Assessing the quality of studies supporting genetic susceptibility and outcomes of ARDS

Marialbert Acosta-Herrera [1,2,3], **Maria Pino-Yanes** [1,2,4], **Lina Perez-Mendez** [1,2], **Jesús Villar** [1,3,5] and **Carlos Flores** [1,2,6]*

[1] CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain
[2] Research Unit, Hospital Universitario N.S. de Candelaria, Santa Cruz de Tenerife, Spain
[3] Research Unit, Hospital Universitario Dr. Negrin, Las Palmas de Gran Canaria, Spain
[4] Department of Medicine, University of California, San Francisco, CA, USA
[5] Keenan Research Center at the Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, ON, Canada
[6] Applied Genomics Group (G2A), Genetics Laboratory, Instituto Universitario de Enfermedades Tropicales y Salud Pública de Canarias, Universidad de La Laguna, Santa Cruz de Tenerife, Spain

The acute respiratory distress syndrome (ARDS) is a severe inflammatory disease manifested as a result of pulmonary and systemic responses to several insults. It is now well accepted that genetic variation influences these responses. However, little is known about the genes that are responsible for patient susceptibility and outcome of ARDS. Methodological flaws are still abundant among genetic association studies with ARDS and here, we aimed to highlight the quality criteria where the standards have not been reached, to expose the associated genes to facilitate replication attempts, and to provide quick-reference guidance for future studies. We conducted a PubMed search from January 2008 to September 2012 for original articles. Studies were considered if a statistically significant association was declared with either susceptibility or outcomes of all-cause ARDS. Fourteen criteria were used for evaluation and results were compared to those from a previous quality assessment report. Significant improvements affecting study design and statistical analysis were detected. However, major issues such as adjustments for the underlying population stratification and replication studies remain poorly addressed.

Keywords: genetic susceptibility, acute respiratory distress syndrome, outcome, genetic factors, population stratification

INTRODUCTION

Acute lung injury (ALI) and its severe form, the acute respiratory distress syndrome (ARDS), are characterized by acute diffuse lung inflammation and non-cardiogenic pulmonary edema resulting from increased capillary-alveolar permeability. While ALI and ARDS terms continue to be used in the medical literature, their definition criteria were recently revised, although a consensus has not been reached (Ranieri et al., 2012; Villar et al., 2013). New definitions support the categorization of ARDS based on the hypoxemia severity under mechanical ventilation, as well as on other physiological and clinical parameters, discouraging the use of ALI as one of the categories. Hereafter, we will refer to this constellation of syndromes using the term ARDS, irrespective of the classification used by the studies reviewed (Bernard et al., 1994). ARDS shows profound incidence variability across countries (Rubenfeld et al., 2005; Villar et al., 2013), and it is unknown whether differences also exist among ethnic groups (Martin et al., 2003; Erickson et al., 2009; Linko et al., 2009; Villar et al., 2011) and the extent to which demographic, cultural, economical, and health system particularities might underlie such differences.

Predisposing genetic factors can interact with the environment to determine the diversity of clinical manifestations, the response to treatment and outcomes among ARDS patients (Cobb and O'Keefe, 2004; Villar et al., 2004; Rahim et al., 2008).

Exposing those genetic factors might reveal therapeutic targets and a foundation to predict ARDS susceptibility and outcomes. Association studies have been widely used for detecting common, low-penetrant, genetic variants that are suggested to contribute to the genetic architecture of complex diseases (Khoury and Yang, 1998), including ARDS (Flores et al., 2008). For ARDS, these studies have mostly focused on particular biological candidates and, only recently, have explored the entire genome (Christie et al., 2012). We have previously assessed the quality of statistically significant associations of genetic variants with ARDS from 1996 to 2008 based on major recommendations that support study robustness (Flores et al., 2008). We hypothesized that, despite this previous evaluation and the availability of well-known standard guidelines (Janssens et al., 2011), many association studies in this field continue to be performed without awareness of minimal standards and that methodological flaws are still abundant. Here, we aimed to identify those quality criteria where the standards have not been reached, to expose the associated candidate genes to facilitate replication studies, and to create a guidance framework for ongoing and future studies. For that, we have critically assessed statistically significant candidate-gene associations with susceptibility or outcome of all-cause ARDS from 2008 to 2012 using 14 major quality control criteria, and compared the updated results with our previous evaluation (Flores et al., 2008).

