Method

Orthologous gene-expression profiling in multi-species models: search for candidate genes

Dmitry N Grigoryev*, Shwu-Fan Ma*, Rafael A Irizarry†, Shui Qing Ye‡, John Quackenbush§ and Joe GN Garcia¶

Addresses: *Center for Translational Respiratory Medicine, Gene Expression Profiling Core, Division of Pulmonary and Critical Care Medicine, Johns Hopkins University School of Medicine, Hopkins Bayview Circle, Baltimore, MD 21224, USA. †Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA. ‡Center for Translational Respiratory Medicine, Johns Hopkins University, Eastern Ave, Baltimore, MD 21224, USA. §The Institute for Genomic Research, Medical Center Drive, Rockville, MD 20850, USA. ¶Center for Translational Respiratory Medicine, Division of Pulmonary and Critical Care Medicine, Johns Hopkins University, East Monument Street, Baltimore, MD 21287, USA.

Correspondence: Joe GN Garcia. E-mail: drgarcia@jhmi.edu

Abstract

Microarray-driven gene-expression profiles are generally produced and analyzed for a single specific experimental model. We have assessed an analytical approach that simultaneously evaluates multi-species experimental models within a particular biological condition using orthologous genes as linkers for the various Affymetrix microarray platforms on multi-species models of ventilator-associated lung injury. The results suggest that this approach may be a useful tool in the evaluation of biological processes of interest and selection of process-related candidate genes.

Background

Mechanical ventilation is a life-saving therapy for numerous critical illnesses. However, it is now recognized that ventilation with excessive tidal volumes, leading to hyperexpansion or excessive mechanical shear, is potentially directly harmful to susceptible patients. The benefits of lower tidal volumes, which reduce lung-cell stretch, have now clearly been established [1]. The clinical presentation of ventilator-associated lung injury (VALI) is identical to that of other causes of acute lung injury (ALI) and is characterized by increased pulmonary edema. Important studies by Parker [2,3] and Webb and Tierney [4] demonstrated changes in microvascular permeability in isolated lung and intact animal models exposed to increased airway pressures, suggesting that these changes in permeability may in large part be attributed to the effects of mechanical stimuli on various cell-signaling pathways [5,6]. Although several studies have suggested a genetic basis for susceptibility to VALI [7-9], few candidate genes have been implicated in this process.

To identify major genes associated with VALI, we examined gene-expression profiles of several *in vivo* models (rat, mouse, and dog) of ventilator-induced ALI. As a main component of ALI is presumed to involve biophysical stress-induced leakage of the pulmonary vasculature [10], we also included human lung vascular endothelial cells exposed to high-level cyclic stretch as a human *in vitro* model of mechanical stress. Gene-expression profiling of these models was performed and analyzed using species-specific Affymetrix GeneChips. The individual analysis of species-specific arrays produced large lists of candidate genes and several challenges, with the most notable being an excessive number of genes (ranging from 548 candidates in the rat to 963 candidates in the human model) for candidate gene selection. While meta-analysis strategies exist for narrowing candidate gene selection from multiple experimental systems [11-13], this analysis can only be applied to the same species cross-platform array comparison. To use this approach for analysis of experiments involving diverse species we speculated that

multi-species gene-expression profiles could be linked using the Eukaryote Gene Orthologs database (EGO [14]).

Orthologs are genes in different species that have evolved from a common ancestral gene by speciation and generally retain a similar function in the course of evolution. We speculated that overlapping responses to mechanical stretch in orthologous genes across species might reveal candidate genes involved in an evolutionarily conserved defense mechanism to lung injury that might be triggered by ventilator-induced lung injury. Previous studies of three-way comparative analysis of human, mouse and dog DNA [15] showed that the majority of highly conserved human-mouse elements are also conserved in the dog. Furthermore, Frazer *et al.* [16] speculated that comparing human sequence with those of multiple species might be an effective approach for distinguishing actively conserved elements from elements that simply result from a shared ancestry. On the basis of these observations, we predicted that a common stimulus (mechanical stretch) across four different species will initiate actively conserved mechanisms that defend the lungs against adverse environment factors or bacterial products. To select genes involved in these defense mechanisms the functionally related genes from different species should be first identified.

Despite the availability of tools for comparing gene-expression data from Affymetrix GeneChip arrays designed for different species [17,18], there are limited resources for simultaneous array-data analysis across multiple species-specific platforms (GeneChip IDs U34, U74, U95, U133). GeneHopper [18,19], which uses the UniGene and HomoloGene databases to provide comparisons between arrays, is useful for linkage of selected genes of interest from different array platforms, but is less suitable for linking expression sequence tags (ESTs) and uncharacterized genes represented on arrays. Moreover, the database for this software is not yet complete, and does not include the widely used HG_U95Av2 GeneChip.

A better alternative is RESOURCERER [17,20], which is based on the TIGR Eukaryotic Gene Ortholog (EGO) database [21], and contains information for all commercially available Affymetrix GeneChips. However, RESOURCERER allows comparison of only two chips simultaneously and cannot be used directly for multi-species analysis. Therefore, we assembled a database using ortholog links (identified by RESOURCERER) between the most commonly used Affymetrix rat, mouse and human GeneChips (U34A, U74A, U95A and U133A) for our multi-species cross-platform gene-expression analysis.

We first calculated gene-expression changes for each tested species and linked expression values obtained for orthologous genes. Orthologous genes exhibiting similar patterns of expression across all species were selected as VALI-related candidates under the assumption that gene-expression

responses conserved across evolutionary history would be most likely to reveal fundamental biological responses to VALI. After normalizing gene-expression values across species, we next identified orthologous genes with statistically significant changes in response to VALI. A biologically significant fold-change in gene-expression level was determined using MAPPFinder [22,23] by linking selected genes to Gene Ontology (GO) biological processes and identifying functional categories that were significantly regulated. This filtering produced a candidate list of 69 genes that were significantly affected by mechanical stretch. A literature search for these genes using PubMatrix [24] identified 12 genes as related to ALI as well as six new VALI-related candidate genes. Our analytical gene ortholog approach also revealed a number of changes in unsuspected GO processes and biological pathways that may provide new insights and potential therapies in ALI. Thus, this technique offers the capacity to identify genes that are likely to be missed by individual species analysis and facilitates application of a meta-analysis approach to multi-species analyses.

Results

To maximize the number of valid cross-species comparisons, we focused our analysis on the human, mouse and rat Affymetrix 'A' GeneChips, which contain the majority of 'named' or functionally classified genes and the least number of unannotated ESTs. Ortholog tables for each pair were generated using RESOURCERER. This software provides a table in which rows contain paired orthologous probe IDs; IDs corresponding to Affymetrix internal controls were ignored in further analysis. Because the U133A GeneChip contained the largest number of probe IDs (22,215) as compared to U95A, U74A, and U34A chips (12,588, 12,422, and 8,740 probe IDs, respectively), the U133A genes were selected as the reference gene set for orthologous comparisons. As anticipated, the total number of reference genes to participate in forming ortholog pairs (identified by RESOURCERER) was always higher than that of corresponding orthologs (Table 1), which justified our selection of the U133A array as the reference platform.

The linkage of all four arrays identified 3,077 genes common to the U133A reference gene ortholog nodes (Figure 1). An example of an ortholog node for the ODC-1 gene is shown in Figure 2a. This ortholog node was missing one link, rendering our ortholog-linked database incomplete. Therefore, we identified all orthologs with missing links (Figure 2b) and connected them as putative orthologs on the basis of homology to the common reference gene.

