Recent advances in genetic predisposition to clinical acute lung injury

Li Gao and Kathleen C. Barnes

Divisions of 1Allergy and Clinical Immunology and 2Pulmonary and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland

Gao L, Barnes KC. Recent advances in genetic predisposition to clinical acute lung injury. Am J Physiol Lung Cell Mol Physiol 296: L713–L725, 2009.—It has been well established that acute lung injury (ALI), and the more severe presentation of acute respiratory distress syndrome (ARDS), constitute complex traits characterized by a multigenic and multifactorial etiology. Identification and validation of genetic variants contributing to disease susceptibility and severity has been hampered by the profound heterogeneity of the clinical phenotype and the role of environmental factors, which includes treatment, on outcome. The critical nature of ALI and ARDS, compounded by the impact of phenotypic heterogeneity, has rendered the amassing of sufficiently powered studies especially challenging. Nevertheless, progress has been made in the identification of genetic variants in select candidate genes, which has enhanced our understanding of the specific pathways involved in disease manifestation. Identification of novel candidate genes for which genetic association studies have confirmed a role in disease has been greatly aided by the powerful tool of high-throughput expression profiling. This article will review these studies to date, summarizing candidate genes associated with ALI and ARDS, acknowledging those that have been replicated in independent populations, with a special focus on the specific pathways for which candidate genes identified so far can be clustered.

l lung injury; acute respiratory distress syndrome; genetic epidemiology; genetic association; polymorphism

MAJOR ADVANCES IN THE EPIDEMIOLOGICAL study of acute lung injury (ALI) and its more severe form, acute respiratory distress syndrome (ARDS), reiterate the considerable heterogeneity that has long been observed in the response to injurious stimuli and their impact on the lung among the critically ill. Associations between genetic polymorphisms and lung injury-related phenotypes in some populations but not others reflect, to a certain extent, limitations related to statistical power and inconsistencies in phenotype definition, but also the profound role that environmental factors play as key determinants in the development (and manifestation) of ALI.

Despite the fact that, in the last few years, there has been a number of studies focusing on the genetics of ALI (11, 29, 30, 37, 73, 81), there remain significant challenges in elucidating the major genes involved in lung injury pathophysiology. First, ALI represents a continuum rather than a discreet group of phenotypes, and therefore lacks unique, readily measurable biomarkers for definitive diagnosis. Second, ALI arises from diverse precipitating factors in a critically ill population and subsequently exhibits a variable clinical course. Because of the phenotypic variance and likely complex gene-environment interactions, significant difficulties persist in elucidating the most likely candidate genes that govern susceptibility and severity of ALI out of the ~20,000 potential genes in the human genome. Well-designed studies examining the association between genetic polymorphisms and susceptibility and severity of ALI should be conducted with careful consideration of selection of candidate genes, phenotype definition, power estimation, and sample size of the study population, selection of appropriate controls, and adjustment, as needed, for confounding factors including population stratification (33, 37, 81).

In this review, we summarize the role of genetic polymorphisms in the pathogenesis of lung injury based on a review of the literature as of 2008, with a special focus on target pathways. We further discuss the complexities of the ALI phenotype and role of microarray-based analyses of genome-wide gene expression in identifying suitable targets for candidate gene studies in ALI. Coassociations between candidate genes with ALI and other inflammatory phenotypes (sepsis and asthma) are presented as a thought-provoking approach toward testing the common variant/multiple disease hypothesis.

Throughout this paper, we highlight strengths and weaknesses of studies conducted to date, for the purpose of underscoring the importance of consideration for ideal study design and attributes associated with the basic principles of genetic epidemiology.

**ALI and ARDS: Complex Etiology and Heterogeneity of the Phenotype**

Both ALI and ARDS are devastating clinical syndromes affecting over 200,000 patients in the United States each year (80). Despite the survival benefit of advances in therapy such as low tidal volumes and lung protective ventilation, mortality remains high, ranging between 25% and 50%, with an average of 38.5% (13, 79, 80). The most common underlying causes of ALI include direct injury to the lung (e.g., “pulmonary” ALI),...
such as pneumonia, aspiration of gastric contents, pulmonary contusion, fat or air emboli, near drowning, inhalation injury, or reperfusion injury, as well as indirect injury (e.g., “extrapulmonary”), such as sepsis, burn, severe trauma with multiple transfusions, cardiopulmonary bypass, drug overdose or toxic ingestion, or acute pancreatitis (94). The many inciting injuries that can lead to ALI, compounded by the multigenic and multifactorial features that are characteristic of a complex trait such as ALI, are the basis for the heterogeneity of disease manifestations and outcomes.

To optimize the dissection of the genetic basis for ALI, a major focus has been on candidate genes that are prominently involved in the primary pathophysiology of the disease, which includes widespread endothelial and epithelial disruption leading to increased permeability, protein-rich pulmonary edema, neutrophilic inflammation, and surfactant dysfunction (67, 95). A unifying theme regardless of the etiological basis for disease is an imbalance between pro- and anti-inflammatory cytokines, oxidants and antioxidants, procoagulants and anticoagulants, neutrophil recruitment/activation and clearance, and proteases and protease inhibitors (94). The pathogenesis of ALI is similar to that of sepsis, as both disease states involve uncontrolled host defense responses that lead to inflammation, endothelial damage, enhanced coagulation, diminished fibrinolysis, and fibroproliferation (58). Sepsis is associated with the highest risk of progression to ALI, ~40% (95). Many stimuli can trigger sepsis, but gram-negative bacteria (containing endotoxin or LPS) are responsible for most clinical sepsis cases (71). Genes and polymorphisms involved in multiple processes of the bacteria-induced cellular response have been associated with susceptibility and severity of the sepsis phenotype, and, therefore, by association, represent potential candidates of relevance for susceptibility to lung injury.