MATERIALS AND METHODS

LITERATURE SEARCH

We have previously assessed the quality of genetic association studies supporting susceptibility and/or outcome in adult ARDS patients of the period of 1996–2008 (Flores et al., 2008). We have now conducted a PubMed search from January 2008 to September 2012 by utilizing the same keyword combinations for querying ("polymorphism" and "acute lung injury," "polymorphism" and "ARDS," and "polymorphism" and "acute respiratory distress syndrome"). Because of the plausibility that a fraction of risk variants for ARDS susceptibility could be also risk factors for outcomes, both possibilities were jointly analyzed. The retrieved references were then manually reviewed. Excluding meta-analysis, those reporting statistically significant associations in adults ($p \leq 0.05$) for any cause of ALI or ARDS irrespective of the type of genetic variants associated, and published in English, were reviewed by three of the authors. We are aware that a number of such reported associations might be false positives. However, this threshold for significance is preferable over a more conservative strategy at this stage of field development (Thomas and Clayton, 2004). Finally, we considered the gene as the unit of replication (Neale and Sham, 2004).

STUDY ASSESSMENT

For simplicity, we focused on the 14 most relevant criteria, previously utilized by us in Flores et al. (2008), modifying the exhaustive list provided by Chanock et al. (2007), scoring each item as present or absent. Chi-squared tests were performed in SPSS (SPSS Inc., Chicago, IL).

GENE COVERAGE IN GENOTYPING ARRAYS

Gene coverage was calculated with the tagger tool (Barrett et al., 2005) for SNPs with minor allele frequency >5% in the gene region captured directly and indirectly by the genome-wide genotyping array utilized (with a multi-marker $r^2 \geq 0.8$).

RESULTS

The PubMed search on the period 2008–2012 allowed a closer review of 27 original articles reporting statistically significant association findings on 31 candidate genes with susceptibility and/or outcomes of all-cause ARDS (Table S1), and the first genome wide association study (GWAS) for this syndrome (Christie et al., 2012). The latter was excluded from the evaluation as its quality control assessment differs substantially from those applied to candidate-gene studies. A complementary search querying for the syndrome name in the HuGeNet Navigator (Yu et al., 2008) gave overlapping results, showing studies for additional genes albeit all reporting statistically non-significant findings. We, therefore, continued the quality assessment based on the PubMed search.

Seventeen studies (63%) provided statistically significant findings with a case-control design and ten (37%) with a cohort. These were based on a median sample size of 251 cases [interquartile range (P_{25}–P_{75}): 84–365] and 288 controls (P_{25}–P_{75}: 190–724) in case-control studies, whereas for cohort studies the median sample size was 145 patients (P_{25}–P_{75}: 118–215).

In this period, almost all studies (96%) appropriately described demographical and clinical data for cases and all had an adequate characterization of the control group (47.1% of them utilized healthy subjects or population-based controls and 52.9% opted to use at risk patients as controls). However, only 50% of the studies explored their power to detect statistically significant findings.

