View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Carolina Digital Repository

REPORT

Gene Expression in Transformed Lymphocytes Reveals Variation in Endomembrane and HLA Pathways Modifying Cystic Fibrosis Pulmonary Phenotypes

Wanda K. O'Neal,[1,*] Paul Gallins,[2] Rhonda G. Pace,[1] Hong Dang,[1] Whitney E. Wolf,[1] Lisa C. Jones,[1] XueLiang Guo,[1] Yi-Hui Zhou,[3] Vered Madar,[4] Jinyan Huang,[5] Liming Liang,[6] Miriam F. Moffatt,[7] Garry R. Cutting,[8] Mitchell L. Drumm,[9] Johanna M. Rommens,[10] Lisa J. Strug,[11] Wei Sun,[2,4] Jaclyn R. Stonebraker,[1] Fred A. Wright,[3,12,13] and Michael R. Knowles[1,13]

Variation in cystic fibrosis (CF) phenotypes, including lung disease severity, age of onset of persistent *Pseudomonas aeruginosa* (*P. aeruginosa*) lung infection, and presence of meconium ileus (MI), has been partially explained by genome-wide association studies (GWASs). It is not expected that GWASs alone are sufficiently powered to uncover all heritable traits associated with CF phenotypic diversity. Therefore, we utilized gene expression association from lymphoblastoid cells lines from 754 p.Phe508del CF-affected homozygous individuals to identify genes and pathways. *LPAR6*, a G protein coupled receptor, associated with lung disease severity (false discovery rate q value = 0.0006). Additional pathway analyses, utilizing a stringent permutation-based approach, identified unique signals for all three phenotypes. Pathways associated with lung disease severity were annotated in three broad categories: (1) endomembrane function, containing p.Phe508del processing genes, providing evidence of the importance of p.Phe508del processing to explain lung phenotype variation; (2) HLA class I genes, extending previous GWAS findings in the HLA region; and (3) endoplasmic reticulum stress response genes. Expression pathways associated with lung disease were concordant for some endosome and HLA pathways, with pathways identified using GWAS associations from 1,978 CF-affected individuals. Pathways associated with age of onset of persistent *P. aeruginosa* infection were enriched for HLA class II genes, and those associated with MI were related to oxidative phosphorylation. Formal testing demonstrated that genes showing differential expression associated with lung disease severity were enriched for heritable genetic variation and expression quantitative traits. Gene expression provided a powerful tool to identify unrecognized heritable variation, complementing ongoing GWASs in this rare disease.

The genetic architecture of phenotypic variability in cystic fibrosis (CF [MIM 219700]) is beginning to be defined,[1–5] but GWASs for CF are limited by numbers of subjects compared to common diseases, where tens of thousands of subjects have been used to identify pathophysiologically relevant pathways.[6–8] Studies of gene expression provide an alternative approach to identify gene modifiers.[9–11] Based upon the established utility of gene expression studies in lymphoblastoid cell lines (LCLs),[12–14] global gene expression was measured from LCLs of a highly phenotyped CF cohort previously used for GWAS analysis[1] and analyzed for association with three distinct CF phenotypes: lung disease severity, age of onset of persistent *Pseudomonas aeruginosa* (*P. aeruginosa*) pulmonary infection, and meconium ileus (MI [MIM 614665]) at birth (Table 1; Figure S1).

Affymetrix Human Exon (1.0 ST) microarray data were collected from RNA isolated from 754 LCLs selected from a cohort of 1,137 samples from *CFTR* (MIM 602421) p.Phe508del European individuals homozygous for the mutation (chr7: 98,809–98,811 delCTT; RefSeq accession number NG_016465.3; c.1521_1523delCTT). These CF-affected individuals were originally obtained for the Genetic Modifiers in CF Lung Disease Study where a GWAS had been performed[1] (Figure S2). Considerable efforts were taken to ensure that high-quality microarray data were utilized and that interpretation would not be confused by known effect of SNPs on probe hybridization kinetics (Figure S2). For the highly polymorphic HLA region, probe set filtering removed 438 of the 797 probe sets. However, because of the concern that probe set filtering might not have been adequate in HLA genes, additional analysis was performed to identify HLA genes whose expression values were probably affected by probe set binding (Figure S3). As a result of this analysis, HLA-DRB1 (MIM 142857) expression values were removed from subsequent

[1]Marsico Lung Institute/UNC CF Research Center, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA; [2]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA; [3]Bioinformatics Research Center and Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA; [4]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA; [5]State Key Laboratory of Medical Genomics, Shanghai Institute of Hematology, Rui Jin Hospital affiliated with Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China; [6]Department of Biostatistics, Harvard University, Boston, MA 02115, USA; [7]National Heart and Lung Institute, Imperial College London, London SW7 2AZ, UK; [8]McKusick-Nathans Institute of Genetic Medicine and Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA; [9]Department of Pediatrics, School of Medicine, Case Western Reserve University, Cleveland, OH 44106, USA; [10]Program in Genetics and Genome Biology, The Hospital for Sick Children and Department of Molecular Genetics, University of Toronto, Toronto, ON M5G 1X8, Canada; [11]Program in Child Health Evaluative Sciences, The Hospital for Sick Children, and Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, ON M5G 1X8, Canada; [12]Department of Biological Sciences, North Carolina State University, Raleigh, NC 27695, USA
[13]These authors contributed equally to this work
*Correspondence: wanda_o'neal@med.unc.edu
http://dx.doi.org/10.1016/j.ajhg.2014.12.022. ©2015 by The American Society of Human Genetics. All rights reserved.

CrossMark

Table 1. Characteristics of Subject Population for Phenotypes

Study Group	Consortium Lung Phenotype (Primary Analysis)[a]					Age of Onset of Persistent *Pseudomonas aeruginosa*			Meconium Ileus (MI)
	Size of Population	Age at Enrollment (year)		No. Males (%)	No. European[b] (%)	Persistent Culture Positive[c] (%)	Age of Onset (year)		Presence of MI[d] (%)
		Mean ± SD	Range				Mean ± SD	Range	
Severe	317	16.5 ± 4.6	8–25	157 (49.5)	317 (100)	208 of 222 (93.7)	5.2 ± 4.3	0.6–19	52 of 301 (17.3)
Mild	437	28.0 ± 9.9	15–58	221 (50.5)	437 (100)	203 of 233 (87.1)	16.8 ± 10.3	0.6–57	54 of 405 (13.3)[e]
Total	754					455			706

[a]Subjects were classified as having either severe or mild lung disease, as defined by the quantitative Consortium lung phenotype (KNoRMA) value of <0.3 or >0.3, respectively.[18]
[b]Based on self-identified ancestry and principal components analysis via SNP genotypes.
[c]Data were obtained at the encounter level (each clinic visit) from the Cystic Fibrosis Foundation (CFF) Patient Registry. Persistent is defined as cultured *P. aeruginosa* in respiratory cultures 2 years in a row, or 2 out of 3 years, unless subjects had at least 5 consecutive years of negative cultures after meeting minimal criteria (2 out of 3 years of positive cultures). Subjects who were above age 7 needed to have a negative culture before the first positive culture to be included into the analysis.[28] There were 14 severe and 30 mild subjects who were negative for *P. aeruginosa* at last culture.
[d]Subjects were confirmed to have MI if a diagnosis at birth was supported by source documents, such as the original surgical or medical report, detailed clinical or admissions note, or verbal confirmation from the subject or the parent with documentation of an abdominal scar. Subjects were removed from the analysis if MI could not be confirmed or if the diagnosis was unclear or unknown.
[e]Presence of MI was 17.6% (36 of 205) for subjects enrolled at 15–25 years of age.

analysis. The study was approved by the biomedical institutional review board of the University of North Carolina and the institutional review board of each participating institution. CF-affected individuals and their parents (if they were a minor) provided written informed consent.