Table 1 summarizes the most current review of literature through 2008 and includes all genes that were identified as a potential ALI locus for susceptibility and/or outcome in conventional association studies. Unfortunately, and due to the fact that non-significant associations are less likely to be published in the scientific arena, it is not possible to determine to what extent these studies have been truly confirmed, and, for every significant association, how many others have failed to observe the same relationship (16, 74). An area of particular discourse (and disagreement) is what constitutes the optimal control group for a reference population of ALI or other severe disease phenotypes such as sepsis (81, 84). The published studies to date on ALI, as summarized in Table 1, have utilized mostly population-based healthy individuals (9 out of 26 studies) or sick controls at risk for development of ALI but whom do not develop ALI (13 out of 26 studies). The fundamental principle behind selection of a control group is to insure that the control group does not have the disease of interest, or any associated intermediate phenotypes (84), and that the control group has had equal opportunity of exposure to the relevant environmental factors that are associated with disease pathogenesis, but are clearly unaffected (70). True associations can also be missed when cases and controls are not matched by important confounders, such as stratification due to admixture, whereby the ethnic make-up of the case group is significantly different than the control group; this confounder, and its relevance in designing genetic epidemiology studies of lung injury, is discussed in detail elsewhere (11). Indeed, as demonstrated in our studies and studies by others, associations with specific polymorphisms may be unique to certain ethnic groups (i.e., MYLK polymorphisms and associations among African Americans as summarized in Refs. 33 and 24), and for this reason, it is not advisable to “mix” ethnic groups in a single case-control or cohort group.

An obvious shortcoming of using healthy controls for studying associations with severe disease is that the significant associations may represent nothing more than “critical illness” genes, not necessarily the specific clinical trait of interest (i.e., ALI). However, how one defines “sick controls” is also problematic because they are potentially heterogeneous. As pointed out by Sapru and colleagues (81), hospital-based at-risk controls run the risk of eliminating the control group that actually contributes to the case group, at least in the context of ALI/ARDS, where the disease is rapidly evolving and misclassification of phenotype is a serious risk. As discussed by Tracey and Warren (90), even controlling for the outcome “severe sepsis” is problematic given that the disease is caused by different microbes affecting different organs but is classified as a single diagnosis. A solution they put forth is to select patients and controls on the basis of clinical and molecular (including genomic) data. However, the practicality of this solution given the a priori difficulties of providing informed consent to severely ill patients (and often surrogate consentees) and the huge costs of screening controls for these parameters before actually enrolling them is questionable. The case-cohort study design, in which a group of at-risk patients is followed for the development of disease (which mitigates selecting a control group), has been advocated (24, 78), but again, this approach is relatively costly as it requires the screening of many more patients to achieve a sufficiently powered case group (15).

In terms of power, this is perhaps one of the greatest concerns of the majority of publications on genetic studies of ALI to date. Sufficient power is established on the basis of a variety of factors, including the effect size of the causative genotype on the trait of interest, frequency of the risk allele, the degree of linkage disequilibrium (LD) between the genetic marker(s) typed and the actual causal allele, and the extent to which the allele frequencies match between the risk allele and genetic marker(s) tested (82). While there is no uniform figure that can be applied across studies, it is generally accepted that well-powered association studies are based on hundreds of samples (and depending on the number of variants to be studied, thousands). As illustrated by Table 1, less than 40% (12 out of 31) of the studies conducted have been performed in study groups where the cases and controls each exceeded 100 or the cohort exceeded 200, and very few of these papers (10, 24, 31, 41) addressed adequacy of power in the Methods section of the manuscript, assuring the reader that power was sufficient for detecting true associations.

Genomic Studies in Lung Injury: A Tool for the Discovery of Novel Candidate Genes

The sporadic nature of ALI has precluded one of the most conventional approaches for elucidating genes in a disease for which a strong genetic basis is suspected but not understood: genome-wide linkage mapping. Linkage mapping is a powerful approach whereby families with both affected and unaffected members are examined for marker loci that segregate with or
are transmitted along with genes that are linked to a disease of interest. Whole genome approaches have the advantage of making no assumptions about the candidacy of particular loci. They require, however, families with multiple individuals having the disease trait, and given the acute nature of lung injury, this approach has not been feasible.

Because of these limitations, candidate gene-based approaches have been the preferred mode of study in the genetic epidemiology of lung injury, and for the practical reasons described above; these types of studies have been conducted among unrelated lung injury cases and controls selected on the basis of not having lung injury. In this approach, the major challenge is choosing the genes most likely to be involved in the trait, because lung injury is such a complex pathophysiological process involving a variety of molecular pathways. Microarray approaches allow the characterization and quantification of the global levels of transcript abundance, or expression profile, of all known genes in the human genome. Similar to the genome-wide linkage approach used in genetic studies, microarray-based analyses of genome-wide gene expression have the potential to provide an unbiased view of multiple molecular pathways simultaneously, thus allowing an integrated view of disease pathogenesis.

Several examples exist to date whereby a previously unknown candidate gene for lung injury was identified using high-throughput expression profiling, and subsequently demonstrated, through genetic association studies, to be relevant to disease (e.g., PBEF1) (97). One example of this approach is that by Grigoryev and colleagues (43), in which they utilized gene expression data derived from multiple animal models of mechanical ventilation and shear stress and identified five key biological process (ontologies), which included inflammation and immune responses, chemotaxis, cell proliferation, and blood coagulation. Using this same approach, it is possible to derive and prioritize a total of 85 candidate genes significantly differentially expressed in mechanical ventilation or LPS-induced ALI models across different species (mouse, canine, and human). Of these, 54 genes are affected by mechanical ventilation, 51 genes are affected by LPS, and 20 candidate genes are commonly differentially expressed in both mechanical ventilation and LPS-induced ALI models (unpublished data). The bioprocesses affected by a particular stimulus were subsequently identified using MAPPFinder software (27), which can link the gene expression data to the Gene Ontology (GO; http://www.geneontology.org) hierarchy. As a proof of concept, this approach has successfully identified gene ontologies previously linked to ALI, as well as novel ones that offer new insights into disease pathogenesis. Twenty major ontologies are listed according to the number of supporting evidences in the literature in Table 2 below.