Gene-expression data for populating our ortholog-linked database was generated by hybridization of total mRNA from rat, mouse and dog lung tissues and human endothelial cell cultures to GeneChips U34A, U74A, U133A and U95A, respectively. All hybridizations were represented by a mini-
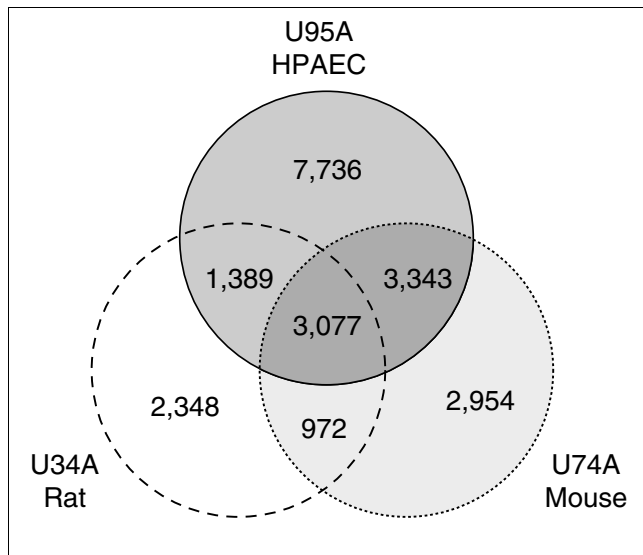
Table 1

Relationship of EGO orthologs between selected Affymetrix GeneChips

GeneChip	Reference genes on U133A	Orthologs for reference genes	Reference gene ortholog pairs
U95A	11,437	9,677	15,545
U74A	7,665	6,013	10,346
U34A	4,872	4,630	7,786

The number of orthologs linked to the U133A arrays is lower than the number of reference genes because of the same ortholog being shared by different reference genes. The number of reference gene ortholog pairs is always higher than the number of reference genes itself; this is attributed to the multiple orthologs for the common reference genes (see Figure 4).

mum of three control and four mechanical stretch-challenged samples, with the exception of the rat model which had two control and two stretch-affected samples (see Materials and methods). The signal intensities produced during hybridization were extracted from hybridization images using Affymetrix software MAS 5.0 and ratios of transcript abundance calls were computed. Rat, mouse, and human array assays produced 51%, 52% and 49% present ($p < 0.04$) and marginal ($p < 0.06$) transcript abundance calls, respectively. In contrast, however, the canine hetero-hybridization to the human U133A GeneChip created only 17% marginal and present calls (Figure 3a). Probe-level analysis revealed that poor cross-species hybridization to a subset of the probe pair sets was responsible for the loss of many present calls from the canine

Figure 1
Overlaps between rat (U34A GeneChip), mouse (U74A GeneChip) and human (U95A GeneChip) Affymetrix array platforms based on the human (U133A GeneChip) ortholog assignments. The sum of numbers inside each circle represents the total number of ortholog pairs formed with reference genes on the U133A GeneChip by corresponding arrays (see also Table 1). The reference genes formed 3,077 pairs with corresponding orthologs that were represented on all depicted arrays.

array data. To address this, we adjusted results of this cross-species hybridization by modifying U133A array probe-set compositions on the basis of differences between dog and human DNA. The poorly performing probes were also identified in the species-specific hybridizations and subsequently masked using masking protocol embedded in MAS 5.0. When modified probe sets were reprocessed by MAS 5.0, the ratio of present calls was increased on average by 25% (Figure 3b). Next, we replaced remaining absent calls with the corresponding chip background value (see Materials and methods), which allowed us to use all available data on each chip. Subsequent statistical analysis was conducted for each experimental system individually and four generated gene lists were later used for comparison with gene lists generated using the ortholog approach.

For statistical analysis of combined cross-platform expression data, we pooled control and mechanical stretch-challenged samples from all tested species into corresponding groups $n_{control} = 11$ ($n_{rat} = 2$, $n_{mouse} = 3$, $n_{canine} = 3$, and $n_{HPAEC} = 3$) and $n_{stretch} = 14$ ($n_{rat} = 2$, $n_{mouse} = 4$, $n_{canine} = 4$, and $n_{HPAEC} = 4$). Because these arrays contain multiple paralogues (similar sequences in a single organism), the multiple orthologs for the same reference gene were identified (Figure 4). Therefore, approximately 62% of formed ortholog groups failed to follow the $n_{control} = 11/n_{stretch} = 14$ pattern. To avoid unequal contribution of each species to the statistical analysis, the expression values of multiple paralogues were averaged and then $n_{control} = 11/n_{stretch} = 14$ set was built. Once the groups for comparison were formed, we used the independent variance double-tailed *t*-test for statistical evaluation of changes in gene expression of reference genes and their orthologs. This analysis identified significant changes in the expression of 141 reference genes and their corresponding orthologs across all experimental systems.

To further refine this list, we established a fold-change cutoff for biologically significant gene-expression changes based on the analysis of the relationship of the biological processes driven by these genes. Starting from the notion that genes coding for proteins involved in the same biological processes are regulated in coordinated manner, and that expression of members of a given bioprocess is more likely to be co-regu-
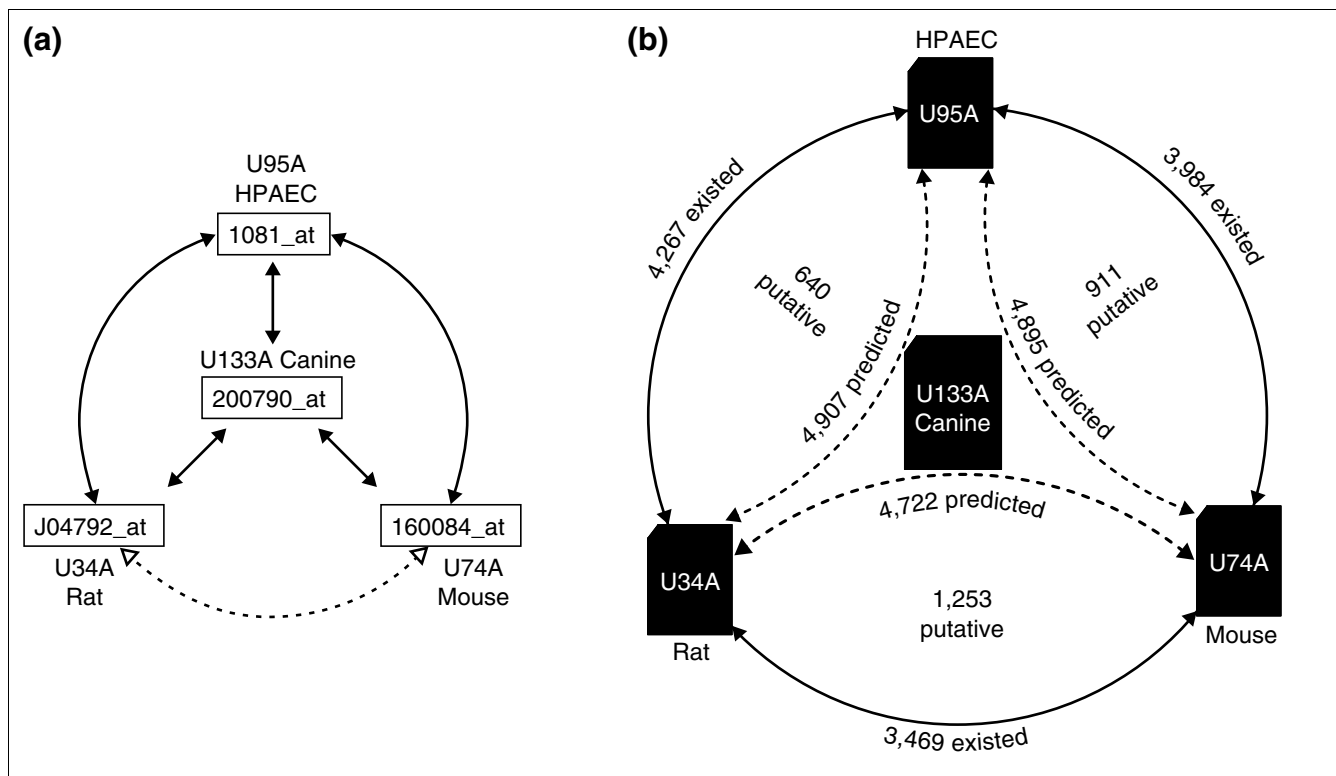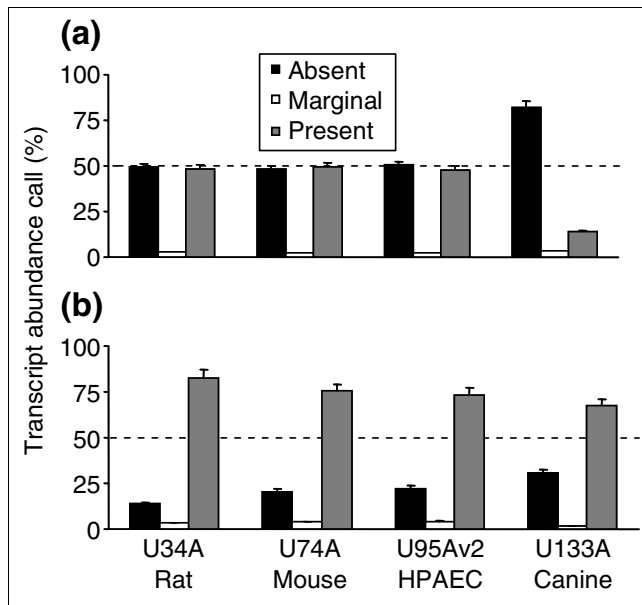
Figure 2
Schema of the centric approach in ortholog-linked database building and putative ortholog detection. **(a)** An example of putative ortholog creation for the ornithine decarboxylase 1 (ODC-1) gene. U74A and U34A probe IDs were EGO orthologs (solid line) for the U133A and U95A ODC-1 gene but were not directly linked (dashed line) either in EGO or in the Affymetrix ortholog table. **(b)** The reference genes common to all arrays (see Figure 1) and their corresponding orthologs for U95A-U74A, U95A-U34A, and U74A-U34A pairs were permutated and all possible combinations counted (dashed lines). EGO combinations were retrieved from RESOURCERER-generated tables for these paired arrays and counted (solid lines). The difference in the predicted and existing pairs represents the number of putative orthologs to be created, based on homology to the common reference gene.

