0.44$) with each other, implying independent effects on IREB2 expression. The IREB2 and CHRNA5 eQTL SNPs were not in LD with each other, suggesting the presence of at least two COPD susceptibility genes on chromosome 15q25.

### Table 3. Genetic association analysis of sputum expression quantitative trait locus (eQTL) single nucleotide polymorphisms (SNPs) with COPD susceptibility.

| SNP        | Chrom | BP      | Gene     | Effect on COPD | COPD risk in IGCM | IGCM p-value | COPDDGene OR | COPDGene p-value | GWAS OR* | GWAS p-value* |
|------------|-------|---------|----------|----------------|--------------------|--------------|---------------|------------------|----------|---------------|
| rs1999261  | 6     | 6515106 | INTERGENIC | Increase       | 0.042              | 0.94         | 0.53          | 1.15             | 0.0080   |               |
| rs1265998  | 6     | 31214156| PSORS1C1  | Increase       | 0.024              | 1.20         | 0.098         | 1.18             | 0.0065   |               |
| rs4750277  | 10    | 12954527| INTERGENIC | Increase       | 0.0062             | 0.97         | 0.80          | 0.83             | 0.0097   |               |
| rs1025607  | 12    | 94884637| AMHD1     | Increase       | 0.011              | 1.12         | 0.26          |                 |          |               |
| rs2347279  | 18    | 2528545 | METTL4    | Decrease       | 0.097              | 1.01         | 0.91          | 0.83             | 0.0025   |               |
| rs1878553  | 18    | 2560155 | NDC80     | Decrease       | 0.030              | 1.03         | 0.83          | 0.83             | 0.0028   |               |
| rs4803481  | 19    | 46758396| INTERGENIC | Decrease       | 0.055              | 0.97         | 0.81          |                 |          |               |
| rs2302188  | 19    | 46777713| CEACAM21  | Decrease       | 0.090              | 0.97         | 0.77          | 0.85             | 0.0075   |               |

SNPs with p-value<0.1 in the International COPD Genetics Network (ICGN) are shown.

Odds Ratio and p-value from combined genomewide association study (GWAS) analysis [11] are shown, except as noted: rs1025607 p = 0.0089 in GWAS meta-analysis.

rs4803481 p = 0.0043 in GWAS meta-analysis.

Odds Ratio and p-value from combined genomewide association study (GWAS) analysis [11] are shown, except as noted: rs1025607 p = 0.0089 in GWAS meta-analysis.

rs4803481 p = 0.0043 in GWAS meta-analysis.

doi:10.1371/journal.pone.0024395.t003

Prior studies have also used eQTL analyses to identify novel genes for complex traits, including age related decline in kidney function [8] and body mass index [55]. In contrast to our study, these papers first found gene transcripts correlated with the phenotype, then tested SNPs in/near these genes for association with expression levels. We performed the cis-eQTL analysis as the initial step, then tested the eQTL SNPs for phenotype association. This limits multiple testing compared to a GWAS, enriching for eQTL SNPs which may be more likely to be associated with disease [56].

This study has several limitations. The sample size of 131 subjects, though adequate for gene expression analyses, may be underpowered to detect all potential eQTL associations. Therefore, we limited the cis-acting analysis to SNPs within 50 kb from the gene, to limit the multiple testing burden. Based
Figure 2. Boxplots of sputum gene expression levels stratified by genotype in 131 Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE) subjects with chronic obstructive pulmonary disease.  

- rs1265098 (PSORS1C3, 238997_at), p = 8.2e-5.
- rs13180 (IREB2, 1555476_at), p = 6.7e-9.
- rs1051730 (CHRNA5, 206533_at), p = 2.2e-4; LD bin 1 (see Table 4).
- rs6495306 (CHRNA5, 206533_at), p = 9.9e-6; LD bin 3 (see Table 4).

doi:10.1371/journal.pone.0024395.g002
on RNA sequencing data, Pickrell et al. estimate that 90% of eQTL SNPs are within 15 kb of a gene [57]. Previous papers have used a 50 kb limit to define cis-acting eQTLs [6]. Using this method, we were able to replicate published eQTL associations from other tissues and were able to identify a set of significant eQTL SNPs to carry forward for COPD association studies. However, our method would be unable to detect cis-eQTLs located >50 kb from the TSS, such as a SNP in an upstream enhancer or in the 3' UTR of a large gene. Due to the sample size, we limited our investigation to cis-acting eQTL SNPs, as a full search for trans-acting regulatory SNPs greatly increases the number of tests performed. The literature suggests that sample sizes under 200 subjects may be inadequate to find true trans-eQTLs [58].