In addition to the identification of candidate genes by using the data to build detailed pathways of genes involved in lung injury, expression profiling using high-throughput microarray technology has the potential to distinguish disease subtypes, to follow the expression of genes across a time- and/or dosage-dependent gradient when exposed to a stimulus to reveal the complex control processes taking place in an individual cell, and it is also useful in validating genetic studies (34). A thorough review of genomic studies in lung injury was recently presented by Wurl et al. (96), and there are increasing numbers of reports demonstrating the utility of this tool.

Despite the enormous promise of this revolutionary technology, there are potential pitfalls and inherent difficulties in interpreting data from microarray-based, genome-wide expression studies. These may include difficulties in defining normal expression because of 1) heterogeneity of disease in the patients from which the samples were selected; 2) heterogeneity of the target tissue; 3) difficulties in establishing statistically valid comparability; and 4) cross-platform variability (50, 93). Human subject research in ALI using high-throughput expression profiling is constrained given that collection of lung biopsies from critically ill patients is not feasible on a large scale, certainly to the extent that collection of material such as bronchoalveolar lavage-derived cell collection is from smokers and patients with other types of lung disease such as scleroderma or acute lung transplant rejection (96).

Established Genetic Associations in Lung Injury: A Pathway-Oriented Approach

With completion of the International HapMap Project (1), there has been a remarkable increase in the number of genetic association studies performed for complex traits, especially common complex lung diseases such as asthma (74, 92). Sadly, this has not been the case for some of the devastating lung diseases, including ALI and ARDS. Consider that, based on an updated review of the literature on genetic associations with lung injury, only 31 genetic association studies have been published, and of those, associations between polymorphisms in only 21 genes and ALI/ARDS phenotypes have been confirmed (11, 30, 37, 51, 73, 81), as summarized in Table 1.

The cellular and molecular mechanisms of ALI involve endothelial injury, epithelial injury, cytokine-mediated inflammation and injury, neutrophil-mediated injury, oxidant-mediated injury, ventilator-induced injury, coagulation pathway and fibrinolysis, and alveolitis (94). Recently, the role of altered iron mobilization and decompartmentalization in ALI was also reported (51). With this backdrop, it is reasonable that many of the candidate genes studied to date fall within these specific pathways. Some of the most compelling associations have been between specific gene polymorphisms and susceptibility and/or outcomes of ALI in genes encoding surfactant proteins (SPs) (39, 54, 64, 77); angiotensin converting enzyme (ACE) (2, 48, 60); proinflammatory cytokines [TNFA/TNFB (40), IL6 (31, 61, 73, 87), IL8 (44), PBEF1 (10, 97), MIF (32), and VEGF (66, 99)]; and anti-inflammatory cytokine IL10 (38, 83). The most replicated genes include IL6, SFTPB, and ACE. Associations observed between variants in the myosin light chain kinase gene (MYLK), which plays a key role in endothelial cell barrier dysfunction, and ALI have been recently replicated in African American patients with ALI after major trauma (24, 33). Several associations have been reported between polymorphisms in genes encoding innate immune receptors (CD14) (91) and signaling molecule mannose-binding lectin (MBL) (41). Associations between transcription regulator NF-κB (NFκB1) and inhibitor NFκBIA and ARDS have also been reported recently (3, 100). After the first gene in regulating oxidant-mediated response in ALI-transcription factor NRF2 was reported (63), two additional genes in oxidant/antioxidant pathways were reported (SOD3/EC-SOD and NQO1) (8, 78). Recently the role of genes in blood coagulation pathway (PLAU and F5) (4, 7) has been reported, and genes involved in
<table>
<thead>
<tr>
<th>Function/Pathway</th>
<th>Gene (Symbol)</th>
<th>Variant</th>
<th>Ref. (Year)</th>
<th>Population</th>
<th>Significant Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiratory gaseous exchange</td>
<td>Surfactant, pulmonary-associated protein B (SFTPB)</td>
<td>Intron 4 repeats</td>
<td>64 (1996)</td>
<td>EA (15:21)/population-based</td>
<td>ARDS: 1580C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T1580C</td>
<td>54 (2000)</td>
<td>EA (52:46:25 “others”)/both</td>
<td>ARDS: 1580C</td>
</tr>
<tr>
<td>Cytokine</td>
<td>IL-6 (IL6)</td>
<td>G-174C</td>
<td>61 (2002)</td>
<td>EA (96:88, 174, 1906)/both</td>
<td>C allele and CC genotype associated with mortality</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 SNPs</td>
<td>73 (2005)</td>
<td>EA (98:84)/population-based</td>
<td>ALI: -174C/1753C/2954G, GG/G, or G/C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6 SNPs</td>
<td>31 (2008)</td>
<td>European (67:96)/population-based</td>
<td>ALI: 3-SNP haplotype</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A-251T</td>
<td>44 (2007)</td>
<td>EA (23:74)/at-risk</td>
<td>ARDS: -251A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8 SNPs</td>
<td>32 (2007)</td>
<td>EA (90 ALI:113 sepsis:85 controls); AA (61 ALI:69 sepsis:88 controls)/population-based</td>
<td>Sepsis-associated ALI: 3’ haplotype</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G-308A (in TNFA)</td>
<td>40 (2005)</td>
<td>(1/2 NcoI)/at-risk</td>
<td>-308A allele conferred increased mortality in young ARDS patients</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C936T</td>
<td>66 (2005)</td>
<td>(117:137:103 respiratory failure)/both</td>
<td>ARDS mortality: 936TFT and haplotype TCT</td>
</tr>
<tr>
<td>Blood pressure regulation/</td>
<td>Angiotensin I converting enzyme 1 (ACE)</td>
<td>DIP</td>
<td>60 (2002)</td>
<td>EA (96:88, 174 and 1906)/both</td>
<td>ARDS: deletion homozygotes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DIP</td>
<td>2 (2007)</td>
<td>EA (84:200)/population-based</td>
<td>ARDS mortality: ACE deletion</td>
</tr>
<tr>
<td>Endothelial barrier regulation</td>
<td>Myosin light chain kinase (MYLK)</td>
<td>28 SNPs in EA; 25 SNPs in AA</td>
<td>33 (2006)</td>
<td>EA (92 ALI:114 sepsis:85 controls); AA (46 ALI:46 sepsis:61 controls)/population-based</td>
<td>ARDS in EA: one SNP and one haplotype ALI in AA: 2 haplotypes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17 tagging SNPs</td>
<td>24 (2008)</td>
<td>EA (48:83); AA (43:99)/at-risk</td>
<td>Trauma-ass. ALI in AA: 3 variants and multiple haplotypes</td>
</tr>
</tbody>
</table>

