

Rare genetic variant risks in patients with sepsis-associated acute respiratory distress syndrome

Received: 15 September 2025

Accepted: 16 February 2026

Published online: 25 February 2026

Cite this article as: Tosco-Herrera E., Rubio-Rodríguez L.A., Muñoz-Barrera A. *et al.* Rare genetic variant risks in patients with sepsis-associated acute respiratory distress syndrome. *Respir Res* (2026). <https://doi.org/10.1186/s12931-026-03588-4>

Eva Tosco-Herrera, Luis A. Rubio-Rodríguez, Adrián Muñoz-Barrera, David Jáspez, Eva Suárez-Pajes, Almudena Corrales, Aitana Alonso-González, Miryam Prieto-González, Aurelio Rodríguez-Pérez, Demetrio Carriedo, Jesús Blanco, Alfonso Ambrós, Leonardo Lorente, María M. Martín, Jordi Solé-Violán, Carlos Rodríguez-Gallego, Elena González-Higueras, Elena Espinosa, Arturo Muriel-Bombin, David Domínguez, Marina Soro, Tamara Hernández-Beeftink, José M. Añón, Jesús Villar, Beatriz Guillén-Guio, Itahisa Marcelino-Rodríguez, José M. Lorenzo-Salazar, Rafaela González-Montelongo & Carlos Flores

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Rare genetic variant risks in patients with sepsis-associated acute respiratory distress syndrome

Eva Tosco-Herrera¹, Luis A. Rubio-Rodríguez², Adrián Muñoz-Barrera², David Jáspez², Eva Suárez-Pajes¹,
 Almudena Corrales^{1,3}, Aitana Alonso-González¹, Miryam Prieto-González⁴, Aurelio Rodríguez-Pérez^{5,6},
 Demetrio Carriedo⁷, Jesús Blanco^{3,8}, Alfonso Ambrós⁹, Leonardo Lorente¹⁰, María M. Martín¹¹, Jordi Solé-
 Violán^{3,12,13}, Carlos Rodríguez-Gallego^{3,6,12,14}, Elena González-Higueras¹⁵, Elena Espinosa¹⁶, Arturo Muriel-
 Bombin⁸, David Domínguez¹⁶, Marina Soro¹⁷, Tamara Hernández-Beeftink^{18,19}, José M. Añón^{3,20}, Jesús
 Villar^{3,21,22,23}, Beatriz Guillén-Guio^{3,18,19}, Itahisa Marcelino-Rodríguez^{24,25}, José M. Lorenzo-Salazar², Rafaela
 González-Montelongo², Carlos Flores^{1,2,3,12,25}

1 Research Unit, Hospital Universitario Nuestra Señora de Candelaria, Instituto de Investigación Sanitaria de Canarias (IISC), Santa Cruz de Tenerife, Spain.

2 Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife, Spain.

3 CIBER de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III, Madrid, Spain.

4 Intensive Care Unit, Complejo Asistencial Universitario de Palencia, Palencia, Spain.

5 Department of Anesthesiology, Hospital Universitario de Gran Canaria Dr. Negrín, Las Palmas de Gran Canaria, Spain.

6 Department of Medical and Surgical Sciences, School of Medicine, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain.

7 Intensive Care Unit, Complejo Hospitalario Universitario de León, León, Spain.

8 Intensive Care Unit, Hospital Universitario Río Hortega, Valladolid, Spain.

9 Intensive Care Unit, Hospital General de Ciudad Real, Ciudad Real, Spain.

10 Intensive Care Unit, Hospital Universitario de Canarias, Santa Cruz de Tenerife, Spain.

11 Intensive Care Unit, Hospital Universitario Nuestra Señora de Candelaria, Santa Cruz de Tenerife, Spain.

12 Department of Clinical Sciences, University Fernando Pessoa Canarias, Las Palmas de Gran Canaria, Spain.

13 Critical Care Unit, Hospital Universitario de Gran Canaria Dr. Negrín, Las Palmas de Gran Canaria, Spain.

14 Department of Immunology, Hospital Universitario de Gran Canaria Dr. Negrín, Las Palmas de Gran Canaria, Spain.

15 Intensive Care Unit, Hospital Virgen de la Luz, Cuenca, Spain.

16 Department of Anesthesiology, Hospital Universitario Nuestra Señora de Candelaria, Santa Cruz de Tenerife, Spain.

17 Anesthesiology and Critical Care Department, Hospital IMED Valencia, Valencia, Spain.

18 Department of Population Health Sciences, University of Leicester, Leicester, United Kingdom.

19 NIHR Leicester Biomedical Research Centre, Leicester, United Kingdom.

20 Intensive Care Unit, Hospital Universitario La Paz, IdiPAZ, Madrid, Spain.

21 Research Unit, Hospital Universitario Dr. Negrín, Las Palmas de Gran Canaria, Spain.

22 Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Canada.

23 Faculty of Health Sciences, Universidad del Atlántico Norte, Las Palmas de Gran Canaria, Spain.

24 Public Health and Preventive Medicine Area, University of La Laguna, Santa Cruz de Tenerife, Spain.

25 Instituto de Tecnologías Biomédicas (ITB), University of La Laguna, Santa Cruz de Tenerife, Spain.

Correspondence: Dr. Carlos Flores, Email: cflores@ull.edu.es

Abstract

Background

Acute respiratory distress syndrome (ARDS) is a complex, heterogeneous, and deadly condition often resulting from pulmonary lesions due to sepsis, among other causes. There is a lack of targeted therapies to specifically treat the patients. Common genetic factors in the population (frequency >1%) have been associated with ARDS susceptibility, but systematic genetic screens of the role of rare genetic variants are lacking. We used the network of known molecular interactions to identify ARDS risks from clusters of biologically related genes containing qualifying variants (QVs) with frequency <1% likely affecting function.

Methods

We conducted whole-exome sequencing in sepsis patients from the GEN-SEP cohort (n=822, of which 272 developed ARDS). A network-based heterogeneity clustering algorithm was used to discover significant gene clusters ($p < 1 \times 10^{-5}$). Gene-set enrichment analysis and logistic regression models aggregating QVs were used for cross-verification to confirm consistency and deepen understanding of the effect sizes of gene clusters.

Results

We identified 19 significant clusters ($p_{\text{lowest}} = 3.29 \times 10^{-10}$), each containing an average of 102 genes (11.6% mean similarity). QVs in nine gene clusters were associated with sepsis-associated ARDS ($p_{\text{lowest}} = 1 \times 10^{-5}$) but were not associated with 28-day survival. Clusters were enriched in several biological pathways, notably the *Toll-like receptor cascades*.

Conclusions

These results support a marked genetic heterogeneity underlying ARDS susceptibility and the presence of rare risk variants involving multiple biological processes that are associated with sepsis outcomes. Particularly, they underscore the importance of rare variants in genes of the *Toll-like receptor cascades* in the risk for sepsis-associated ARDS.

Introduction

Acute respiratory distress syndrome (ARDS) is a heterogeneous and complex lung condition that develops after pulmonary or systemic insults, including trauma, severe pneumonia, and sepsis [1]. The diagnosis is based on acute onset, bilateral opacities on chest radiography or CT scan (of non-cardiac origin), and a PaO₂/FiO₂ ratio of less than 300 mm Hg according to the latest clinical guidelines [2–4]. ARDS affects almost 25% of all critically ill adult patients requiring mechanical ventilation. The overall in-hospital mortality rate exceeds 40% [5,6], while it increased in periods of the coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [7]. ARDS imposes a significant economic burden on adult intensive care unit (ICU) worldwide. However, no specific pharmacotherapy is effective for ARDS treatment [8–11]. Current patient management is based on life-supportive therapies for maintaining gas exchange and reduce inflammation.

ARDS develops from a hyperinflammatory state causing alveolar-capillary membrane dysregulation [1], leading to severe impairment of gas exchange, non-cardiogenic pulmonary edema, resulting in hypoxemia and acute respiratory failure [12]. Host conditions such as age, sex, comorbidities, and exposure to airborne pollutants are known risk factors for ARDS [13–15]. However, these factors and the clinical status poorly predict ARDS risk among patients. It is thus crucial to understand the genetic and molecular mechanisms that govern host responses to severe infections and ARDS development. This knowledge is essential for patient risk stratification and for identifying novel treatments. Host genetic factors have been increasingly recognized as affecting the severity of responses to infections. To a certain point, the genetic influence on infection severity and outcomes could be exemplified in patients with inborn errors of immunity (IEI) [16–18]. Genetic studies in COVID-19 patients show that there are both common [19,20] and rare variants affecting gene function [21,22] associated with severe SARS-CoV-2 infections and life-threatening disease.