Figure 3. Detailed analysis of the chromosome 15q25 chronic obstructive pulmonary disease (COPD) locus. a) Association between single nucleotide polymorphisms (SNPs) in the chromosome 15q25 COPD locus and expression levels of IREB2 (1555476_at), CHRNA5 (206533_at) and CHRNA3 (211587_x_at) in sputum samples from 131 Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE) subjects. SNP rs numbers are listed in Table 4. b) Linkage disequilibrium r² values between SNPs in the chromosome 15q25 COPD locus (listed in Table 4) in 131 ECLIPSE subjects.

doi:10.1371/journal.pone.0024395.g003
Table 4. Single nucleotide polymorphism (SNP) associations with expression of *IREB2* (1555476_at) and *CHRNA5* (206533_at) in induced sputum samples from COPD subjects.

| SNP        | Position | *IREB2* eQTL pvalue | *CHRNA5* eQTL pvalue | LD bin | r² with locus tag | r² with rs13180 |
|------------|----------|---------------------|----------------------|--------|-------------------|-----------------|
| rs4243082  | 78682076 | 0.88                | 0.89                 |        |                   |                 |
| rs11635084 | 78686823 | 0.066               | 0.82                 |        |                   |                 |
| rs2869040  | 78688834 | 0.26                | 0.68                 |        |                   |                 |
| rs6495296  | 78696492 | 0.32                | 0.71                 |        |                   |                 |
| rs4887052  | 78703631 | 0.26                | 0.88                 |        |                   |                 |
| rs7171749  | 78708715 | 0.41                | 0.44                 |        |                   |                 |
| rs12913946 | 78709937 | 0.30                | 0.72                 |        |                   |                 |
| rs7183034  | 78710766 | 0.16                | 0.66                 |        |                   |                 |
| rs4887053  | 78712699 | 0.0012              | 0.11                 |        |                   |                 |
| rs1394371  | 78724469 | 0.0028              | 0.031                |        |                   |                 |
| rs12903150 | 78724645 | 0.014               | 0.0014               |        |                   |                 |
| rs12899131 | 78726885 | 0.71                | 2.8E-05              | 3a     | 0.57              | <0.01           |
| rs10519198 | 78742754 | 0.73                | 3.3E-05              | 3a     | 0.57              | <0.01           |
| rs2656069  | 78745707 | 4.0E-04             | 0.11                 |        | *IREB2*           | 0.44            |
| rs13180    | 78789488 | 6.7E-09             | 0.74                 |        | *IREB2*           | 0.36            |
| rs3743079  | 78791061 | 3.9E-04             | 0.063                |        | *IREB2*           | 0.99            |
| rs1062980  | 78792527 | 7.7E-09             | 0.78                 |        | *IREB2*           | 0.99            |
| rs8034191  | 78806023 | 0.010               | 3.8E-04              | 1      | 0.90              |                 |
| rs3885951  | 78825917 | 0.025               | 0.20                 |        |                   |                 |
| rs2036534  | 78826948 | 0.0015              | 0.043                | 2      | 0.79              |                 |
| rs12915366 | 78831753 | 0.83                | 3.7E-07              | 3      | 0.80              |                 |
| rs2292117  | 78834689 | 0.80                | 6.4E-07              | 3      | 0.82              |                 |
| rs6495306  | 78865893 | 0.79                | 9.9E-06              | 3      |                   | <0.01           |
| rs608024   | 78871288 | 0.79                | 9.9E-06              | 3      | 0.99              |                 |
| rs621849   | 78872861 | 0.79                | 9.9E-06              | 3      | 0.99              |                 |
| rs578776   | 78888400 | 0.016               | 0.70                 | 2      | 0.72              |                 |
| rs12910984 | 78891627 | 0.0061              | 0.45                 | 2      | 0.99              |                 |
| rs1051730  | 78894339 | 0.016               | 2.2E-04              | 1      |                   | 0.19            |
| rs3743077  | 78894896 | 0.93                | 6.7E-05              | 3      | 0.98              |                 |
| rs936682   | 78896547 | 0.0061              | 0.45                 | 2      | 0.26              |                 |
| rs12914385 | 78898723 | 0.014               | 8.0E-04              | 1      | 0.81              |                 |
| rs8042374  | 78908032 | 0.0064              | 0.53                 | 2      | 0.98              |                 |
| rs3743075  | 78909452 | 0.76                | 2.0E-04              | 3b     | 0.76              | 0.01            |
| rs6495309  | 78915245 | 0.088               | 0.64                 | 2      | 0.90              |                 |
| rs1948     | 78917399 | 0.16                | 4.2E-04              | 3c     | 0.60              | 0.03            |
| rs11636753 | 78928946 | 0.91                | 0.0023               |        |                   |                 |
| rs12441998 | 78929372 | 0.14                | 0.85                 |        |                   |                 |
| rs1316971  | 78930510 | 0.083               | 0.89                 |        |                   |                 |
| rs12594247 | 78946633 | 0.63                | 0.62                 |        |                   |                 |
| rs17487514 | 78953785 | 0.10                | 0.068                |        |                   |                 |
| rs1996371  | 78956806 | 0.025               | 0.025                |        |                   |                 |
| rs6495314  | 78960529 | 0.025               | 0.025                |        |                   |                 |
| rs922691   | 78963994 | 0.14                | 0.0060               |        |                   |                 |
| rs8032156  | 78964498 | 0.51                | 0.66                 |        |                   |                 |
| rs8038920  | 78974545 | 0.059               | 0.011                |        |                   |                 |
| rs4887077  | 78978364 | 0.049               | 0.029                |        |                   |                 |
| rs1163872  | 78983559 | 0.049               | 0.029                |        |                   |                 |