Continued
Table 1.—Continued

<table>
<thead>
<tr>
<th>Function/Pathway</th>
<th>Gene (Symbol)</th>
<th>Variant</th>
<th>Ref. (Year)</th>
<th>Population (Cases:Controls/Control Type)</th>
<th>Significant Association</th>
<th>Susceptibility</th>
<th>Severity/Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxidant/antioxidant</td>
<td>Nuclear factor (erythroid-derived 2)-like 2 (NFE2L2 or NRF2)</td>
<td>C-617A</td>
<td>63 (2007)</td>
<td>EA: AA = 1:1 (30 cases &amp; 60 matched controls; 2 controls for each case subject)</td>
<td>Trauma-assoc. ALI: −617A</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Extracellular superoxide dismutase 3 (SOD3 or EC-SOD)</td>
<td>28 SNPs</td>
<td>8 (2008)</td>
<td>EA (251:179; 157:179)/population-based</td>
<td>Infection-associated</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NAD(P)H:quinine oxidoreductase 1 (NQO1)</td>
<td>3 SNPs</td>
<td>78 (2008)</td>
<td>EA and AA (93:185)/at-risk</td>
<td>Trauma-assoc. ALI: protective AC genotype at position −1221</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inmate immune response</td>
<td>Mannose-binding lectin (protein C) 2 (MBL2)</td>
<td>Codon −221, 52, 54, and 57</td>
<td>41 (2007)</td>
<td>EA (212:42)/at-risk</td>
<td>Infection-associated</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nuclear factor of κ light polypeptide gene enhancer in B cells 1 (NFKB1)</td>
<td>−94ins/del ATTG</td>
<td>3 (2007)</td>
<td>European (103 ARDS)</td>
<td>ARDS severity:  D allele (homozygote deletion or heterozygote)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iron ion homeostasis</td>
<td>Ferriin light polypeptide (FTL), heme oxygenase (decycling) 2 (HMOX2)</td>
<td>3 SNPs in FTL; 5 SNPs in HMOX2</td>
<td>52 (2008)</td>
<td>EA (104:193)/population-based</td>
<td>Risk extrapulmonary ARDS: FTL −3381 GG; decreased risk pulmonary ARDS: HMOX2 haplotype</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood coagulation</td>
<td>Urokinase (PLAU)</td>
<td>6 SNPs</td>
<td>7 (2008)</td>
<td>EA (252:175)/population-based</td>
<td>ARDS mortality: 6-SNP haplotype</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ALI, acute lung injury; ARDS, acute respiratory distress syndrome; DIP, deletion/insertion polymorphism in intron 16 of ACE gene defined by the absence (deletion, D) or presence (insertion, I) of a 287-bp repeat; EA, European American; AA, African American.
Table 2. Major biological processes identified by MAPPFinder based on 20 common ALI candidate genes and number of relevant literature references

<table>
<thead>
<tr>
<th>GO Name</th>
<th>Acute lung injury</th>
<th>ARDS</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transport</td>
<td>1,760</td>
<td>690</td>
<td>2,395</td>
</tr>
<tr>
<td>Development</td>
<td>1,476</td>
<td>692</td>
<td>2,122</td>
</tr>
<tr>
<td>Inflammatory</td>
<td>829</td>
<td>289</td>
<td>1,153</td>
</tr>
<tr>
<td>Nitric oxide</td>
<td>266</td>
<td>92</td>
<td>358</td>
</tr>
<tr>
<td>Phospholipid</td>
<td>243</td>
<td>78</td>
<td>321</td>
</tr>
<tr>
<td>Blood coagulation</td>
<td>232</td>
<td>78</td>
<td>330</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>246</td>
<td>57</td>
<td>303</td>
</tr>
<tr>
<td>Cell motility</td>
<td>244</td>
<td>54</td>
<td>299</td>
</tr>
<tr>
<td>Transcription</td>
<td>222</td>
<td>50</td>
<td>272</td>
</tr>
<tr>
<td>Cell adhesion</td>
<td>195</td>
<td>58</td>
<td>253</td>
</tr>
<tr>
<td>Signal transduction</td>
<td>187</td>
<td>36</td>
<td>223</td>
</tr>
<tr>
<td>Chemotaxis</td>
<td>156</td>
<td>32</td>
<td>188</td>
</tr>
<tr>
<td>Cell communication</td>
<td>150</td>
<td>29</td>
<td>179</td>
</tr>
<tr>
<td>Collagen catabolism</td>
<td>124</td>
<td>36</td>
<td>160</td>
</tr>
<tr>
<td>Immune response</td>
<td>118</td>
<td>32</td>
<td>150</td>
</tr>
<tr>
<td>Heart development</td>
<td>82</td>
<td>28</td>
<td>110</td>
</tr>
<tr>
<td>Phagocytosis</td>
<td>91</td>
<td>16</td>
<td>107</td>
</tr>
<tr>
<td>Arginine metabolism</td>
<td>80</td>
<td>13</td>
<td>93</td>
</tr>
<tr>
<td>Muscle contraction</td>
<td>56</td>
<td>18</td>
<td>74</td>
</tr>
<tr>
<td>Cell cycle</td>
<td>60</td>
<td>11</td>
<td>71</td>
</tr>
</tbody>
</table>

Shown are the number of references found by PubMed (http://pubmed.ncbi.nlm.nih.gov) citing biological processes and relevant terms.