While roughly a third of studies (35%) focused on a single variant of the gene under study, the majority (65%) analyzed several polymorphisms attaining appropriate gene coverage of common variation by means of linkage disequilibrium (LD)-based methods. In most cases (74%), the studies allowed to unambiguously identify the genomic location of the associated variant(s) on public resources. Similarly, most studies declared that Hardy-Weinberg equilibrium expectations were assessed (93%), and that further genotyping error checks were implemented during the study (59%). Almost half of the studies (48%) stated that genotyping was performed blind to the disease status of samples.
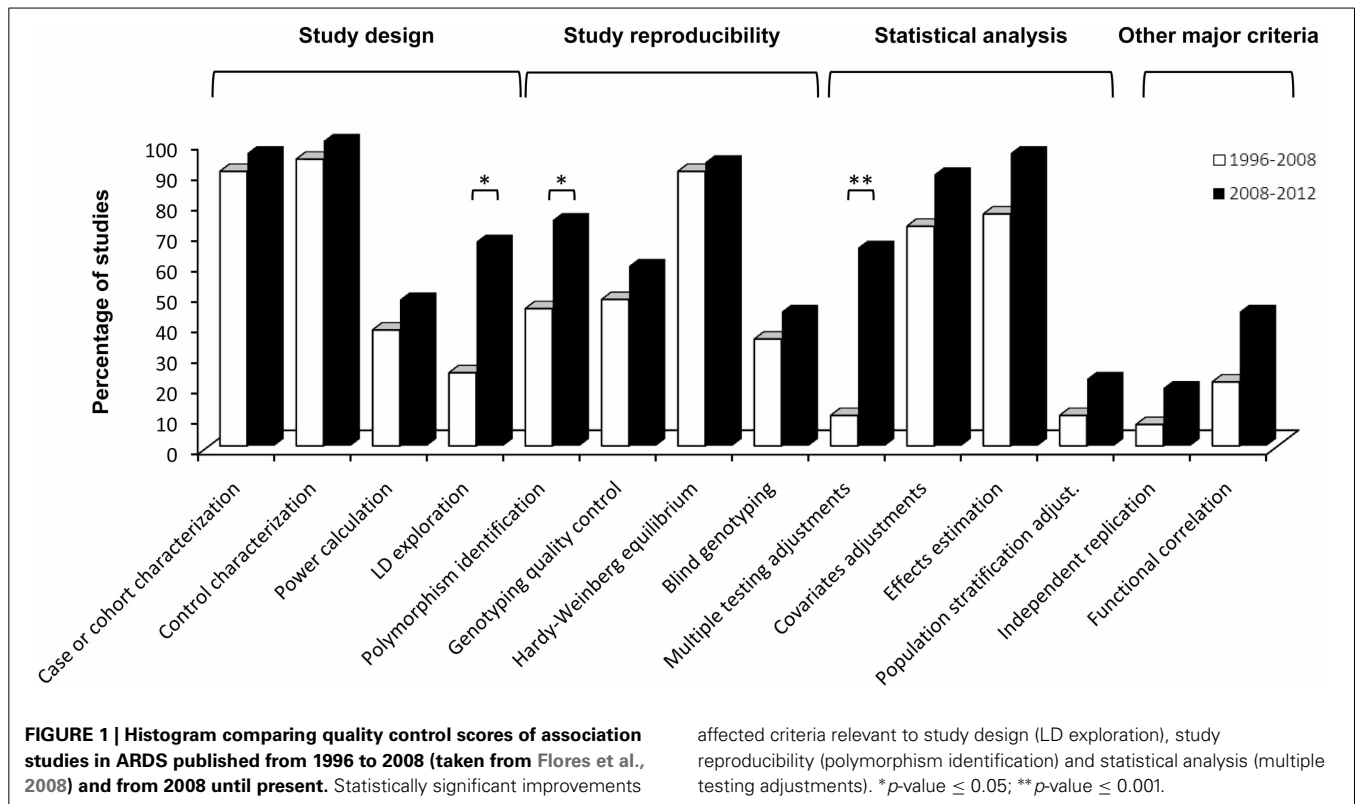
Focusing on the statistical analyses, 65% of the studies that needed to control type-I error due to multiple hypothesis testing did so, and 89% included covariates in the regression analyses. The magnitude of effects was appropriately reported in terms of hazard ratios (HRs) or odds ratios (ORs) in almost all reviewed studies (96%) (Table S1). The adjustment for population stratification and replication, in at least an independent study sample, were declared only in 22 and 19% of the studies, respectively, two major issues that has not improved over the years (Flores et al., 2008) (**Figure 1**). Similarly, almost half of the studies (44%) pursued the functional significance of associated variants.

On a side-by-side comparison of the two periods reviewed to date (i.e., 1996–2008 reviewed by Flores et al., 2008 and this one from 2008 to 2012), significant improvements in the quality of the published studies were observed in the most recent period (**Figure 1**) affecting study design, study reproducibility, and statistical analysis. These improvements were due to an increase of studies exploiting the available tools for LD exploration to efficiently select the genetic variants (from 24 to 67%, chi-squared $p = 0.003$); controlling type-I error by incorporating multiple testing adjustments on the analyses (from 10 to 65%, chi-squared $p = 0.0003$); and accurately identifying the genomic location of the associated variant(s) (from 45 to 74%, chi-squared $p = 0.033$).

DISCUSSION

We have assessed the evidence obtained during 2008–2012 from ARDS candidate-gene association studies and compared them with our previous assessment to objectively evaluate the evolution of the field, especially in light of the methodology applied in genetic susceptibility studies. In total, including the evidence accumulated before 2008 (Flores et al., 2008), 56 studies on 41 candidate genes reported statistically significant associations with susceptibility or outcomes of all-cause ARDS (**Figure 2**).

We detected significant improvements affecting the exploitation of resources for LD exploration, the inclusion of multiple testing adjustments, and the way studies identified the associated variants by established recommendations. This was also extensible to sample sizes for case-control designs, as these have roughly

FIGURE 1 | Histogram comparing quality control scores of association studies in ARDS published from 1996 to 2008 (taken from Flores et al., 2008) and from 2008 until present. Statistically significant improvements affected criteria relevant to study design (LD exploration), study reproducibility (polymorphism identification) and statistical analysis (multiple testing adjustments). *p-value ≤ 0.05; **p-value ≤ 0.001.

doubled their median sample by group compared to studies published before 2008. Despite this improvement, replications in independent studies are needed to improve the association reliability. Worth noting, the diversity of samples has increased over the years, so that across all published studies a few have focused on African-Americans (6.6%), while the majority continues to use Europeans (66.7%), East Asians (15%), or multiethnic samples (11.7%). While all these improvements are stimulating, a downside continues to be recognized on the adjustment for population stratification and replication attempts, as these were conducted in less than a fifth of all reviewed reports.