Linear regression was utilized to establish association of gene expression with phenotypes. Gene expression values meeting a minimal threshold of expression above 6.03 (on the Affymetrix RMA standard log2 scale) were utilized, based on the 95th percentile of mean "expression" in females for genes on the Y chromosome, because this threshold was considered to reliably represent true signal above background. All genes meeting this criterion (12,033 out of 17,868 annotated genes; 67.3%) were included in the linear regression analysis, including genes whose probes overlaid SNPs with high minor allele frequency (MAF), but these genes were "flagged" so that potentially important interpretive issues could be considered later. The covariates used for all analyses are listed in Table S1. The genotype PCs used as covariates were calculated with Eigenstrat[15] and available genotype data from the previously conducted GWASs.[1] The surrogate variables of gene expression data were calculated with the "sva" package in Bioconductor in R.[16] The Q-Q plots for all three phenotypes suggested that the covariates included were appropriate to control for population stratification or technical factors that could potentially lead to false positives (Figure S4).

The expression of lysophosphatidic acid receptor 6 (*LPAR6* [MIM 278150]) achieved transcriptome-wide significance for association with lung disease (false discovery rate q value = 0.0006, p value = 5.35×10^{-8}), using both standard and alternative probe annotation (ANNMAP, formerly known as X:MAP),[17] with higher levels of *LPAR6* being associated with worse lung function. Array-based *LPAR6* expression was technically validated by TaqMan quantitative real-time PCR ($p < 0.0001$ between 36 low-expressing and 40 high-expressing LCL samples from CF-affected individuals). *CHMP4C* ($p = 1.05 \times 10^{-5}$ [MIM 610899]), *SSBP2* ($p = 2.60 \times 10^{-5}$ [MIM 607389]), and *P2RX4* ($p = 8.03 \times 10^{-5}$ [MIM 600846]) were suggestive for association (Table S2; Figure S5; see Table S5 for complete list).

As explicitly accounted for by the Consortium lung phenotype,[18] older surviving CF-affected individuals have milder lung disease, reflecting high mortality in CF (Table 1). To investigate a possible relationship between age and gene expression in the CF cohort, but unrelated to CF lung disease, we examined three large external studies of LCL gene expression. These included a childhood asthma (MIM 600807) cohort evaluated on the Affymetrix platform,[19] available data from the Cholesterol and Pharmacogenomics (CAP) trial (available on ArrayExpress),[20] and the Multiple Tissue Human Expression Resource (MuTHER) study.[21] No correspondence emerged between differentially expressed genes for the Consortium lung phenotype and those associated with age in these three non-CF populations (Figure S6), although *LPAR6* was nominally associated with age (not corrected for multiple comparison) in older women (age ~59 years) in the MuTHER study.[21] Consequently, we conclude that the associations seen in our study reflect CF lung disease severity and not aging.

Rigorous "pathway" (gene set) analysis was conducted via a permutation-based approach (Significance Analysis of Function and Expression; SAFE), which accounts for gene expression correlation structures and allows testing of both standard and custom-derived pathways.[22] Pathway analysis was conducted by SAFE in R (v.3.0) and annotation databases (available at Bioconductor) hugene10stprobeset.db and GO.db (Gene Ontology annotation maps). Multiple pathways with q values < 0.15 were found to

Table 2. Gene Expression Pathways Significantly Associated with Consortium Lung Phenotype

Pathway		Genes			Statistics			
ID	Name	Number	↑[a]	↓[b]	Trend[c]	p Value[d]	q Value[e]	Genes with Gene-Level p Value < 0.05 (Ordered by p Value)[f]
GO Cellular Component Pathways								
0001673	male germ cell nucleus	14	0	14	down	0.0001	0.0164	*TNP1*; *REC8*; *TCFL5*
0012507	ER to Golgi transport vesicle membrane	25	23	2	up	0.0003	0.0481	*HLA-E*; *MCFD2*; *TMED7*; *HLA-F*
0043073	germ cell nucleus	17	1	16	down	0.0004	0.0442	*TNP1*; *REC8*; *TCFL5*
0042470; 0048770	melanosome; pigment granule	78	52	26	up	0.0007	0.0582	*SLC3A2*; *TPP1*; *CTSD*; *ANXA2*; *STOM*; *HSPA5*; *BSG*
0030134	ER to Golgi transport vesicle	29	25	4	up	0.0011	0.0737	*HLA-E*; *MCFD2*; *TMED7*; *HLA-F*
0030176	integral to endoplasmic reticulum membrane	85	64	21	up	0.0024	0.1181	*TTC35*; *HLA-E*; *EDEM1*; *TAP1*; *SELS*; *HLA-F*; *HSPA5*; *MMGT1*
0031301	integral to organelle membrane	171	113	58	up	0.0026	0.1181	*TTC35*; *HLA-E*; *EDEM1*; *TAP1*; *SELS*; *ST6GALNAC6*; *HLA-F*; *A4GALT*; *ARMCX3*; *P2RX7*; *LARGE*; *HSPA5*; *MMGT1*
0000421	autophagic vacuole membrane	13	11	2	up	0.0028	0.1181	*WIPI1*; *ATG9A*
0031227	intrinsic to endoplasmic reticulum membrane	95	70	25	up	0.0031	0.1181	*TTC35*; *HLA-E*; *EDEM1*; *TAP1*; *SELS*; *HLA-F*; *HSPA5*; *MMGT1*
0031300	intrinsic to organelle membrane	184	121	63	up	0.0036	0.1231	*TTC35*; *HLA-E*; *EDEM1*; *TAP1*; *SELS*; *ST6GALNAC6*; *HLA-F*; *A4GALT*; *ARMCX3*; *P2RX7*; *LARGE*; *HSPA5*; *MMGT1*
0030658	transport vesicle membrane	49	33	16	up	0.0039	0.1231	*HLA-E*; *MCFD2*; *TMED7*; *HLA-F*; *NCALD*
GO Biological Process Pathways								
0006518	peptide metabolic process	64	46	18	up	0.0001	0.0837	*GSTK1*; *DNPEP*; *PSEN2*; *TPP1*
0072384	organelle transport along microtubule	24	21	3	up	0.0001	0.0837	*PRKCZ*; *COPG*
0006925	inflammatory cell apoptotic process	10	10	0	up	0.0003	0.1107	none
0006944	cellular membrane fusion	61	42	19	up	0.0003	0.1107	*CD9*; *PLDN*; *ANXA2*; *BET1*
0007030	golgi organization	38	28	10	up	0.0003	0.1107	*GCC2*; *BHLHA15*; *GOLGB1*; *PLK3*; *COG1*; *TMED2*
0043603	cellular amide metabolic process	101	65	36	up	0.0003	0.1107	*GSTK1*; *DNPEP*; *PSEN2*; *TPP1*; *PRKCD*
0034067	protein localization to Golgi apparatus	14	13	1	up	0.0004	0.1166	*GOLGA4*; *GCC2*; *ATG9A*
0045684	positive regulation of epidermis development	11	10	1	up	0.0004	0.1166	none
GO Molecular Function Pathways								
0050839	cell adhesion molecule binding	33	15	18	two sided	0.0004	0.1181	*P2RX4*; *MLLT4*; *CD1D*;[g] *CTNNA1*; *PVRL1*[g]
0042287	MHC protein binding	15	9	6	two sided	0.0006	0.1191	*TAP1*; *LAG3*; *MARCH8*
MSigDB Pathways								
	ATAAGCT.MIR.21	81	45	36	two sided	0.0001	0.0387	*BAHD1*; *BTBD3*;[g] *C5orf41*; *STK40*; *UBR3*; *NF2*;[g] *SSFA2*; *JAG1*; *PPARA*; *PELI1*; *RHOB*; *CREBL2*
	V.HMGIY_Q6	158	70	88	two sided	0.0006	0.1499	*ZNF675*;[g] *LMO4*; *TNFSF11*;[g] *PLAGL2*; *POLD3*;[g] *SLC7A1*; *UBE2E2*;[g] *TAZ*; *UBR3*; *MRC2*;[g] *TNFSF4*; *IKZF2*[g]
MetaMiner Cystic Fibrosis Specific Pathways[h]								
	cholesterol and sphingolipids transport/recycling to plasma membrane in lung (normal and CF)	14	9	5	two sided	0.0036	0.0597	*ABCG1*[g]
	normal wtCFTR traffic/sorting endosome formation	14	11	3	up	0.0052	0.0621	none