lated rather than inversely regulated [25], we speculated that an increased ratio of inversely regulated bioprocesses at low fold-change cutoff values (Figure 5a,b) is due to the contribution of spurious (false-positive) changes in gene expression assigned to low fold-change values. As shown in Figure 5a for inflammatory response bioprocess at 1.1- and 1.15-fold-change cutoffs, this process was classified as inversely regulated. However, with a 1.2-fold-change cutoff, this becomes a co-regulated pathway. In contrast, the DNA-dependent regulation of transcription bioprocess (Figure 5b) is classified as an inversely regulated through all tested fold-change cutoffs. Although most low fold-change genes in this process were eliminated, the ratio of upregulated and downregulated genes remained constant and was stabilized beyond the 1.3-fold-change cutoff. From these observations we propose that the point at which sharp changes in the number of genes involved in GO bioprocesses subsides could be considered as biologically meaningful fold-change cutoff.

The bioprocesses affected by mechanical stretch were identified using MAPPFinder [13] software designed by BayGenomics PGA group for dynamic linkage of gene-expression

data to the GO [26] hierarchy. When we analyzed the gene pool that included genes with slight changes in their expression (1.1-fold), the MAPPFinder identified 432 bioprocesses, with 288 activated and 147 suppressed bioprocesses. Of these 432 bioprocesses, a total of 54 bioprocesses were common to both groups and, therefore, were classified as inversely regulated (shared) bioprocesses (Figure 5). To identify the point at which the number of the shared bioprocesses will approach the monotonic phase at which only real inversely regulated pathways will survive, we tested our gene list by gradually increasing the stringency of the fold-change cutoff. The fold-change cutoff of ±1.3 and ±1.35 satisfied this condition for inversely regulated and co-regulated bioprocesses, respectively (Figure 5). Using this filtering strategy and applying ±1.3-fold-change cutoff, we further refined our gene list to 69 genes (see Additional data files) which comprised 61 upregulated and 8 downregulated genes.
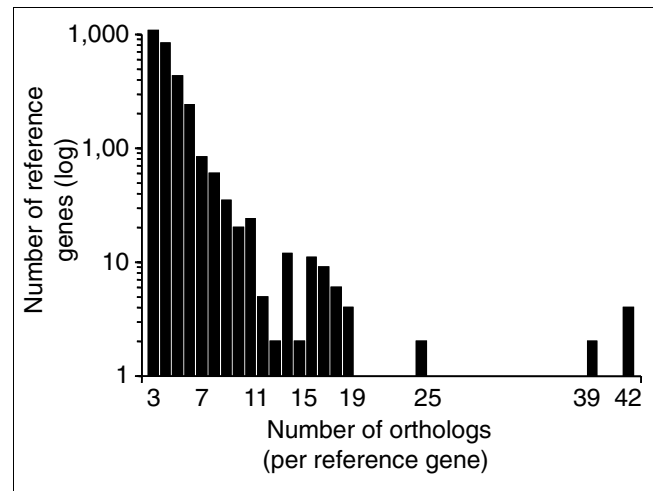
We next matched these 69 genes against the PubMed database using the PubMatrix [24] software tool. This analysis identify 12 genes that were extensively linked to lung-injury-related articles, with six of these genes also linked to mechan-

Figure 3
Experimental data used for populating the ortholog-link database. **(a)** Using Affymetrix MAS 5.0 software, absent (black), marginal (white) and present (gray) transcript-abundance calls were counted for each experimental dataset and the values obtained expressed as a percentage of all calls. **(b)** By masking poorly performing probes for U95A, U74A and U34A, the present call ratio for these GeneChips was increased by 25%. As dog mRNA was hybridized to the human U133A chip, the present call ratio for this hetero-hybridization was much lower than that in other experiments. We therefore corrected U133A probe sets for differences in gene sequence between human and canine, which increased the present call ratio by more than 50%.

Figure 4
Overall distribution of orthologs among reference genes. Most of the reference genes (1,088) had only one ortholog on each of the U95A, U74A and U34A arrays used in these studies. The first bar shown here represents the number of reference genes that had three orthologs. The majority of remaining reference genes had two orthologs on one of the studied arrays. Overall, about 62% of reference genes had at least one multiple ortholog set.

ical ventilation-related articles, a finding that indirectly validates our approach (Table 2). Given the pre-eminent importance of the vascular component in ALI pathogenesis, our primary trait in selecting candidate genes was their expression in vascular endothelium. The PubMatrix output identified a number of genes linked to articles that included lung, endothelium, and even pulmonary endothelium terms in their context, which again facilitated our selection of new gene candidates for further studies.

We also investigated whether our gene list might reveal unsuspected biological processes and pathways activated or suppressed by VALI. To address this we linked the available GenMAPP [27] biological GO processes [23] to our gene list. The resulting picture of the biological processes affected by mechanical stretch in our models is shown in Table 3 with 'Immune Response,' 'Inflammatory Response,' 'Blood Coagulation,' and 'Cell Cycle Arrest' biological processes identified as the most significantly upregulated by mechanical stretch. As our gene list had only eight downregulated genes, the MAPPFinder output for downregulated pathways did not allow filtering (see Materials and methods). The complete list

of genes and GO processes identified by our procedure is provided in our supplemental data files.