Genome-wide screens in cohorts of ARDS patients are still scarce in the literature [23]. Genome-wide association studies (GWAS) in ARDS have revealed common genetic risk factors (>1% frequency) in genes involved in cellular adhesion [24], platelet activation [25], vascular permeability [26], and immune-inflammatory processes [27]. Sequencing-based studies evaluating the spectrum of variant frequencies (<1%) enriched for deleterious effects have focused on small cohorts of fewer than 100 patients or on familial cases affected by IEI due to genetic defects in type I interferon signaling [28–30]. The latter has also evidenced some degree of genetic overlap between IEI and ARDS [23,31].

Given that multiple biological processes are involved in ARDS, it becomes necessary to use approaches that allow aggregating genetic information from multiple pathways to capture the genetic heterogeneity. In this study, we aimed to identify rare genetic risk factors for sepsis-associated ARDS using a gene-clustering approach. We leveraged the largest exome sequencing dataset analyzed so far for this condition and the information from biological processes for adapting to patient heterogeneity.

Materials and methods

Patients and ethical considerations

The study examined a cohort of patients with sepsis from different foci and etiological agents from the GEN-SEP study. These patients were enrolled between 2002 and 2019 from a network of Spanish ICUs. Except for 323 patients analyzed for the first time in this study, the rest were included in previous GWAS [26,32]. Sepsis was clinically defined according to the Third International Consensus Definitions for Sepsis [33]. The clinical and demographic data are available in **Additional file 1: Supplementary Table 1**.

Initially, 995 patients were included in the cohort, but 32 patients were excluded for having incomplete clinical data, and 141 were excluded due to the DNA quantity or integrity requirements. Consequently, 822 patients remained for whom complete whole-exome sequencing (WES) data were considered for the analyses. From those, 272 patients met ARDS criteria [2]. These 272 patients were used as cases and 550 sepsis patients without ARDS during their hospitalization were considered as controls (**Additional file 1: Supplementary Table 1**).

The study was reviewed by The University Hospital of the Canary Islands Review Board (CHUNSC_2021-40) and the Research Ethics Committees of participating centers, and was conducted in accordance with the Declaration of Helsinki. Written informed consents were obtained from all participants or their representatives.

Library preparation, whole-exome sequencing, and variant calling

Genomic DNA was purified from peripheral blood samples using a commercial column-based DNA extraction kit (GE Healthcare, Chicago, IL). We measured DNA concentration using the Qubit 3.0 fluorimeter with the Qubit dsDNA High Sensitivity Assay kit (Thermo Fisher Scientific, Waltham, MA). Genomic libraries were prepared using either the DNA Prep with Enrichment kit (Illumina Inc., San Diego, CA) as described elsewhere [34] or the SureSelect XT HS2 DNA Reagent Kit (Agilent Technologies, Santa Clara, CA). We then assessed library sizes

on a 4200 TapeStation (Agilent Technologies) and determined their concentration using the Qubit dsDNA HS Assay kit (Thermo Fisher Scientific).

Sequencing of libraries was conducted at the Instituto Tecnológico y de Energías Renovables (ITER, Santa Cruz de Tenerife, Spain). Sequencing was performed in parallel along with 1% of a PhiX control V3 (Illumina Inc.) to an average depth of at least 100X on NextSeq 550, HiSeq 4000, or NovaSeq 6000 Sequencing Systems (Illumina Inc.) using 75 base pairs (bp) or 100 bp paired-end reads as recommended. We preprocessed the sequence reads using bcl2fastq v2.18 for demultiplexing, BWA-MEM 0.7.15 [35] for read alignment to the GRCh37/hg19 reference, SAMtools v1.3 [36] for alignment sorting and Picard v2.10.10 [37] for marking of duplicate reads as detailed elsewhere [26]. The GATK HaplotypeCaller v3.8 [38] was used following the Best Practices recommendations for variant calling of single-nucleotide variants (SNVs) and small insertions and deletions (indels<50 bp) with a padding of 100 bp.

Variant quality controls were applied using BCFtools v1.16 [39] to retain only high-confidence variants in the final dataset. Variants with a PASS, depth of coverage (DP) ≥ 10 , genotype quality (GQ) ≥ 20 , and missingness rate (FMISS) < 0.05 were included in the study. Detected variants were annotated for population allele frequency (AF) according to gnomAD Exome v2.1 [40], RefSeq-based annotation, functional consequences, and pathogenic potential based on ClinVar and the Gene Damage Index (GDI) [41]. Pathogenicity prediction scores, such as the Combined Annotation-Dependent Depletion (CADD) score v1.6, were annotated with Ensembl Variant Effect Predictor (VEP) v.105 [42] and ANNOVAR v18.04.16 [43]. The in-house pipeline is fully described online [44]. All analyses were carried out using the TeideHPC Supercomputing facility [45] (see **Additional file 1: Supplementary Methods**).

Network-based heterogeneity clustering and biological enrichment analyses

We used the Network-based heterogeneity clustering v3 (NHC) to systematically aggregate biologically proximate genes from WES data based on a predefined protein-protein interaction network [46]. We followed recommendations provided by Zhang et al., (2021) to select qualifying variants (QVs) for analysis. These QVs had an AF ≤ 0.01 in non-Finnish European populations, a CADD ≥ 10 , a GDI ≤ 10 , and that were annotated for protein truncation prediction according to VEP (non-synonymous, missense, start/stop gain/loss, frameshift, or splicing variants). Thereafter, we refer to these as QVs-CADD10. We obtained HUGO Gene Nomenclature Committee names for QVs and used them as the input for NHC on a per-sample basis. To control for genetic heterogeneity in the cohort, the first three principal components (PCs) of common variation derived from paired

SNP genotyping array data (**Additional file 1: Supplementary Methods and Supplementary Figure 1**) were also included in the NHC analysis.

Following recommendations [46], gene clusters differing between cases and controls were considered significant at $p \leq 1 \times 10^{-5}$. A similarity matrix based on gene-level overlap was calculated among significant clusters using the Jaccard distance. Gene interaction graphs within significant clusters were represented using Cytoscape v3.10.2 [47] to integrate the known functional interactions between gene products and the count of ARDS cases with QVs. Finally, we assessed biological pathway enrichment of each significant cluster of genes using REACTOME [48]. In this discovery stage, we retained enriched significant pathways ($p \leq 1 \times 10^{-5}$) and selected the most significant as the top pathway to be representative for each cluster [46].

Qualifying variant carrier associations with ARDS

To confirm consistency of associations and deepen understanding of the effect sizes contributed by gene clusters, we then used the QVs to identify carriers among the GEN-SEP patients, provided that at least one QV was identified in any gene from a particular cluster. However, since many of the QVs are expected to be non-pathogenic in this case, a stricter CADD score threshold ≥ 15 was used and thereafter referred to as QVs-CADD15. We restricted the analysis to significant gene clusters having minimal gene-level overlap. For this, we relied on a cutoff similarity $< 20\%$ (and choosing the most significant cluster when compared clusters had $> 20\%$) to have a balance in the overlap between clusters while limiting the prioritization of clusters enriched in pathways that were functionally broad. This approach avoids an overly conservative statistical penalty due to gene-level similarity between clusters. Since only 10 of the significant clusters satisfied this condition, we applied a $p < 5 \times 10^{-3}$ ($p = 0.05/10$) threshold to declare that QVs carriers of a specific cluster were significantly associated with ARDS. For those 10 significant clusters, the association of QVs carriers with sepsis-associated ARDS was tested using binomial logistic regressions adjusted for sex, age, and Acute Physiology And Chronic Health Evaluation II (APACHE II) scores as in previous GWAS [32,49] using R v4.3.2 [50]. Based on these, we also computed the Nagelkerke's pseudo R^2 approximation to estimate the genetic variance explained by the QVs in the different clusters.

A schematic overview of the analysis workflow is presented in **Additional file 1: Supplementary Figure 2**.