(*Continued on next page*)

Table 2. *Continued*

Pathway		Genes			Statistics			
ID	Name	Number	↑[a]	↓[b]	Trend[c]	p Value[d]	q Value[e]	Genes with Gene-Level p Value < 0.05 (Ordered by p Value)[f]
	F508-CFTR traffic/ER-to-Golgi in CF; Normal wtCFTR traffic/ER-to-Golgi	22	20	2	up	0.0075	0.0621	*COPG*; *COPZ2*
	mucin expression in CF via TLRs, EGFR signaling pathways	48	34	14	up	0.0116	0.0770	*JUN*; *PRKCD*
PFAM Pathways								
00035	double-stranded RNA binding motif	17	2	15	down	0.0001	0.0135	*STRBP*; *STAU2*
07716	basic region leucine zipper	11	7	4	two sided	0.0002	0.0276	*DDIT3*; *CREBL2*; *CEBPB*
03953	tubulin C-terminal domain	15	2	13	down	0.0009	0.0804	*TUBB2B*
CF Relevant Custom Pathways								
	ER stress response	169	127	42	up	0.0005	0.0106	*DNAJB9*; *EDEM1*; *CISD2*; *TANK*; *DDIT3*; *SERP1*; *FDPS*; *LONP1*; *NANS*; *SSR4*; *JUN*; *GADD45A*; *LY9*; *PGM3*; *HSPA5*; *ARF4*; *IER3IP1*; *BTG2*; *CEBPB*; *CNIH*; *MANF*; *PDIA6*
	XBP1 target genes	13	10	3	two sided	0.0079	0.1165	*DNAJB9*; *EDEM1*; *SERP1*; *PDIA6*
HLA-Specific Pathways								
	class I	3	3	0	up	0.0221	0.0261	*HLA-E*; *HLA-F*
	class II	8	7	1	up	0.0868	0.0580	none
	class I and class II	11	10	1	up	0.0299	0.0261	*HLA-E*; *HLA-F*

Pathways limited to those with ≥10 but ≤200 genes. SAFE analysis utilized 10,000 permutations to establish significance thresholds. CF Relevant Custom Pathways developed primarily as described for mice[46] using human gene counterparts (Table S8).
[a]Number of genes in pathway with increased expression.
[b]Number of genes in pathway with decreased expression.
[c]Up (increased) or down (decreased) differential expression of genes in the pathways associated with milder lung disease. Two-sided indicates pathways that contained both increased and decreased differentially expressed genes that contributed significantly to the signal.
[d]Determined by 10,000 permutations in the SAFE package.[22]
[e]Benjamini-Hochberg false-discovery for pathways testing within each pathway set; q values < 0.15 were included.
[f]See Table S6 (tab A) for the inclusive list of genes for these pathways; "none" indicates that no individual genes within the pathway had a p value less than 0.05; see Table S5 for gene MIM numbers.
[g]For the two-sided "Trend," these genes have a "down" trend.
[h]MetaMiner CF Specific Pathways represent a version of Thomson Reuters' (formerly GeneGo) MetaDiscovery suite that is enriched with content specific for cystic fibrosis.

associate with lung disease severity (Table 2; Table S6, tab A). Of the 35 pathways listed (Table 2), 16 were related to the endomembrane system for synthesis and post-translational modification of membrane proteins (membranes, vesicle traffic, and Golgi/endoplasmic reticulum [ER]) and two pathways were related to ER stress response, which also could represent a subset of endomembrane processes. Of the 11 Gene Ontology (GO) Cellular Component pathways, 7 contained HLA class I genes, and custom-derived pathways consisting exclusively of HLA genes were also highly significant (Table 2). Importantly, although the HLA genes clearly contributed to the significance of the endomembrane pathways, these same pathways also contained *TTC35* (Table 2 [MIM 607722]) and *TMEM85* (Table S6, tab A; p value = 0.06), which are the human homologs of yeast genes *EMC2* and *EMC4*, respectively, known to modulate yeast homolog of p.Phe508del processing.[23] MetaMiner Cystic Fibrosis Specific Pathways not containing HLA genes also supported association with p.Phe508del processing (Table 2). We conclude that three important pathophysiological signals have emerged: HLA class I, p.Phe508del processing, and the ER stress response. The significance of the miR21 (miRNA-21 [MIM 611020]) pathway is also relevant given the expanding role of this microRNA (miRNA) in pulmonary biology.[24] Most pathways trended in the "up" direction (increased expression of genes in the pathways associated with milder lung disease), with two pathways (annotated to germ cell nuclei) trending "down."

We hypothesized that the expression pathways would be concordant with pathways identified with GWAS associations of genotype data with lung disease severity. We used a gene- and pathway-testing approach[25] (GeneSetScan v.0.01) applied to GWAS data from the previously genotyped cohort of US and Canadian CF-affected individuals[1] (n = 1,978, including the CF-affected individuals from the expression study) to provide resampling-based multiple-comparison corrected p values for the numbers of pathways tested. Three significant (corrected p < 0.05) pathways were identified (Table 3; Table S6, tab B).