Finally, we compared our list of candidate genes with the genes obtained from four individual experimental systems using the same filtering conditions (± 1.3-fold-change cutoff and $p < 0.05$). As shown in Table 4, analysis of gene expression in canine, human, mouse and rat models identified 9, 7, 13, and 15 genes out of our 69 candidates, respectively. The total of 28 genes (~40%) successfully identified by our ortholog approach did not survive selection by individual species analysis, and included well known ALI-related candidate genes such as IL1β, COX-2, PAI-1, BTG1, and FGA. The linkage of orthologous genes from different arrays increased the statistical power of our gene-expression analysis and allowed us to identify candidate genes that would otherwise remain unnoticed. A small fraction (~15%) of known ALI-related genes [7,28,29] were identified by individual species analysis but not detected by our bioinformatics approach (Table 4). This is to be anticipated, as differences exist in gene representation on multiple array platforms. For example, genes coding for the ALI candidates interleukin-8 and tumor necrosis factor-alpha were not presented on the rodent arrays, and therefore were excluded from our analysis. The tissue-specific gene expression also contributed to this false-negative gene fraction. The gene coding for surfactant C, which is mainly expressed in epithelial cells, was identified during analysis of stretched canine lung tissues but was excluded by our orthol-
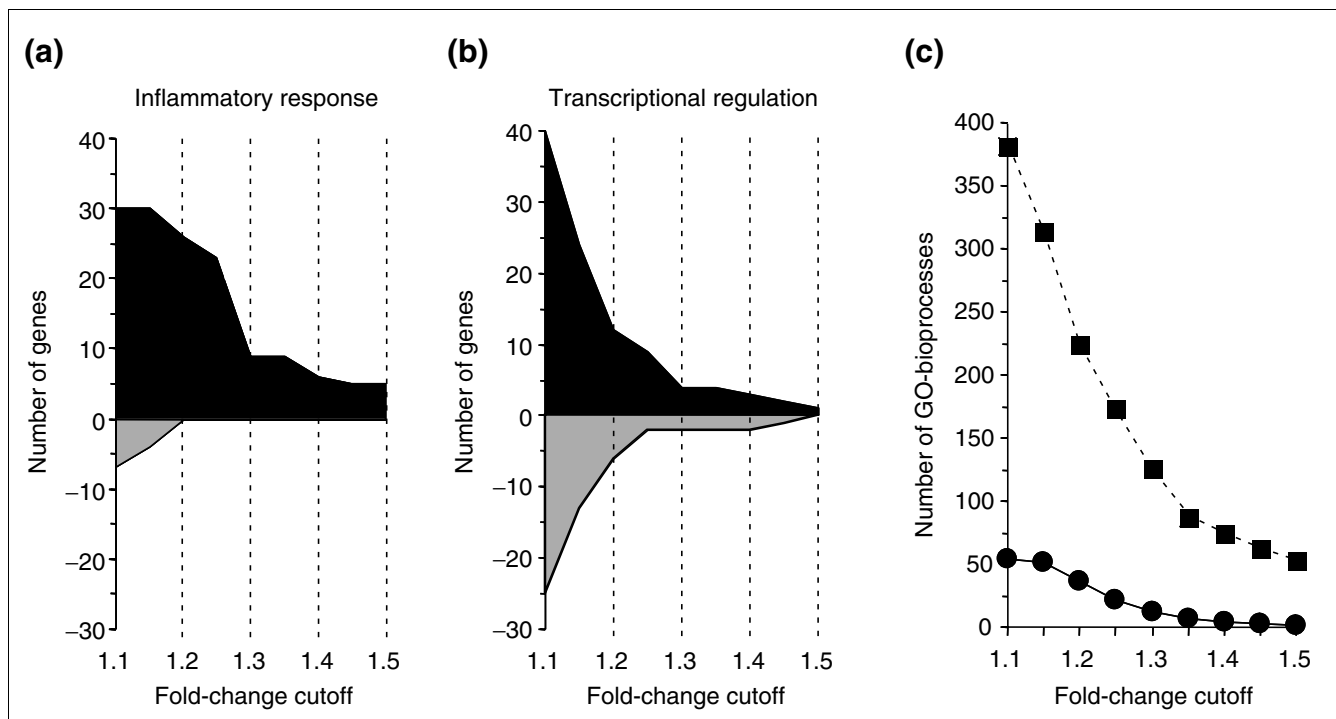
Figure 5
Distribution of co-regulated and inversely regulated biological bioprocesses identified by linkage to GO. **(a)** Genes involved in a co-regulated bioprocess (inflammatory response; GO 6954) and **(b)** an inversely regulated bioprocess (DNA-dependent regulation of transcription; GO 6355). Solid areas under the curve represent upregulated genes and gray areas under the curve represent downregulated genes. **(c)** A summary of all co-regulated (top curve) and inversely regulated (bottom curve) GO bioprocesses identified by MAPPFinder corresponding to the increment in the fold-change cutoff.

ogous method because of the virtual absence of expression in stretched endothelial cells (Table 4).

Discussion

The procedure we have described presents a complementary and potentially useful approach in searching for candidate genes involved in specific biological processes of interest. General trends in the expression of common groups of genes in response to a specific stimulus in diverse species might relate unsuspected evolutionarily conserved responses triggered by this stimulus. At the same time, known biological pathways and genes, either activated or suppressed by a selected stimulus, can be used as a validation of this approach. In this study, we investigated the response of four different biological systems (rat, mouse, dog, and human cell culture) to levels of mechanical stretch relevant to ALI. Our ortholog approach and filtering algorithm allowed us to identified 12 VALI candidate genes previously linked to ALI, five of which went undetected using a common analytical approach. We also selected six novel endothelium-related candidate genes that warrant further investigation (Table 2).

The most commonly cited upregulated ALI genes in our list were those for IL-1β and interleukin-6 (IL-6), which were

cited as lung-injury-related proteins in 287 and 173 references, respectively. Importantly, IL-1β did not survive standard selection as a candidate gene and was undetected by the same-species analytical approach. IL-6 had the highest number of links (75 citations) to mechanical ventilation (Table 2). Clinical studies showed that IL-1β and IL-6 concentrations in broncho-alveolar lavage fluid (BALF) from patients with established adult respiratory distress syndrome (ARDS) were higher than in BALF from normal volunteers [30]. Moreover, IL-1β was self-sufficient in causing ALI when overexpressed in mouse lungs [31] and was directly related to VALI in another mouse model [32]. IL-6 levels in ALI patients correlated with the mode of mechanical ventilation, as low tidal volume was associated with lower IL-6 and elevated tidal volume with high IL-6 concentrations [33].