Sensitivity analyses

Alternative models testing the association of QVs with ARDS risk were also conducted to assess the impact of adjusting for relevant demographic and clinical variables, by adjusting for the first two main principal components of genetic heterogeneity, by lowering the AF threshold ($AF \leq 0.001$), or by relying on lenient CADD scores (QVs-CADD10) as those used in the discovery stage.

Loss-of-function variants in 15 genes of the Toll-like receptor (TLR)3- and TLR7-dependent interferon response (*IFNAR1*, *IFNAR2*, *IRF3*, *IRF7*, *IRF9*, *IKBKG*, *STAT1*, *STAT2*, *TBK1*, *TICAM1*, *TRAF3*, *UNC93B1*, *TYK2*, *TLR3*, and *TLR7*) are known to cause life-threatening COVID-19 [22]. We tested if ARDS risk could be explained by QVs-CADD15 in these genes by modelling two scenarios based on binomial logistic regressions adjusted by sex, age, and APACHE II scores: i) with $AF \leq 0.01$; and ii) with $AF \leq 0.001$.

To rule out that associations could be due to residual confounding by genetic heterogeneity, we devised a null model by testing the association of rare synonymous variants ($AF \leq 0.01$) with ARDS risk for the 10 significant clusters.

Finally, a Cox proportional hazards model was employed with a subset of 665 patients for which follow-up survival data was available (152 deaths and 513 survivors) to test associations between QVs-CADD15 carriers at those 10 gene clusters and 28-day survival using the *survival* v3.7-0 R package. We did this to rule out the possibility that the QVs carrier status was associated with mortality rather than ARDS. Two alternative survival models were used for this: i) model 1 adjusted for sex, age, and APACHE II; and ii) model 2 adjusted for sex, age, APACHE II, and ARDS status of the patient. The proportionality of Hazard Ratios (HR) assumption across covariates was verified using scaled Schoenfeld residuals. For gene clusters that did not fulfill this assumption, the *coxphw* v4.0.3 R package was applied to perform a weighted estimation of average HR and avoid the need for HR proportionality across time.

Results

NHC analysis of exomes in ARDS patients compared to sepsis controls

The NHC analysis revealed 19 gene clusters significantly different between ARDS cases and sepsis controls (p -value range= 3.29×10^{-10} to 5.68×10^{-6}) (Table 1, Figure 1). Significant clusters contained an average of 102 genes

and had a mean gene-level similarity of 11.6% with other clusters (standard deviation [SD]= 22.6%) (**Additional file 1: Supplementary Figure 3**). None of the genes identified in the previous sequencing-based studies of ARDS cohorts [51–53] were present in the significant gene clusters.

The most frequent pathways among the clusters (**Supplementary Table 2**) were *Integrin cell surface interactions* (clusters 24, 156, 165, and 186) (**Supplementary Figure 4**), *TLR cascades* (clusters 170 and 177) (**Supplementary Figure 5**), and the *Krebs cycle and the respiratory electron transport* (clusters 126 and 181) (**Supplementary Figure 6**). Less frequently enriched pathways were *Transcriptional regulation of white adipocyte differentiation* (cluster 169) (**Supplementary Figure 7**), *Interferon signaling* (cluster 153) (**Supplementary Figure 8**), and *mTOR signaling* (cluster 182) (**Supplementary Figure 9**). Additional information can be found in the **Additional file 1: Supplementary Figures 10-13**.

We identified 208 genes (or “nodes” in the networks) exhibiting QVs in ≥ 10 cases (average of 15.6 [SD=6.79] cases) in the 19 gene clusters (**Additional file 1: Supplementary Table 3**). One of those genes was *FLT1* that was linked to cluster 186 (enriched in *Integrin cell surface interactions*) (**Additional file 1: Supplementary Table 4**), which we identified in the cohort as the main association signal based on a previous GWAS [26]. This gene list included 25 (12.02%) that are considered causes of IEIs according to the 2024 classification update [54]. The top 10 genes accumulating QV carriers among ARDS cases were: *CFTR* (57 cases), *MPDZ* (56 cases), *SRRM2* (38 cases), *BRCA2* (33 cases), *NOTCH2* (32 cases), *PLCG2* (32 cases), *CGN* (32 cases), *PLXNA4* (32 cases), *NRP2* (31 cases), and *TJP2* (31 cases). Most of these genes were part of more than one significant cluster (*CFTR* in 13/19, *NOTCH2* and *CGN* in 4/19, *SRRM2*, *PLXNA4*, *BRCA2* and *NRP2* in 2/19). *MPDZ*, *PLCG2*, and *TJP2* were each part of one significant cluster (**Additional file 1: Supplementary Table 5**).

Qualifying variant carriers and sepsis-associated ARDS risk

As a cross-verification of results, we then tested the association between carrier status for the QVs-CADD15 in the 10 gene clusters having the lowest pairwise similarity with ARDS risk using regression models. Carrier status in nine of the gene clusters was associated with increased risk of sepsis-associated ARDS (**Table 2**), reflecting the broad systemic response involved in ARDS development. These results were robust to model adjustments for the first two main PCs, and for relevant demographic and clinical variables (**Supplementary Table 6**). The models based on QVs-CADD10 or on QVs-CADD15 with $AF < 0.001$ provided consistent results, although showing larger effect sizes or being non-significant, reflecting the limited sample size (**Supplementary Table 7**). Null models

with the synonymous variants for the 10 gene clusters at the same AF threshold showed no association with ARDS (**Supplementary Table 8**).

The most significant association was found for carriers of QVs in genes from cluster 170, where *TLR cascades* was the top-most enriched pathway (Odds Ratio [OR]=3.72; 95% Confidence Interval [CI]=2.05-6.71; $p=1 \times 10^{-5}$). However, the largest effect size was found for carriers of QVs in genes from cluster 12, in which *Amino acids regulate mTORC1* was the top enriched pathway (OR=6.57; 95% CI=2.75-15.68; $p=2 \times 10^{-5}$). A similar large effect was also detected for carriers of QVs in genes from cluster 186, where *Integrin cell surface interactions* was the most enriched pathway (OR=6.10; 95% CI=2.15-17.29; $p=6.6 \times 10^{-4}$). Despite that, all clusters provided consistent genetic variance estimates of the QVs in the range of 7.8%-10.9%. We warn that such estimates should be taken with caution given the potential of model overfitting. Associations were not significant for models that only considered QV carriers in the TLR3- and TLR7-dependent interferon response genes underlying life-threatening COVID-19 at $AF \leq 0.01$ (OR=0.81; 95% CI=0.54-1.22; $p=0.33$) or at $AF \leq 0.001$ (OR=0.79; 95% CI=0.47-1.35; $p=0.39$).

Finally, to rule out that QVs carrier status in these gene clusters was merely reflecting higher mortality, associations between QVs carriers and 28-day survival were tested using two alternative models. Models for all but one gene cluster (cluster 182) passed the proportionality of hazards test and none of the models were significant (**Table 3**). This evidenced that associations of carrier status with ARDS risk were not confounded by mortality.

Discussion

Studies performed over the last 30 years have revealed that a fraction of severe infectious disease by a particular microorganism can be etiologically explained by human genetic and immunological determinants in a growing proportion of patients [55]. Although most of this evidence is based on familial analyses or isolated cases, cohort studies also support this idea, and the COVID-19 pandemic has accelerated the acceptance of this view by a wider audience [17]. This is consistent with the idea that there is a strong genetic influence on the risk of fatal outcomes from infections, as shown by family, twin, and adoption studies [56,57]. Here, we show that genetic variation affecting biologically proximal gene clusters is associated ARDS risk among sepsis patients. This provides novel, actionable evidence for specific therapies targeting host dysregulation. We identified QVs with predicted impact in gene function in a cohort of unrelated sepsis patients. We used a network-based approach to identify variant

enrichment in high-confidence, biologically connected gene clusters. The most salient genetic association was observed for variants in gene clusters enriched in the *TLR cascades*. We also show that there is no overlap between the genetic risk for ARDS and patient survival, likely indicating that sepsis severity and mortality may be governed by distinct genetic factors, as proposed for COVID-19 [58,59]. To our knowledge, this is the first study reporting that rare genetic variation is associated with sepsis-associated ARDS.