Table 3. GWAS Data Pathways Significantly Associated with Consortium Lung Phenotype

ID	Name	Genes (n)	Corrected p Value[a]	Z Score	Genes with Gene-Level p Value < 0.05 (Ordered by p Value)
Analyses Included All Available Pathways[b]					
KEGG Pathways					
05320	autoimmune thyroid disease	49	0.002	4.469	*HLA-DRB1; HLA-DQA1; HLA-DRA; IFNA13; IFNA8; IFNA2; IFNA16; IFNA17; IFNA6; IFNA10; IFNA1; IFNA7; IFNA14; IFNA4; HLA-DQB1; IFNA5; IFNA21; HLA-E; HLA-DQA2*
04672	intestinal immune network for IgA production	45	0.018	3.822	*HLA-DRB1; HLA-DQA1; HLA-DRA; TNFRSF17; HLA-DQB1; CXCL12; HLA-DQA2*
MSigDB Pathways					
	TGCAAAC.MIR-452	106	0.049	3.676	*SAV1; TRPS1; ATL1; ZIC1; NIN; SH3BGRL; SYN3; XPNPEP1; XPO4; RAB8B*
Analyses Confined to Pathways Significant for Differential Expression[c]					
GO Cellular Component Pathways					
0030134	ER to Golgi transport vesicle	39	0.001	4.271	*HLA-DRB1; HLA-DQA1; HLA-DRA; MCFD2; HLA-DQB1; SEC24B; HLA-E; SREBF1; HLA-DQA2*
0012507	ER to Golgi transport vesicle membrane	33	0.001	4.218	*HLA-DRB1; HLA-DQA1; HLA-DRA; MCFD2; HLA-DQB1; SEC24B; HLA-E; SREBF1; HLA-DQA2*
CF Relevant Custom Pathways					
	ER stress response	248	0.029	1.962	*FGFR4; HMOX1; UBE2L6; TAX1BP1; CREB3; ARL1; UBXN4; RNF5; ATF6B; USO1; GADD45A; NIPSNAP1; XBP1; AMFR; TOR1A; SREBF1; PSMB2*
HLA-Specific Pathways					
	class I	7	0.0382	1.842	*HLA-E*
	class II	11	0.0254	2.285	*HLA-DRB1; HLA-DQA1; HLA-DRA; HLA-DQB1; HLA-DQA2*
	class I and class II	18	0.0101	2.771	*HLA-DRB1; HLA-DQA1; HLA-DRA; HLA-DQB1; HLA-E; HLA-DQA2*

Default parameters with 1,000 resamples were used, and pathways were limited to those that contained ≥10 but ≤200 genes. GeneSetScan utilizes a mapping of SNPs to genes based upon a reference panel from HapMap release 23a and mapped genes to pathways based upon a March 2011 release of GO and KEGG pathways. The number of permutations for the HLA-specific analysis was increased to 10,000 to improve precision for this relevant group of genes. See Table S6 (tab B) for the inclusive list of genes for these pathways. See Table S5 for gene MIM numbers.
[a]Multiple comparison corrected p.
[b]Gene level p values calculated using family-wise (all SNPs, genes, and pathways tested) as provided by GeneSetScan.
[c]Pathways listed in Table 2 were evaluated for association to genotype. Pathways are listed in this table if p value < 0.05.

Two of these pathways (KEGG05320, autoimmune thyroid disease; and KEGG04672, intestinal immune network for IgA production) were enriched in HLA class II genes, congruent with GWAS signal previously reported at chromosome 6p.[1] A third pathway indicated a role for miR452 (miRNA-452), a miRNA associated with epithelial development[26] and alveolar macrophage function.[27] Additional analysis of the GWAS data confined specifically to pathways significant for differential gene expression (Table 2) identified two overlapping ER-to-Golgi HLA-enriched pathways and a pathway containing ER stress response genes (Table 3; Table S6, tab B). Thus, concordance between GWAS and gene expression was observed in biologically relevant pathways. However, the top-ranked genes within the pathways were different between GWAS and expression datasets; thus, the concordance might represent distinct genetic signals.

Analyses of expression associated with age of onset of persistent *P. aeruginosa* lung infection and MI phenotypes were also performed on subsets of the CF cohort (Table 1; Table S1; Figures S1B and S1C). Significant pathways were also associated with each of these two phenotypes (Table 4; Table S3; Tables S5 and S6, tabs C and D). After adjusting for lung disease severity, which correlates with older age of onset of persistent *P. aeruginosa* infection in CF-affected individuals with milder lung disease[28] (Table 1; Figure S1B), multiple pathways were associated with age of onset of persistent *P. aeruginosa* (Table 4). KEGG and GO pathways associated with *P. aeruginosa* were enriched for pathways containing HLA class II genes (5 of 12 total listed pathways; Table 4; Figure 1; Table S4) rather than the class I genes that were prevalent in expression association with lung disease (Table 2; Figure 1; Table S4).

As seen in association with lung disease severity (Table 2), GO Cellular Component and MetaMiner CF Specific Pathways annotated to endomembrane function (lysosomes, vacuolar, and endosomes) were associated with age of onset of persistent *P. aeruginosa* infection. Additionally, congruent with the lung function associations, pathways consisting exclusively of HLA genes were significant for the *P. aeruginosa* phenotype (Table 4 bottom, HLA-Specific Pathways), but the HLA class II genes contributed to a larger

Table 4. Gene Expression Pathways Significantly Associated with Age of Onset of Persistent *Pseudomonas aeruginosa*

