Lung eQTLs to Help Reveal the Molecular Underpinnings of Asthma

Ke Hao1,2,3*, Yohan Bossé4,5*, David C. Nickle1,2,3*, Peter D. Paré6,7, Dirkje S. Postma8, Michel Laviolette5, Andrew Sandford6,7, Tillie L. Hackett5, Denise Daley6,7, James C. Hogg6,9, W. Mark Elliott5, Christian Couture5, Maxime Lamontagne5, Corry-Anke Brandsma10, Maarten van den Berge6, Gerard Koppelman11, Alise S. Reicin2, Donald W. Nicholson2, Vladislav Malkov2, Jonathan M. Derry12, Christine Suver12, Jeffrey A. Tsou2, Amit Kulkarni2, Chunsheng Zhang2, Rupert Vessey2, Greg J. Opiteck2, Sean P. Curtis2, Wim Timens10*, Don D. Sin6,7**

1 Merck Research Laboratories, Boston, Massachusetts, United States of America, 2 Merck, Rahway, New Jersey, United States of America, 3 Genetics, Rosetta Inpharmatics, Merck, Seattle, Washington, United States of America, 4 Department of Molecular Medicine, Laval University, Quebec City, Canada, 5 Institut Universitaire de Cardiologie et de Pneumologie de Québec, Laval University, Quebec City, Canada, 6 The University of British Columbia James Hogg Research Laboratory, St Paul’s Hospital, Vancouver, Canada, 7 Respiratory Division, Department of Medicine, University of British Columbia, Vancouver, Canada, 8 Department of Pulmonology, University Medical Center Groningen, GRIAC Research Institute, University of Groningen, Groningen, The Netherlands, 9 Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, Canada, 10 Department of Pulmonology, University Medical Center Groningen, GRIAC Research Institute, University of Groningen, Groningen, The Netherlands, 11 Department of Pediatric Pulmonology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, 12 Sage Bionetworks, Seattle, Washington, United States of America

Abstract

Genome-wide association studies (GWAS) have identified loci reproducibly associated with pulmonary diseases; however, the molecular mechanism underlying these associations are largely unknown. The objectives of this study were to discover genetic variants affecting gene expression in human lung tissue, to refine susceptibility loci for asthma identified in GWAS studies, and to use the genetics of gene expression and network analyses to find key molecular drivers of asthma. We performed a genome-wide search for expression quantitative trait loci (eQTL) in 1,111 human lung samples. The lung eQTL dataset was then used to inform asthma genetic studies reported in the literature. The top ranked lung eQTLs were integrated with the GWAS on asthma reported by the GABRIEL consortium to generate a Bayesian gene expression network for discovery of novel molecular pathways underpinning asthma. We detected 17,178 cis- and 593 trans- lung eQTLs, which can be used to explore the functional consequences of loci associated with lung diseases and traits. Some strong eQTLs are also asthma susceptibility loci. For example, rs3859192 on chr17q21 is robustly associated with the mRNA levels of GSDMA (P = 3.55 × 10−15). The genetic-gene expression network identified the SOCS3 pathway as one of the key drivers of asthma. The eQTLs and gene networks identified in this study are powerful tools for elucidating the causal mechanisms underlying pulmonary disease. This data resource offers much-needed support to pinpoint the causal genes and characterize the molecular function of gene variants associated with lung diseases.


Editor: Scott M. Williams, Dartmouth College, United States of America

Received March 2, 2012; Accepted August 14, 2012; Published November 29, 2012

Copyright: © 2012 Hao et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Twelve authors are from Merck, which paid for almost all molecular data acquisition and provided the compute cluster that performed much of the analyses. However, Merck as an institution had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

*e-mail: david_nickle@merck.com

* These authors also contributed equally to this work.

**These authors contributed equally to this work.

Introduction

Recent genome-wide association studies (GWAS) have identified loci that harbor susceptibility genes for asthma and other pulmonary conditions [1–16]. Many of the genes at these loci have unknown function and have not previously been considered biologically plausible candidates for disease pathogenesis. Extensive linkage disequilibrium (LD) within these loci makes it difficult to identify the actual susceptibility genes, let alone which genetic variants are responsible for altered expression or function of their protein products. Moreover, the associated polymorphisms can only explain a relatively small proportion of the variability of the phenotype in the population [9] and of its heritability [17].

