Gene association studies in acute lung injury: replication and future direction

Michelle N. Gong

Division of Pulmonary and Critical Care Medicine, Department of Medicine, Mount Sinai School of Medicine, New York, New York

ACUTE LUNG INJURY AND ACUTE RESPIRATORY DISTRESS SYNDROME (ALI/ARDS) usually develops in response to a major insult such as sepsis, trauma, pneumonia, and multiple transfusions. Our current understanding of why some patients develop and die from ALI while others do not is incomplete. Although genetic determinants to the development of and outcome in ALI/ARDS have only recently been investigated, there has been a sudden explosion of studies in this field in the last few years.

In their review, Gao and Barnes (4) make their latest contribution by summarizing the recent advances in the genetic susceptibility to ALI/ARDS. Their review highlights how genetic association studies in ALI/ARDS can add to our current understanding of this devastating condition. Yet it also makes it clear that the investigation into the genetic susceptibility to ALI/ARDS is still in its infancy and much still needs to be done.

One vital need in this field is more replication studies in the future. A causal genotype-phenotype relationship cannot be established with one initial report. As such, replication studies are the backbone to the genetic epidemiology of complex diseases (3). In one meta-analysis, 20% of these replication studies showed a significant result confirming only 44% of these genetic associations (10). But there is debate as to what constitutes a successful replication or refutation of a genetic association in complex diseases. Interpretations of replication studies are challenged by differences in genotyping techniques, insufficient power, and heterogeneity in the study population.

Some early studies in ALI/ARDS, like many other early genetic epidemiology studies of complex diseases, focused on reportedly functional variants of, most commonly, single nucleotide polymorphisms (SNPs). Others have examined haplotypes or haplotype-tagging SNPs, which more comprehensively describes the variation in the gene. However, the ability of replicate studies to confirm the original association depends on whether the genotyping method provides adequate coverage of the genetic variation and the linkage disequilibrium (LD) between the true causal locus and variants genotyped in different studies. In addition, multiple susceptibility loci may exist on the same genetic region (13). Thus, the unit of replication should be the gene, and association with any variant or haplotype on the gene should be considered confirmatory (12).

In their review, Gao and Barnes (4) pointed out the increasingly recognized “flip-flop” phenomenon, seen in up to 20% of replicate studies (10). This occurred more frequently than would be expected by chance and cannot be explained completely by publication bias. Such a phenomenon can occur in the setting of different LD, different haplotype frequencies, and different gene-gene interactions (9). With conditions as heterogeneous as ALI/ARDS, interaction between the gene and other clinical risk factors may also play a role. The true extent of this phenomenon in ALI/ARDS and other complex conditions is unknown, but replicate studies that show associations that are opposite from the original reports should not be dismissed reflexively. Rather, additional studies will be needed to determine whether the results confirm or refute the original finding.

Another common reason for failure of replication studies is insufficient power. Of 25 genetic associations reported in the literature, 8 (32%) were replicated only after pooling the effect of multiple false-negative replication studies (10). Why are so many replication studies underpowered to confirm the original report? It seems that 64–92% of initial reports of genetic associations is larger than the true estimate (10). “Winner’s curse” refers to a tendency for the odds ratio in an initial study to be biased upward due to pressure on investigators to be the first to report a significant result and on journals to publish a novel finding (14). Thus, replicate studies need to be powered sufficiently to detect a considerably smaller effect size than the original report (5).

Heterogeneity in the study population can complicate the interpretation of replicate studies in ALI/ARDS. Heterogeneity can occur with differences in case definition and control selection, population stratification, and gene-environment interaction. Most use the American European Consensus Committee definition for ALI/ARDS, but the outcome is ALI/ARDS development in some studies and ALI/ARDS mortality in others. Additional studies will be needed to determine whether genetic variants in ALI/ARDS are equally important in susceptibility and outcomes.

As pointed out by Gao and Barnes (4), about one-third of the published studies in ALI/ARDS have used healthy controls, whereas the rest used critically ill patients at risk for ALI/ARDS. Choice of controls affects the ability to replicate genetic associations and the sample size needed to demonstrate replication. If a risk allele is associated with both the development of ALI/ARDS and the critical illness predisposing to lung injury, replication with at-risk controls will require a larger sample size than with healthy controls.

Race and population stratification is another source of heterogeneity. A true disease locus should be associated with the disease in all racial groups. But the frequency of the disease genotype and the size of the effect may vary with different racial groups due to differences in the LD between the variant genotyped and the true disease locus, frequency of modifying genes that interact with the risk variant, and differences in clinical factors (such as diabetes and alcohol abuse) that may modify any effect of the at-risk genotype. Thus, failure to replicate a genetic association in a racial population different from the original report does not necessarily negate the importance of the initial finding (3). Within racial groups, it is not clear whether undetected population stratification poses a significant risk of false-positive findings (1, 15).
Gene-environment interaction also contributes to heterogeneity between study populations. Gene-environment interactions have been noted for several genetic associations in ALI/ARDS. Lack of adjustment for such gene-environmental interaction is now known to contribute to inconsistent genetic association studies for complex disease (8).

Of course, the future in the genetic epidemiology of ALI/ARDS does not only involve repeating the past with replication studies, no matter how important they may be. There has been much interest in future genome-wide association studies (GWAS) in ALI/ARDS. By simultaneously examining large numbers of SNPs distributed throughout the genome, GWAS has the ability to detect disease associations with known and previously unsuspected genes. The technical aspect of GWAS is becoming more realistic as the cost of high throughput genotyping is decreasing.

However, limitations in GWAS open up other potential approaches to the study of genetic determinants in ALI/ARDS. First, because of the large number of multiple comparisons, GWAS requires very large sample sizes of as many as 1,000 cases and 1,000 controls to detect modest estimates (11). If we consider that in ALI/ARDS, gene-environment interactions are likely, the sample size requirements will increase further. One potential solution is the candidate pathway approach demonstrated so powerfully by Gao and Barnes (4). This approach is hypothesis driven and allows for examination of gene-gene interaction while reducing the number of genes to be tested. An alternative approach is a two-stage design whereby the first stage of analyses narrows down the number of SNPs to be examined in the second stage (7).

Most GWAS platforms are mainly geared towards genotyping common SNPs. Copy number variants (CNVs) and rare variants are now being recognized as important contributors to disease susceptibility. CNVs consist of variable numbers of insertions or deletions of DNA, which account for substantial variability in the human genome and may be better associated with disease than SNPs. While some platforms are now capable of indirectly genotyping CNVs, only a small proportion of CNVs are captured by these methods.

GWAS or other current indirect genotyping approaches relies on the common disease common variant hypothesis. Alternatively, the rare variant hypothesis argues that most genetic risk to common disease will be due to multiple rare variants (frequency 0.1% to 2–3%) on a risk gene. Accumulation of these rare variants have been found to be associated with colorectal cancer and other common diseases with odds ratios (ORs) that are much higher (2–3) than those found for common variants (OR 1.2–1.4) (2). The higher penetrance indicates greater impact on determining individual risk, which is consistent with the susceptibility to common disease.

REFERENCES