Polymorphisms in cellular messenger genes. Cytokines are peptides produced by diverse types of cells (including immune, endothelial, and nervous cells), which act as cell-to-cell messengers to regulate diverse processes, including immunity, inflammation, and cellular response to adverse conditions (85). Genetic variations in the proinflammatory cytokines TNFA, TNFB, MBL, and IL6 and the anti-inflammatory cytokine IL10 are the most extensively studied polymorphisms in relation to sepsis (75). A number of these polymorphisms have also been implicated in ALI (Table 1). Pre-B cell colony-enhancing factor (PBEF1), also known as visfatin, was identified as a novel biomarker in ALI by extensive analysis of microarray-based gene expression profiling data, as described above (97), and has since been shown to exert three distinct activities related to cellular energetics and innate immunity. PBEF1 can be strongly induced by inflammatory stimuli in cells involved in innate immunity, specifically neutrophils, monocytes, and macrophages, as well as in endothelial and epithelial cells (57). Novel PBEF1 variants were identified through resequencing of PBEF1 exons (including exon-intron boundaries) and 2 kb of the 5′ UTR (97). Subsequent genotyping was conducted in a European American case-control population focusing on two promoter SNPs (T-1001G and C-1543T) that were physically close to mapped transcription factor binding sites. The T-1001G polymorphism was associated with an increased risk of sepsis-associated ALI, whereas the variant C-1543T polymorphism was associated with a protective effect. These findings were eventually replicated in an independent European American ARDS population (10), and since the initial report of an association between PBEF1 polymorphisms and disease, genetic associations have also been reported for other diseases that share an inflammation basis, such as type 2 diabetes (9).

As suggested above, associations with other cytokine genes are summarized in Table 1, indicating replication of association at both the gene level, and, in several cases, at the level of the variant. Upon close examination, however, seemingly contradictory results have been reported for several of these variants, rendering it difficult to determine whether true replication has been provided or not, and underscoring a phenomenon that is not unique to genetics studies of ALI. For example, Gong et al. (38) reported that the high IL-10-producing −1082GG genotype was associated with the development of ARDS, but only in the presence of a significant interaction between the −1082GG genotype and age. Elsewhere, Schreuder et al. (83) reported in a study from Europe that carriers of the same −1082GG genotype were less likely to develop acute respiratory failure after major trauma. One explanation for this phenomenon was attributed to “ascertainment bias” (23), owed to the manner by which subjects were selected into the study, interactions with clinical variables (in this case, age) that potentially modify genetic associations, or because of the diverse etiological mechanisms (i.e., sepsis, pneumonia, trauma, massive transfusion) that predisposed patients to the condition under study. Certainly, as in the case of these two studies, replication of a disease-marker association where the risk allele is reversed in the follow-up study, also referred to as a “flip-flop” association (53), has been increasingly observed as more and more association studies are performed in general, and have raised the question of whether these associations are confirmations or contradictions. Lin and colleagues (53) argue that, certainly in the case of opposite alleles for the same disease in the same ethnic group, particularly for a genuine causal variant, a flip-flop association is unlikely. However, for studies performed in different populations (which is usually the case in replication studies, and certainly the case for the Gong et al. vs. the Schreuder et al. studies on IL10 G-1082A), flip-flop associations may highlight heterogeneity due to differences in the genetic background of different populations, which is discussed in greater detail in the next section. The Gong et al. study was a nested case-control study focused on cases of those of European descent living in Boston with at least one defined risk factor for ARDS (sepsis, pneumonia, trauma, massive transfusion, or aspiration) and intensive care unit-based controls with the same risk factors but without ARDS; the primary outcome was organ failure and 60-day mortality. The Schreuder et al. study was a relatively small cohort study of those of European ancestry living in Berlin, Germany, with multiple trauma, where the primary outcome measure was acute respiratory failure. Moreover, the variant of interest is not functional and is therefore likely to be in linkage disequilibrium with the true (causal) variant. Thus, the differences in associations observed in this particular example cannot be reconciled without further efforts toward replication of the IL10 G-1082A variant in other independent populations selected for the same phenotypes.

Polymorphisms in endothelial-mediated lung injury-related genes. Both in vivo and in vitro studies have demonstrated that lung endothelial cell barrier dysfunction plays a key role in the pathophysiology of ALI. The nonmuscle endothelial cell myosin light chain kinase (EC MLCK), encoded by the myosin light chain kinase (MYLK) gene, is centrally involved in...
cytoskeleton rearrangement participating in apoptosis, inflammation, cell trafficking, cell division, and angiogenesis (28). Lung edema formation in ALI, which is caused by injury to the lung microvascular endothelium, also involves the MLCK-dependent cytoskeletal rearrangement. We first explored MYLK as a novel candidate gene in sepsis and sepsis-associated ALI because of its pivotal role in the disease process, and through sequencing, polymorphisms were identified for which significant associations were observed for both severe sepsis and sepsis-associated ALI (33). A novel observation in our initial report was the difference in risk-conferring haplotypes according to ethnicity in two case-control populations (African Americans vs. European Americans) ascertained under identical protocols. Upon close examination, we observed striking differences in genetic structure, or LD, of the two populations according to the 28 SNPs that were genotyped, and considered these differences, as well as potential interactions with other (untyped) genetic markers or environmental risk factors, as an explanation for the greater risk of ALI in the African American group.

The association between MYLK variants and the development of ALI was recently replicated in a study of trauma-induced ALI patients, although the risk allele for each of the three functional variants (Thr335Thr, Pro147Ser, His21Pro) was reversed from our original report. In comparisons between associations in the combined ethnic groups vs. the individual ethnic groups, Christie and colleagues (24) noted that associations appeared to be driven by the African Americans, a finding that supported our original report. Although modestly powered to evaluate each ethnic group independently (i.e., less than 100 individuals in each case or control group), it is important to highlight the group’s own contention that differences in the pathophysiology of ALI between major trauma and sepsis as an etiological factor could account for the flip-flop effect described in the section above. As described elsewhere, the flip-flop effect is not expected in the event that the association is for a genuine causal variant and in samples ascertained from a common population in the same manner (53); however, if we consider that 1) even though the three MYLK variants are functional variants, there is no evidence as of yet that they are “causal”; 2) although the populations showing the strongest association in both studies are African American, there is known heterogeneity among African Americans in the U.S. dependent on where they reside (76); and 3) the outcome of interest in the two studies was distinctly different (i.e., sepsis-associated vs. trauma-associated ALI), it is perhaps not surprising that inconsistent patterns of association are observed across the studies.

