Disease-Specific Gene Expression Profiling in Multiple Models of Lung Disease

Christina C. Lewis, Jean Yee Hwa Yang, Xiaozhu Huang, Suman K. Banerjee, Michael R. Blackburn, Peter Baluk, Donald M. McDonald, Timothy S. Blackwell, Vijaya Nagabhushanam, Wendy Peters, David Voehringer, David J. Erle

Online Data Supplement

ONLINE DATA SUPPLEMENT

METHODS

Microarray experiments

Spotted 70-mer oligonucleotide microarrays were produced at the UCSF NHLBI Shared Microarray Facility using the Operon Mouse Genome Oligo set 2.0, supplemented by some additional oligonucleotides (1). RNA was purified from whole lungs, reverse transcribed to amino allyl-modified cDNA, and labeled with Cy3 or Cy5 as previously described (1). Two cDNA preparations were hybridized simultaneously to each array. One cDNA preparation (labeled with Cy5) was derived from a single experimental or control mouse and the other cDNA preparation (labeled with Cy3) was made from a lung reference pool RNA. The lung reference pool RNA was generated from equivalent numbers of 5-7 week-old male C57BL/6J, BALB/CJ, C3H, and FVB/NJ mice (Jackson Laboratory, Bar Harbor, ME). Microarray images were analyzed using Bioconductor software (http://www.bioconductor.org) for quality control. In order to minimize experimental noise, no background subtraction was used. This method tends to underestimate fold-difference in some cases since signals below the threshold of detection are treated as if expression was at the level of detection (2). A complete description of the arrays, the experimental protocols, and the raw array data are available from the Gene Expression Omnibus website (http://www.ncbi.nlm.nih.gov/geo, accession number GSE4231).

Microarray Data Analysis

A linear model was used to estimate relative transcript expression (fold-difference) for each experimental group compared to the corresponding control group. The model can be written as $y_{ijk} = \mu + \alpha_{ij} + \varepsilon_{ijk}$ where $i = 1,...12$ and j = 0 or 1. y is defined as the log-ratio of the expression data between an experimental or control mouse and the lung reference pool; ε represents the error associated with the log-ratio measurements; and k represents the number of replications. The subscript i refers to the 12 different mouse models. The subscript j = 0 refers to the control condition and j = 1 refers to the experimental condition within each mouse model. For each model, the comparisons of experimental to control ($\alpha_{i1} - \alpha_{i0}$) were made for each gene and log fold differences, moderated T-statistics and p-values (adjusted using Holm method) were calculated. We considered the differences in transcript expression between each experimental group and the corresponding control group to be statistically significant when the adjusted p-value was less than 0.05. Genes with missing values were excluded and hierarchical clustering performed, as previously described (3), with Acuity software (Axon Instruments, Union City, CA) using the Pearson correlation distance metric and the complete linkage method.

PubMed literature database mining

We used a systematic approach to determine whether genes identified in the microarray data analysis had been previously associated with lung diseases of interest. First, the NCBI Gene database was used to obtain synonyms for each gene. The official gene symbol and all unambiguous synonyms were used to make a list of search terms for PubMatrix (4) (http://pubmatrix.grc.nia.nih.gov/). The terms "asthma," "allergy,"

"bacterial infection," "lipopolysaccharide," "pulmonary fibrosis," and "bleomycin" were used as modifier terms. PubMatrix determined the number of references containing associations of interest by querying the PubMed database (e.g., "(CCL2 OR MCP-1) AND asthma"). If no references were identified by PubMatrix, all PubMed citations for the gene of interest were reviewed manually to identify any relevant publications not detected using PubMatrix. In total, PubMed citations for 192 genes identified in the microarray experiments were analyzed during January of 2007; of these, 103 were found to have relevant lung disease associations using PubMatrix and 7 more were found to have relevant associations by manual review of PubMed.

Real-time PCR

PCR primer sequences were selected from the PrimerBank database (5) (http://pga.mgh.harvard.edu/primerbank) and the qPrimerDepot database (6) (*http://mouseprimerdepot.nci.nih.gov/*) and synthesized by Sigma-Genosys (The Woodlands, TX). Genes analyzed by the SYBR-Green method were Ear11 (PrimerBank probe pair 31981460a1), Sprr2a (PrimerBank probe pair 31560549a1), Guca1a (PrimerBank probe pair 6680131a3), Arg1 (PrimerBank probe pair 7106255a1), Matk (PrimerBank probe pair 6754646a3), BC021513 (PrimerBank probe pair 21450275a1), Fxyd4 (PrimerBank probe pair 16258807a1), Mxd1 (PrimerBank probe pair 6754604a2), Sfn (PrimerBank probe pair 9055338a1), Il1a (qPrimerDepot probe pair for RefSeqID NM_010554), Il1b (qPrimerDepot probe pair for RefSeqID NM_008361), Tnf (qPrimerDepot probe pair for RefSeqID NM_013693), Nfkb2 (qPrimerDepot probe pair for RefSeqID NM_019408), Nfkbiz (qPrimerDepot probe pair for RefSeqID

