Genomic insights into acute inflammatory lung injury

Joe G. N. Garcia and Liliana Moreno Vinasco
Pulmonary and Critical Care Medicine, Department of Medicine, Pritzker School of Medicine, University of Chicago, Chicago, Illinois

Garcia, Joe G. N., and Liliana Moreno Vinasco. Genomic insights into acute inflammatory lung injury. Am J Physiol Lung Cell Mol Physiol 291: L1113–L1117, 2006. First published July 28, 2006; doi:10.1152/ajplung.00266.2006.—Acute lung injury (ALI) is a devastating syndrome (usually associated with sepsis) that represents a major healthcare burden in the United States. We have focused our studies on unraveling the genetic underpinnings of this syndrome utilizing a candidate gene approach to identify novel genes for ALI susceptibility. Two novel genes identified by this approach include pre-B cell colony-enhancing factor (PBEF) and the gene for myosin light chain kinase (MLCK). PBEF protein levels were elevated in human bronchoalveolar lavage and serum samples from patients with ALI, and DNA sequencing identified two single nucleotide polymorphisms in the PBEF promoter (T-1001G, C-1543T) that were overrepresented in patients with sepsis-induced ALI. More recently, we found MLCK single nucleotide polymorphisms and haplotypes to be associated with human ALI with unique variants observed in African-Americans with ALI. Thus genomic and genetic approaches represent powerful strategies in the identification of novel candidate genes and potential targets for ALI therapies.

ACUTE LUNG INJURY (ALI) and acute respiratory distress syndrome (ARDS) are inflammatory lung syndromes characterized by diffuse alveolar infiltration, hypoxemia, respiratory failure, and deaths due to multiorgan failure. Mortality rates in ARDS, the most severe ALI clinical scenario, range from 34 to 58% (13) with ~150,000–200,000 ALI cases per year in the United States (27) and an incidence of 17–34 cases/100,000 people per year in Europe, Australia, and other developed countries. Thus ALI and ARDS constitute a major healthcare burden due to the intensive and often prolonged intensive care unit (ICU) hospitalizations. In addition to these epidemiologic studies, race and gender differences in ARDS deaths in the United States over the past several decades have clearly demonstrated an increase in incidence and mortality due to sepsis and ALI in African-Americans when compared with Caucasians (16).

ALI is usually caused by sepsis, acid inhalation, or trauma with mechanical ventilation, an intervention strategy commonly used in the ICU to treat ALI, potentially exacerbating ALI pathophysiology and reducing survival if excessive (22, 24, 10). The hallmarks for ALI, i.e., cellular and spatial heterogeneity, profound high permeability, leukocyte influx, and lung edema, are often augmented by mechanical ventilation in animal models of ALI (30) and are contributing factors for death due to multiorgan failure (23).

Genetics/genomics. Over the past several decades, a lingering issue of concern for critical care physicians has been the enormous heterogeneity in the outcomes observed in the care of the ALI/ARDS patient. Why do some conventionally ventilated ARDS patients recover quickly with rapid extubation, whereas other ARDS patients progress to multiorgan failure and death? What underlies the now recognized healthcare disparity observed in ALI and sepsis outcomes in African-Americans? Although the answers to these queries are incomplete, recent genetic associations strongly suggest that genetic variations, or gene polymorphisms, contribute to ALI susceptibility and severity in a racial- and ethnic-specific manner (20, 25, 28).

Search for susceptibility genes in ALI Ventilation-associated lung injury. There have been many challenges in the search for susceptibility genes on ALI and ventilation-associated lung injury (VALI), such as the heterogeneity in the phenotype (susceptibility and outcome), complex gene-environment interactions, possible incomplete penetrance, and locus heterogeneity. In addition, and most importantly, ALI does not affect families, and pedigrees are not available. Linkage mapping, a strategy involving the scanning of entire family genomes (11) and a method proven useful in asthma and chronic obstructive pulmonary disease (COPD) genetic epidemiology studies, has the advantage of not requiring prior knowledge of the biology underlying an illness; this is a feature especially important in complex disorders such as sepsis and ALI. However, as noted above, this approach has the disadvantage of requiring large families with both affected and unaffected individuals. Thus the use of linkage studies for identifying those genetic variants (alleles) that are shared by affected family members more frequently than would be expected based on chance are significantly limited in human ALI populations.