Integrative genomics is a promising new approach to identify true causal genes and variants. By using gene expression as a phenotype and examining how DNA polymorphisms contribute to both gene expression (expression quantitative trait loci – eQTLs) and disease phenotype, true causal relationships can be discovered [18–20]. In the present study we performed genome-wide genotyping and lung-specific gene expression on a large dataset of lung tissue (1,111 human subjects) to explore effects of genetic variation on gene expression and their joint relationship...
Author Summary

Recent genome-wide association studies (GWAS) have identified genetic variants associated with lung diseases. The challenge now is to find the causal genes in GWAS-nominated chromosomal regions and to characterize the molecular function of disease-associated genetic variants. In this paper, we describe an international effort to systematically capture the genetic architecture of gene expression regulation in human lung. By studying lung specimens from 1,111 individuals of European ancestry, we found a large number of genetic variants affecting gene expression in the lung, or lung expression quantitative trait loci (eQTL). These lung eQTLs will serve as an important resource to aid in the understanding of the molecular underpinnings of lung biology and its disruption in disease. To demonstrate the utility of this lung eQTL dataset, we integrated our data with previous genetic studies on asthma. Through integrative techniques, we identified causal variants and genes in GWAS-nominated loci and found key molecular drivers for asthma. We feel that sharing our lung eQTLs dataset with the scientific community will leverage the impact of previous large-scale GWAS on lung diseases and function by providing much needed functional information to understand the molecular changes introduced by the susceptibility genetic variants.

to asthma. Although the lung tissue came from a heterogeneous group of subjects, the discovery of lung tissue eSNPs could elucidate the causal molecular pathways in a variety of pulmonary disorders.

Results

Demographic characteristics of study participants

Demographic and clinical data for the 1,111 patients who passed clinical, genotyping and gene expression quality control assessments are summarized in Table 1 according to the three sites of recruitment, Laval University, University of British Columbia, and University of Groningen (henceforth referred to as Laval, UBC, and Groningen, respectively). The majority of the subjects were smokers or former-smokers. There were some differences in the clinical characteristics across the sites including age, lung function and smoking status. To account for this heterogeneity between sites, we performed a meta-analysis rather than a pooled analysis for eQTL discovery. The overall analysis workflow is illustrated in Figure 1.

Genome-wide association analysis for expression traits (eQTL)

We identified cis- and trans-acting eQTLs using established methods [19]. For simplicity we assumed that each transcript could have at most one cis-eQTL. We considered all association signals from SNPs within 1 Mb up and downstream, of the transcription probe set as a single cis-eQTL. Trans-eQTLs were defined as association signals from SNPs located greater than 1 Mb from the probe set. The eSNP was identified as the SNP that explained more than 30% of the variance in the transcript and 15.3% of the variance for the probe sets. For each expression trait, statistical significance was assessed using a permutation test (10,000 permutations) to calculate the P value.

We identified 17,049 cis-eQTLs and 534 trans-eQTLs in the Laval, Groningen and UBC cohorts, respectively. There were also variations in the number of detectable trans-eQTLs (Table 2). The P values reported for the trans-eQTLs are lower than those for cis-eQTLs, owing to the higher statistical threshold required to achieve statistical significance. In general, a single SNP (eQTL peak) explained between 10 and 20% of the transcript's expression variance for cis- and trans-eQTLs, respectively (Table S1). However 6.8% of the cis acting SNPs explained more than 30% of the variance in the transcript and 15.3% of the trans acting SNPs explained more than 30% of a transcript's variance (Figure S1).

Further, we investigated the consistency of the eQTLs derived from the cohorts. Rather than simply comparing the existence of eQTLs, we adopted replication criteria similar to GWAS. A successful replication was defined as an eQTL in which the relationship between the same SNP and gene expression (with the same sign) was observed in more than one cohort at a P value of ≤1×10⁻⁴. Using this criterion, 54.5% of the Laval cis-eQTLs were replicated in at least one other cohort and 79.5% of the Laval trans-eQTLs were replicated in another cohort. With respect to the Groningen eQTLs, the replication rates of the cis- and trans-eQTLs were 72.3% and 86.0%, respectively. The replication rates of the UBC cis- and trans-eQTLs were 65.4% and 90.5%, respectively. The eQTLs with higher R² values (variance of gene expression level explained) were more likely to be replicated than those with lower R² values. There was a higher replication rate with trans-eQTLs than with cis-eQTLs.