Analyses accounting for aggregated variants of genes and pathways, as opposed to traditional GWAS frameworks with variant- or gene-level analyses, have shown clear power improvements. Despite the existing network data are incomplete limiting our ability to understand biological systems, these enable capturing the genetic architecture of complex traits, particularly affected by genetic heterogeneity and incomplete penetrance, and for providing biological explanations of disease risk and novel therapeutic targets [46,60]. Similar concepts underlie genetic screenings in Mendelian diseases when prioritizing the genetic etiology from WES data based on protein interaction networks combined with phenotype information [61]. Our study demonstrates that these approaches are powerful for assessing the heterogeneous genetic risks of sepsis-associated ARDS. We have identified that aggregated gene variant effects of *TLR cascades* are strongly associated. While the way the QVs are selected influenced the effect size estimates, previous studies have found significant discrepancies of effect size estimates derived from models using predicted vs. biochemically ascertained loss-of-function variants [62]. Thus, we warn that the aggregated estimated effect sizes and the associated genetic variance explained by the models in this study should be interpreted with caution and require validation in independent studies.

Common genetic variation in stress and immune response-related genes, particularly those involved in pathogen recognition and signaling [17], influence individual responses to infection and organ dysfunction [63]. TLR-mediated activation is intrinsically related to immune response to pathogens. Experimental animal models of sepsis support that *TLR cascades* are determinants of survival [64]. Myd88 or Tlr4 knock-outs, as well as neutralization of TLR-4, protects mice from lethal Gram-negative sepsis [65], suggesting that excessive inflammation mediated by *TLR cascades* contributes to sepsis aggravation. This is in line with the prevailing view that sepsis and sepsis-associated organ dysfunction are due to a dysregulated immune response dependent on different factors, including innate immune activation [66]. Variants from several of the genes participating in the *TLR cascades* have been associated with differential cytokine responses to pathogen associated molecular patterns in humans [67]. Patients with inherited deficiencies in MyD88 and IRAK-4, which are crucial for canonical TLR signaling, are particularly prone to specific invasive bacterial infections. Remarkably, clinical and laboratory signs of inflammation develop slowly in these patients, even during invasive bacterial disease [68]. Common variants

in *TLRI* have been linked to increased risk of ARDS, organ dysfunction, and TLR1-mediated production of inflammatory cytokines in healthy volunteers. Those variants have also been associated with sustained pro-inflammatory responses in sepsis patients [67,69]. One of the main etiological factors revealed so far of life-threatening viral infections is the type-I interferon response genes [22,70]. There is evidence that common and rare loss-of-function variants of genes involved in TLR-3, TLR-7, and IRF7-dependent type I interferon immunity are key components associated with severe COVID-19 [19,71], and patients with TLR3 deficiency are particularly susceptible to herpes simplex encephalitis and prone to critical influenza pneumonia [55]. In COVID-19, rare variant aggregated effects have shown very strong (OR >20) effect sizes in independent cohorts [22,62]. Despite models focusing on TLR3- and TLR7-dependent interferon response genes underlying life-threatening COVID-19 did not explain ARDS risk in our study, our results support a highly heterogeneous genetic component associated with rare variants affecting genes of the *TLR cascades* and with strong aggregated effect sizes in ARDS among non-viral sepsis patients given that most infections with pathogens identified in the cohort (>90%) were due to bacteria.

Our results also identified associations between other gene clusters and pathways that seem to play a key role in ARDS risk. However, their pathogenic links are less frequently explored in the literature and are further blurred by the pervasive genetic pleiotropy underlying complex diseases [72], the important genetic heterogeneity of ARDS, and the profound systemic nature of the underlying condition [3]. The gene cluster enriched in *Integrin cell surface interactions* is significant at pathophysiological level in the detrimental fibroproliferative response occurring in ARDS [73]. This view is supported by the shared genetic component, particularly of genes in integrin and cadherin signaling pathways, between idiopathic pulmonary fibrosis and severe COVID-19 [58]. Carriers of rare variants affecting *CFTR* gene function, who do not develop cystic fibrosis, have a higher risk of severe COVID-19 symptoms and death [74]. *CFTR* heterozygous carriers are at higher risk of respiratory infections [75,76], and a common intronic risk variant in *CFTR* is associated with respiratory failure due to infections [77]. *CFTR* was found in 13 out of the 19 significant gene clusters revealed in our study. Interestingly, it has been recognized as a possible therapeutic target in ARDS due to severe pneumonia [78] based on findings in experimental animal models [79]. Increasing evidence shows that cell-cell signaling mechanisms play an important role during inflammation [80]. On the other hand, the acute and energetic early host defense associated with ARDS is known to lead to metabolic failure, immune cell dysfunction [81], and immunosuppression [66], where mitochondrial dysfunction is central [82]. The reasons are diverse and related to the key role of mitochondria in supporting energy-dependent processes through cellular respiration, linked to the *Krebs cycle and*

respiratory electron transport, and their metabolic influence on immune response [83]. These and the other significant gene clusters revealed, which are enriched in intertwined biological processes, suggest the possibility that multiple and biologically related genetic factors are acting in combination contributing to ARDS risk.

We acknowledge several limitations in the study. First, we support the biological relevance of our findings with the published literature, but we lack data from similar sequencing studies in independent sepsis-associated ARDS patients that can validate our findings. Most published sequencing studies to date have focused on patients with COVID-19 or on kindreds with affected family members. Second, there is a limitation in how our findings generalize to other populations given that the study mainly involved individuals of European genetic ancestry. Further steps to overcome these limitations will need data from whole-exome or whole-genome sequencing studies in cohorts of patients of diverse genetic ancestries with clinically curated ARDS cases to avoid phenotypic misclassification that can dilute effect sizes as have been observed in other diseases [84]. Third, given the large heterogeneity of ARDS, we relied on a network-based approximation to identify multiple genes that are biologically connected and that could be collectively associated with the disease risk. Thus, while our analyses were unrestricted to candidate genes or inheritance models and offers an unbiased genome-wide assessment, we cannot draw conclusions at variant- or gene-level data. Connected to this, the underlying gene networks of NHC were based on incomplete protein interaction data, which could have affected our inferences. Lastly, our analyses relied on QVs that were kept in the study based on their frequency and potential protein truncation prediction. We do not provide biochemical evidence of QVs effects in gene function, which could have improved the prioritization of perturbed gene clusters and refined the most salient pathways associated with ARDS.

Conclusions

Our findings highlight the complexity of the genetic factors involved in sepsis-associated ARDS risk. These factors affect diverse biological processes, including the *TLR cascades*, that may reconcile with its complex pathophysiology. They also support an underlying landscape of genetic heterogeneity in the spectrum of rare variants influencing disease susceptibility. It is likely that some of these genetic factors could be actionable, offering a possibility to identify therapeutic targets for further research to mitigate the risk of severe outcomes among sepsis patients.

Declarations

Ethics approval and consent to participate

The study was reviewed by The University Hospital of the Canary Islands Review Board (CHUNSC_2021-40) and the Research Ethics Committees of participating centers, and was conducted in accordance with the Declaration of Helsinki. Written informed consents were obtained from all participants or their representatives.

Consent for publication

Not applicable.

Availability of data and materials

Raw genotype or phenotype data cannot be made publicly available due to restrictions imposed by the ethics approval. Variant calls from whole-exome sequencing have been deposited in EGA (EGAS50000001119). Details for accessing the data can be found at <https://github.com/genomicsITER/HGinfections>

Competing interests

The authors declare no conflict of interest.

Funding

This work was supported by Instituto de Salud Carlos III (CB06/06/1088, PI17/00610, PI20/00876, COV20_01333, COV20_01334, and PI23/00980), co-financed by the European Regional Development Funds, “A way of making Europe” from the EU; by ERA PerMed (JTC_2021) through the contract AC21_2/00039 with Instituto de Salud Carlos III and funds from Next Generation EU as part of the actions of the Recovery Mechanism and Resilience (MRR); by Agencia Estatal de Investigación (RTC-2017-6471-1); by the agreement with Instituto Tecnológico y de Energías Renovables (OA17/008 and OA23/043); by Cabildo Insular de Tenerife (CGIEU0000219140 and “Apuestas científicas del ITER para colaborar en la lucha contra la COVID-19”); by Wellcome Trust grant 221680/Z/20/Z; by Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC22/27); by Grupo DISA (OA18/017); by Fundación MAPFRE Guanarteme (OA19/072); and by Agencia Canaria de Investigación, Innovación y Sociedad de la Información de la Consejería de Economía, Conocimiento y Empleo del Gobierno de Canarias, co-funded by the European Union, Programa Operativo Integrado de

Canarias 2014-2020, Axis 3 Priority Topic 74 (85%), & European Social Fund (ESF) “Canarias Avanza con Europa” (TESIS2022010042 and TESIS2021010046). This research was partially supported by the National Institute for Health Research (NIHR) Leicester Biomedical Research Centre.