ALI is characterized by overlapping phases: exudative (early/acute) and fibroproliferative (subacute/late phase) (58, 94). Animal studies of ALI suggested that circulating endothelial precursor cells are critical in the repair of alveolar endothelium in the late fibroproliferative phase of ALI (14). Higher lung levels of VEGF, a mitogen that specifically acts on endothelial cells and has various effects, are associated with recovery from ALI (89). VEGF is also a chemotractant for endothelial precursors. Medford et al. (66) first reported that the variant allele of the CT SNP at +936 of the 3’ UTR of the gene associated with lower VEGF plasma levels and conferred increased risk for the development of the disease, whereas Zhai (99) et al. did not find any increased risk of ALI associated with the C936T polymorphism but reported only an association with the risk of mortality among patients with ARDS. These results provide grounds for future research aimed at investigating whether or not host genetic factors also influence the recovery of ALI.

Polymorphisms in innate immune receptor and signaling molecule genes. Sepsis describes a complex clinical syndrome as a result of a systemic inflammatory response to live bacteria and/or bacterial products, and sepsis is the most common clinical risk factor for indirect lung injury (95). During the past decade, functional and association studies involving genetic polymorphisms in essential host defense genes, including innate immune receptors and signaling molecules, cytokines and coagulation factors, have provided important insights into the mechanisms involved in the pathogenesis of sepsis and its induced organ dysfunction (6). A number of genetic associations have been reported between sepsis and polymorphisms in innate immune receptor genes, including TLR4 (5, 18, 55), CD14 (36, 86), and MBL (35, 42, 86). In several of these studies, the specific polymorphism associated with the trait has also been associated with ALI/ARDS, i.e., the variant B allele of codon 54 of MBL gene (41). TLRs function as the pattern recognition receptors through which the innate immune system recognizes microbial pathogens, and TLR4 is one of the most extensively studied. Genetic association between two coding SNPs, Asp299Gly and Thr399Ile, and increased susceptibility to bacterial infection and severe sepsis/septic shock have been reported (5, 18, 55). Similarly, CD14 is a ubiquitous pattern recognition receptor, specific for LPS and a variety of other ligands, and is expressed primarily on immune and endothelial cells. CD14 also acts as a coreceptor for TLR4 (86). A soluble form (sCD14) is shed by monocytes to facilitate LPS signaling for all other cells (49). The relatively common C-260T variant (initially identified as C-159T) functions by decreasing the affinity of Sp protein binding and subsequently enhancing transcriptional activity, and is associated with increased sCD14, which is markedly increased in bronchoalveolar lavage of patients with ARDS (62). In a Canadian study, the T allele was associated with increased prevalence of gram-negative infections and sepsis (86), and in a French multicenter study, an association with sepsis mortality was reported (36). MBL is a circulating serum protein that recognizes and binds carbohydrate structures on the surface of infectious agents and activates the complement system through MBL-associated serine proteases (88). Evidence for association with sepsis was found in several studies (35, 42, 86), and recently Gong and colleagues (41) reported that patients homozygous for the variant codon 54B allele (54BB) had increased odds of ARDS when compared with heterozygotes and homozygotes for the wild-type allele, especially among patients with septic shock. Given the prominent role that sepsis plays in lung injury etiology, it is not surprising that genes associated with sepsis and severity of sepsis are likely candidates for ALI.

Polymorphisms in oxidant-mediated injury genes. Studies have shown that the transcription factor NRF2 appears to be an essential regulatory element in response to oxidant injury (19–21). Positional cloning studies in inbred mice identified Nrf2 as a candidate susceptibility gene for hyperoxia-induced lung inflammation and injury, a model of ALI (19, 20), Marzec and coworkers (63) performed resequencing to scan the entire coding region, and the proximal 1 kb of the promoter and genotyping was performed on two promoter SNPs (~617 and ~651); they subsequently found that patients who were ho-
mozygous for the −617 minor “A” allele had a significantly higher risk for developing ALI after major trauma relative to patients with the wild type (−617 CC). Although this is a relatively new group of candidate genes in the field of lung injury genetic epidemiology, there has been a tremendous number of genetic associations between genes associated with defense against oxidant stress and other complex lung diseases, including glutathione synthesis-related genes [e.g., GCLC and cystic fibrosis (65)] and superoxide dismutase (SOD) genes [e.g., SOD3 and COPD (98)]. Recent work by Arcaroli et al. (8) confirmed the association between haplotypes in extracellular superoxide dismutase 3 gene (SOD3 or EC-SOD) and mortality in patients with infection-associated ALI, and another report suggested that the A-1221C variant of the NAD(P)H: quinine oxidoreductase 1 (NQO1), a phase II antioxidant gene, conferred higher risk for trauma-associated ALI (78).

Polymorphisms in blood coagulation pathway genes. Polymorphisms in genes encoding proteins of the coagulation pathways [i.e., protein C and plasminogen activator inhibitor 1 (PAI-1)] are associated with altered levels of their respective mediators and phenotypes of disordered coagulation (81). Recently, Arcaroli et al. (7) reported the association between haplotypes in urokinase gene (PLAU), which encodes a serine protease that cleaves plasminogen to generate plasmin, a potent mediator of fibrinolysis, and outcomes (both 60-day mortality and ventilator-free days) from ALI. Growing evidence also supported that a factor V Leiden (F5) genotype influenced survival in patients with ARDS (4). Thus, this group of candidates represents a novel area in the field of lung injury genetics.