The meta-analysis revealed many more eQTLs owing to the increased sample size (Table 2). We identified 17,049 cis-eQTLs and 534 trans-eQTLs and the details of all eQTLs are listed in Table S2. 68.7% of these cis-eQTLs and 31.7% of trans-eQTLs were identified in at least one cohort, suggesting that the meta-analysis detected many additional eQTLs. The 'non-combinability' among cohorts is quantified in terms of Cochran's Q and the τ statistics (Table S2). Cochran's Q was calculated as the weighted sum of squared differences among individual study effects and the pooled effect across studies, with the weights proportional to the inverse of variance of the eQTL effect. Q follows a chi-square statistic with k-1 degrees of freedom (k is the number of studies), and by these means, we calibrated the p value (termed Q.pvalue). We detected modest degrees of heterogeneity among the three cohorts. 16% of the QTLs had Q.pvalue<0.05 and 2% of the

| Table 1. Demographic and clinical characteristics of study cohort. |
|-------------------|-------------------|-------------------|
|                   | Laval      | Groningen  | UBC       |
|                   | n = 409   | n = 363   | n = 339   |
| Quantitative traits, mean ± SD |
| Age (years) *     | 63.3±10.0  | 51.8±15.2  | 60.0±14.3 |
| PreBD FEV₁ % predicted * | 80.5±18.9 | 60.5±30.0 | 80.4±22.4 |
| PreBD FVC % predicted * | 89.8±16.1 | 75.0±26.5 | 88.9±19.7 |
| Categorical traits, % |
| Male              | 56.0       | 53.6       | 53.4       |
| Smoking Status *  |
| Non-smoker        | 8.7        | 28.6       | 8.8        |
| Ex-smoker         | 69.0       | 54.9       | 56.8       |
| Current-smoker    | 22.2       | 16.5       | 34.4       |

PreBD, before bronchodilator use; FEV₁, forced expiratory volume in one second; FVC, forced vital capacity.

*significantly different among groups at p<0.01 level (Kruskal Wallis and χ² tests were applied to quantitative and categorical traits, respectively). doi:10.1371/journal.pgen.1003029.0001
eQTLs showed $Q_p < 0.001$. We also present the t statistic, which is the moment-based estimate of the between-study variance and is not dependent on the number of studies. Both the fixed effect and random effect meta-analysis results are presented in Table S2, and we used the fix effect results in the downstream analysis. We previously showed that the number of eQTLs detectable at a certain FDR was related directly to the sample size (i.e. statistical power) [21]. The meta-analysis yielded a greater relative increase in the number of trans-activating eQTLs than of cis-eQTLs, suggesting that sample size is particularly important for detecting trans-eQTLs. Among the 51,627 non-control probesets on our array, 33.0% showed a cis-eQTL. 74.0%, 64.4%, 56.4% of the 51,627 probesets were “present” in at least 20%, 50%, and 80% of the tissue samples. Using these three cutoffs, 40.2%, 42.5%, and 43.4% of the “present” transcripts were found to be controlled by cis-eSNPs, respectively. Accordingly we believe that we have identified the majority of the strong lung cis-eQTLs.

As reported previously [22,23], eQTLs were often found with the expression of more than one gene underlying a GWAS signal (Table 3). For example, the rs7216389-T allele on chromosome 17q was reported to increase asthma risk with an odds ratio (OR) of 1.45 [10]. This SNP was significantly associated with the expression levels of four genes, ORMDL3, GSDMA, GSDMB, and CRKRS (Figure 2a). Although ORMDL3 was originally suggested [10] to be the gene mediating rs7216389-T’s association with asthma, this gene demonstrated the weakest eQTL signal of the four genes in our dataset. Rs7216389-T was positively associated with expression levels of ORMDL3 ($P = 1.6 \times 10^{-7}$), CRKRS ($P = 1.76 \times 10^{-5}$) and GSDMB ($P = 4 \times 10^{-15}$), consistent with the results of Moffat et al. [10] However, this same allele was inversely related to the level of expression of GSDMA ($P = 8.78 \times 10^{-25}$) (Figure 2b). The strongest eSNP in the 17q asthma susceptibility locus was rs3859192 located in intron 6 of the GSDMA gene governing the expression levels of this gene ($P = 3.55 \times 10^{-151}$) (Figure 2b).