The identification of genuine gene associations with ARDS relies on conducting more replication studies, albeit without sacrificing study robustness, as only a few associated genes have been replicated to date (**Figure 2**). Among those genes, *ACE* was associated several times and a meta-analysis was recently published (Matsuda et al., 2012). Although results should be taken with caution because of power limitations, they revealed variable effects of an *ACE* polymorphism with ARDS mortality, present in East Asians but lacking in Europeans. This illustrates the growing evidence supporting that genetic risks may be population-specific, either because of gene-gene or gene-environment interactions or because of frequency effects (Need and Goldstein, 2009). Given that we are far from having a complete list of ARDS genes, and that an incomplete overlap of genetic risks between populations is expected, the study of samples of diverse ancestry should be encouraged in future studies. It must be noted that across all reviewed studies, genetic associations with ARDS susceptibility or outcomes with opposite effects in different ancestry groups

were absent, despite differences by the ARDS triggering insult have been detected (Christie et al., 2008). One major issue that is determinant of the robustness of association studies with unrelated individuals is the assessment and adjustment of results for the underlying (sometimes cryptic) population stratification, which is usually based on data from independent genetic polymorphisms (Price et al., 2006). Still today, more than 80% of the published association studies in ARDS did not apply such an approach, despite few dozen of very informative genetic variants (termed AIMs) have demonstrated their utility in specific populations (Pino-Yanes et al., 2011; Galanter et al., 2012). As the studies that focus on particular genomic regions will continue to be relevant in the field (Chanock et al., 2007), population stratification effects should be minimized in future association studies, irrespective of the study population being assessed. Therefore, it becomes essential to develop efficient and straightforward methods that: (1) could be applied to different populations and be universally used, and (2) could assist researchers to easily select a reduced set of AIMs to accurately assess ancestry maintaining affordable costs. Such tools would be useful to validate study robustness as well as to address the biological differences between populations, and whether these may trigger disparities in ARDS susceptibility or outcomes. It must be noted; however, that population stratification also introduces non-genetic effects that will not be addressed by these methods. It is expected that analyses of these effects and interactions will bring new opportunities and challenges in the field (Rotimi and Jorde, 2010).