NM_030612), Slfn4 (qPrimerDepot probe pair for RefSeqID NM_011410), Mlp (qPrimerDepot probe pair for RefSeqID NM_010807), Hoxa5 (qPrimerDepot probe pair for RefSeqID NM_010453), Eln (qPrimerDepot probe pair for RefSeqID NM_007925), Lama1 (qPrimerDepot probe pair for RefSeqID NM_008480), Col5a2 (qPrimerDepot probe pair for RefSeqID NM_007737), Ltbp2 (qPrimerDepot probe pair for RefSeqID NM_013589), Phlda3 (qPrimerDepot probe pair for RefSeqID NM_013750), Loxl2 (qPrimerDepot probe pair for RefSeqID NM_033325), and Wnt7b (qPrimerDepot probe pair for RefSeqID NM_009528). Cycle thresholds for each of these genes were normalized using the mean cycle threshold for three housekeeping genes, Ppia (PrimerBank probe pair 6679439a1), Gapdh (PrimerBank probe pair 6679937a1), Actb (PrimerBank probe pair 6671509a1), and Rps14 (qPrimerDepot probe pair for RefSeqID NM_020600). All other genes that underwent validation by real-time PCR were analyzed using the TaqMan method and a set of existing primers and probes that have been described previously (1). TaqMan cycle thresholds were normalized using TaqMan primers and probes for the same three housekeeping genes used for SYBR green PCR. First-strand cDNA synthesis and PCR were conducted using the ABI Prizm 7700 Sequence Detection System (Applied Biosystems, Foster City, CA). Gene expression differences were calculated from the mean values for experimental and control groups using the ΔΔCt method (7, 8). Samples from at least three mice were analyzed in all experimental and control groups, except for the control group for the bleomycin 21 day pulmonary fibrosis model, where only 2 samples were available.

Supplemental Table E1. Other Gene Expression Changes Characteristic of Bacterial Infection Models†

Description	Symbol	LPS aerosolized	LPS IP	*P. aeruginosa*	*M. tuberculosis*	*M. pulmonis*
‡RIKEN cDNA 2310014H01 gene	2310014H01Rik	2.1*	1.8*	3.3*	2.3*	2.4*
‡guanylate binding protein 6	Gbp6	3.3*	5.9*	2	7.9*	3.1*
‡cDNA sequence BC004022	BC004022	1.8*	3.3*	1.7*	2.9*	2.7*
‡cDNA sequence BC006779	BC006779	2.2*	5.7*	2.6*	2.0*	1.6
B-cell translocation gene 1, anti-proliferative	Btg1	2.1*	2.9*	2.9*	1.3	2.3*
‡CCR4 carbon catabolite repression 4-like (S. cerevisiae)	Ccrn4l	2.6*	6.7*	5.6*	1.1	3.1*
DNA segment, human D4S114	D0H4S114	-2.4*	-2.4*	-1.3	-1.9*	-1.9*
‡ets homologous factor	Ehf	2.2*	1.6*	2.1*	1.2	1.5*
‡elongation factor RNA polymerase II 2	Ell2	1.7*	1.9*	2.2*	0.9	1.7*
fibrinogen-like protein 2	Fgl2	3.5*	3.7*	3.4*	5.3*	2.8*
fos-like antigen 2	Fosl2	3.2*	5.2*	8.7*	1.8	2.8*
G protein-coupled receptor 109A	Gpr109a	2.3*	3.3*	2.9*	2.5*	10.3*
‡GRAM domain containing 1A	Gramd1a	2.8*	3.0*	2.1*	1.2	1.6*
‡glutamate receptor, ionotropic, N-methyl D-asparate-associated protein 1 (glutamate binding)	Grina	2.0*	1.9*	2.0*	2.5*	3.6*
Hect domain and RLD 5	Herc5	3.4*	10.0*	2.3	4.1*	4.1*
homeo box A5	Hoxa5	-2.0*	-2.5*	-1.2	-2.4*	-2.5*
‡Immediate early response 5	Ier5	2.0*	2.4*	3.7*	1.4	2.1*
‡Immunoglobulin superfamily, member 6	Igsf6	1.6	2.2*	2.3*	2.0*	3.6*
‡MAX dimerization protein 1	Mxd1	3.2*	7.3*	5.3*	4.1*	13.7*
matrix metallopeptidase 15	Mmp15	2.7*	4.1*	4.5*	1.6*	4.7*
‡opioid growth factor receptor	Ogfr	2.5*	3.0*	1.3	1.8*	1.6*
‡poly (ADP-ribose) polymerase family, member 14	Parp14	3.2*	9.9*	2.2	3.2*	3.3*
6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	Pfkfb3	3.1*	2.2*	2.6*	0.7	2.5*
‡placenta-specific 8	Plac8	3.5*	8.2*	2.0	7.4*	8.6*
‡plectin 1	Plec1	2.7*	4.4*	1.2	2.2*	2.6*
purine-nucleoside phosphorylase	Pnp	1.7*	3.6*	1.3	3.1*	1.6*
proline-serine-threonine phosphatase-interacting protein 2	Pstpip2	1.7	2.0*	3.1*	1.8*	2.7*
schlafen 1	Slfn1	4.3*	8.8*	4.8*	3.5*	10.5*
schlafen 4	Slfn4	7.8*	36.5*	6.1*	4.6*	21.8*
‡sorting nexin 10	Snx10	3.2*	6.5*	1.5	2.8*	4.9*
‡syntaxin 11	Stx11	3.4*	4.0*	2.1	2.4*	2.6*
transporter 2, ATP-binding cassette, sub-family B (MDR/TAP)	Tap2	1.9*	3.5*	1.2	4.4*	2.3*
‡transmembrane protein 37	Tmem37	2.1*	2.2*	4.0*	2.1*	2.0*
‡uridine phosphorylase 1	Upp1	3.3*	8.2*	1.4	3.4*	5.7*