Linkage disequilibrium (LD) bins for *CHRNA5* associations are defined as in Saccone et al. [37]. SNPs tags for LD bins are shown in bold. Additional SNPs added to bins based on LD in ECLIPSE genomewide association study dataset are shown in italics.

doi:10.1371/journal.pone.0024395.t004
Several groups have compared eQTLs in different tissues from the same individual, finding both overlapping and tissue-specific eQTLs [59–61]. Multiple tissues are known to be important in COPD biology, including large and small airways, lung parenchyma and immune cells. By only surveying sputum, we may have missed significant eQTLs for COPD genes that are expressed in other tissues. Multiple cell types may be present in sputum, yet neutrophils have been shown to be the predominant cell type in the sputum samples from COPD subjects in ECLIPSE [16]. Despite these limitations, sputum is a clinically important tissue in COPD and is more accessible for genomic and biomarkers studies than lung tissue. Studying diseased individuals may be advantageous to identify eQTL SNPs for potential disease genes, which may only be expressed, or may be expressed at higher levels, in patients compared to healthy controls.

In conclusion, we combined genomewide SNP genotyping with genomewide expression profiling from a relevant tissue in well-characterized subjects with a common chronic disease. Using this strategy, we were able to gain insights into the functional role of SNPs previously associated through GWAS, as well as identify a potential novel disease susceptibility gene which would have been missed using standard GWAS analysis. Previous eQTL studies have provided important information about genetic control of human gene expression. Integrative genomics studies in relevant tissue from well-phenotyped individuals, as we have performed, will be required to apply this knowledge to human disease.

Supporting Information

Figure S1 Principal components plot of RMA expression values, demonstrating lack of batch effects based on hybridization dates or other systematic effects. (TIF)

Data S1 Microarray Quality Control. (DOC)

Table S1 The top cis-expression quantitative trait loci (eQTLs) in sputum samples from 131 ECLIPSE COPD subjects. Single nucleotide polymorphism (SNP)-probe set associations with FDR-adjusted p<0.05 are shown. (XLS)

Table S2 Cis-expression quantitative trait locus (eQTL) single nucleotide polymorphisms (SNPs) from sputum samples from 131 ECLIPSE COPD subjects associated with COPD case-control status in combined ECLIPSE, NETT-NAS, and Norway GWAS (Cho et al. 2010). SNPs associated at p<0.01 are shown. (DOC)