Authors' Contributions

Conceptualization, C.F.; Resources, M.P.-C., A.R.-P., D.C., J.B., A.A., L.L., M.M.M., J.S.-V., C.R.-G., E.G.-H., E.E., A.M.-B., D.D., M.S., J.M.A., J.V., C.F.; Data curation, E.T.-H., L.R.-R., A.M.-B., D.J., E.S.-P., A.C., A.A.-G., T.H.-B., B.G.-G., R.G.-M., J.M.L.-S., C.F.; Visualization, E.T.-H., D.J., C.F.; Writing—original draft preparation, E.T.-H., I.M.-R., C.F.; Writing—review and editing, all authors; Supervision, C.F.; Funding acquisition, C.F. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

We would like to acknowledge Dr. Peng Zhang for his availability and assistance with cluster network visualization and variant-to-gene mapping, integrated into the NHC v3 update. The authors would like to thank the TeideHPC Supercomputer facility team (<https://teidehpc.iter.es/en/home>) for their support, as well as to the members from the Genomics Division at the Instituto Tecnológico y de Energías Renovables (ITER) for their continuous support. E.T.-H., L.A.R.-R. and E.S.-P. acknowledge the training support provided by the University of La Laguna.

References

1. Bos LDJ, Ware LB. Acute respiratory distress syndrome: causes, pathophysiology, and phenotypes. *The Lancet* [Internet]. 2022 [cited 2024 Mar 27];400:1145–56. Available from: [http://doi.org/10.1016/S0140-6736\(22\)01485-4](http://doi.org/10.1016/S0140-6736(22)01485-4)
2. ARDS Definition Task Force. Acute Respiratory Distress Syndrome: The Berlin Definition. *JAMA* [Internet]. 2012 [cited 2020 Nov 27];307. Available from: <http://doi.org/10.1001/jama.2012.5669>
3. Grasselli G, Calfee CS, Camporota L, Poole D, Amato MBP, Antonelli M, et al. ESICM guidelines on acute respiratory distress syndrome: definition, phenotyping and respiratory support strategies. *Intensive Care Med* [Internet]. 2023 [cited 2024 Apr 14];49:727–59. Available from: <http://doi.org/10.1007/s00134-023-07050-7>
4. Matthay MA, Arabi Y, Arroliga AC, Bernard G, Bersten AD, Brochard LJ, et al. A New Global Definition of Acute Respiratory Distress Syndrome. *Am J Respir Crit Care Med* [Internet]. 2024 [cited 2024 Mar 8];209:37–47. Available from: <http://doi.org/10.1164/rccm.202303-0558WS>

5. Beitler JR, Thompson BT, Baron RM, Bastarache JA, Denlinger LC, Esserman L, et al. Advancing precision medicine for acute respiratory distress syndrome. *Lancet Respir Med* [Internet]. 2022 [cited 2023 May 10];10:107–20. Available from: [http://doi.org/10.1016/S2213-2600\(21\)00157-0](http://doi.org/10.1016/S2213-2600(21)00157-0)
6. Bellani G, Laffey JG, Pham T, Fan E, Brochard L, Esteban A, et al. Epidemiology, Patterns of Care, and Mortality for Patients With Acute Respiratory Distress Syndrome in Intensive Care Units in 50 Countries. *JAMA* [Internet]. 2016 [cited 2024 Mar 27];315:788. Available from: <http://doi.org/10.1001/jama.2016.0291>
7. Papazian L, Pauly V, Hamouda I, Daviet F, Orleans V, Forel J-M, et al. National incidence rate and related mortality for acute respiratory distress syndrome in France. *Anaesth Crit Care Pain Med* [Internet]. 2021 [cited 2024 Apr 14];40:100795. Available from: <http://doi.org/10.1016/j.accpm.2020.100795>
8. Matthay MA, Zemans RL, Zimmerman GA, Arabi YM, Beitler JR, Mercat A, et al. Acute respiratory distress syndrome. *Nat Rev Dis Primer* [Internet]. 2019 [cited 2023 May 11];5:18. Available from: <http://doi.org/10.1038/s41572-019-0069-0>
9. Xia F, Chen H, Liu Y, Huang L, Meng S, Xu J, et al. Development of genomic phenotype and immunophenotype of acute respiratory distress syndrome using autophagy and metabolism-related genes. *Front Immunol* [Internet]. 2023 [cited 2024 Apr 14];14:1209959. Available from: <http://doi.org/10.3389/fimmu.2023.1209959>
10. Xu H, Sheng S, Luo W, Xu X, Zhang Z. Acute respiratory distress syndrome heterogeneity and the septic ARDS subgroup. *Front Immunol* [Internet]. 2023 [cited 2024 Nov 29];14:1277161. Available from: <http://doi.org/10.3389/fimmu.2023.1277161>
11. Aribindi K, Lim M, Lakshminrusimha S, Albertson T. Investigational pharmacological agents for the treatment of ARDS. *Expert Opin Investig Drugs* [Internet]. 2024 [cited 2024 Apr 14];33:243–77. Available from: <http://doi.org/10.1080/13543784.2024.2315128>
12. Zayed Y, Askari R. Respiratory Distress Syndrome. *StatPearls* [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 [cited 2024 Mar 27]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK538311/>
13. Lemos-Filho LB, Mikkelsen ME, Martin GS, Dabbagh O, Adesanya A, Gentile N, et al. Sex, Race, and the Development of Acute Lung Injury. *Chest* [Internet]. 2013 [cited 2024 Apr 14];143:901–9. Available from: <http://doi.org/10.1378/chest.12-1118>
14. Keskinidou C, Vassiliou A, Dimopoulou I, Kotanidou A, Orfanos S. Mechanistic Understanding of Lung Inflammation: Recent Advances and Emerging Techniques. *J Inflamm Res* [Internet]. 2022 [cited 2024 Apr 14];Volume 15:3501–46. Available from: <http://doi.org/10.2147/JIR.S282695>
15. Zhang J, Guo Y, Mak M, Tao Z. Investigational pharmacological agents for the treatment of ARDS. *J Transl Med* [Internet]. 2024 [cited 2024 Apr 14];22:25. Available from: <http://doi.org/10.1186/s12967-023-04828-7>
16. Kwok AJ, Mentzer A, Knight JC. Host genetics and infectious disease: new tools, insights and translational opportunities. *Nat Rev Genet* [Internet]. 2021 [cited 2024 Apr 14];22:137–53. Available from: <http://doi.org/10.1038/s41576-020-00297-6>
17. Casanova J-L, Abel L. From rare disorders of immunity to common determinants of infection: Following the mechanistic thread. *Cell* [Internet]. 2022 [cited 2024 Apr 14];185:3086–103. Available from: <http://doi.org/10.1016/j.cell.2022.07.004>
18. Schmidt A, Groh AM, Frick JS, Vehreschild MJGT, Ludwig KU. Genetic predisposition and the variable course of infectious diseases. *Dtsch Arztebl Int* [Internet]. 2022 [cited 2024 Apr 14]; Available from: <http://doi.org/10.3238/arztebl.m2022.0105>
19. Pairo-Castineira E, Rawlik K, Bretherick AD, Qi T, Wu Y, Nassiri I, et al. GWAS and meta-analysis identifies 49 genetic variants underlying critical COVID-19. *Nature* [Internet]. 2023 [cited 2025 Jan 13];617:764–8. Available from: <http://doi.org/10.1038/s41586-023-06034-3>