* Significantly different from controls by *t* test.

† Includes additional genes not shown in Table 3 that were significantly increased in at least 4 of the 5 bacterial infection models but not in any of the other 7 models. Values represent fold-change compared with controls.

‡ This gene was not found to be associated with bacterial infection in the PubMed database.

Supplemental Table E2. Gene Expression Decreases Characteristic of Bleomycin-Induced Lung Disease.†

Description	Symbol	Bleomycin 7 d	Bleomycin 21 d	Other Model with decreased expression
‡RIKEN cDNA 1110018M03 gene	1110018M03Rik	-1.8*	-2.2*	LPS intraperitoneal -3.0*
‡acyl-Coenzyme A oxidase-like	Acoxl	-1.9*	-3.2*	---
‡acyl-CoA synthetase short-chain family member 2	Acss2	-1.6*	-1.5*	LPS intraperitoneal -1.6*
adenylate cyclase 8	Adcy8	-1.6*	-1.7*	---
‡chromobox homolog 7	Cbx7	-2.0*	-1.9*	---
‡cysteine dioxygenase 1, cytosolic	Cdo1	-1.5*	-1.8*	LPS intraperitoneal -1.8*
‡chordin-like 1	Chrdl1	-1.7*	-2.1*	*M. pulmonis* -1.8*
‡cytochrome P450, family 2, subfamily d, polypeptide 22	Cyp2d22	-1.8*	-2.0*	*M. tuberculosis* -2.2*
‡Fas apoptotic inhibitory molecule 2	Faim2	-2.0*	-3.1*	---
‡forkhead box A1	Foxa1	-1.6*	-1.6*	---
‡homocysteine-inducible, endoplasmic reticulum stress-inducible, ubiquitin-like domain member 1	Herpud1	-1.8*	-1.6*	Ovalbumin BALB/c -1.5*
‡LIM and senescent cell antigen-like domains 1	Lims1	-1.8*	-2.0*	---
‡lecithin-retinol acyltransferase (phosphatidylcholine-retinol-O-acyltransferase)	Lrat	-2.3*	-3.5*	*Aspergillus* extract -2.5*
‡methyl CpG binding protein 2	Mecp2	-1.7*	-1.6*	*N. brasiliensis* -1.5*
‡matrix metalloproteinase 15	Mmp15	-1.8*	-1.7*	*N. brasiliensis* -1.6*
‡membrane-spanning 4-domains, subfamily A, member 4D	Ms4a4d	-2.1*	-2.2*	*Aspergillus extract* -2.1*
‡musculoskeletal, embryonic nuclear protein 1	Mustn1	-1.7*	-1.5*	LPS intraperitoneal -2.1*
nuclear receptor subfamily 4, group A, member 1	Nr4a1	-2.3*	-2.3*	---
‡phospholipase A2, group IB, pancreas	Pla2g1b	-2.4*	-1.8*	Ovalbumin BALB/c -1.5*
‡ring finger protein 167	Rnf167	-1.6*	-1.6*	LPS intraperitoneal -1.6*
‡wingless related MMTV integration site 10b	Wnt10b	-1.6*	-2.0*	---

* Significantly different from controls by *t*- test.