Establishing the association of genes with ARDS susceptibility is only the beginning of a process that should continue with
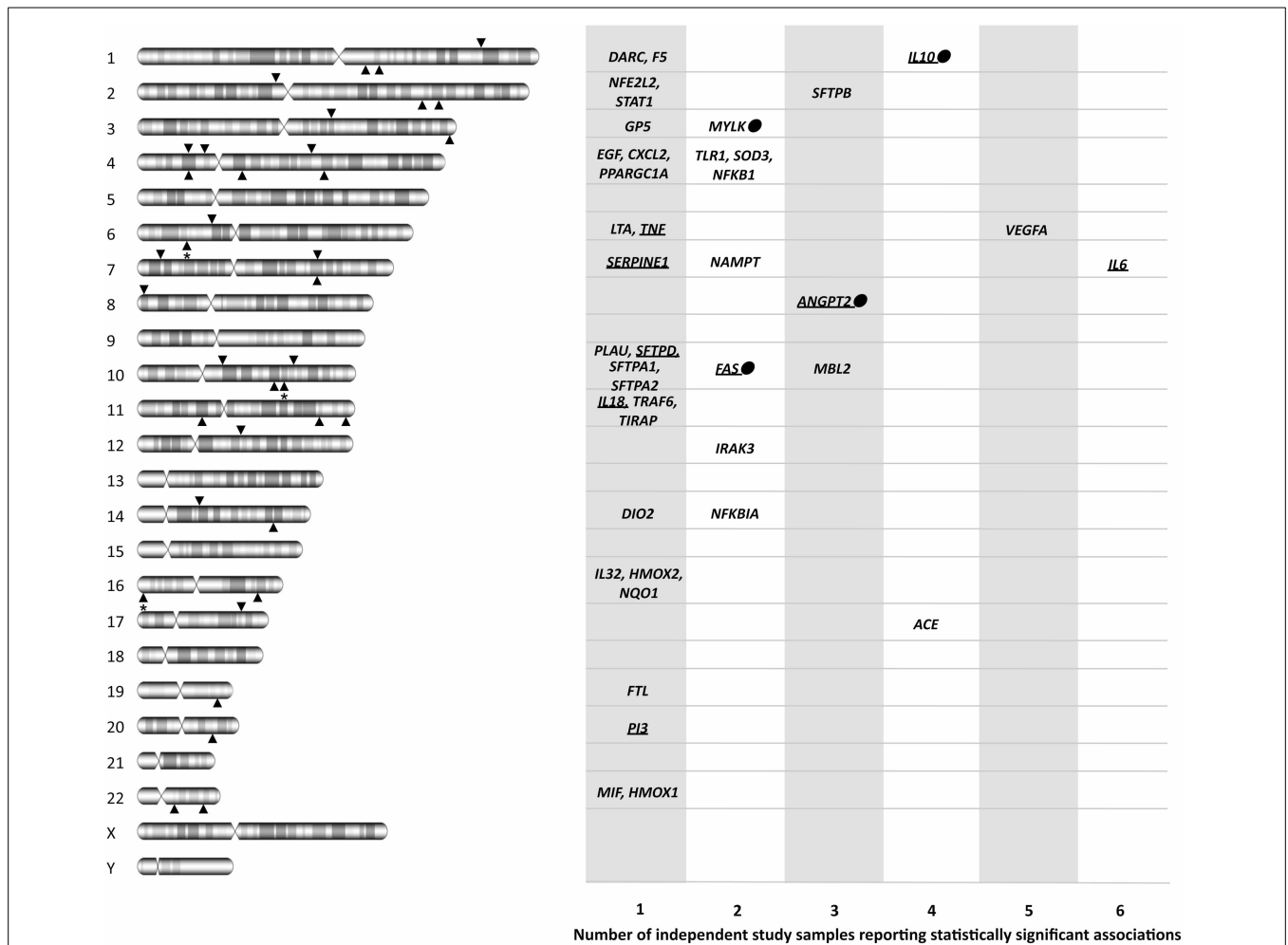
FIGURE 2 | Diagram showing the official gene symbols for the 41 candidate genes associated with ARDS susceptibility and outcomes, depicting both chromosome locations and the number of study samples with statistically significant associations. For each chromosome, lower arrowheads indicate the location of genes with a single sample association, and upper arrowheads indicate the location of genes with statistically significant association findings in at least two study samples. Arrowheads with asterisk indicate more than one gene in that region. Dots denote that the gene was replicated in the only GWAS of ARDS published to date. Underlined gene names indicate that the product has been suggested as a biomarker for ARDS or its progression in at least one study. *ACE*, angiotensin I converting enzyme; *ANGPT2*, angiopoietin 2; *CXCL2*, chemokine (C-X-C motif) ligand 2; *DARC*, duffy blood group, chemokine receptor; *DIO2*, deiodinase, iodothyronine, type II; *EGF*, epidermal growth factor; *F5*, coagulation factor V (proaccelerin, labile factor); *FAS*, TNF receptor superfamily, member 6; *FTL*, ferritin, light polypeptide; *GP5*, glycoprotein V (platelet); *HMOX1*, heme oxygenase 1; *HMOX2*, heme oxygenase 2; *IL6*,

interleukin 6; *IL10*, interleukin 10; *IL18*, interleukin 18; *IL32*, interleukin 32; *IRAK3*, interleukin-1 receptor-associated kinase 3; *LTA*, lymphotoxin alpha; *MBL2*, mannose-binding lectin 2; *MIF*, macrophage migration inhibitory factor; *MYLK*, myosin light chain kinase; *NAMPT*, nicotinamide phosphoribosyltransferase; *NFE2L2*, nuclear factor (erythroid-derived 2)-like 2; *NFKB1*, nuclear factor of kappa light polypeptide gene enhancer in B-cells 1; *NFKBIA*, nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor alpha; *NQO1*, NAD(P)H dehydrogenase, quinone 1; *PI3*, peptidase inhibitor 3; *PLAU*, plasminogen activator, urokinase; *PPARGC1A*, peroxisome proliferator-activated receptor gamma, coactivator 1 alpha; *SFTPA1*, surfactant protein A1; *SERPINE1*, serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1; *SFTPA2*, surfactant protein A2; *SFTPB*, surfactant protein B; *SFTPD*, surfactant protein D; *SOD3*, superoxide dismutase 3; *STAT1*, signal transducer and activator of transcription 1, 91kDa; *TIRAP*, toll-interleukin 1 receptor (TIR) domain containing adaptor protein; *TLR1*, toll-like receptor 1; *TNF*, tumor necrosis factor; *TRAF6*, TNF receptor-associated factor 6; *VEGFA*, vascular endothelial growth factor A.