20. The COVID-19 Host Genetics Initiative, Leadership, Kanai M, Andrews SJ, Cordioli M, Stevens C, et al. A second update on mapping the human genetic architecture of COVID-19. *Nature* [Internet]. 2023 [cited 2025 Jan 13];621:E7–26. Available from: <http://doi.org/10.1038/s41586-023-06355-3>
21. Zhang D, Povysil G, Kobeissy PH, Li Q, Wang B, Amelotte M, et al. Rare and Common Variants in *KIF15* Contribute to Genetic Risk of Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* [Internet]. 2022 [cited 2022 Nov 2];206:56–69. Available from: <http://doi.org/10.1164/rccm.202110-2439OC>
22. Matuozzo D, Talouarn E, Marchal A, Zhang P, Manry J, Seeleuthner Y, et al. Rare predicted loss-of-function variants of type I IFN immunity genes are associated with life-threatening COVID-19. *Genome Med* [Internet]. 2023 [cited 2024 Sep 19];15:22. Available from: <http://doi.org/10.1101/2022.10.22.22281221>
23. Suarez-Pajes E, Tosco-Herrera E, Ramirez-Falcon M, Gonzalez-Barbuzano S, Hernandez-Beeftink T, Guillen-Guio B, et al. Genetic Determinants of the Acute Respiratory Distress Syndrome. *J Clin Med* [Internet]. 2023 [cited 2023 Jul 7];12:3713. Available from: <http://doi.org/10.3390/jcm12113713>
24. Christie JD, Wurfel MM, Feng R, O’Keefe GE, Bradfield J, Ware LB, et al. Genome wide association identifies PPF1A1 as a candidate gene for acute lung injury risk following major trauma. *PLoS One* [Internet]. 2012 [cited 2025 Jan 13];7. Available from: <http://doi.org/10.1371/journal.pone.0028268>
25. Bime C, Pouladi N, Sammani S, Batai K, Casanova N, Zhou T, et al. Genome-Wide Association Study in African Americans with Acute Respiratory Distress Syndrome Identifies the Selectin P Ligand Gene as a Risk Factor. *Am J Respir Crit Care Med* [Internet]. 2018 [cited 2023 Jan 24];197:1421–32. Available from: <http://doi.org/10.1164/rccm.201705-0961OC>
26. Guillen-Guio B, Lorenzo-Salazar JM, Ma S-F, Hou P-C, Hernandez-Beeftink T, Corrales A, et al. Sepsis-associated acute respiratory distress syndrome in individuals of European ancestry: a genome-wide association study. *Lancet Respir Med* [Internet]. 2020 [cited 2020 Nov 27];8:258–66. Available from: [http://doi.org/10.1016/S2213-2600\(19\)30368-6](http://doi.org/10.1016/S2213-2600(19)30368-6)
27. Du M, Garcia JGN, Christie JD, Xin J, Cai G, Meyer NJ, et al. Integrative omics provide biological and clinical insights into acute respiratory distress syndrome. *Intensive Care Med* [Internet]. 2021 [cited 2023 Jan 24];47:761–71. Available from: <http://doi.org/10.1007/s00134-021-06410-5>
28. Ciancanelli MJ, Huang SXL, Luthra P, Garner H, Itan Y, Volpi S, et al. Life-threatening influenza and impaired interferon amplification in human IRF7 deficiency. *Science* [Internet]. 2015 [cited 2023 Jan 31];348:448–53. Available from: <http://doi.org/10.1126/science.aaa1578>
29. Hernandez N, Melki I, Jing H, Habib T, Huang SSY, Danielson J, et al. Life-threatening influenza pneumonitis in a child with inherited IRF9 deficiency. *J Exp Med* [Internet]. 2018 [cited 2023 Jan 31];215:2567–85. Available from: <http://doi.org/10.1084/jem.20180628>
30. Lim HK, Huang SXL, Chen J, Kerner G, Gilliaux O, Bastard P, et al. Severe influenza pneumonitis in children with inherited TLR3 deficiency. *J Exp Med* [Internet]. 2019 [cited 2023 Jan 31];216:2038–56. Available from: <http://doi.org/10.1084/jem.20181621>
31. Quintanilla E, Diwa K, Nguyen A, Vu L, Toby IT. A Data Report on the Curation and Development of a Database of Genes for Acute Respiratory Distress Syndrome. *Front Genet* [Internet]. 2021 [cited 2023 Jan 24];12:750568. Available from: <http://doi.org/10.3389/fgene.2021.750568>
32. Hernandez-Beeftink T, Guillen-Guio B, Lorenzo-Salazar JM, Corrales A, Suarez-Pajes E, Feng R, et al. A genome-wide association study of survival in patients with sepsis. *Crit Care* [Internet]. 2022;26:341. Available from: <http://doi.org/10.1186/s13054-022-04208-5>
33. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* [Internet]. 2016 [cited 2020 Nov 27];315:801. Available from: <http://doi.org/10.1001/jama.2016.0287>
34. Díaz-de Usera A, Lorenzo-Salazar J, Rubio-Rodríguez L, Muñoz-Barrera A, Guillen-Guio B, Marcelino-Rodríguez I, et al. Evaluation of Whole-Exome Enrichment Solutions: Lessons from the High-End of the Short-

- Read Sequencing Scale. *J Clin Med* [Internet]. 2020 [cited 2022 Nov 2];9:3656. Available from: <http://doi.org/10.3390/jcm9113656>
35. Li H. Burrow-Wheeler Aligner for short-read alignment (see minimap2 for long-read alignment): lh3/bwa [Internet]. 2019 [cited 2019 May 27]. Available from: <https://github.com/lh3/bwa>
36. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *GigaScience* [Internet]. 2021 [cited 2023 Jan 16];10:giab008. Available from: <https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giab008/6137722>
37. Picard Tools - By Broad Institute [Internet]. [cited 2025 Jun 5]. Available from: <https://broadinstitute.github.io/picard/>
38. HaplotypeCaller – GATK [Internet]. [cited 2025 Jun 5]. Available from: <https://gatk.broadinstitute.org/hc/en-us/articles/360037225632-HaplotypeCaller>
39. samtools/bcftools [Internet]. samtools; 2025 [cited 2025 Jun 5]. Available from: <https://github.com/samtools/bcftools>
40. Gudmundsson S, Singer-Berk M, Watts NA, Phu W, Goodrich JK, Solomonson M, et al. Variant interpretation using population databases: Lessons from gnomAD. *Hum Mutat* [Internet]. 2022 [cited 2025 Jun 5];43:1012–30. Available from: <http://doi.org/10.1002/humu.24309>
41. Itan Y, Shang L, Boisson B, Patin E, Bolze A, Moncada-Vélez M, et al. The human gene damage index as a gene-level approach to prioritizing exome variants. *Proc Natl Acad Sci* [Internet]. 2015 [cited 2024 Feb 23];112:13615–20. Available from: <http://doi.org/10.1073/pnas.1518646112>
42. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol* [Internet]. 2016 [cited 2025 Mar 24];17:122. Available from: <http://doi.org/10.1186/s13059-016-0974-4>
43. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* [Internet]. 2010 [cited 2020 Mar 1];38:e164–e164. Available from: <http://doi.org/10.1093/nar/gkq603>
44. benchmarking/WES at master · genomicsITER/benchmarking · GitHub [Internet]. [cited 2025 Jun 5]. Available from: <https://github.com/genomicsITER/benchmarking/tree/master/WES>
45. Teide HPC | [Internet]. [cited 2025 Jun 5]. Available from: <https://teidehpc.iter.es/en/home/>
46. Zhang P, Cobat A, Lee Y-S, Wu Y, Bayrak CS, Boccon-Gibod C, et al. A computational approach for detecting physiological homogeneity in the midst of genetic heterogeneity. *Am J Hum Genet* [Internet]. 2021 [cited 2025 Jan 30];108:1012–25. Available from: <http://doi.org/10.1016/j.ajhg.2021.04.023>
47. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* [Internet]. 2003 [cited 2024 Feb 22];13:2498–504. Available from: <http://doi.org/10.1101/gr.1239303>
48. Milacic M, Beavers D, Conley P, Gong C, Gillespie M, Griss J, et al. The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res* [Internet]. 2024 [cited 2025 Jun 5];52:D672–8. Available from: <http://doi.org/10.1093/nar/gkad1025>
49. Guillen-Guio B, Hernández-Beeftink T, Marcelino-Rodríguez I, Rodríguez-Pérez H, Lorenzo-Salazar JM, Espinilla-Peña M, et al. Admixture mapping of asthma in southwestern Europeans with North African ancestry influences. *Am J Physiol-Lung Cell Mol Physiol* [Internet]. 2020 [cited 2020 Jun 27];318:L965–75. Available from: <http://doi.org/10.1152/ajplung.00344.2019>
50. R Studio Team. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>. 2020.

51. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal Unified Approach for Rare-Variant Association Testing with Application to Small-Sample Case-Control Whole-Exome Sequencing Studies. *Am J Hum Genet* [Internet]. 2012 [cited 2023 Jan 20];91:224–37. Available from: <http://doi.org/10.1016/j.ajhg.2012.06.007>
52. Shortt K, Chaudhary S, Grigoryev D, Heruth DP, Venkitachalam L, Zhang LQ, et al. Identification of Novel Single Nucleotide Polymorphisms Associated with Acute Respiratory Distress Syndrome by Exome-Seq. Zhao Y-Y, editor. *PLoS ONE* [Internet]. 2014 [cited 2023 Jan 20];9:e111953. Available from: <http://doi.org/10.1371/journal.pone.0111953>
53. Xu J-Y, Liu A-R, Wu Z-S, Xie J-F, Qu X-X, Li C-H, et al. Nucleotide polymorphism in ARDS outcome: a whole exome sequencing association study. *Ann Transl Med* [Internet]. 2021 [cited 2023 Feb 14];9:780–780. Available from: <https://doi.org/10.21037/atm-20-5728>
54. Poli MC, Aksentijevich I, Bousfiha AA, Cunningham-Rundles C, Hambleton S, Klein C, et al. Human inborn errors of immunity: 2024 update on the classification from the International Union of Immunological Societies Expert Committee. *J Hum Immun* [Internet]. 2025 [cited 2025 Apr 16];1:e20250003. Available from: <http://doi.org/10.70962/jhi.20250003>
55. Casanova J-L, Abel L. The Microbe, the Infection Enigma, and the Host. *Annu Rev Microbiol* [Internet]. 2024 [cited 2025 Mar 7];78:103–24. Available from: <http://doi.org/10.1146/annurev-micro-092123-022855>
56. Petersen L, Andersen PK, Sørensen TIA. Genetic Influences on Incidence and Case-Fatality of Infectious Disease. Brandstaetter A, editor. *PLoS ONE* [Internet]. 2010 [cited 2020 Nov 27];5:e10603. Available from: <http://doi.org/10.1371/journal.pone.0010603>
57. Miller JE, Carter KW, De Klerk N, Burgner DP. The familial risk of infection-related hospitalization in children: A population-based sibling study. Ramagopalan SV, editor. *PLOS ONE* [Internet]. 2021 [cited 2025 Jan 30];16:e0250181. Available from: <http://doi.org/10.1371/journal.pone.0250181>
58. Guillen-Guio B, Marcelino-Rodriguez I, Lorenzo-Salazar JM, Leavy OC, Allen RJ, Pompa-Mera EN, et al. Polygenic risk of idiopathic pulmonary fibrosis and COVID-19 severity. *ERJ Open Res* [Internet]. 2024 [cited 2025 Jan 30];00978–2024. Available from: <http://doi.org/10.1183/23120541.00978-2024>
59. Minnai F, Biscarini F, Esposito M, Dragani TA, Bujanda L, Rahmouni S, et al. A genome-wide association study for survival from a multi-centre European study identified variants associated with COVID-19 risk of death. *Sci Rep* [Internet]. 2024 [cited 2025 Jan 30];14:3000. Available from: <http://doi.org/10.1038/s41598-024-53310-x>
60. Smith SP, Shahamatdar S, Cheng W, Zhang S, Paik J, Graff M, et al. Enrichment analyses identify shared associations for 25 quantitative traits in over 600,000 individuals from seven diverse ancestries. *Am J Hum Genet* [Internet]. 2022 [cited 2025 Jan 30];109:871–84. Available from: <http://doi.org/10.1016/j.ajhg.2022.03.005>
61. Smedley D, Jacobsen JOB, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat Protoc* [Internet]. 2015 [cited 2022 Aug 11];10:2004–15. Available from: <http://doi.org/10.1038/nprot.2015.124>
62. Boos J, Van Der Made CI, Ramakrishnan G, Coughlan E, Asselta R, Löscher B-S, et al. Stratified analyses refine association between TLR7 rare variants and severe COVID-19. *Hum Genet Genomics Adv* [Internet]. 2024 [cited 2025 Jan 30];5:100323. Available from: <http://doi.org/10.1016/j.xhgg.2024.100323>
63. Burnham KL, Milind N, Lee W, Kwok AJ, Cano-Gamez K, Mi Y, et al. eQTLs identify regulatory networks and drivers of variation in the individual response to sepsis. *Cell Genomics* [Internet]. 2024 [cited 2025 Jan 30];4:100587. Available from: <http://doi.org/10.1016/j.xgen.2024.100587>
64. Beutler B, Hoebe K, Du X, Ulevitch RJ. How we detect microbes and respond to them: the Toll-like receptors and their transducers. *J Leukoc Biol* [Internet]. 2003 [cited 2025 Jan 30];74:479–85. Available from: <http://doi.org/10.1189/jlb.0203082>

65. Roger T, Froidevaux C, Le Roy D, Reymond MK, Chanson A-L, Mauri D, et al. Protection from lethal Gram-negative bacterial sepsis by targeting Toll-like receptor 4. *Proc Natl Acad Sci* [Internet]. 2009 [cited 2025 Jan 30];106:2348–52. Available from: <http://doi.org/10.1073/pnas.0808146106>
66. Meyer NJ, Prescott HC. Sepsis and Septic Shock. Hardin CC, editor. *N Engl J Med* [Internet]. 2024 [cited 2025 Jan 30];391:2133–46. Available from: <http://doi.org/10.1056/NEJMra2403213>
67. Wurfel MM, Gordon AC, Holden TD, Radella F, Strout J, Kajikawa O, et al. Toll-like Receptor 1 Polymorphisms Affect Innate Immune Responses and Outcomes in Sepsis. *Am J Respir Crit Care Med* [Internet]. 2008 [cited 2025 Jan 30];178:710–20. Available from: <http://doi.org/10.1164/rccm.200803-462OC>
68. Picard C, von Bernuth H, Ghandil P, Chrabieh M, Levy O, Arkwright PD, et al. Clinical features and outcome of patients with IRAK-4 and MyD88 deficiency. *Medicine (Baltimore)* [Internet]. 2010;89:403–25. Available from: <http://doi.org/10.1097/MD.0b013e3181fd8ec3>
69. Pino-Yanes M, Corrales A, Casula M, Blanco J, Muriel A, Espinosa E, et al. Common Variants of TLR1 Associate with Organ Dysfunction and Sustained Pro-Inflammatory Responses during Sepsis. Morty RE, editor. *PLoS ONE* [Internet]. 2010 [cited 2020 Nov 27];5:e13759. Available from: <http://doi.org/10.1371/journal.pone.0013759>
70. Casanova J-L, Abel L. The human genetic determinism of life-threatening infectious diseases: genetic heterogeneity and physiological homogeneity? *Hum Genet* [Internet]. 2020 [cited 2023 May 16];139:681–94. Available from: <http://doi.org/10.1007/s00439-020-02184-w>
71. Zhang Q, Bastard P, Liu Z, Le Pen J, Moncada-Velez M, Chen J, et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* [Internet]. 2020 [cited 2020 Nov 30];370:eabd4570. Available from: <http://doi.org/10.1126/science.abd4570>
72. Gratten J, Visscher PM. Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. *Genome Med* [Internet]. 2016 [cited 2025 Jan 30];8:78. Available from: <http://doi.org/10.1186/s13073-016-0332-x>
73. Burnham EL, Janssen WJ, Riches DWH, Moss M, Downey GP. The fibroproliferative response in acute respiratory distress syndrome: mechanisms and clinical significance. *Eur Respir J* [Internet]. 2014 [cited 2025 Jan 30];43:276–85. Available from: <http://doi.org/10.1183/09031936.00196412>
74. Baldassarri M, Fava F, Fallarini C, Daga S, Benetti E, Zguro K, et al. Severe COVID-19 in Hospitalized Carriers of Single CFTR Pathogenic Variants. *J Pers Med* [Internet]. 2021 [cited 2025 Jan 30];11:558. Available from: <http://doi.org/10.3390/jpm11060558>
75. Polgreen PM, Brown GD, Hornick DB, Ahmad F, London B, Stoltz DA, et al. CFTR Heterozygotes Are at Increased Risk of Respiratory Infections: A Population-Based Study. *Open Forum Infect Dis* [Internet]. 2018 [cited 2025 Jan 30];5:ofy219. Available from: <http://doi.org/10.1093/ofid/ofy219>
76. Miller AC, Comellas AP, Hornick DB, Stoltz DA, Cavanaugh JE, Gerke AK, et al. Cystic fibrosis carriers are at increased risk for a wide range of cystic fibrosis-related conditions. *Proc Natl Acad Sci* [Internet]. 2020 [cited 2025 Jan 30];117:1621–7. Available from: <http://doi.org/10.1073/pnas.1914912117>
77. Chen H-H, Shaw DM, Petty LE, Graff M, Bohlender RJ, Polikowsky HG, et al. Host genetic effects in pneumonia. *Am J Hum Genet* [Internet]. 2021;108:194–201. Available from: <http://doi.org/10.1016/j.ajhg.2020.12.010>
78. Frick M. CFTR-a (novel) target in ARDS : Commentary on “Loss of endothelial CFTR drives barrier failure and edema formation in lung infection and can be targeted by CFTR potentiation”, Erfinanda et al., *Sci. Transl. Med.* 14, eabg8577 (2022). *Pflugers Arch* [Internet]. 2023;475:417–9. Available from: <http://doi.org/10.1007/s00424-023-02800-4>
79. Honrubia JM, Gutierrez-Álvarez J, Sanz-Bravo A, González-Miranda E, Muñoz-Santos D, Castaño-Rodríguez C, et al. SARS-CoV-2-Mediated Lung Edema and Replication Are Diminished by Cystic Fibrosis Transmembrane