Approach for identifying disease-modifying genes. For the reasons cited above, we chose to utilize a candidate gene approach, a strategy that can be performed using unrelated cases and controls, to identify ALI disease-modifying genes (Fig. 1). Complementing the linkage studies described above, the candidate gene approach generates a disease-specific candidate gene list via extensive microarray gene profiling as well as by the analysis of relevant pathophysiological pathways. Microarray gene profiling does not require prior knowledge about disease pathogenesis, whereas, in contrast, the analysis of known pathways obviously requires an element of prior knowledge of disease pathogenesis for candidate gene identification. Candidate gene variants or polymorphisms are next identified via sequencing of the gene or by interrogation of
Fig. 1. Potential strategies to build a candidate gene list in complex disorders. Linkage analysis is a powerful approach to uncover the genes involved in complex diseases. However, it requires large pedigrees with affected and unaffected members, which are not available for acute lung injury (ALI) due to the heterogeneity of the syndrome. A promising alternative is to use a candidate gene-based case-control association study, by sampling unrelated individuals from the population of patients and controls, and assess differences in frequency of variants in the genes of interest. This strategy has become the most extended tool for dissecting the genetic factors influencing ALI susceptibility and outcome. A crucial step in such studies is to know which genes are the relevant ones for the development of the disease. In this respect, extensive expression analysis in animal models as well as in patient samples has been successfully applied to identify which genes are relevant in ALI. PEBF, pre-B cell colony-enhancing factor; MYLK, myosin light chain kinase; PROCR, protein C receptor, endothelial.

publicly available databases of single nucleotide polymorphisms (SNPs). The importance of these gene variants in the disease susceptibility or severity is then validated via utilization of genetically engineered mice or by consomic rodent approaches. Finally, the association between gene variants in a certain gene or allele and a specific disease can be studied by defining the frequency of the target variant allele in a population of affected patients and comparing this with the frequency in controls. Studying an association between one or more SNPs and a disease can help researchers to focus on certain areas of DNA and potentially identify additional candidate genes.

Sample selection and microarray analysis. Our studies employed human samples recruited from patients with ALI and human healthy controls (bronchoalveolar lavage and blood samples) as well as samples (bronchoalveolar lavage, lung tissue, or blood) obtained from mice, rats, or canine in established models of ALI (endotoxin exposure, Refs. 17, 15, 19; mechanical ventilation/VALI, Refs. 15, 18). Human and animal protocols were approved by the Institutional Review Boards. Lung samples were processed for microarray analysis and bronchoalveolar lavage/serum for protein analysis as we have previously described (6) with the signal intensity fluorescent images converted to Gene Chip Cell files (CEL) using the Affymetrix software (MAS 5.0). The analysis of the probe level data described by our group (12) used the Bioconductor Affy Package for background correction, across any array global normalization, and extraction of the probe level data. Gene ontologies are generated by the Gene Ontology Consortium (http://www.geneontology.org) using GenMAPP and MAPPFinder (1, 2). Genes with change in expression of 20% or higher and a false discovery rate of <0.05 were selected from microarray data (at least 4 replicates), a filtering approach previously successfully applied for selection of candidate genes (9). The slight increase in false discovery rate was applied equally throughout all gene ontologies and did not affect individual ontology selection.

Validation: RT-PCR, Western blot, and real-time PCR. Semiquantitative RT-PCR, Western blot, and real-time PCR methods were used for validation of the gene expression from animal tissues/samples or in human bronchoalveolar lavage or blood samples (12) with 18S ribosomal RNA (rRNA) used as internal standard. Specific primers were synthesized for each gene of interest based on GenBank sequences database. Optimal conditions for RT-PCR reaction were standardized, and the data were analyzed by densitometry, normalizing to rRNA. For real-time RT-PCR, the different samples from each experimental group were run with the ABI Prism 7700 Sequence Detector following instructions from the manufacturer (Perkin Elmer-Applied Biosystems). Changes in gene expression level were compared with controls, and specific mRNA transcript levels were expressed as fold difference (12).