Pathway		Genes			Statistics			
ID	Name	Number	↑[a]	↓[b]	Trend[c]	p Value[d]	q Value[e]	Genes with Gene-Level p Value < 0.05 (Ordered by p Value)[f]
KEGG Pathways								
05213	endometrial cancer	50	36	14	up	0.0014	0.1479	*PIK3R3*; *CTNNA1*; *GSK3B*; *CTNNB1*; *LEF1*; *CASP9*
05323	rheumatoid arthritis	50	38	12	up	0.0024	0.1479	*TNF*; *HLA-DMA*; *HLA-DPA1*; *HLA-DOB*; *ATP6V0E1*; *HLA-DMB*; *CD80*; *IL15*; *CSF1*; *HLA-DRA*; *HLA-DOA*
04940	type I diabetes mellitus	24	20	4	up	0.0026	0.1479	*TNF*; *HLA-E*; *HLA-DMA*; *HLA-DPA1*; *LTA*; *HLA-DOB*; *HLA-DMB*; *CD80*; *FAS*; *HLA-DRA*; *HLA-DOA*
GO Molecular Function Pathways								
0004620	phospholipase activity	46	34	12	up	0.0003	0.0857	*PLA2G12A*; *SMPD2*; *LIPH*; *MGLL*; *SMPD1*
0001614; 0016502	purinergic nucleotide receptor activity; nucleotide receptor activity	19	6	13	two sided	0.0007	0.1392	*GPR15*; *P2RY1*; *GPR109A*; *GPR18*[g]
GO Biological Process Pathways								
0001556	oocyte maturation	11	4	7	two sided	0.0002	0.1391	*PTK2B*; *BRCA2*;[g] *CDC25B*;[g] *RPS6KA2*; *TRIP13*;[g] *INSL3*;[g] *ANG*;[g] *CCNB1*[g]
0006865	amino acid transport	79	43	36	two sided	0.0002	0.1391	*TNF*; *CLN8*; *PRKCD*; *CACNB4*; *ARL6IP5*; *SLC25A26*; *PSEN1*; *CPT2*; *ACACB*;[g] *SLC38A10*;[g] *SLC1A1*; *SLC25A32*;[g] *SLC9A3R1*;[g] *CPT1A*;[g] *SERINC1*
0007163	establishment or maintenance of cell polarity	85	51	34	two sided	0.0002	0.1391	*ARHGEF2*; *ARHGEF11*; *PTK2B*; *NDC80*;[g] *CKAP5*;[g] *DLG1*; *CAP1*; *CYTH1*; *CTNNA1*; *ERBB2IP*; *DST*; *ACTR3*; *GNB2L1*; *NCKAP1L*; *CENPA*;[g] *ZW10*;[g] *CLASP1*; *CYTH3*; *GSK3B*; *SCRIB*;[g] *MAP4*; *PRKCZ*; *CDK5RAP2*;[g] *RAB11FIP2*;[g] *EZR*
GO Cellular Component Pathways								
0000794	condensed nuclear chromosome	54	10	44	down	0.0004	0.0617	*SMC1A*; *CHEK1*; *BUB1B*; *PLK1*; *BUB1*; *NDC80*; *TOP2A*; *AURKB*; *NCAPD3*; *INCENP*; *CENPA*; *H2AFX*; *NEK2*; *SUV39H1*; *SMC3*; *RAD21*; *CCNB1*; *ADD3*
0005765	lysosomal membrane	109	78	31	up	0.0008	0.1064	*CD74*; *CD1D*; *HLA-DPA1*; *HLA-DMA*; *OSTM1*; *HLA-DOB*; *VPS39*; *HLA-DMB*; *ARL8B*; *PSEN1*; *TMEM63A*; *ARL8A*; *RAB7A*; *HLA-DRA*; *HLA-DOA*; *AP1S1*; *AP1S3*; *LAMP3*
0044437	vacuolar part	184	129	55	up	0.0010	0.1064	*CD74*; *CD1D*; *HLA-DPA1*; *HLA-DMA*; *ATG16L1*; *OSTM1*; *DAPK2*; *TPP1*; *GLB1*; *HLA-DOB*; *VPS39*; *GM2A*; *HLA-DMB*; *ARL8B*; *PSEN1*; *TMEM63A*; *ARL8A*; *VPS41*; *IDUA*; *RAB7A*; *WIPI2*; *CTSF*; *HLA-DRA*; *HLA-DOA*; *AP1S1*; *SMPD1*; *AP1S3*; *HEXA*; *LAMP3*
0005774	vacuolar membrane	141	99	42	up	0.0015	0.1211	*CD74*; *CD1D*; *HLA-DPA1*; *HLA-DMA*; *ATG16L1*; *OSTM1*; *HLA-DOB*; *VPS39*; *HLA-DMB*; *ARL8B*; *PSEN1*; *TMEM63A*; *ARL8A*; *VPS41*; *RAB7A*; *WIPI2*; *HLA-DRA*; *HLA-DOA*; *AP1S1*; *AP1S3*; *LAMP3*
MetaMiner Cystic Fibrosis Specific Pathways[h]								
	normal wtCFTR traffic/sorting endosome formation	14	13	1	up	0.0022	0.0352	*STX12*; *VPS45*; *RAB7A*
	F508-CFTR traffic/sorting endosome formation in CF	20	16	4	up	0.0042	0.0455	*STX12*; *STAM2*; *VPS45*; *RAB7A*
PFAM Pathways								
00017	SH2 domain	79	60	19	up	0.0003	0.0275	*PIK3R3*; *LYN*; *FER*; *TXK*; *BLK*; *HCK*; *SOCS6*; *INPP5D*; *LCP2*; *SH3BP2*; *FGR*
00788	Ras association (RalGDS/AF-6) domain	26	22	4	up	0.0003	0.0275	*RGL2*; *RASSF5*; *MYO9B*; *RGL1*; *ARAP2*
00225	kinesin motor domain	31	10	21	two sided	0.0003	0.0418	*KIF23*;[g] *KIF4A*;[g] *KIF14*;[g] *KIF11*;[g] *KIF15*;[g] *KIF18A*;[g] *KIF20B*;[g] *KIF22*;[g] *KIF3A*; *KIF2A*

(*Continued on next page*)

Table 4. *Continued*

Pathway		Genes			Statistics			
ID	Name	Number	↑[a]	↓[b]	Trend[c]	p Value[d]	q Value[e]	Genes with Gene-Level p Value < 0.05 (Ordered by p Value)[f]
HLA-Specific Pathways								
	class I	3	3	0	up	0.0960	0.0645	*HLA-E*
	class II	8	8	0	up	0.0117	0.0105	*HLA-DPA1*; *HLA-DMA*; *HLA-DOB*; *HLA-DMB*; *HLA-DRA*; *HLA-DOA*
	class I and class II	11	11	0	up	0.0065	0.0086	*HLA-E*; *HLA-DPA1*; *HLA-DMA*; *HLA-DOB*; *HLA-DMB*; *HLA-DRA*; *HLA-DOA*

Pathways were limited to those with ≥10 but ≤200 genes. The SAFE analysis utilized 10,000 permutations to establish significance thresholds. CF Relevant Custom Pathways were developed primarily as described for mice[46] using human gene counterparts (Table S8).
[a]Number of genes in pathway with increased expression.
[b]Number of genes in pathway with decreased expression.
[c]Up (increased) or down (decreased) differential expression of genes in the pathways associated with milder lung disease. Two-sided indicates pathways that contained both increased and decreased differentially expressed genes that contributed significantly to the signal.
[d]Determined by 10,000 permutations in the SAFE package.[22]
[e]Benjamini-Hochberg false-discovery for pathways testing within each pathway set; q values < 0.15 were included.
[f]See Table S6 (tab C) for the inclusive list of genes for these pathways; see Table S5 for gene MIM numbers.
[g]For the two-sided "Trend," these genes have a "down" trend.
[h]MetaMiner CF Specific Pathways represent a version of Thomson Reuters' (formerly GeneGo) MetaDiscovery suite that is enriched with content specific for cystic fibrosis.

proportion of the signal for the *P. aeruginosa* phenotype (Figure 1; Table S4). The direction of the association effect was consistent with the hypothesis that increased expression of genes in the HLA class II pathways provides protection against persistent infection with *P. aeruginosa* early in life (Table 4). Other pathways exhibiting an "up" signal included genes involved in endometrial cancer, phospholipase activity, and membrane-bound organelle function/transport. Pathways associated with MI did not represent HLA-dominated or endomembrane signatures, but instead reflected oxidative phosphorylation and overall pointed to variation in mitochondrial function (Figure 1; Tables S3 and S4).

Genotype and gene expression data were entered into Matrix eQTL to establish local eQTL associations under false discovery control.[29] SNPs with a minor allele frequency of < 0.01 were removed from the analyses. Expression quantitative trait loci (eQTLs) were abundant in our sample set, with a preponderance of significant ($q < 0.05$) local eQTLs within 1 Mb of the target gene (Table S7). Many of the genes most strongly associated with the Consortium lung phenotype also had highly significant eQTLs (Table S2). To evaluate the hypothesis that differential expression associated with the Consortium lung phenotype is at least partially driven by constitutional genetics, rather than treatment or other factors unrelated to etiology, we performed a global test of concordance for expression association rank with the lung phenotype versus eQTL rank, which was highly significant ($p = 1.25 \times 10^{-8}$). As a further test of this hypothesis, we examined the heritability values reported in the twin transcriptomic peripheral blood study from the Netherlands Twin Registry (NTR),[30] among the largest eQTL studies yet reported ($n = 2{,}752$ twins), which obtained estimates of total additive heritability via a classical twin design. By using a described permutation approach[30] and evaluating genes expressed in LCLs, we found significant correlation ($p = 6.4 \times 10^{-6}$) between ranked p values of differential expression for lung disease severity in our samples. With a multiple regression approach as reported for the NTR heritability study,[30] which corrects for additional genomic factors such as gene conservation and sequence context (see all factors considered as predictors in Table 2 of Wright et al.[30]), the overlap remained significant ($p = 2.9 \times 10^{-7}$). We conclude that heritable factors underlie our differential expression results, but individual components of the signal were not strong enough to be identified as significant. The analogous results for correlation between gene heritability and differential expression with respect to age of onset of persistent *P. aeruginosa* infection or MI, based on a smaller sample sizes, and GWAS association to lung disease severity, showed no significant overlap with twin heritability (data not shown).