Predictably, we identified several genes encoding enzymes that are highly conserved throughout evolution, including the ALI-related enzyme prostaglandin-endoperoxide synthase 2/ cyclooxygenase-2 (PTGS-2/COX-2). COX-2 is involved in eicosanoid synthesis and appears to be important to both edemagenesis and the pattern of pulmonary perfusion in experimental ALI. Gust *et al.* showed that the effect of endotoxin on pulmonary perfusion in ALI could be, in part, the result of activation of inducible COX-2 [34]. Upregulation of the COX-

Table 2

Genes showing significant changes in expression throughout all biological systems tested

Gene symbol	Fold change	*p* value ventilation vs control	Lung	Lung injury	Mechanical ventilation	Endothelium	Pulmonary endothelium
ALI related							
IL-1B*	1.53	0.023	1536	287	3	1759	70
IL-6	1.84	0.015	963	173	75	527	30
F3	1.52	0.017	411	54	4	713	18
PAI-1*	1.47	0.003	201	31	3	7367	22
COX2*	1.79	0.011	257	28	1	206	9
IL-13	1.30	0.013	327	21	0	62	2
AQP-1	-1.30	0.038	49	9	0	35	6
PLAUR*	1.47	0.033	83	8	0	113	5
FGA*	1.30	0.024	22	4	10	189	12
C/EBP	1.40	0.013	27	3	0	9	2
CCL2	2.00	0.024	11	1	0	9	2
ADMR	-1.35	0.022	19	1	0	16	2
VALI candidates							
CXCR4	1.62	0.007	26	0	0	79	4
GJA-1	1.33	0.026	7	0	0	30	1
IL1R2*	1.88	0.026	11	0	0	7	0
GADD45A	1.71	0.004	21	0	0	4	0
BTG-1*	1.38	0.007	2	0	0	1	0
TFF-2*	-1.32	0.045	9	0	0	1	0

ADMR, adrenomedullin receptor; AQP-1, aquaporin 1; BTG-1, B-cell translocation gene 1; CCL2, cell chemokine 2; C/EBP, CCAAT/enhancer-binding protein; COX2, prostaglandin G/H synthase and cyclooxygenase 2; CXCR4, chemokine (C-X-C motif) receptor 4; F3, coagulation factor III (thromboplastin, tissue factor); FGA, fibrinogen alpha; GADD45A, growth arrest and DNA-damage-inducible, alpha; GJA-1, gap junction protein, alpha 1 (connexin 43); IL-1B, interleukin 1 beta; IL1R2, interleukin 1 receptor, type II; IL-6, interleukin 6; IL-13, interleukin 13; PAI-1 - plasminogen activator inhibitor type 1; PLAUR, plasminogen activator, urokinase receptor; TFF-2, trefoil factor 2 (spasmolytic protein 1). *Not detected by common single-experiment analysis.

2 gene is also linked to increased pulmonary microvascular permeability in a sheep model of combined burn and smoke inhalation injury [35].

We also showed that the lung-specific surfactant protein regulation transcription factor, CCAAT enhancer-binding protein (C/EBP), was upregulated in all VALI models. C/EBP has an important role in the regulation of expression of surfactant proteins A and D, which are heavily involved in pulmonary host defense and innate immunity [36], with increased gene expression in patients with ALI [37,38]. Upregulation of C/EBP by severe lung injury [39] is highly correlated with our findings (1.4-fold increase in C/EBP expression, *p* = 0.013, Table 2). As endothelium does not generate surfactant, it will be of interest to identify the molecular targets of C/EBP in endothelium; these may include

inter-leukin-13 (IL-13) [40] and cell chemokine 2 (CCL2) [41]. These genes belong to the 'Inflammatory Response' GO biological process that was rated by MAPPFinder as highly upregulated (Table 3).

The second most highly represented ontology in the ALI-related genes bioprocess was 'Blood Coagulation' (Table 3), a finding consistent with previous reports of increased levels of coagulation factor III (thromboplastin, tissue factor, F3) and plasminogen activator inhibitor type 1 (PAI-1) in patients with ALI [42-44] or VALI [45,46]. Fibrinogen A (FGA) and plasminogen activator - the urokinase receptor (PLAUR) - are involved in IL-1β signaling and regulation, respectively. Fibrinogen indirectly activates transcription of IL-1β [47], which in turn increases expression of the urokinase receptor [48]. Interestingly, this bioprocess was identified by

Table 3

MAPPFinder results for significantly upregulated genes throughout all species tested

GO ID	GO name	Number of genes with FC >1.3	Number of measured genes	Number of genes in GO	Percent changed genes	Percent present genes	z-score
6955	Immune response	17	91	559	18.7	16.3	6.617
6954	Inflammatory response	13	55	161	23.6	34.2	6.076
8285	Negative regulation of cell proliferation	6	33	120	18.2	27.5	5.484
7050	Cell cycle arrest	4	12	53	33.3	22.6	5.001
7596	Blood coagulation	4	25	68	16.0	36.8	4.091
6960	Antimicrobial humoral response	4	25	92	16.0	27.2	4.091
6917	Induction of apoptosis	3	17	60	17.6	28.3	4.091
7267	Cell-cell signaling	12	98	284	12.2	34.5	3.303
6935	Chemotaxis	5	40	97	12.5	41.2	3.255

MAPP-Finder solely on the basis of data generated by our ortholog algorithm, as in a single-species analysis, three out of four genes related to the blood coagulation bioprocess did not survive statistical filtering (Table 4).

The interconnection of coagulation and inflammation is well recognized in that inflammation leads to increased coagulation, relevant to ALI (for a review see [8]) and a likely link is vascular endothelium. There is some evidence that the 'crosstalk' between coagulation and inflammation could be reversed. Blood coagulation *in vitro* stimulates release of inflammatory mediators from neutrophils and endothelial cells [49,50]. On the basis of these findings and data generated by our cross-species analysis of VALI, we speculate that mechanical stretch may produce either injury or activation of the pulmonary endothelium with activation of a coagulation cascade that may involve platelet aggregation. Procoagulation genes are therefore key participants in the early stages of VALI. Given that a multitude of inflammatory cytokines produce upregulation of the coagulation cascade, further studies of the time-course analysis of expression patterns of selected candidate genes in response to VALI are needed to clarify this paradigm.

In summary, our findings indicate that alterations in gene expression in response to mechanical ventilation alone can be detected by microarray techniques applied across diverse biological systems. Our data suggest that ortholog-link gene-expression analysis of multi-species VALI-simulating experimental systems is a useful tool in selecting candidate genes involved in this pathobiological process, with clear advantages over single-species analysis. We anticipate that predicted drawbacks such as incompleteness of gene representation on different array platforms and tissue-specific gene expression can be overcome by careful selection of array platforms and experimental models, respectively, as well as further improvements or refinements in the Affymetrix platform itself.

The ortholog gene-expression approach promotes application of the meta-analysis of multi-species gene-expression profiles in diverse human pathologic conditions and facilitates the selection of candidate genes of interest, with the emphasis on actively evolutionarily conserved genes.

Materials and methods
Animal models of acute lung injury (ALI)
Rats were anesthetized with 0.4 mL of etomidate (2 mg/ml) by intraperitoneal injection before cannulating the trachea for ventilation. Rats were then placed in heated water-jacketed chambers and core body temperature was adjusted to 37°C. The experimental group of rats ($n = 2$) was mechanically ventilated (12 ml/kg tidal volume, 150 breaths/min) while the control group ($n = 2$) breathed spontaneously. After 5 h ventilation the lungs were rapidly excised, snap frozen and

Table 4

Comparison of candidate gene list generated by multi-species cross-platform analysis with that obtained using a single-experiment analysis