the discovery of the causal genetic variants. The challenge continues to be the validation of existing and novel ARDS associations via robust studies, and future and ongoing studies should amend the critical issues here recognized. In this effort, new technologies are allowing a faster field development by means of genome-wide studies, either using genotyping arrays or exome/whole genome sequencing. GWAS are as efficient as candidate-gene studies for

detecting weak effect risks, not requiring a previous hypothesis of the biological processes related to the trait. They have allowed to identify new disease genes never anticipated and led to new hypothesis and perspectives about disease pathogenesis (Marchini and Howie, 2010). Despite that, GWAS have major limitations including high costs, usually impacting on the sample size, the statistical burden and the gene coverage. In addition,

most commercial platforms may offer less coverage for the gene(s) of interest compared to that achieved in optimal candidate-gene studies, which can substantially impact study power (Voight et al., 2012). The first GWAS of ARDS was recently published by Christie et al. (2012), revealing *PPFIA1* as a novel susceptibility gene involved in cell adhesion and cell-matrix interactions, and suggesting many others with putative functional roles. This study also replicated the association of four candidate genes including *IL10*, *MYLK*, *ANGPT2*, and *FAS*. This may suggest that all other candidate gene associations should be considered false discoveries. However, one explanation for this inconsistency could be also the insufficient GWAS coverage of the non-associated candidate genes (average ≈57%; Table S2). Whatever the case, commercial platforms will only allow studying a fraction of the millions of existing genetic variants (Abecasis et al., 2012), and it is anticipated that the associations to be revealed will only explain a small component of the disease (Manolio et al., 2009). Only complete re-sequencing of individual genomes will guarantee the analysis of all genetic variation.

Here we have shown that the field still faces several methodological challenges, and in the clinical arena there are key issues to be improved in order to fully understand the genetic processes underlying ARDS. Misclassification of phenotypes can lead to significant reduction in statistical power to detect true genetic associations, therefore it becomes necessary a better and more homogeneous patient classification. This could be achieved by combining the clinical information with different integrative approaches, those based on the determination of the causal microorganisms by means of metagenomics (Lysholm et al., 2012) or performing gene expression profiling among patients (Hu et al., 2012), to name a few. As a proof of concept, in a recent study by O'Mahony et al. (2012), only when the samples were restricted to the more severe phenotype, new associations were revealed and previous findings were replicated. Furthermore, quantitative phenotypes could be utilized for association testing, such as ventilator-free days (Kangelaris et al., 2012) or ideally other traits that are closer to the genotype. This possibility has been explored in the field with striking (Wurfel et al., 2008) and replicable results (Pino-Yanes et al., 2010). Additionally, the selection of the control samples remains a challenge; it is not an easy task and not a single design is free of bias. The use of either healthy subjects or at-risk individuals is common among the reviewed studies. An alternative solution can be the utilization of both types of controls to reduce selection biases and be able to confidentially assess the quality of the genotypic data. This strategy has been used (Song et al., 2010), and will surely reduce the chances that risk variants reported are causally associated with a confounder and not with ARDS.

In summary, the methodology for assessing genetic risks in complex diseases is under development. For ARDS, we conclude that the main challenge continues to be in providing an analytically rigorous methodology (adjusting for population stratification, relatedness, and technical quality) accompanied by independent replication and mechanistic explanations for the results provided. Still today, the evidence supporting the genetic associations with ARDS susceptibility or outcomes is at best uncertain, given the limited statistical power of most studies and the effects expected for genetic variants involved in complex traits. To guarantee proper and high quality studies on genetic susceptibility and outcomes, we strongly encourage the use of large and well-defined collection of samples. Consequently, a shift toward the establishment of international consortia will be necessary.