To determine the cellular source and relative expression of GSDMB and GSDMA in human lung we performed real-time PCR, and Western blots on primary normal airway epithelial cells and immunohistochemistry on formalin-fixed normal lung tissue (see Text S1 for details). Figure 3 shows abundant mRNA and protein expression of GSDMA but little GSDMB from primary human lung epithelial cells (Figure 3a and 3b respectively). Figure 3c shows that GSDMA is expressed in both apical and basal airway epithelial cells in the conducting airways.
<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene Symbol</th>
<th>ProbeSet</th>
<th>eQTL / (P value)</th>
<th>Meta P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Laval</td>
<td>Groningen</td>
</tr>
<tr>
<td>rs7216389</td>
<td>Moffatt et al., Nature</td>
<td>100127152_TGI_at</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>C T</td>
<td>ORMDL3</td>
<td>100138145_TGI_at</td>
<td>−0.43 (7.60E–11)</td>
<td>−0.41 (4.38E–08)</td>
</tr>
<tr>
<td></td>
<td>GSDMA</td>
<td>100144966_TGI_at</td>
<td>NA</td>
<td>0.39 (6.42E–08)</td>
</tr>
<tr>
<td></td>
<td>CRKRS</td>
<td>100152237_TGI_at</td>
<td>−0.35 (4.28E–07)</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>CRKRS</td>
<td>100305614_TGI_at</td>
<td>−0.35 (3.61E–07)</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>CRKRS</td>
<td>100307813_TGI_at</td>
<td>−0.31 (6.95E–06)</td>
<td>NA</td>
</tr>
<tr>
<td>rs2305480</td>
<td>Moffatt et al., NEJM</td>
<td>A G</td>
<td>ORMDL3</td>
<td>GSDMA</td>
</tr>
<tr>
<td>rs3859192</td>
<td>Moffatt et al., NEJM</td>
<td>C T</td>
<td>GSDMA</td>
<td>NA</td>
</tr>
<tr>
<td>rs2244012</td>
<td>Li et al., JACI</td>
<td>A G</td>
<td>RAD50</td>
<td>100132153_TGI_at</td>
</tr>
<tr>
<td>rs9273349</td>
<td>Moffatt et al., NEJM</td>
<td>C T</td>
<td>HLA-DRB5</td>
<td>AGPAT1</td>
</tr>
<tr>
<td>rs1295686</td>
<td>Moffatt et al., NEJM</td>
<td>C T</td>
<td>CRKRS</td>
<td>100307813_TGI_at</td>
</tr>
<tr>
<td>rs1343182</td>
<td>Moffatt et al., NEJM</td>
<td>G A</td>
<td>GSDMA</td>
<td>NA</td>
</tr>
<tr>
<td>rs2786098</td>
<td>Sleiman et al., NEJM</td>
<td>T C</td>
<td>CRKRS</td>
<td>100307813_TGI_at</td>
</tr>
</tbody>
</table>

*Beneath the SNP ID in column 1, the reference allele and risk allele are shown (e.g. for rs7216389, C is the reference allele and T the risk allele).

**Moffatt et al. [9] reported that rs3859192 in the 17q region was associated with asthma (P = 3.55 × 10−15). SNP rs3859192 was the most significant eSNP at this locus – strongly associated with GSDMA levels (P = 3.55 × 10−15) but weakly associated with GSDMB levels (P = 3.86 × 10−14).
While most GWAS publications only report the top signals, the GABRIEL study [9] released the asthma association results on all SNPs investigated, allowing an in-depth analysis. For this analysis, we did not limit the list of eSNPs to the most significant SNP for each probe set; all cis eSNPs which passed the 10% FDR for any probe-set were considered. Using this strategy, 60,530 of the 567,589 GABRIEL SNPs were eSNPs in lung tissue. These 60,530 eSNPs were enriched for significant association with asthma in the GABRIEL study (Figure 4). This is consistent with previous studies showing that SNPs associated with complex traits are more likely to be eQTLs [24].