Gene ID	U133		U95		U74		U34	
	FC	pV	FC	pV	FC	pV	FC	pV
TCF21	x	x	X		x	x		
ADPRH	x	x	X				X	
AREG	x	x			X		X	
S100A9	x	x			X		X	
EST	x	x			X		X	
EST	x	x			X			
IL-13	x	x					**X**	
GAPD	x	x					X	
LGALS3	x	x					X	
YWHAZ	x		x	x			X	
KCNJ6			x	x	X		X	
CXCR4			x	x	**X**		**X**	
CCL2			x	x	**X**		**X**	
BHLHB2			x	x	X			
GJA-1			x	x				
AQP-1			x	x				
GADD45A	**X**				x	x	x	x
MAT2A			**X**		x	x	x	
ADMR					x	x	x	x
CDKN1A			X		x	x		
PRKAR2A			X		x	x		
GABRD					x	x	X	
TNFSF6					x	x	X	
F3					x	x	**X**	
IFRD1					x	x	X	
GCH1					x	x	X	
CEBPD					x	x		
NPY1R					x	x		
EST		X	X		X		x	x
SLC2A3	X		X				x	x
DNAJA1	X				X		x	x
BAIAP3			X		X		x	x
IL-6			**X**		**X**		x	x
PSMB5	X						x	x
GBP2					X		x	x
CYCS					X		x	x
CD2					X		x	x
EIF2S1					X		x	x
CD79B					X		x	x
ENSA					X		x	x
LAT							x	x
RPLP2	X		X				X	
ARPC4	X		X				X	
NOS1		X	X				X	
EGFR			X		X		X	
P4HA1			X		X		X	

Table 4 *(Continued)*

Comparison of candidate gene list generated by multi-species cross-platform analysis with that obtained using a single-experiment analysis

PAI-1			X		X		X	
ATF3			X		X		X	
RAB5A			X		X		X	
IL1R2			X		X		X	
BTG1	X				X			
HSPA8	X				X			
CEBPB	X				X			
ARG2		X			X			
MYBPH			X		X			
COX1			X		X			
MST1			X		X			
FGA			X				X	
ACADL			X				X	
TFF2					X		X	
PLCG2					X		X	
PLAA					X		X	
PDHB					X		X	
BTG2					X		X	
IL-1B					X		X	
AIF1					X		X	
XCL1					X		X	
PLAUR					X			
GCLC					X			
Other ALI-related genes								
SP-C	x	x						
ODC-1	X				X		x	x
SP-B	X							
ACE							X	
SP-D			NA	NA				
TGFA							NA	NA
IL-8					NA	NA	NA	NA
TNFA					NA	NA	NA	NA

FC, fold change in gene expression; pV, *p*-value produced by variance-independent double-tailed *t*-test of mechanical stretch vs control. X or x denotes changes in gene expression greater than 1.3-fold or *p* < 0.05; lower-case x represent genes that satisfied both filtering conditions. NA (not available) represents genes that were not present on the array. Rows in bold depict genes presented in Table 2.

stored at -80°C until processed for RNA isolation. Mice were anesthetized by intraperitoneal injection of ketamine (150 mg/kg) and acetylpromazine (15 mg/kg). The endotracheal intubation was performed and mice (*n* = 4) were exposed to high tidal volume (15 ml/kg; breathing rate = 92/min) ventilation for 2 h using a small animal mechanical ventilator; a control group (*n* = 3) was not ventilated. The excised lungs were snap-frozen and stored at -80°C.

Dogs were anesthetized, intubated, and the lungs were lavaged and either ventilated for 5 h (*n* = 4) or collected immediately following the lavage procedure (*n* = 3) as control tissues. Lungs were snap-frozen and stored at -40°C. All experimental protocols were approved by the Johns Hopkins University Animal Care Committee.

Human HPAEC cells (Clonetics), passages 6-8, grown on flexible, bottomed collagen I-coated BioFlex plates in the

Table 5

Data sets and analytical tools sources

Affymetrix gene arrays	Rat. RG_U34A profiles	http://microarray.cnmcresearch.org/pgadatatable.asp
	Mouse. MG_U74A profiles	http://www.hopkins-genomics.org/ali/ali008/index.html
		http://www.hopkins-genomics.org/ali/ali007/index.html
	Dog. HG_U133A profiles	http://www.hopkins-genomics.org/research/researchALI.html
	HPAEC. HG-U95Av2 profiles	http://www.hopkins-genomics.org/ali/ali015/index.html
Software and tools	RESOURCERER	http://pga.tigr.org/tigr-scripts/magic/r1.pl
	NetAffx	http://www.affymetrix.com/analysis/index.affx
	GenMAPP and MAPPFinder	http://www.GenMAPP.org
	PubMatrix	http://www.hopkins-genomics.org/resources/resourcesSoftware.html
	Python 2.2	http://www.python.org
	Stats Module for Python	http://www.nmr.mgh.harvard.edu/Neural_Systems_Group

presence of complete culture medium (20% FCS) were exposed to cyclic stretch (25 cycles/min, 18% elongation) for 48 h ($n = 4$) as we have described [10] using FlexerCell Tension Plus T-4000 cell culture stretch system or remained static ($n = 3$).

The time-course of the experiments was selected according to the manifestation of the defining feature of ALI - vascular leakage.

RNA isolation and hybridization
Smaller frozen tissues (~50 mg) were directly solubilized in chaotropic solubilization buffer using a Brinkman Polytron tissue disruptor. Larger tissue fragments (>100 mg) were pulverized into frozen powder with a mortar and pestle, pre-chilled to liquid nitrogen temperature, and the frozen powder solubilized with the Polytron. RNA was purified using Trizol LS (Life Technologies) and an additional RNA purification step was conducted using the RNAeasy purification kit (Qiagen). Approximately 10 μg of purified, total RNA was used for analyses. HPAEC total RNA was purified using Trizol LS and an additional RNA clean-up step was conducted using the RNAeasy purification kit. Purified total RNA was reverse transcribed to first-strand cDNA using a hybrid primer consisting of oligo(dT) and T7 RNA polymerase promoter sequences. The single-stranded cDNA was then converted to double-stranded cDNA. Complementary DNA corresponding to 5-10 μg total RNA was used in a cRNA amplification step using T7 RNA polymerase and two biotinylated nucleotide precursors. The resulting biotinylated cRNA was fragmented to a size of approximately 50 bp. Approximately 20-30 μg of the biotinylated canine, mouse, rat and HPEAC cRNA was hybridized to U133A, U74A, U34A, and U95A GeneChips (Affymetrix), respectively. The bound cRNA was visualized by binding of streptavidin/phycoerythrin conjugates to the hybridized GeneChip, followed by laser scanning of bound phycoerythrin. These scan results are available on the Hop-Gene website (Table 5).

Building the ortholog-linked database
Probe IDs of U74A, U34A, and U95A GeneChips were linked to their orthologous counterparts on the U133A using RESOURCERER (Table 5). This linkage identified 3,077 genes common to all array ortholog nodes (Figures 1, 2a), which were built around 2,887 reference genes from the U133A chip. The actual number of reference genes was lower, owing to the fact that in some cases multiple orthologs for the same reference gene are represented on the arrays (Figure 4). Any identified missing links between members of a node (Figure 2b) were filled, and the newly linked node members were coined as putative orthologs on the basis of their homology to the common reference gene. In total, the final ortholog-linked database contained 2,887 reference probe sets from U133A, and 2,631, 2,365 and 2,848 ortholog probe sets from U95A, U74A and U34A, respectively. The unequal numbers are due to the sharing of the same ortholog by different reference genes (Table 1).

Expression-data analysis
The signal intensity fluorescent images produced during Affymetrix GeneChip hybridizations were read using the Agilent Gene Array Scanner and converted into GeneChip Cell files (CEL) using MAS 5.0 software (Affymetrix). The analysis of the probe level data (available in the .CEL files) was performed using the Bioconductor affy package [24,51]. In particular, the package was used to extract the probe-level data and convert into expression measures of individual probe pairs. Various strategies have been used which in general involve three steps: background correction, across-array normalization, and summarization. For the analysis presented

here, we utilized the mas5 module of the affy package [52]. This probe-level analysis was conducted in all species tested and poorly performing probe pairs thus revealed were masked before converting CEL files into GeneChip Sequence files (CHP) using MAS 5.0. The signal intensity values obtained for U95Av2, U74A and U34A were used directly for analysis, and those of the hetero-hybridized U133A arrays were adjusted on the basis of differences in canine and human gene sequences (the detailed procedure and validation of probe-level analysis and probe-set modification will be described elsewhere). This approach increased the present call of a transcript ($p < 0.04$) by 25% on average, compared to unadjusted probe-set processing (Figure 4). The remaining absent calls ($p > 0.06$) for transcript abundance were assumed to be a result of undetectable message concentration (<1 pM [53]) rather than technical or detection errors. Therefore, absent calls for each GeneChip were averaged and all absent calls for a given chip were replaced with this average value. This modified dataset was used for further analysis.