AUTHOR CONTRIBUTIONS

All authors contributed equally in the assessment design and read and approved the final manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fgene.2014.00020/abstract

REFERENCES

Abecasis, G. R., Auton, A., Brooks, L. D., Depristo, M. A., Durbin, R. M., Handsaker, R. E., et al. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65. doi: 10.1038/nature11632

Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265. doi: 10.1093/bioinformatics/bth457

Bernard, G. R., Artigas, A., Brigham, K. L., Carlet, J., Falke, K., Hudson, L., et al. (1994). The american-european consensus conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination. *Am. J. Respir. Crit. Care Med.* 149, 818–824. doi: 10.1164/ajrccm.149.3.7509706

Chanock, S. J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D. J., Thomas, G., et al. (2007). Replicating genotype-phenotype associations. *Nature* 447, 655–660. doi: 10.1038/447655a

Christie, J. D., Ma, S. F., Aplenc, R., Li, M., Lanken, P. N., Shah, C. V., et al. (2008). Variation in the myosin light chain kinase gene is associated with development of acute lung injury after major trauma. *Crit. Care Med.* 36, 2794–2800. doi: 10.1097/CCM.0b013e318186b843

Christie, J. D., Wurfel, M. M., Feng, R., O'Keefe, G. E., Bradfield, J., Ware, L. B., et al. (2012). Genome wide association identifies PPFIA1 as a candidate gene for acute lung injury risk following major trauma. *PLoS ONE* 7:e28268. doi: 10.1371/journal.pone.0028268

Cobb, J. P., and O'Keefe, G. E. (2004). Injury research in the genomic era. *Lancet* 363, 2076–2083. doi: 10.1016/S0140-6736(04)16460-X

Erickson, S. E., Shlipak, M. G., Martin, G. S., Wheeler, A. P., Ancukiewicz, M., Matthay, M. A., et al. (2009). Racial and ethnic disparities in mortality from acute lung injury. *Crit. Care Med.* 37, 1–6. doi: 10.1097/CCM.0b013e31819292ea

Flores, C., Pino-Yanes Mdel, M., and Villar, J. (2008). A quality assessment of genetic association studies supporting susceptibility and outcome in acute lung injury. *Crit. Care* 12, R130. doi: 10.1186/cc7098

Galanter, J. M., Fernandez-Lopez, J. C., Gignoux, C. R., Barnholtz-Sloan, J., Fernandez-Rozadilla, C., Via, M., et al. (2012). Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. *PLoS Genet.* 8:e1002554. doi: 10.1371/journal.pgen.1002554

Hu, P., Wang, X., Haitsma, J. J., Furmli, S., Masoom, H., Liu, M., et al. (2012). Microarray meta-analysis identifies acute lung injury biomarkers in donor lungs that predict development of primary graft failure in recipients. *PLoS ONE* 7:e45506. doi: 10.1371/journal.pone.0045506

Janssens, A. C., Ioannidis, J. P., Van Duijn, C. M., Little, J., and Khoury, M. J. (2011). Strengthening the reporting of genetic risk prediction studies: the GRIPS statement. *Eur. J. Epidemiol.* 26, 255–259. doi: 10.1007/s10654-011-9552-y

Kangelaris, K. N., Sapru, A., Calfee, C. S., Liu, K. D., Pawlikowska, L., Witte, J. S., et al. (2012). The association between a Darc gene polymorphism and clinical outcomes in African American patients with acute lung injury. *Chest* 141, 1160–1169. doi: 10.1378/chest.11-1766

Khoury, M. J., and Yang, Q. (1998). The future of genetic studies of complex human diseases: an epidemiologic perspective. *Epidemiology* 9, 350–354. doi: 10.1097/00001648-199805000-00023

Linko, R., Okkonen, M., Pettila, V., Perttila, J., Parviainen, I., Ruokonen, E., et al. (2009). Acute respiratory failure in intensive care units. FINNALI: a prospective cohort study. *Intensive Care Med.* 35, 1352–1361. doi: 10.1007/s00134-009-1519-z

Lysholm, F., Wetterbom, A., Lindau, C., Darban, H., Bjerkner, A., Fahlander, K., et al. (2012). Characterization of the viral microbiome in patients with severe lower respiratory tract infections, using metagenomic sequencing. *PLoS ONE* 7:e30875. doi: 10.1371/journal.pone.0030875

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., et al. (2009). Finding the missing heritability of complex diseases. *Nature* 461, 747–753. doi: 10.1038/nature08494

Marchini, J., and Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* 11, 499–511. doi: 10.1038/nrg2796

Martin, G. S., Mannino, D. M., Eaton, S., and Moss, M. (2003). The epidemiology of sepsis in the United States from 1979 through 2000. *N. Engl. J. Med.* 348, 1546–1554. doi: 10.1056/NEJMoa022139

Matsuda, A., Kishi, T., Jacob, A., Aziz, M., and Wang, P. (2012). Association between insertion/deletion polymorphism in angiotensin-converting enzyme gene and acute lung injury/acute respiratory distress syndrome: a meta-analysis. *BMC Med. Genet.* 13:76. doi: 10.1186/1471-2350-13-76