Current ALI/VALI biomarkers. VALI samples from mice placed on mechanical ventilation for 6 h revealed several genes that were differentially regulated, with a total of 41 genes that were significantly upregulated and 7 genes that were significantly downregulated (12). Filtering of these genes using MAPPFinder (1) revealed eight biological processes, including four pathways (inflammation, blood coagulation, immune response, and apoptosis). To further characterize VALI-related gene candidates, a clustering analysis was performed (6 gene clusters) with cluster 1 (consisting of 48 genes) showing 8/14 genes previously identified and associated with VILI including: IL-6, serine proteinase inhibitor [plasminogen activator inhibitor type 1 (PAI-1)], amphiregulin, chemokine C-C ligand 2, coagulation factor III, growth arrest and DNA damage-inducible 45a, urokinase plasminogen activator receptor, and prostaglandin-endoperoxide synthase 2. The Mig-6 gene (along with heat shock protein-70 and IL-1β) has also been highly expressed early in LPS-induced lung injury (12), potentially mediating coagulation processes and stress responses (26), indicating that common pathways affected are conserved across species. Analysis of the highly differentiated genes between the apex and the base of the lungs revealed several genes were associated with ALI (VEGF, Toll-like receptor,
PAI-1, transforming growth factor-β, etc). Interestingly, the gene with the highest level of increased gene expression was pre-B cell colony-enhancing factor (PBEF), a relatively new gene encoding a proinflammatory cytokine originally implicated in the maturation of B cell precursors, but never described in the lungs. Thus, our studies revealed novel expression of PBEF in lung tissues, a process validated by RT-PCR (4, 29).

**PBEF.** We first demonstrated that the gene is highly expressed in samples from humans and animals with ALI (29) and cloned canine PBEF gene to generate antibody that proved useful across all species (14). Gene expression profiles and validation were done by Affymetrix GeneChip microarray system and with RT-PCR, respectively, by using lung tissues from animals and human bronchoalveolar lavage from patients with ALI and healthy controls. The spatial localization of PBEF expression was investigated by immunofluorescence, and genotyping of the PBEF gene promoter SNPs was identified by isolating leukocyte DNA from humans with sepsis-associated ALI, sepsis alone, and healthy controls. A small cohort of individuals per group underwent direct DNA sequencing (36 individuals, 12/group) and revealed that SNPs T-1001G and C-1543T were overrepresented in both Caucasian and African-American subjects (restriction site polymorphism assay and by service-SNP genotyping method; Fig. 2).

Thus PBEF represents an excellent success story for the application of genomic strategies such as the candidate gene approach, in human disease. The gene was first identified through extensive gene expression studies (when there were no reports relating the gene to lung dysfunction) and validated through a number of complementary strategies (RT-PCR, real-time PCR, immunofluorescence, and by demonstration of immunoreactive PBEF in bronchoalveolar lavage and blood from animal models and humans with ALI). Finally, genetic analysis showed novel association with ALI vs. SNPs in the PBEF promoter. Recently, PBEF has been renamed visfatin, a new adipokine described to regulate metabolic syndrome and diabetes (7). PBEF is also a cytosolic enzyme involved in NAD biosynthesis, with its expression in neutrophils, it is crucial in

---

**Fig. 2.** Validation of PBEF as a candidate gene. A: regional differences (apex and base) in gene expression from canine lung tissue with ALI showed significant gene expression changes in apex compared with base. Ontologies were generated using MAPPFinder software. Folded changes showed PBEF as the one gene with the highest expression. B: protein and RNA expression levels in lungs and bronchoalveolar lavage (BAL) across species (canine, murine, and human) with ALI. mRNA levels of PBEF expression are shown in canine, murine lungs, and human BAL from patients with ALI. Analysis was done by RT-PCR, and PBEF protein levels were obtained by Western blots from blood and BAL samples from animal models of ALI and human BAL samples. PBEF expression levels were significantly elevated compared with uninjured controls. C: common variants were found in human (African-Americans and Caucasians) patients with ALI. The minor allele frequencies of the two PBEF promoter single nucleotide polymorphisms (SNPs) T-1001G and 1543C transversions were assessed in cytokine-stimulated endothelium, where measurements of T-1543C or C-1543C-pGL-3 vectors were done in transfected human microvascular endothelial cells. The CC genotype in the T-1543C promoter of PBEF reduced luciferase activity (right), whereas the T-1001G promoter did not show any change. Thus C-1543T SNP altered PBEF, indicating potential protection for ALI in Caucasians (29).
delaying neutrophil apoptosis in patients with sepsis (8). Additional studies are ongoing to more clearly understand the role of PBEF in ALI susceptibility and severity.