**Key Driver Analyses (KDA) of networks**

One of the drawbacks of GWAS is the reliance on a large number of statistical tests, which puts the threshold for significance at an extremely low level thereby increasing the chance of missing real associations. Given that asthma is a pulmonary disorder, it is reasonable to assume that important molecular drivers are expressed in lung tissue. Therefore, instead of filtering primarily by P values to identify loci/genes that explain asthma, we filtered the loci in the GABRIEL [9] dataset by their status as a cis acting eSNP. Specifically, all SNPs from the GABRIEL study associated with asthma with \( P < 0.01 \) were translated into genes via our lung eSNP list. A total of 7,613 SNPs were linked to the expression of 739 unique genes.

Using Bayesian networks constructed on gene expression data we attempted to find the molecular underpinnings of asthma. For this discussion we define a Bayesian network as a probabilistic graphical model that represents a set of random variables (gene expression in this case) and their conditional dependencies (edges) via a directed acyclic graph (see Text S1). The workflow is illustrated in Figure S2. We first identified each gene from the GABRIEL GWAS dataset and collected all genes within 3 edges of that gene on the Bayesian Networks and found the largest coherent sub-network (i.e. the largest network with an observable path between every gene in the network) that contained the highest proportion of GABRIEL genes (Figure S3). We choose to use 3 edges so that we had enough genes to lead to a reasonable biological annotation of the genes in the network, i.e. limiting to 2 edges is too restrictive and results in a small number of relevant genes; on the other hand, using 4 edges leads to broad gene sets and makes it hard to pinpoint pathways. From this sub-network, we performed Key Driver Analyses (KDA) to capture key regulators of asthma [25,26]. KDA is a method that captures the “hub nodes” conditioning on the direction that one gene influences over other genes and isolating those genes with the strongest influence over the entire network. KDA determines the regulatory components in a directed network for a particular set of genes (i.e. the genes whose expression levels were control by GABRIEL GWAS SNPs). With this analysis, we identified well documented asthma candidate genes [27] that drive genes discovered in GABRIEL. These genes are denoted as canonical asthma genes. Figure 5 shows the top 6 key driver genes (yellow nodes) that control the asthma canonical genes (blue nodes) as well as other genes that are one edge away from the key driver genes (grey nodes). The network when annotated with GeneGo intuitively is described as “Immune Response” (the number of overlap genes is 25, from a 1,346 genes set, with the background being 16,606, \( P = 3.53 \times 10^{-20} \)). The KDA data in Table S3 shows the six genes that remain significant after adjustments for multiple testing.

**Network biology**

Networks provide a means for data integration. Above we used the networks to define a central node of the network that drives asthma by integrating lung Bayesian networks, asthma GWAS and literature (canonical genes) to arrive at set of genes that drive the asthma genes in the lung Bayesian network. Similarly we can use the Bayesian networks to rank competing genetic hypotheses around the molecular basis of asthma. It has been shown that the neighborhood milieu of proteins in a network can predict the probable function of a protein for which no function is known [28]. Following the same logic we assessed the sub-network surrounding each candidate causal gene for asthma. The gene with the highest connection to other asthma genes is most likely a causal asthma gene. Chromosomes 2q12 and 17q21 both have strong GWAS hits for asthma. However, these loci harbor a number of genes that could potentially explain these hits [10,11]. Following the workflow shown in Figure S4 [27], we drew inference on causal genes underlying asthma GWAS signals on these two chromosomes (Table 4). Clearly IL1RL1’s subnetwork is enriched for canonical asthma genes \( (P = 1.8 \times 10^{-07}) \) making it the most likely gene driving the asthma association on 2q12. The result for 17q21 is less clear, but still points away from ORMDL3, the gene first suggested as causal by Moffatt et al. [10]. The sub-network around GSDMA is significantly enriched for canonical genes \( (P = 0.005) \) suggesting that GSDMA is the most likely driver behind the asthma association on 17q21.