Neale, B. M., and Sham, P. C. (2004). The future of association studies: gene-based analysis and replication. *Am. J. Hum. Genet.* 75, 353–362. doi: 10.1086/423901

Need, A. C., and Goldstein, D. B. (2009). Next generation disparities in human genomics: concerns and remedies. *Trends Genet.* 25, 489–494. doi: 10.1016/j.tig.2009.09.012

O'Mahony, D. S., Glavan, B. J., Holden, T. D., Fong, C., Black, R. A., Rona, G., et al. (2012). Inflammation and immune-related candidate gene associations with acute lung injury susceptibility and severity: a validation study. *PLoS ONE* 7:e51104. doi: 10.1371/journal.pone.0051104

Pino-Yanes, M., Corrales, A., Basaldua, S., Hernandez, A., Guerra, L., Villar, J., et al. (2011). North African influences and potential bias in case-control association studies in the Spanish population. *PLoS ONE* 6:e18389. doi: 10.1371/journal.pone.0018389

Pino-Yanes, M., Corrales, A., Casula, M., Blanco, J., Muriel, A., Espinosa, E., et al. (2010). Common variants of TLR1 associate with organ dysfunction and sustained pro-inflammatory responses during sepsis. *PLoS ONE* 5:e13759. doi: 10.1371/journal.pone.0013759

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi: 10.1038/ng1847

Rahim, N. G., Harismendy, O., Topol, E. J., and Frazer, K. A. (2008). Genetic determinants of phenotypic diversity in humans. *Genome Biol.* 9:215. doi: 10.1186/gb-2008-9-4-215

Ranieri, V. M., Rubenfeld, G. D., Thompson, B. T., Ferguson, N. D., Caldwell, E., Fan, E., et al. (2012). Acute respiratory distress syndrome: the Berlin Definition. *JAMA* 307, 2526–2533. doi: 10.1001/jama.2012.5669

Rotimi, C. N., and Jorde, L. B. (2010). Ancestry and disease in the age of genomic medicine. *N. Engl. J. Med.* 363, 1551–1558. doi: 10.1056/NEJMra0911564

Rubenfeld, G. D., Caldwell, E., Peabody, E., Weaver, J., Martin, D. P., Neff, M., et al. (2005). Incidence and outcomes of acute lung injury. *N. Engl. J. Med.* 353, 1685–1693. doi: 10.1056/NEJMoa050333

Song, Z., Tong, C., Sun, Z., Shen, Y., Yao, C., Jiang, J., et al. (2010). Genetic variants in the TIRAP gene are associated with increased risk of sepsis-associated acute lung injury. *BMC Med. Genet.* 11:168. doi: 10.1186/1471-2350-11-168

Thomas, D. C., and Clayton, D. G. (2004). Betting odds and genetic associations. *J. Natl. Cancer Inst.* 96, 421–423. doi: 10.1093/jnci/djh094

Villar, J., Blanco, J., Anon, J. M., Santos-Bouza, A., Blanch, L., Ambros, A., et al. (2011). The ALIEN study: incidence and outcome of acute respiratory distress syndrome in the era of lung protective ventilation. *Intensive Care Med.* 37, 1932–1941. doi: 10.1007/s00134-011-2380-4

Villar, J., Maca-Meyer, N., Perez-Mendez, L., and Flores, C. (2004). Bench-to-bedside review: understanding genetic predisposition to sepsis. *Crit. Care* 8, 180–189. doi: 10.1186/cc2863

Villar, J., Perez-Mendez, L., Blanco, J., Anon, J. M., Blanch, L., Belda, J., et al. (2013). A universal definition of ARDS: the PaO2/FiO 2 ratio under a standard ventilatory setting-a prospective, multicenter validation study. *Intensive Care Med.* 39, 583–592. doi: 10.1007/s00134-012-2803-x

Voight, B. F., Kang, H. M., Ding, J., Palmer, C. D., Sidore, C., Chines, P. S., et al. (2012). The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.* 8:e1002793. doi: 10.1371/journal.pgen.1002793

Wurfel, M. M., Gordon, A. C., Holden, T. D., Radella, F., Strout, J., Kajikawa, O., et al. (2008). Toll-like receptor 1 polymorphisms affect innate immune responses and outcomes in sepsis. *Am. J. Respir. Crit. Care Med.* 178, 710–720. doi: 10.1164/rccm.200803-462OC

Yu, W., Gwinn, M., Clyne, M., Yesupriya, A., and Khoury, M. J. (2008). A navigator for human genome epidemiology. *Nat. Genet.* 40, 124–125. doi: 10.1038/ng0208-124

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.