In summary, this study of gene expression in LCLs from a large cohort of p.Phe508del homozygotes is one of the largest clinical expression studies to date and provides findings complementary to previous CF GWAS results. It validates the concept that gene expression associated with biologically and pathophysiologically relevant heritable genomic variation contributes to phenotypic variation in CF. *LPAR6* is the single gene that achieved transcriptome-wide significant association with lung disease. LPAR6 is a recently described G protein coupled receptor with no known link to CF; however, LPAR6 is known to be expressed in pulmonary endothelial cells[31] and belongs to a family of LPA-activated receptors that mediate signaling involved in multiple biological functions, including epithelial cell apoptosis, lung fibrosis, and wound healing.[32] Because no SNPs in *LPAR6* were even nominally significant in previous GWAS analysis,[1] epigenetic variation

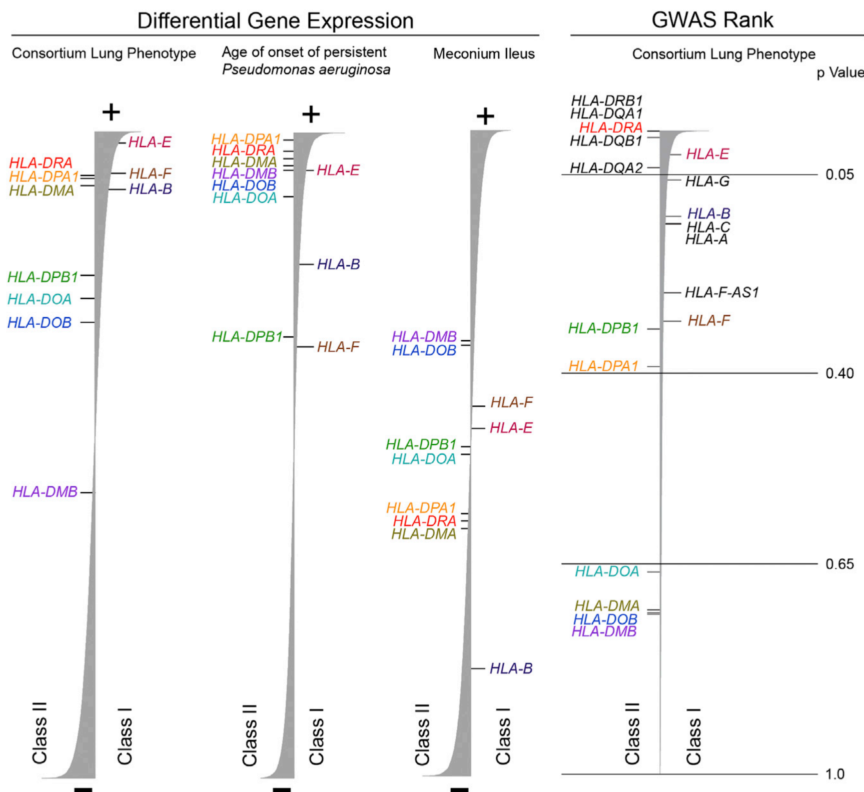

Figure 1. HLA Region Consistently Associated with Consortium Lung Phenotype and Age of Onset of Persistent *P. aeruginosa* Phenotype across Multiple Analyses, but not MI Phenotype

Differential gene expression (left). HLA class I and II genes (listed on the right and left sides of each vertical bar, respectively) whose mean expression values are above the cutoff of expressed genes were ranked according to the association strength (t-statistic, negative [−; bottom] or positive [+; top]; see Table S5 for details) in the expression data (Consortium lung phenotype, n = 754; age of onset, n = 455; MI, n = 706).

GWAS rank (right). HLA genes represented in the GWAS panel (n = 1,978) are depicted according to strength of association to Consortium lung phenotype. For GWAS rank, p values are provided as a reference to aid interpretation. The width of the vertical bar represents the relative strength of the association finding. Individual genes are color coded for convenience. GeneSetScan software was utilized to provide gene ranks for the HLA genes.

or complex unrecognized genetic variation (copy-number variation, insertions/deletions) might be involved at the *LPAR6* locus to explain this finding.

Our most striking discoveries emerged from expression pathway analysis. Identified pathways, related to the endomembrane system for synthesis and post-translational modification of membrane proteins (membranes, vesicles, ER/Golgi) and the ER stress response, have been previously implicated in CF pathophysiology, including p.Phe508del processing.[23,33–35] Although the ER stress response is related to endomembrane function, the two signals (post-translational processing and ER stress response) are both pathologically relevant and are probably acting independently to modify lung disease phenotype. The strength of the genomic signatures in our results supports continued research directed at these processes for CF.

The HLA-associated pathways (some overlapping endomembrane pathways) point toward antigen processing and signaling through membrane-bound organelles as key mediators defining susceptibility to lung disease progression and bacterial infection. These results support previous reports implicating HLA alleles in CF-relevant phenotypes[1,36–38] and add to the growing list of disease phenotypes associated with the major histocompatibility complex.[39] HLA genes have strong eQTL signatures in various datasets,[7,40] including in our own LCL data, and it is probable that expression and allele type both interact to define the final functional consequences of this complex genetic region.[41,42] The strong linkage disequilibrium across the HLA region suggests that previously reported HLA allelic associations with phenotypes should be re-evaluated to consider effects of genetic variation on gene expression as a mechanistic contributor. Additionally, the GO pathways annotated to contain HLA genes are also endomembrane pathways containing genes previously associated with p.Phe508del processing, and which probably reflect genetic variation responsible for producing residual CFTR function as a result of low-level CFTR processing to the membrane in some CF individuals.[43,44]

The use of transformed LCLs for these expression studies provides an opportunity to rigorously explore gene expression in a well-characterized CF cohort previously utilized for GWASs. Given that LCLs are transformed and not the proximal cell type in the CF lung, their use in this study has some limitations. Transformation itself can significantly alter gene expression,[45] and genes and pathways not expressed in LCLs that would be expressed in airway cell populations cannot be queried. Nonetheless, significant gene expression signatures associated with lung phenotypes were identified. The ability to grow LCLs in large numbers and under highly controlled culture conditions are major advantages, and the high power that was achieved allowed for significant findings to emerge. The results add to the expanding knowledge supporting genetic modifier and systems biology studies for CF. Moving forward, gene expression studies in more proximal cell types (airway epithelium, freshly isolated cells) should prove especially powerful. Pathways identified in this

study should be considered in on-going and future mechanistic studies focused on CF biology.

Accession Numbers

The GEO accession number for the expression data and covariates reported in this paper is GSE60690.

Supplemental Data

Supplemental Data include six figures and eight tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2014.12.022.