**Discussion**

One of the primary goals of genetic studies is to identify genes and pathways which contribute to susceptibility to human diseases and towards this goal many GWA studies have discovered SNPs in novel genes [1–16]. However, the mechanism by which these SNPs lead to disease susceptibility cannot be directly inferred from GWAS. The discovery of eQTLs has been shown to be a powerful tool in addressing this gap [18–20]. Although eQTL analysis of readily available tissue types such as peripheral blood leukocytes and lymphoblastoid cell lines has contributed to the understanding of how genes modulate risk of respiratory diseases, the discovery of lung-specific eQTL is probably more revealing in understanding the pathogenesis of respiratory disease. In this paper, we determined genetic variation and interrogated gene expression in a large collection \( (n = 1,111) \) of human lung samples to systematically characterize the genetic architecture of gene expression in this tissue. Owing to the large sample size, we discovered over 17,000 eQTLs. These eQTLs are gene variants that directly or indirectly govern gene expression in the lung and the data set represents a unique resource which we have made available for the use of lung researchers.

To illustrate the power of the resource we merged these data with SNPs known to be strongly associated with asthma and were able to identify the most likely causal gene variants. Chromosome 17q21 is the most consistent locus associated with asthma [9,10,29–36]. In the original GWAS study [10], the SNPs

---

**Table 3. cont**

***Moffatt et al. [9] reported that SNPs at the IL1RL1 locus were associated with asthma and the GWAS peak was located at rs3771166. The latter SNP was not a strong lung eSNP (failed to pass 10% FDR). Another SNP at the IL1RL1 locus, rs13431828, was strongly associated with asthma (Moffatt et al. [9], \( P = 1 \times 10^{-15} \)) and IL1RL1 expression level \( (P = 7.48 \times 10^{-9}) \). The KDA data in Table S3 shows the six genes that remain significant after adjustments for multiple testing.***
Figure 2. Lung eQTLs found on chromosome 17q21. (a) The upper panel is a region of chromosome 17q21 from the UCSC browser showing the genes located in this region. The next panel shows the results from the GABRIEL consortium [9] (each bar represents a SNP and the y-axis shows $-\log_{10} P$ values for association with asthma). The following three panels show the lung eQTL results for expression levels of ORM DL3 (blue bars), GSDMB (green bars), and GSDMA (red bars), respectively. The y-axis represents $-\log_{10} P$ values derived from the meta-analysis of gene expression. The black horizontal lines are drawn at $P=0.05$. (b) Boxplots of lung gene expression levels for GSDMA according to genotype groups for SNPs rs3859192 and rs7216389 in 1,111 subjects. 

doi:10.1371/journal.pgen.1003029.g002
associated with asthma in the 17q21 susceptibility locus were also associated with transcript levels of ORMDL3 in lymphoblastoid cell lines, suggesting that ORMDL3 was the causal gene. However, more refined analysis of the same samples, exploiting data generated by the 1000 genomes project, suggested that GSDMB, in close proximity to ORMDL3, could be the causal gene [9]. Another eQTL study performed in white blood cell RNA samples suggested that many SNPs in the 17q21 regions are associated with transcript levels of both ORMDL3 and GSDMB [32]. Verlaan et al. [37] showed that SNPs in the region demonstrate domain-wide cis-regulatory effects suggesting long-range chromatin interactions and they found allele-specific differences in nucleosome distribution and binding of the insulator protein CTCF. A recent study also showed that asthma risk alleles on 17q21 were associated with increased ORMDL3 and GSDMA gene expression and elevated IL-17 secretion in cord blood mononuclear cells [38]. To refine the causal gene/variant, we have interrogated our eQTL results within this locus. Interestingly, the SNPs most strongly associated with the expression of ORMDL3 are located in the promoter (rs4794820) and intron 1 (rs12603332) of the ORMDL3 gene, but the P values are not convincing (most significant SNP rs4794820 $P=1.0\times10^{-7}$). In addition the SNP