References

1. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, et al. (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422: 297–302.
2. Cheung VG, Conlin AK, Weber TM, Morley M, et al. (2005) Natural variation in human gene expression assessed in lymphoblastoid cells. Nat Genet 33: 422–423.
3. Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, et al. (2005) Mapping determinants of human gene expression by regional and genome-wide association study of global gene expression. Nat Genet 39: 1202–1207.
4. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, et al. (2007) Population genomics of human gene expression. Nat Genet 39: 1217–1224.
5. Hart CS, Bolstad BM, Irizarry RA, Speed TP (2006) ANOVA for cDNA microarray data. Genome Biol 7: R96.
6. Dudoit S, Fridlyand J (2002) A note on reweighting in diffusion maps for gene expression data. Genome Biol 3: research0021.1–0021.14.
7. Xia Y, Xi L, Shi Y, Kong J (2006) Statistical issues in microarray data analysis for disease classification. Stat Med 25: 851–873.
8. Wheeler HE, Metter EJ, Tanaka T, Alsheh D, Higgins J, et al. (2009) Sequential use of transcriptional profiling, expression quantitative trait mapping, and gene expression implicates MMP20 in human kidney aging. PLoS Genet 5: e1000685.
9. Hersh CP, DeMeo DL, Silverman EK (2005) Chronic Obstructive Pulmonary Disease. In: Silverman EK, Shapiro SD, Lomas DA, Weiss ST, eds. Respiratory Disease. In: Silverman EK, Shapiro SD, Lomas DA, Weiss ST, eds. Respiratory Genetics. New York: Hodder Arnold. pp 253–296.
10. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, et al. (2007) Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 176: 532–555.
11. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, et al. (2010) Variants in FAM13A are associated with chronic obstructive pulmonary disease. Nat Genet 42: 200–202.
12. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, et al. (2009) A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS Genet 5: e1000421.
13. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, et al. (2009) A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS Genet 5: e1000421.
14. Wheeler HE, Metter EJ, Tanaka T, Alsheh D, Higgins J, et al. (2009) Sequential use of transcriptional profiling, expression quantitative trait mapping, and gene expression implicates MMP20 in human kidney aging. PLoS Genet 5: e1000685.
15. Hersh CP, DeMeo DL, Silverman EK (2005) Chronic Obstructive Pulmonary Disease. In: Silverman EK, Shapiro SD, Lomas DA, Weiss ST, eds. Respiratory Genetics. New York: Hodder Arnold. pp 253–296.
16. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, et al. (2007) Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 176: 532–555.
17. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, et al. (2010) Variants in FAM13A are associated with chronic obstructive pulmonary disease. Nat Genet 42: 200–202.
18. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, et al. (2009) A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS Genet 5: e1000421.
14. Vesho J, Anderson W, Coxson HO, Crim C, Davber F, et al. (2008) Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). Eur Respir J 31: 869–873.
15. Lomas DA, Silverman EK, Edwards LD, Miller BE, Coxson HO, et al. (2008) Evaluation of serum CC-16 as a biomarker for COPD in the ECLIPSE cohort. Thorax 63: 1058–1063.
16. Singh D, Edwards L, Tal-Singer R, Rennard S (2010) Sputum neutrophils as a biomarker in COPD: findings from the ECLIPSE study. Respir Res 11: 77.
17. Singh D, Fee SM, Tal-Singer R, Pambij B, Bates S, et al. (2011) Induced sputum genes associated with spirometric and radiological disease severity in COPD ex-smokers. Thorax 66: 499–495.
18. Zhu G, Warren L, Aponte J, Gulsvik A, Bakke P, et al. (2007) The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two large populations. Am J Respir Crit Care Med 176: 167–173.