Acknowledgments

The work described in this paper was funded by the US National Heart, Lung, and Blood Institute (R01HL095396, M.R.K. and F.A.W.); the US National Institute of Diabetes and Digestive and Kidney Diseases (P30DK065988, W.K.O.); the US Cystic Fibrosis Foundation (RDP-026, W.K.O.); the Canadian Institutes of Health Research (MOP-258916, L.J.S.); and CF Canada (L.J.S.). Enrollment and sample collection was funded by the US Cystic Fibrosis Foundation (KNOWLE00A0, M.R.K.) and the US NIH (HL068890, M.R.K.), with additional analysis support from MH101819 (F.A.W.). Funds were provided through Aetna/U.S. Healthcare Chair (G.R.C.). Additionally, funds for genome-wide genotyping were generously provided by the US Cystic Fibrosis Foundation (CFF). The authors would like to thank the Cystic Fibrosis Foundation for the use of CF Foundation Patient Registry data to conduct this study. Additionally, we would like to thank the CF-affected individuals and their families, care providers, and clinic coordinators at CF Centers throughout the United States for their contributions to the CF Foundation Patient Registry. The authors would also like to thank the Canadian Consortium for CF genetic studies, the University of North Carolina DNA Laboratory, and the following for their contributions: for manuscript preparation, Syanne Olson; for recruitment and data entry, Sonya Adams, Colette Bucur, Leia Charnin, John Dunn, Patricia Miller, Sarah A. Norris, and Sally D. Wood; for genotyping, Rodney Gilmore; for data analysis, Anthony T. Dang, Michael V. Patrone, Clayton W. Commander, Evan J. Hawbaker, and Aaron Webel; and for bioinformatics, Hemant Kelkar, Tom Randall, and Annie Xu.

Received: September 26, 2014
Accepted: December 23, 2014
Published: January 29, 2015

Web Resources

The URLs for data provided herein are as follows:

Annmap Genome Browser (formerly X:MAP), http://annmap.cruk.manchester.ac.uk/

ArrayExpress – E-GEOD-36868, http://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-36868/

Bioconductor – GO.db, http://www.bioconductor.org/packages/release/data/annotation/html/GO.db.html

Bioconductor – hugene10stprobeset.db, http://www.bioconductor.org/packages/devel/data/annotation/html/hugene10stprobeset.db.html

Bioconductor – sva, http://www.bioconductor.org/packages/release/bioc/html/sva.html

Ensembl Genome Browser, http://www.ensembl.org/index.html

Gene Expression Omnibus (GEO), http://www.ncbi.nlm.nih.gov/geo/

GeneSetScan, http://www.mayo.edu/research/documents/gss-manual/DOC-20088346

HapMap downloads, http://hapmap.ncbi.nlm.nih.gov/downloads/frequencies/2010-08_phaseII+III/

OMIM, http://www.omim.org/

R statistical software, http://www.r-project.org/

RefSeq, http://www.ncbi.nlm.nih.gov/RefSeq

References

1. Wright, F.A., Strug, L.J., Doshi, V.K., Commander, C.W., Blackman, S.M., Sun, L., Berthiaume, Y., Cutler, D., Cojocaru, A., Collaco, J.M., et al. (2011). Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2. Nat. Genet. *43*, 539–546.
2. Green, D.M., Collaco, J.M., McDougal, K.E., Naughton, K.M., Blackman, S.M., and Cutting, G.R. (2012). Heritability of respiratory infection with *Pseudomonas aeruginosa* in cystic fibrosis. J. Pediatr. *161*, 290–295.e1.
3. Emond, M.J., Louie, T., Emerson, J., Zhao, W., Mathias, R.A., Knowles, M.R., Wright, F.A., Rieder, M.J., Tabor, H.K., Nickerson, D.A., et al.; National Heart, Lung, and Blood Institute (NHLBI) GO Exome Sequencing Project; Lung GO (2012). Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis. Nat. Genet. *44*, 886–889.
4. Li, W., Soave, D., Miller, M.R., Keenan, K., Lin, F., Gong, J., Chiang, T., Stephenson, A.L., Durie, P., Rommens, J., et al. (2014). Unraveling the complex genetic model for cystic fibrosis: pleiotropic effects of modifier genes on early cystic fibrosis-related morbidities. Hum. Genet. *133*, 151–161.
5. Sun, L., Rommens, J.M., Corvol, H., Li, W., Li, X., Chiang, T.A., Lin, F., Dorfman, R., Busson, P.F., Parekh, R.V., et al. (2012). Multiple apical plasma membrane constituents are associated with susceptibility to meconium ileus in individuals with cystic fibrosis. Nat. Genet. *44*, 562–569.
6. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al.; RACI consortium; GARNET consortium (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature *506*, 376–381.
7. Bønnelykke, K., Matheson, M.C., Pers, T.H., Granell, R., Strachan, D.P., Alves, A.C., Linneberg, A., Curtin, J.A., Warrington, N.M., Standl, M., et al.; Australian Asthma Genetics Consortium (AAGC); EArly Genetics and Lifecourse Epidemiology (EAGLE) Consortium (2013). Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. Nat. Genet. *45*, 902–906.
8. Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., et al.; International IBD Genetics Consortium (IIBDGC) (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature *491*, 119–124.
9. Cookson, W., Liang, L., Abecasis, G., Moffatt, M., and Lathrop, M. (2009). Mapping complex disease traits with global gene expression. Nat. Rev. Genet. *10*, 184–194.

10. Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S., et al. (2008). Genetics of gene expression and its effect on disease. Nature *452*, 423–428.
11. Nica, A.C., Montgomery, S.B., Dimas, A.S., Stranger, B.E., Beazley, C., Barroso, I., and Dermitzakis, E.T. (2010). Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genet. *6*, e1000895.
12. Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., and Cox, N.J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. *6*, e1000888.
13. Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C., et al. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science *315*, 848–853.
14. Zhang, W., Duan, S., Kistner, E.O., Bleibel, W.K., Huang, R.S., Clark, T.A., Chen, T.X., Schweitzer, A.C., Blume, J.E., Cox, N.J., and Dolan, M.E. (2008). Evaluation of genetic variation contributing to differences in gene expression between populations. Am. J. Hum. Genet. *82*, 631–640.
15. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.
16. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics *28*, 882–883.
17. Yates, T., Okoniewski, M.J., and Miller, C.J. (2008). X:Map: annotation and visualization of genome structure for Affymetrix exon array analysis. Nucleic Acids Res. *36*, D780–D786.
18. Taylor, C., Commander, C.W., Collaco, J.M., Strug, L.J., Li, W., Wright, F.A., Webel, A.D., Pace, R.G., Stonebraker, J.R., Naughton, K., et al. (2011). A novel lung disease phenotype adjusted for mortality attrition for cystic fibrosis genetic modifier studies. Pediatr. Pulmonol. *46*, 857–869.
19. Liang, L., Morar, N., Dixon, A.L., Lathrop, G.M., Abecasis, G.R., Moffatt, M.F., and Cookson, W.O. (2013). A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. Genome Res. *23*, 716–726.
20. Yu, C.Y., Theusch, E., Lo, K., Mangravite, L.M., Naidoo, D., Kutilova, M., and Medina, M.W. (2014). HNRNPA1 regulates HMGCR alternative splicing and modulates cellular cholesterol metabolism. Hum. Mol. Genet. *23*, 319–332.
21. Glass, D., Viñuela, A., Davies, M.N., Ramasamy, A., Parts, L., Knowles, D., Brown, A.A., Hedman, A.K., Small, K.S., Buil, A., et al.; UK Brain Expression consortium; MuTHER consortium (2013). Gene expression changes with age in skin, adipose tissue, blood and brain. Genome Biol. *14*, R75.
22. Barry, W.T., Nobel, A.B., and Wright, F.A. (2005). Significance analysis of functional categories in gene expression studies: a structured permutation approach. Bioinformatics *21*, 1943–1949.
23. Louie, R.J., Guo, J., Rodgers, J.W., White, R., Shah, N., Pagant, S., Kim, P., Livstone, M., Dolinski, K., McKinney, B.A., et al. (2012). A yeast phenomic model for the gene interaction network modulating CFTR-ΔF508 protein biogenesis. Genome Med *4*, 103.
24. Kumarswamy, R., Volkmann, I., and Thum, T. (2011). Regulation and function of miRNA-21 in health and disease. RNA Biol. *8*, 706–713.
25. Schaid, D.J., Sinnwell, J.P., Jenkins, G.D., McDonnell, S.K., Ingle, J.N., Kubo, M., Goss, P.E., Costantino, J.P., Wickerham, D.L., and Weinshilboum, R.M. (2012). Using the gene ontology to scan multilevel gene sets for associations in genome wide association studies. Genet. Epidemiol. *36*, 3–16.
26. Sheehy, N.T., Cordes, K.R., White, M.P., Ivey, K.N., and Srivastava, D. (2010). The neural crest-enriched microRNA miR-452 regulates epithelial-mesenchymal signaling in the first pharyngeal arch. Development *137*, 4307–4316.
27. Graff, J.W., Powers, L.S., Dickson, A.M., Kim, J., Reisetter, A.C., Hassan, I.H., Kremens, K., Gross, T.J., Wilson, M.E., and Monick, M.M. (2012). Cigarette smoking decreases global microRNA expression in human alveolar macrophages. PLoS ONE *7*, e44066.
28. Pittman, J.E., Calloway, E.H., Kiser, M., Yeatts, J., Davis, S.D., Drumm, M.L., Schechter, M.S., Leigh, M.W., Emond, M., Van Rie, A., and Knowles, M.R. (2011). Age of *Pseudomonas aeruginosa* acquisition and subsequent severity of cystic fibrosis lung disease. Pediatr. Pulmonol. *46*, 497–504.
29. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353–1358.
30. Wright, F.A., Sullivan, P.F., Brooks, A.I., Zou, F., Sun, W., Xia, K., Madar, V., Jansen, R., Chung, W., Zhou, Y.H., et al. (2014). Heritability and genomics of gene expression in peripheral blood. Nat. Genet. *46*, 430–437.
31. Ren, Y., Guo, L., Tang, X., Apparsundaram, S., Kitson, C., Deguzman, J., Fuentes, M.E., Coyle, L., Majmudar, R., Allard, J., et al. (2013). Comparing the differential effects of LPA on the barrier function of human pulmonary endothelial cells. Microvasc. Res. *85*, 59–67.
32. Shea, B.S., and Tager, A.M. (2012). Role of the lysophospholipid mediators lysophosphatidic acid and sphingosine 1-phosphate in lung fibrosis. Proc. Am. Thorac. Soc. *9*, 102–110.
33. Farinha, C.M., Matos, P., and Amaral, M.D. (2013). Control of cystic fibrosis transmembrane conductance regulator membrane trafficking: not just from the endoplasmic reticulum to the Golgi. FEBS J. *280*, 4396–4406.
34. Blohmke, C.J., Mayer, M.L., Tang, A.C., Hirschfeld, A.F., Fjell, C.D., Sze, M.A., Falsafi, R., Wang, S., Hsu, K., Chilvers, M.A., et al. (2012). Atypical activation of the unfolded protein response in cystic fibrosis airway cells contributes to p38 MAPK-mediated innate immune responses. J. Immunol. *189*, 5467–5475.
35. Rymut, S.M., Harker, A., Corey, D.A., Burgess, J.D., Sun, H., Clancy, J.P., and Kelley, T.J. (2013). Reduced microtubule acetylation in cystic fibrosis epithelial cells. Am. J. Physiol. Lung Cell. Mol. Physiol. *305*, L419–L431.
36. Muro, M., Mondejar-López, P., Moya-Quiles, M.R., Salgado, G., Pastor-Vivero, M.D., Lopez-Hernandez, R., Boix, F., Campillo, J.A., Minguela, A., Garcia-Alonso, A., et al. (2013). HLA-DRB1 and HLA-DQB1 genes on susceptibility to and protection from allergic bronchopulmonary aspergillosis in patients with cystic fibrosis. Microbiol. Immunol. *57*, 193–197.
37. Aron, Y., Polla, B.S., Bienvenu, T., Dall'ava, J., Dusser, D., and Hubert, D. (1999). HLA class II polymorphism in cystic fibrosis. A possible modifier of pulmonary phenotype. Am. J. Respir. Crit. Care Med. *159*, 1464–1468.
38. Laki, J., Laki, I., Németh, K., Ujhelyi, R., Bede, O., Endreffy, E., Bolbás, K., Gyurkovits, K., Csiszér, E., Sólyom, E., et al. (2006). The 8.1 ancestral MHC haplotype is associated with delayed onset of colonization in cystic fibrosis. Int. Immunol. *18*, 1585–1590.

39. Trowsdale, J., and Knight, J.C. (2013). Major histocompatibility complex genomics and human disease. Annu. Rev. Genomics Hum. Genet. *14*, 301–323.
40. Weidinger, S., Willis-Owen, S.A., Kamatani, Y., Baurecht, H., Morar, N., Liang, L., Edser, P., Street, T., Rodriguez, E., O'Regan, G.M., et al. (2013). A genome-wide association study of atopic dermatitis identifies loci with overlapping effects on asthma and psoriasis. Hum. Mol. Genet. *22*, 4841–4856.
41. Wissemann, W.T., Hill-Burns, E.M., Zabetian, C.P., Factor, S.A., Patsopoulos, N., Hoglund, B., Holcomb, C., Donahue, R.J., Thomson, G., Erlich, H., and Payami, H. (2013). Association of Parkinson disease with structural and regulatory variants in the HLA region. Am. J. Hum. Genet. *93*, 984–993.
42. Armstrong, D.L., Zidovetzki, R., Alarcón-Riquelme, M.E., Tsao, B.P., Criswell, L.A., Kimberly, R.P., Harley, J.B., Sivils, K.L., Vyse, T.J., Gaffney, P.M., et al. (2014). GWAS identifies novel SLE susceptibility genes and explains the association of the HLA region. Genes Immun. *15*, 347–354.
43. Thomas, S.R., Jaffe, A., Geddes, D.M., Hodson, M.E., and Alton, E.W.F.W. (1999). Pulmonary disease severity in men with deltaF508 cystic fibrosis and residual chloride secretion. Lancet *353*, 984–985.
44. Sermet-Gaudelus, I., Vallée, B., Urbin, I., Torossi, T., Marianovski, R., Fajac, A., Feuillet, M.N., Bresson, J.L., Lenoir, G., Bernaudin, J.F., and Edelman, A. (2002). Normal function of the cystic fibrosis conductance regulator protein can be associated with homozygous (Delta)F508 mutation. Pediatr. Res. *52*, 628–635.
45. Hansen, K.D., Sabunciyan, S., Langmead, B., Nagy, N., Curley, R., Klein, G., Klein, E., Salamon, D., and Feinberg, A.P. (2014). Large-scale hypomethylated blocks associated with Epstein-Barr virus-induced B-cell immortalization. Genome Res. *24*, 177–184.
46. Saini, Y., Dang, H., Livraghi-Butrico, A., Kelly, E.J., Jones, L.C., O'Neal, W.K., and Boucher, R.C. (2014). Gene expression in whole lung and pulmonary macrophages reflects the dynamic pathology associated with airway surface dehydration. BMC Genomics *15*, 726.