Figure 3. Relative expression of GSDMA and GSDMB. Primary human airway epithelial cells in monolayer culture were analyzed for; (a) the relative abundance of mRNA for GSDMA and GSDMB (n = 7) and (b) Protein expression of GSDMA and GSDMB by Western blot normalized to expression of β-tubulin (n = 10). (c) Representative image of GSDMA expression in human conducting airway by immunohistochemistry, which shows expression of GSDMA in both basal and apical cells within the airway epithelium.

doi:10.1371/journal.pgen.1003029.g003
rs4794820 is more strongly associated with mRNA expression levels of GSDMB ($P = 9.3 \times 10^{-12}$) and GSDMA ($P = 2.2 \times 10^{-46}$). Rs4794820 is located between ORMDL3 and GSDMA on 17q21 (see Figure 2a for the location of genes). Considering the transcription orientation of ORMDL3, GSDMA, and GSDMB, rs4794820 lies in the promoter of the three genes and is possibly a pleiotropic regulatory variant or, more likely, in LD with such a regulatory element. We observed a concordance between SNPs associated with transcript levels of ORMDL3, GSDMB, and GSDMA (Figure 2a). These genes are likely co-regulated, as proposed previously [32,37]. However, the most compelling eQTL on 17q21 was observed in the association between the asthma susceptibility SNP (rs3859192 – $p = 1.11 \times 10^{-12}$ for association with asthma) and the level of expression of GSDMA ($P = 3.55 \times 10^{-151}$, Figure 2b). In fact, many SNPs on 17q21 were strongly associated with the expression of GSDMA (Figure 2a). Thus our data strongly suggest that the SNPs across this whole locus are associated with asthma because they modulate GSDMA expres-
sion in the lung, clearly showing the value of our lung eQTL dataset to refine previous GWAS hits for asthma. It should be noted that correlations between GWAS and eQTL results must be interpreted with caution. Here we related eQTLs in the lung to GWAS studies for asthma which is an airway disease but which is known to involve other tissue types including immune cells. The genetic control of gene expression is tissue specific. Regulatory variants for GSDMA identified in the lung may regulate a different gene(s) in another relevant tissue or cell-type for asthma. More functional work in multiple tissues and cell-types measuring gene expression with or without various stimuli will be required to confirm the causal asthma gene on 17q21.

The case for GSDMA as the susceptibility gene underlying the 17q GWAS signal is supported by biologic plausibility. GSDMA is a member of the gasdermin family of proteins first identified in the mouse [39]. They have been reported as being expressed in the upper gastrointestinal tract and skin where they are involved in the regulation of apoptosis and act as tumor suppressors [40]. Further support for GSDMA as an asthma susceptibility gene is our observation that GSDMA is robustly expressed in human airway epithelium (Figure 3). In addition its network neighborhood is enriched in genes involved in immune responses (Table S3). This network of genes is annotated as inflammatory response

<table>
<thead>
<tr>
<th>Chr</th>
<th>Genes</th>
<th>x</th>
<th>k</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>GSDMA</td>
<td>7</td>
<td>130</td>
<td>0.0049</td>
</tr>
<tr>
<td></td>
<td>GSDMB</td>
<td>4</td>
<td>94</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>ORM6L2</td>
<td>3</td>
<td>204</td>
<td>0.2208</td>
</tr>
<tr>
<td>2</td>
<td>IL1RAP</td>
<td>16</td>
<td>381</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td>IL1RL1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

m = Canonical list [27] (119 genes).

n = Bayesian – Canonical list (7000 genes).

k = Genes within 3 edges of QTL gene.

doi:10.1371/journal.pgen.1003029.0004

In summary, the present study reports on a comprehensive set of lung eQTLs that complement previous large scale studies of eQTLs in other tissue types [18,19] and which can be used to shed light on GWAS findings in lung diseases. Using the results of the largest asthma GWA study as an example we show how the lung tissue eQTL dataset can be used to identify the most likely causal genes and pathways. This dataset constitutes an invaluable tool to provide new insights into the pathogenesis of other lung diseases such as chronic obstructive pulmonary disease, lung cancer and cystic fibrosis.