Coassociations between ALI Candidate Gene Variants and Lung Inflammatory Phenotypes: The Common Variant/Multiple Disease Hypothesis

Sepsis is one of the main risk factors leading to the development of ALI, and, as demonstrated above, polymorphisms in genes previously implicated in sepsis function as strong candidate genes for lung injury. As summarized, several sepsis-associated genes, such as ACE, IL6, TNFA/TNFB, and MBL, have already been associated with ALI susceptibility. Additional sepsis-associated genes, such as coagulation pathway-related polymorphisms, were also involved in ALI (4, 7, 81). In contrast, genes of relevance for risk of ALI (e.g., MYLK and MIF) may also constitute potential candidates for susceptibility to other distinct, but possibly related, clinical syndromes, such as allergic asthma (34). Macrophage migration inhibitory factor (MIF) is a proinflammatory cytokine central to the response to endotoxemia, and a putative biomarker in ALI. As a pleiotropic lymphocyte and macrophage cytokine, MIF is likely to play an important role in innate immunity. Haplotypes in the 3′ end of the MIF gene were associated with severe sepsis and sepsis-associated ALI (32). Evidence for significant association between the functional MIF gene promoter SNP (−173G/C) and −794 [CATT]5–8 repeat polymorphisms and atopy was similarly reported in a Japanese population (46), and a significant association between asthma and the 5′-CATT MIF allele was subsequently confirmed by Mizue et al. (68). Collectively, these findings support the “common variant/multiple disease” hypothesis, which is an extension of the common disease/common variant hypothesis, and suggests that certain disease genes may not be disease specific and may contribute to related clinical phenotypes (34). Indeed, these sorts of examples of genetic association of the same gene (or allele) to multiple, related disorders abound (http://geneticassociationdb.nih.gov). Given the prominent role that certain pathways, such as host defense, play across various complex lung diseases, it is likely that many more coassociations will be observed in the near future.

General Insights into the Pathophysiological Pathways Related to Lung Injury

For complex inflammatory diseases such as ALI, it is anticipated that many genes will be involved and multiple variants will contribute to the alteration of gene function and expression. Because the effects of these genes/polymorphisms are typically relatively modest, it is generally assumed that variants in multiple genes cooperate in an additive or synergistic manner to impact disease risk. This phenomenon is referred to as “epistasis.” One approach toward characterizing potential gene-gene interactions and systematically evaluating the role of ALI genes/polymorphisms in disease susceptibility is pathway analysis, whereby hierarchical clustering of genes is performed considering their functional closeness in a pathway or risks imparted independently from other genes in the pathway. Using the program Ingenuity Pathways Analysis (Ingenuity Systems: www.analysis.ingenuity.com), Loza and co-workers (56) identified 1,027 key inflammation-related genes targeting various phases of inflammation responses and associated 17 subpathways that can further facilitate pathway-focused genetic association analyses for inflammatory diseases. The Ingenuity Pathways knowledge base is a web-based entry tool developed by Ingenuity Systems to characterize genes according to the predefined canonical pathway(s) into which they fit, and also to investigate the extent to which genes identified in a particular study (e.g., microarray-based studies) are in shared networks and may cooperate in a synergistic or an additive manner to impact risk of disease.

As a proof of concept, we evaluated the 21 genes that have been well established as ALI/ARDS genes and summarized in Table 1 using the Ingenuity Pathway Analysis. Consequently, two major networks were revealed (Table 3). Twelve genes (ACE, IL6, IL8, IL10, TNF, LTA, NFKB1, NFKBIA, SFTPB, VEGFA, PLAU, F5) that have previously been associated with lung injury were clustered into a major network related to 1) “immune and lymphatic system development and function”; 2) tissue morphology; and 3) cellular movement. Moreover, an additional 24 related genes/molecules belong to the same network. Using the multiplex literature search tool, PubMatrix (http://pubmatrix.grc.nia.nih.gov), we found evidence supporting the involvement of 9 out of the 24 related genes (Ap1, DBT, DEFB4, F2, Histone h3, MICA, RNA polymerase II, STAT3, and TNFSF4) in ALI/ARDS. Furthermore, another nine genes previously associated with ALI [FIL, HMOX2, MBL2, MIF, MYLK, PBEF1 (NAMPT), NFE2L2, SOD3, NQO1] were in the second major network related to “cellular compromise, molecular transport, hematological system development and function.” Among the 26 related genes in the same network, evidence was found in the literature for eight genes (C4A, ENC1, Ho, JUN, MINA, MYC, NQO, PPCS) supporting their involvement in ALI/ARDS pathogenesis. The top 5 “diseases
and disorders” and “molecular and cellular functions” suggested by the two networks derived from 21 ALI genes were also summarized in Table 3. Notably, the top molecular and cellular functions represented in current association studies fit within the major biological processes underlying ALI development suggested by the common ALI candidate genes derived from diverse across species microarray experiments (Table 2). While this list is not intended to be a comprehensive summary of all genes potentially related to ALI pathology, it serves as an example of the power of this approach in selecting optimal candidates for genetic association studies. Figure 1 illustrates the overlay of the two major networks that are summarized in Table 3. Other tools such as MetaCore, a pathway analysis and data mining tool provided by GeneGo (http://www.genego.com), can serve a similar purpose.