Selecting significant gene-expression changes using a single experiment or orthologous approaches
The data from each orthologous group (reference gene and its orthologs) was pooled in four species-specific groups. For statistical analysis, each species-specific group contributed three control and four mechanical stretch-challenged samples, except for rat, which had two control and two challenged lung samples. Therefore, the dataset for each ortholog group was comprised of 11 control and 14 stress-challenged samples. However, in more than 60% of analyzed array probes, multiple paralogs existed for each species-specific group (Figure 3). To maintain three-control/four-condition input from these groups, we averaged expression data of multiple paralogues on the chipwise basis (the only data for paralogs from the same chip was averaged). At the same time, the data were scaled using raw-wise average normalization. A two-tailed unequal variance independent *t*-test was performed on ortholog-generated expression datasets (11 controls vs 14 mechanical stretches) or individual experiments (three controls vs four mechanical stretch samples for canine, HPAEC and mouse models, and two control vs two mechanical stretch samples for the rat model) and changes in gene expression with $p < 0.05$ was used as a cutoff to produce preliminary lists of candidate genes. The fold-change ratio was computed from the mean values of control and mechanical stretch sets produced by *t*-test.

Gene ontology (GO) analysis
The MAPPFinder software is not yet compatible with the U133A probe sets. To overcome this incompatibility we substituted the U133A probe IDs with corresponding MAPP-Finder-compatible U95A probe IDs. To link our gene list to corresponding GenBank accession numbers and consequently to GO terms, we utilized GenMAPP software and linked 2,278 genes out of 2,887 reference genes to GO and repeatedly (nine cycles) analyzed these results by MAPP-

Finder using different settings of fold-change limit from ±1.1 to ±1.5 with increment of ±0.05 (Figure 5). GO biological process assignments were selected and filtered by Z-score (>0) and the number of hits in the first GO node (>0). GO terms simultaneously identified as both down- and upregulated bioprocesses were selected using Microsoft Access 2000 and considered to be inversely regulated biological processes. If the results represented unique down- or upregulated GO bioprocesses, these processes were considered to be co-regulated (Figure 5). The point at which the number of shared GO terms became constant was selected as a threshold fold-change cutoff (±1.3). The analysis depicted in Table 3 was conducted using a candidate set of 69 genes, using GO and local MAPPs with fold-change set at 1.3 or higher and $p < 0.05$.

Journal articles that referenced lung, lung injury, mechanical ventilation, endothelium, or pulmonary endothelium and our candidate genes at the same time were obtained using the PubMatrix tool [24]. URLs for statistical tools and analytical software employed in our analysis are listed in Table 5.

Additional data files
The following additional data files are available with the online version of this article: a final list of reference genes (Additional data file 1), a list of gene candidates (Additional data file 2), the full GO results (Additional data file 3), and data for Table 4 (Additional data file 4).

References
1.   The Acute Respiratory Distress Syndrome Network: **Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome.** *N Engl J Med* 2000, **342:**1301-1308.
2.   Parker JC: **Inhibitors of myosin light chain kinase and phosphodiesterase reduce ventilator-induced lung injury.** *J Appl Physiol* 2000, **89:**2241-2248.
3.   Parker JC, Ivey CL, Tucker A: **Phosphotyrosine phosphatase and tyrosine kinase inhibition modulate airway pressure-induced lung injury.** *J Appl Physiol* 1998, **85:**1753-1761.
4.   Webb HH, Tierney DF: **Experimental pulmonary edema due to intermittent positive pressure ventilation with high inflation pressures. Protection by positive end-expiratory pressure.** *Am Rev Respir Dis* 1974, **110:**556-565.
5.   Parker JC, Ivey CL, Tucker JA: **Gadolinium prevents high airway pressure-induced permeability increases in isolated rat lungs.** *J Appl Physiol* 1998, **84:**1113-1118.
6.   Khimenko PL, Moore TM, Wilson PS, Taylor AE: **Role of calmodulin and myosin light-chain kinase in lung ischemia-reperfusion injury.** *Am J Physiol* 1996, **271:**L121-L125.
7.   Villar J, Flores C, Mendez-Alvarez S: **Genetic susceptibility to acute lung injury.** *Crit Care Med* 2003, **31(4 Suppl):**S272-S275.
8.   Russell JA: **Genetics of coagulation factors in acute lung injury.** *Crit Care Med* 2003, **31(4 Suppl):**S243-S247.

9.   Garcia J, Nonas S, Kawkitinarong K, Barnes K: **SNPs, chips and functional genomics in acute lung injury.** *Am J Respir Cell Mol Biol* in press.

10. Birukov KG, Jacobson JR, Flores AA, Ye SQ, Birukova AA, Verin AD, Garcia JG: **Magnitude-dependent regulation of pulmonary endothelial cell barrier function by cyclic stretch.** *Am J Physiol Lung Cell Mol Physiol* 2003, **285:**L785-L797.

11. Ramaswamy S, Ross KN, Lander ES, Golub TR: **A molecular signature of metastasis in primary solid tumors.** *Nat Genet* 2003, **33:**49-54.

12. Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM: **Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer.** *Cancer Res* 2002, **62:**4427-4433.

13. Ghosh D, Barette TR, Rhodes D, Chinnaiyan AM: **Statistical issues and methods for meta-analysis of microarray data: a case study in prostate cancer.** *Funct Integr Genomics* 2003, **3:**180-188.

14. **Eukaryotic gene orthologs**   [http://www.tigr.org/tdb/tgi/ego]

15. Dubchak I, Brudno M, Loots GG, Pachter L, Mayor C, Rubin EM, Frazer KA: **Active conservation of noncoding sequences revealed by three-way species comparisons.** *Genome Res* 2000, **10:**1304-1306.

16. Frazer KA, Sheehan JB, Stokowski RP, Chen X, Hosseini R, Cheng JF, Fodor SP, Cox DR, Patil N: **Evolutionarily conserved sequences on human chromosome 21.** *Genome Res* 2001, **11:**1651-1659.

17. Tsai J, Sultana R, Lee Y, Pertea G, Karamycheva S, Antonescu V, Cho J, Parvizi B, Cheung F, Quackenbush J: **RESOURCERER: a database for annotating and linking microarray resources within and across species.** *Genome Biol* 2001, **2:**software0002.1-0002.4.

18. Svensson BA, Kreeft AJ, van Ommen GJ, den Dunnen JT, Boer JM: **GeneHopper: a web-based search engine to link gene-expression platforms through GenBank accession numbers.** *Genome Biol* 2003, **4:**R35.

19. **GeneHopper**   [http://lgtc.nl/GeneHopper]

20. **RESOURCERER**   [http://pga.tigr.org/tigr-scripts/magic/r1.pl]

21. Lee Y, Sultana R, Pertea G, Cho J, Karamycheva S, Tsai J, Parvizi B, Cheung F, Antonescu V, White J *et al.*: **Cross-referencing eukaryotic genomes: TIGR Orthologous Gene Alignments (TOGA).** *Genome Res* 2002, **12:**493-502.