19. Brogger J, Steen VM, Eiken H, Gulsvik A, Bakke P (2006) Genetic association between COPD and polymorphisms in TNF, ADRB2 and EPHX1. Eur Respir J 27: 602–608.
20. The National Emphysema Treatment Trial Research Group (1999) Rationale and design of the National Emphysema Treatment Trial: a prospective randomized trial of lung volume reduction surgery. Chest 116: 1750–1761.
21. Fishman A, Martinez F, Naunheim K, Piantadosi S, Wise R, et al. (2003) A interdisciplinary and longitudinal study of health and aging. Aging Hum Dev 5: 5–17.
22. Bell B, Rose CI, Damon H (1972) The Normative Aging Study: a 20-year follow-up. Science 176: 143–149.
23. Hersh CP, Pillai SG, Zhu G, Lomas DA, Bakke P, et al. (2010) Multistudy fine mapping of chromosome 2q identifies XRCC3 as a chronic obstructive pulmonary disease susceptibility gene. Am J Respir Crit Care Med 182: 605–613.
24. Patel BD, Coxson HO, Pillai SG, Agasti AG, Calverley PM, et al. (2008) Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 178: 500–505.
25. Regan EA, Hakansson JE, Murphy JR, Make B, Lynch DA, et al. (2010) Genetic epidemiology of COPD (COPDGene) study design. COPD 7: 32–43.
26. Storm N, Darshofer-Patel B, van den Boom D, Rudi CP (2005) MALDI-TOF mass-spectrometry-based SNP genotyping. Methods Mol Biol 212: 241–262.
27. Livak KJ (1999) Allelic discrimination using fluorescent probes and the 5’ nucleic acid nuclease assay. Genet Anal 14: 143–149.
28. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4: 249–264.
29. Wan ES, Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, et al. (2010) Genome-Wide Association Analysis of Body Mass in Chronic Obstructive Pulmonary Disease. Am J Respir Cell Mol Biol press.
30. Carey VJ, Davis AR, Lawrence MF, Gentleman R, Raby BA (2009) Data structures and algorithms for analysis of genetics of gene expression with Bioconductor: GGtools 3.x. Bioinformatics 25: 1447–1448.
31. Fishman A, Martinez F, Naunheim K, Piantadosi S, Wise R, et al. (2003) A interdisciplinary and longitudinal study of health and aging. Aging Hum Dev 5: 5–17.
32. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, et al. (2010) Multiple common variants for celiac disease influencing immune gene expression. Nat Genet 42: 295–302.
33. Fransen K, Visschedijk MC, van Sommeren S, Fu JY, Franke L, et al. (2010) Analysis of SNPs with an effect on gene expression identifies UBE2L3 and BCL3 as potential new risk genes for Crohns disease. Hum Mol Genet 19: 3402–3410.
34. Chang YT, Chou CT, Shiao YM, Lin MW, Yu CW, et al. (2006) Psoriasis vulgaris in Chinese individuals is associated with PSORS1C3 and CDSN genes. Br J Dermatol 155: 663–669.
35. Holm SJ, Sanchez F, Carden LM, Mallbris L, Stabile M, et al. (2005) HLA-Cw0602 associates more strongly to psoriasis in the Swedish population than variants of the novel 6p21.3 gene PSORS1C3. Acta Derm Venereol 85: 2–8.
36. Nair RP, Stuart PE, Nistor I, Hirmagarelo R, Chia NV, et al. (2006) Sequence and haplotype analysis supports HLA-C as the psoriasis susceptibility 1 gene. Am J Hum Genet 82: 427–435.
37. Dreijer J, Weitzman D, Shaprio J, Davidovici B, Cohen AD (2008) Psoriasis and chronic obstructive pulmonary disease: a case-control study. Br J Dermatol 159: 956–960.
38. Setty AR, Curhan G, Choi HK (2007) Smoking and the risk of psoriasis in women. Nurses’ Health Study II. Am J Med 112: 953–959.
39. Kauffmann F, Klessbauer JP, Cambon-De-Mouzon A, Mercier P, Constans J, et al. (1993) HLA-A and -B antigens in chronic bronchitis. J Clin Epidemiol 46: 1413–1416.
40. Frank I, Jansen RC (2009) eQTL analysis in humans. Methods Mol Biol 573: 311–326.
41. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, et al. (2008) Genetics of gene expression and its effect on disease. Nature 423: 432–438.
42. Heinzen EL, Ge D, Cronin KD, Maia JM, Shianna KV, et al. (2008) Tissue-specific genetic control of splicing: Implications for the Study of Complex Traits. PLoS Biology 6: e17.