Myosin light chain kinase polymorphism in ALI. Myosin light chain kinase (MLCK) was a gene first cloned by our laboratory (5) and determined to be involved in cell motility, vascular regulation of inflammation, permeability, and apoptosis. Murine studies show deletion of MLCK in mice subjected to the endotoxin-induced ALI (nmMLCK−/−), which were protected from LPS-mediated lung vascular permeability, whereas MLCK-overexpressing transgenic mice demonstrate increased vascular permeability, indicating that MLCK expression is a major determinant of vascular leak in the acutely inflamed murine lung. Given that endothelial/epithelial barrier dysfunction and vascular leak are hallmarks of ALI, we speculated that on the basis of our understanding of pathways involved in permeability regulation, the gene MYLK, encoding for human MLCK, located on chromosome 3q21, may represent a potential ALI candidate gene. Direct sequencing of the MLCK gene (217.6 kb, 32 exons) in 36 subjects (European-American and African-Americans with sepsis, sepsis-associated ALI, or 12 healthy controls) focused on the 32 exon-intron boundaries, and the 2 kb of the 5′-untranslated region identified 57 genetic variations and 51 polymorphic base substitutions where 10 were located in the exons. Five of these ten MYLK SNPs conferred an amino acid change including Pro21His, Pro147Ser, Val261Ala, Ser131Pro, and Arg1450Gln, and four novel polymorphisms (MYLK_002, 007, 033, and 044) (3). Genotyping studies showed several MYLK SNPs to be overrepresented in Caucasians (single locus analysis, \( P = 0.01 \) to 0.002) with several SNPs overrepresented in African-Americans. Novel, ethnic-specific haplotypes were identified conferring susceptibility to sepsis vs. ALI (odds ratio, OR-2.2-4.9). In addition, specific ethnic-specific haplotypes were identified conferring susceptibility to ALI only (OR 2.4-5.1; Fig. 3).

The capacity for high-throughput technologies (like sequencing) was advantaged by the Human Genome Project, which heralded additional revolutionary technological breakthroughs (i.e., rapid, high-throughput sequencing, gene expression profiling, genotyping). Despite the multiple challenges for studying the genetics of ALI or VALI, gene expression and gene ontology analyses coupled to novel bioinformatic approaches have shed new light on ALI, with an impact in unraveling the genetic basis of ALI and in generating new therapeutic targets. Our genomic studies identified two novel genes, PBEF and MYLK, via the candidate gene approach. PBEF expression was significantly unregulated by both mechanical force and inflammatory stimuli, known key risk factors implicated in the increased lung vascular permeability characteristic of ALI and a cardinal feature of inflammation. MYLK was found to be a potential candidate for sepsis and ALI associated to sepsis, indicating its role as a potential therapeutic modulator of the inflammatory response present not only in ALI but in other

Fig. 3. Haplotype analysis of MYLK variants in African-American and Caucasian patients with ALI. Linkage disequilibrium (LOD) between 17 common SNPs in European-Americans and African-Americans estimated from 170 and 120 chromosomes, respectively. The strength of linkage disequilibrium between respective SNP pairs is noted by progression of the color. For all with LOD >2, the color moves from red to blue, and for LOD <2, it is represented by white boxes (3).
diseases such as asthma and COPD. Further analysis of select candidate genes by additional SNP discovery and mid- or high-throughput genotyping will undoubtedly provide important insights into the genetic basis for ALI susceptibility and severity. The era of molecular medicine, with the availability of high-throughput technologies and molecular phenotyping signatures, represents the capacity to bring clinicians, clinician-scientists, and basic biomedical biologists together to deliver customized care of the ICU patient and improve the survival of patients with critical illness.

ACKNOWLEDGMENTS

We thank Carlos Flores for critical review and Nicholas Shank and Sangeetha Devender for art graphic assistance.

REFERENCES