22. Doniger SW, Salomonis N, Dahlquist KD, Vranizan K, Lawlor SC, Conklin BR: **MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data.** *Genome Biol* 2003, **4:**R7.

23. **GenMAPP homepage**   [http://www.genmapp.org]

24. **HOPGENE   resources**   [http://www.hopkins-genomics.org/resources/resourcesSoftware.html]

25. Lagreid A, Hvidsten TR, Midelfart H, Komorowski J, Sandvik AK: **Predicting gene ontology biological process from temporal gene expression patterns.** *Genome Res* 2003, **13:**965-979.

26. **Gene Ontology Consortium**   [http://www.geneontology.org]

27. Dahlquist KD, Salomonis N, Vranizan K, Lawlor SC, Conklin BR: **GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways.** *Nat Genet* 2002, **31:**19-20.

28. Kurdowska AK, Geiser TK, Alden SM, Dziadek BR, Noble JM, Nuckton TJ, Matthay MA: **Activity of pulmonary edema fluid interleukin-8 bound to alpha(2)-macroglobulin in patients with acute lung injury.** *Am J Physiol Lung Cell Mol Physiol* 2002, **282:**L1092-L1098.

29. Elsayed NM, Ellingson AS, Tierney DF, Mustafa MG: **Effects of ozone inhalation on polyamine metabolism and tritiated thymidine incorporation into DNA of rat lungs.** *Toxicol Appl Pharmacol* 1990, **102:**1-8.

30. Park WY, Goodman RB, Steinberg KP, Ruzinski JT, Radella F 2nd, Park DR, Pugin J, Skerrett SJ, Hudson LD, Martin TR: **Cytokine balance in the lungs of patients with acute respiratory distress syndrome.** *Am J Respir Crit Care Med* 2001, **164:**1896-1903.

31. Kolb M, Margetts PJ, Anthony DC, Pitossi F, Gauldie J: **Transient expression of IL-1beta induces acute lung injury and chronic repair leading to pulmonary fibrosis.** *J Clin Invest* 2001, **107:**1529-1536.

32. Chess PR, Johnston CJ, Wright TW, Jordan HD: **Murine model of mechanical ventilation-induced lung injury.** *Am J Respir Crit Care Med* 2003, **167:**A625.

33. Takala A, Jousela I, Takkunen O, Kautiainen H, Jansson SE, Orpana A, Karonen SL, Repo H: **A prospective study of inflammation markers in patients at risk of indirect acute lung injury.** *Shock* 2002, **17:**252-257.

34. Gust R, Kozlowski JK, Stephenson AH, Schuster DP: **Role of**

cyclooxygenase-2 in oleic acid-induced acute lung injury. *Am J Respir Crit Care Med* 1999, **160:**1165-1170.

35. Enkhbaatar P, Murakami K, Shimoda K, Chandra A, Traber L, Hawkins H, Herndon D, Traber D: **Ketorolac, a nonsteroidal anti-inflammatory agent prevents burn/smoke-induced acute lung injury in sheep by inhibiting nitric oxide.** *Am J Respir Crit Care Med* 2003, **167:**A665.

36. Cheng IW, Ware LB, Greene KE, Nuckton TJ, Eisner MD, Matthay MA: **Prognostic value of surfactant proteins A and D in patients with acute lung injury.** *Crit Care Med* 2003, **31:**20-27.

37. He Y, Crouch E: **Surfactant protein D gene regulation. Interactions among the conserved CCAAT/enhancer-binding protein elements.** *J Biol Chem* 2002, **277:**19530-19537.

38. Rosenberg E, Li F, Reisher SR, Wang M, Gonzales LW, Ewing JR, Malek S, Ballard PL, Notarfrancesco K, Shuman H, Feinstein SI: **Members of the C/EBP transcription factor family stimulate expression of the human and rat surfactant protein A (SP-A) genes.** *Biochim Biophys Acta* 2002, **1575:**82-90.

39. Sugahara K, Sadohara T, Sugita M, Iyama K, Takiguchi M: **Differential expression of CCAAT enhancer binding protein family in rat alveolar epithelial cell proliferation and in acute lung injury.** *Cell Tissue Res* 1999, **297:**261-270.

40. Corne J, Chupp G, Lee CG, Homer RJ, Zhu Z, Chen Q, Ma B, Du Y, Roux F, McArdle J *et al.*: **IL-13 stimulates vascular endothelial cell growth factor and protects against hyperoxic acute lung injury.** *J Clin Invest* 2000, **106:**783-791.

41. Rose CE Jr, Sung SS, Fu SM: **Significant involvement of CCL2 (MCP-1) in inflammatory disorders of the lung.** *Microcirculation* 2003, **10:**273-288.

42. Groeneveld AB, Kindt I, Raijmakers PG, Hack CE, Thijs LG: **Systemic coagulation and fibrinolysis in patients with or at risk for the adult respiratory distress syndrome.** *Thromb Haemost* 1997, **78:**1444-1449.

43. Prabhakaran P, Ware LB, White KE, Cross MT, Matthay MA, Olman MA: **Elevated levels of plasminogen activator inhibitor-1 in pulmonary edema fluid are associated with mortality in acute lung injury.** *Am J Physiol Lung Cell Mol Physiol* 2003, **285:**L20-L28.

44. Ruf W, Riewald M: **Tissue factor-dependent coagulation protease signaling in acute lung injury.** *Crit Care Med* 2003, **31(4 Suppl):**S231-S237.

45. Schultz M, Millo J, Levi M, Garrard C, van der Poll T: **Local activation of coagulation and inhibition of fibrinolysis in the lung during ventilator-associated pneumonia.** *Am J Respir Crit Care Med* 2003, **167:**A861.

46. Schmidt B, Davis P, La Pointe H, Monkman S, Coates G, deSa D: **Thrombin inhibitors reduce intrapulmonary accumulation of fibrinogen and procoagulant activity of bronchoalveolar lavage fluid during acute lung injury induced by pulmonary overdistention in newborn piglets.** *Pediatr Res* 1996, **39:**798-804.

47. Perez RL, Ritzenthaler JD, Roman J: **Transcriptional regulation of the interleukin-1beta promoter via fibrinogen engagement of the CD18 integrin receptor.** *Am J Respir Cell Mol Biol* 1999, **20:**1059-1066.

48. Hasegawa T, Sorensen L, Dohi M, Rao NV, Hoidal JR, Marshall BC: **Induction of urokinase-type plasminogen activator receptor by IL-1 beta.** *Am J Respir Cell Mol Biol* 1997, **16:**683-692.

49. Johnson K, Aarden L, Choi Y, De Groot E, Creasey A: **The proinflammatory cytokine response to coagulation and endotoxin in whole blood.** *Blood* 1996, **87:**5051-5060.

50. Johnson K, Choi Y, DeGroot E, Samuels I, Creasey A, Aarden L: **Potential mechanisms for a proinflammatory vascular cytokine response to coagulation activation.** *J Immunol* 1998, **160:**5130-5135.

51. Irizarry R, Gautier L, Cope L: **An R package for analyses of Affymetrix oligonucleotide arrays.** In *The Analysis of Gene Expression Data: Methods and Software* Edited by: Parmigiani G, Garrett ES, Irizarry RA, Zeger SL. New York: Springer; 2003.

52. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP: **Summaries of Affymetrix GeneChip probe level data.** *Nucleic Acids Res* 2003, **31:**e15.

53. Chudin E, Walker R, Kosaka A, Wu SX, Rabert D, Chang TK, Kreder DE: **Assessment of the relationship between signal intensities and transcript concentration for Affymetrix GeneChip arrays.** *Genome Biol* 2002, **3:**research0005.1-0005.10.

54. **HOPGENE homepage**   [http://www.hopkins-genomics.org]