Materials and Methods

Ethics statement

All lung tissue samples were obtained in accordance with Institutional Review Board guidelines at the three sites: Laval University (Quebec, Canada), University of British-Columbia (Vancouver, Canada) and Groningen University (Groningen, The Netherlands). All patients provided written informed consent and the study was approved by the ethics committees of the Institut universitaire de cardiologie et de pneumologie de Quebec and the UBC-Providence Health Care Research Institute Ethics Board for Laval and UBC, respectively. The study protocol was consistent with the Research Code of the University Medical Center Groningen and Dutch national ethical and professional guidelines (“Code of conduct; Dutch federation of biomedical scientific societies”; http://www.federa.org).

Lung tissues collection

Lung tissue was collected from patients who underwent lung resectional surgery at the three participating sites; Laval University, University of British Columbia, and University of Groningen. Lung specimens from the Laval site were taken from the Respiratory Health Network Tissue Bank of the FRQS (www.rs.chs.qc.ca). Subjects’ enrollment and lung tissues processing at the three sites are described in Text S1.

Genome-wide gene expression and genotyping

Gene expression profiles were obtained using a custom Affymetrix array (see GEO platform GPL10379) testing 51,627 non-control probesets. DNA samples were genotyped on the Illumina Human1M-Duo BeadChip array. 409, 363, and 339 patients had both genotyping and gene expression data that passed standard quality controls (see Text S1) in Laval, Groningen, and UBC, respectively. These final datasets (total n = 1,111) were used to discover eQTLs and to identify eSNPs.
Statistical analysis

The normalized expression data were adjusted for age, gender and smoking status in a robust linear model to accommodate potential outliers in expression level (Figure 1). In parallel, we performed genotype QC to exclude SNPs of low call rate (<0.9) and deviating from Hardy-Weinberg equilibrium (P<1x10^-5). Genotype imputations were based on the cleaned sets. In the end, assayed and imputed genotypes were used to identify cis and trans acting expression quantitative trait loci (eQTLs) following a method similar to that previously described in Schadt et al. [19]. The strategy for assembling of the asthma candidate gene list (Table S5) and for constructing Bayesian networks, and co-expression modules are given in Text S1.

Supporting Information

**Figure S1** Distribution of the R² values for the relationships between eSNPs and gene expression using a 10% false discovery rate for cis acting eQTLs (upper panel) and trans acting eQTLs (lower panel). (PDF)

**Figure S2** Workflow to identify key drivers of the lung Bayesian networks. We filtered the asthma GWAS SNPs down only to those eSNP filtered genes where used to identify the asthma subnetwork from the larger Bayesian network from which KDA was performed. (PDF)

**Figure S3** The largest lung expression subnetwork that has the highest proportion of GABRIEL genes. (PDF)

**Figure S4** Workflow to identify the most likely gene that drives the GWAS asthma association on chromosomes 2 and 17. The logic is that a gene that is surrounded by known asthma genes is more likely to be an asthma gene itself. (PDF)

**Table S1** Mean (Median) of regression r² in expressional QTLs. (PDF)

**Table S2** (a) cis-eQTLs at 10% FDR. (b) Trans-eQTLs at 10% FDR. (XLSX)

**Table S3** Six Key Driver Analyses (KDA) genes that remains significant after correction for multiple testing. (PDF)

**Table S4** Biological functions of the Bayesian network surrounding GSDMA. (PDF)

**Table S5** Asthma candidate genes. (PDF)

**Text S1** Supporting Information. (DOCX)

Acknowledgments

The authors would like to thank Christine Racine and Sabrina Bizard at the IUCPQ site of the Respiratory Health Network Tissue Bank of the FRQS for her valuable assistance. We thank Rémy Thériault and the information technology team at the Institut Universitaire de Cardiologie et de Pneumologie de Québec for the secure server-based data repository. We also acknowledge the staff at the Consortium Laval, Université du Québec, McGill, and Eastern Québec (CLUMEQ) for IT support with the high performance computer clusters. At the Groningen UMCg site Marjuk Jonker is thanked for his support in selecting, handling, and sending of the tissues.

Author Contributions


References


PLOS Genetics | www.plosgenetics.org 11 November 2012 | Volume 8 | Issue 11 | e1003029