† Includes selected genes that were significantly increased in both Bleomycin models and only 1 of the other 10 models. Values represent fold-change compared with controls.

‡ This gene was not found to be associated with bleomycin or pulmonary fibrosis in the PubMed database.

Supplemental Table E3. Real-time PCR Verification of Selected Transcript Expression Changes in 8 Models.†

Bacterial infection models						
	LPS aerosolized		**LPS IP**		***M. pulmonis***	
Symbol	array	PCR	array	PCR	array	PCR
IL1a	2.5	2.0	1.9	7.9	11.5	151.0
IL1b	7.2	44.3	15.3	18.8	20.7	176.5
TNFa	5.3	9.0	5.8	11.2	29.7	30.6
NFkb2	4.0	2.1	4.3	2.1	2.8	3.1
NFkbiz	3.9	242.3	8.7	91.6	6.5	7.0
Slfn4	7.8	18.4	36.5	13.0	21.8	64.6
Mxd1	3.2	5.1	7.3	1.9	13.7	13.5
Marcksl1	4.0	18.6	10.0	6.6	4.5	10.7

Bleomycin-induced lung disease models				
	Acute lung injury (7d)		**Fibrosis (21 d)**	
Symbol	array	PCR	array	PCR
Eln	4.8	5.2	4.1	3.7
Lama1	1.8	8.9	3.1	7.6
Col5a2	2.9	3.1	2.2	2.9
Ltbp2	1.8	2.0	3.1	3.0
Phlda3	3.4	1.6	2.2	1.7
Sfn	2.6	4.2	2.3	4.3

Allergic models						
	Ovalbumin BALB/C		**Aspergillus extract**		**IL-13 overexpression**	
Symbol	array	PCR	array	PCR	array	PCR
Arg1	9.0	126.0	8.9	64.3	4.6	119.7
Sprr2a	8.5	33.9	6.8	102.6	5.8	398.2
Fxyd4	2.7	5.6	12.5	211.5	4.1	61.3
Tff2	4.3	2.6	1.7	6.9	5.0	1.0
Agr2	3.9	5.3	2.5	8.4	4.4	33.5
Muc5b	3.3	3.0	2.5	4.4	4.6	31.6
Reg3g	5.5	4.3	2.7	3.8	3.7	5.2
Guca1a	10.2	2.7	3.5	3.6	2.4	1.9
Muc5ac	13.0	48.5	3.8	16.1	4.2	60.0
Matk	2.8	9.2	2.8	21.7	1.6	5.1
Zgpat	2.1	-2.0	2.8	-1.4	3.6	1.0
Alox15	4.6	4.5	3.8	5.0	1.0	1.0
Chia	2.6	1.3	6.2	14.2	8.5	15.4
Atp2a1	11.2	-2.0	6.0	-33.3	6.7	-10.0
Ear11	20.5	260.3	6.7	53.3	11.3	460.1
Clca3	59.9	306.0	62.9	4628.3	34.8	3831.7

† Values represent fold-change compared with controls. The descriptive names of genes can be found in in Table 3 and Supplemental Table 1 (bacterial infection models), Table 4 (bleomycin models), and Table 5 (allergic models).

REFERENCES

E1. Kuperman, D.A., Lewis, C.C., Woodruff, P.G., Rodriguez, M.W., Yang, Y.H., Dolganov, G.M., Fahy, J.V., and Erle, D.J. 2005. Dissecting asthma using focused transgenic modeling and functional genomics. *J Allergy Clin Immunol* 116:305-311.

E2. Scharpf, R.B., Iacobuzio-Donahue, C.A., Sneddon, J.B., and Parmigiani, G. 2007. When should one subtract background fluorescence in 2-color microarrays? *Biostatistics*. [Epub ahead of print].

E3. Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95:14863-14868.

E4. Becker, K.G., Hosack, D.A., Dennis, G., Jr., Lempicki, R.A., Bright, T.J., Cheadle, C., and Engel, J. 2003. PubMatrix: a tool for multiplex literature mining. *BMC Bioinformatics* 4:61.

E5. Wang, X., and B., S. 2003. A PCR primer bank for quantitative gene expression analysis. *Nucleic Acids Res* 31:1-8.

E6. Cui, W., Taub, D.D., and Gardner, K. 2007. qPrimerDepot: a primer database for quantitative real time PCR. *Nucleic Acids Res* 35:D805-809.

E7. Livak, K.J. 1998. Sequence detector user bulletin #2. *Perkin Elmer Applied Biosystems, Inc.*

E8. Livak, K.J., and Schmittgen, T.D. 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2^{-ddCT} Method. *Methods* 25:402-408.