**Methodology To Date for Genetic Association Studies in ALI**

In the past three decades, substantial progress has been made in understanding the mechanisms that underlie ALI/ARDS. Although there has indeed been an increase in the number of published genetic association studies focusing on lung injury over the past few years, with increasing reliance on novel and potentially powerful tools such as microarray-based approaches and computationally intensive techniques for interpretation of findings, the field has far to go. Improvements in the methodology used in genetic association studies, including population selection, phenotype definition, determination of appropriate sample size, genotype data quality control, statistical methodologies to adjust confounding factors such as multiple testing and population stratification and study replication will greatly facilitate discovery of true associations important in the care of ALI/ARDS patients. From recently published studies, we do see increased reproducibility with 8 out of the 21 genes associated with all-cause ALI susceptibility and/or outcome were replicated. Another major advance is using haplotype tagging SNP (htSNP) to capture gene-wide common variations. The premise behind this approach is that in regions of high LD, and because many SNPs are frequently inherited together, only certain SNPs are selected to fully represent a candidate gene. Several htSNP methods exist, and the selection of which method to use is somewhat determined by the structure of the population under study (12, 17, 25). However, as highlighted in Table 1, of the studies published so far, very few have focused on non-white patient populations. Because ALI is typical of other complex lung disorders whereby ethnic disparities exist in terms of morbidity and mortality (11, 72), a greater focus on diverse populations is warranted. The role of ethnicity is only one component of the observed heterogeneity of disease manifestations and outcomes. ALI is characterized by overlapping phases: exudative (early/acute) and fibroproliferative (subacute/late phase) (58, 94). Some patients recover during the acute phase, but most progress to the subacute phase, which renders ALI/ARDS one of the major causes of pulmonary and nonpulmonary morbidity in patients after discharge (26, 58). It is reasonable that the variation in recovery may be related in part to the underlying genetic characteristics of the individual patient. The identification of genetic factors associated with susceptibility, severity, and outcome/recovery constitutes a necessary step toward personalized treatment. Finally, and importantly, of the published association studies in ALI summarized in Table 1, many studies have been underpowered to detect true false-negative results that could be one of the potential confounding factors. To be fair, this is understandable given the profound challenges in amassing large datasets of what is in actuality an uncommon disease. The use of derivation and validation subsets of patients is one solution to these limitations and a failure to reproduce findings to date (15). Flores et al. (29) evaluated the quality of 29 published positive associations for 16 genes in ALI based on current recommendations suggested for genetic association studies (16). Using a 10-point quality scoring system derived from 14 criteria, they found the studies had an intermediate quality score of 4.62, suggesting that more and better designed studies with larger sample sizes are needed, including the replication of previous findings, and should be more inclusive of population groups other than European descent (29).

Genome-wide association studies offer a potential solution to the limitation in candidate gene approaches; that is, that each causal variant in each candidate gene will only make a modest contribution to overall heritability (45). By screening dense panels of markers representing the most common genetic variation across the genome, it is assumed that many more risk variants will be identified than what could be accomplished using conventional, candidate gene-based studies. High-den-
sity genome-wide association (GWA) studies to date have demonstrated the success and robustness of this approach in identifying genetic variants or regions associated with nearly 40 complex diseases or traits (59), including complex lung diseases such as asthma (69). Although this approach has yet to be applied to ALI or sepsis, the limitations of HapMap-based GWA studies discovered to date from these preliminary successes should be carefully considered (22, 47, 59). GWA

Fig. 1. Networks revealed through the Ingenuity Pathways Analysis based on 21 ALI genes established in acute lung injury genetic association studies. The major network of 36 genes (the 12 published ALI genes are highlighted in blue) was merged with the second network of 35 genes (the 9 genes previously associated with ALI are highlighted in green). Molecules highlighted in gray represent molecules that belong to a group (multiple specific molecules with a similar function) or a protein complex made up of multiple protein subunits in the network. Genes/molecules are connected according to their relationships, and the newly added relationships that connect the two networks are also highlighted in orange. The top 5 diseases/disorders related to genes/molecules in the two networks include 1) organismal injury and abnormalities; 2) immunological disease; 3) hematological disease; 4) respiratory disease; and 5) inflammatory disease. Top molecular and cellular functions suggested by the networks are cellular movement, cell death, lipid metabolism, small molecule biochemistry, and cell-to-cell signaling and interaction.
studies often require very large numbers of samples (i.e., thousands of cases and controls needed to obtain sufficient statistical power to identify the alleles of reasonable effect size); conversely, the GWA design has not proven successful for identifying rare and structural variants. Because of the manner by which the chips are designed (i.e., random “tagging” SNPs are selected as a priority over design of markers of causal variants), the correlations identified with disease do not typically implicate causation, and follow-up studies to identify the specific mutation are not always straightforward. Extensive follow-up and replication of findings in the discovery population are essential to rule out false positives. Given that ALI is not as prevalent as other complex traits such as asthma, and given the complex nature and heterogeneity of the phenotype, there are certainly obstacles with GWA studies in ALI. Nevertheless, it can be anticipated that more carefully designed primary and replication studies focused on candidate genes identified in known cellular pathways implicated in ALI pathogenesis, or genes supported by plausible biological hypothesis, will allow for alternatives to the GWA design and lead to identification of gene–gene and gene–environment interactions (i.e., infection, mechanical ventilation). It is reasonable that collaborative efforts towards establishing large DNA banks from ALI/ARDS patients will ultimately satisfy the requirements for successful GWA studies and integration of data across studies from “omics” (i.e., genomics and proteomics) to facilitate identification of novel targets and cellular pathways specific to the disease.

Conclusion

In summary, additional association studies based on novel candidate genes and a deeper appreciation for the relevant pathways involved in lung injury have advanced the field of genetic epidemiology of ALI and ARDS. The integration of novel tools for elucidating candidate genes, such as microarray-based approaches and computational approaches including pathway analysis, has broadened our understanding of the different groups of genes involved in disease manifestation and outcome. Originally designed for extracting new biological insight from high-throughput genomic studies of human disease, pathway analysis can also be a powerful approach to facilitate the detection of gene–gene interactions and pathway-focused genetic analysis.

Recent advances in genotyping approaches have already been utilized, and the most robust approach to date, genome-wide association studies, is yet to be tested in either ALI or sepsis. Ultimately, a clearer understanding of the genetic determinants conferring risk to ALI, ARDS, and other associated phenotypes, will pave the way toward a more personalized approach of therapy and resolution of disease.

ACKNOWLEDGMENTS

Special thanks to Dr. Dmitry Grigoryev for his work on the expression profiling data. Pat Oldewurtel provided technical assistance in the submission of this manuscript.

GRANTS

This review represents the efforts of investigators participating in the National Heart, Lung, and Blood Institute-funded SCCOR on “Molecular Approaches to Ventilator-Associated Lung Injury” (P50-HL-073994) and the projects entitled “Genetic Modifiers in Acute Lung Injury” and “Genomics and Genotyping Core.” K. C. Barnes was supported in part by the Mary Beryl Patch Turnbull Scholar Program.

REFERENCES

Review