Conductance Regulator Modulators. Subbarao K, editor. mBio [Internet]. 2023 [cited 2024 Jul 8];14:e03136-22. Available from: <http://doi.org/10.1128/mbio.03136-22>

80. Aman J, Margadant C. Integrin-Dependent Cell–Matrix Adhesion in Endothelial Health and Disease. Circ Res [Internet]. 2023 [cited 2024 Jul 9];132:355–78. Available from: <http://doi.org/10.1161/CIRCRESAHA.122.322332>

81. Cheng S-C, Scicluna BP, Arts RJW, Gresnigt MS, Lachmandas E, Giamarellos-Bourboulis EJ, et al. Broad defects in the energy metabolism of leukocytes underlie immunoparalysis in sepsis. Nat Immunol [Internet]. 2016 [cited 2025 Jan 30];17:406–13. Available from: <http://doi.org/10.1038/ni.3398>

82. Supinski GS, Schroder EA, Callahan LA. Mitochondria and Critical Illness. Chest [Internet]. 2020 [cited 2025 Jan 30];157:310–22. Available from: <http://doi.org/10.1016/j.chest.2019.08.2182>

83. Nedel W, Strogulski NR, Kopczyński A, Portela LV. Assessment of mitochondrial function and its prognostic role in sepsis: a literature review. Intensive Care Med Exp [Internet]. 2024 [cited 2025 Jan 30];12:107. Available from: <http://doi.org/10.1186/s40635-024-00694-9>

84. Leavy OC, Allen RJ, Kraven LM, Morgan AD, Tobin MD, Quint JK, et al. The use of genetic information to define idiopathic pulmonary fibrosis in UK Biobank. Chest [Internet]. 2023 [cited 2022 Aug 25];163(2):362-365. Available from: <http://doi.org/10.1016/j.chest.2022.07.027>

Supplementary information

Additional file 1 (.docx). Includes supplementary methods and references describing workflows for WES and SNP array data, supplementary tables 1-6 and figures 1-13. **Supplementary Table 1** Demographics and clinical characteristics of the patients from the GEN-SEP cohort included in this study; **Supplementary Table 2** Enriched significant pathways ($p < 1 \times 10^{-5}$) in the significant gene clusters; **Supplementary Table 3** Genes exhibiting qualifying variants in at least 10 ARDS cases in the study; **Supplementary Table 4** Qualifying variants found in *FLT1* in the GEN-SEP cohort patients; **Supplementary Table 5** Qualifying variants in the ten most frequent genes observed in the ARDS cases. The number of carriers in ARDS cases and in sepsis controls, along with standardized identifiers, are indicated; **Supplementary Table 6** Odds ratios (95% confidence interval, p-values) resulting from sensitivity analysis using logistic regressions of patient-recorded variables for the ARDS risk-associated gene clusters with minimal similarity. Results from the stricter model ($CADD \geq 15$), adjusted for sex, age, and APACHE II scores, excluding and including first 2 principal components, and from models that did not adjust for any covariate are included for reference; **Supplementary Table 7** Results from sensibility analyses of QVs with $AF \leq 0.01$ and $AF \leq 0.001$ in gene clusters with minimal similarity. The models were adjusted for sex, age and APACHE II scores; **Supplementary Table 8** Results of the null model based on logistic regression analysis of rare ($AF \leq 0.01$) synonymous variants in gene clusters with minimal similarity. The models were adjusted for sex, age and APACHE II scores. **Supplementary Fig. 1** Plot of the first two principal components (PCs) of genetic variation of GEN-SEP patients and of data from the 1000 Genomes Project from African (AFR), American (AMR), East Asian (EAS), European (EUR), and South Asian (SAS). GEN-SEP cases and controls are

indicated with differentiated symbols. The percentages represent the explained variability of each axis; **Supplementary Fig. 2** Schematic overview of the sequencing workflow, filtering for qualifying variant identification, and the network-based heterogeneity clustering (NHC) analysis; **Supplementary Fig. 3** Similarity matrix, as the inverse of the Jaccard distance, reflecting the percentage of gene-level overlap between significant gene clusters; **Supplementary Fig. 4** Network representation of Cluster 186, the most significant among the significant clusters with the *Integrin cell surface interactions* as the top-most significant associated pathway; **Supplementary Fig. 5** Network representation of Cluster 170, the most significant among the significant clusters with the *Toll-like receptor cascades* as the top-most significant associated pathway; **Supplementary Fig. 6** Network representation of Cluster 181, the most significant among the significant clusters with the *Krebs cycle and respiratory electron transport* as the top-most significant associated pathway; **Supplementary Fig. 7** Network representation of Cluster 169, the only significant cluster with *Transcriptional regulation of white adipocyte differentiation* as the top-most significant associated pathway; **Supplementary Fig. 8** Network representation of Cluster 153, the most significant among the significant clusters with the *Interferon signaling* as the top-most significant associated biological pathway; **Supplementary Fig. 9** Network representation of Cluster 182, the only significant cluster with *mTOR signaling* as the top-most significant associated pathway; **Supplementary Fig. 10** Network representation of Cluster 12, the only significant cluster with *Amino acids regulate mTORC1* as the top-most significant associated pathway; **Supplementary Fig. 11** Network representation of Cluster 19, the only significant cluster with *Protein-protein interactions at synapses* as the top-most significant associated pathway; **Supplementary Fig. 12** Network representation of Cluster 46, the most significant among the significant clusters with the *mRNA splicing* as the top significant associated pathway; **Supplementary Fig. 13** Network representation of Cluster 97, the most significant among the significant clusters with the *DNA repair* as the top-most significant associated pathway.

Table 1 Gene clusters significantly differing between ARDS cases and sepsis controls.

Cluster ID	Genes	Cases included	Cluster <i>p</i> -value	Top pathway enriched in the cluster	Top pathway <i>p</i> -value
12	110	272	3.29x10 ⁻¹⁰	Amino acids regulate mTORC1	2.02x10 ⁻¹²
170	114	271	2.58x10 ⁻⁹	Toll-like receptor cascades	8.02x10 ⁻²¹
169	99	271	3.34x10 ⁻⁹	Transcriptional regulation of white adipocyte differentiation	6.02x10 ⁻³⁵
186	110	272	8.81x10 ⁻⁸	Integrin cell surface interactions	2.17x10 ⁻¹⁵
92*	84	266	1.13x10 ⁻⁷	Signaling by receptor tyrosine kinases	8.62x10 ⁻¹⁸
153	88	268	1.21x10 ⁻⁷	Interferon signaling	9.36x10 ⁻²⁷
196*	146	272	3.07x10 ⁻⁷	Sumoylation	7.12x10 ⁻¹⁵
97	91	270	4.21x10 ⁻⁷	DNA repair	6.50x10 ⁻²³
19	80	270	4.97x10 ⁻⁷	Protein-protein interactions at synapses	5.48x10 ⁻¹²
182	102	271	6.97x10 ⁻⁷	mTOR signaling	1.59x10 ⁻¹⁴
165*	89	269	9.83x10 ⁻⁷	Integrin cell surface interactions	8.06x10 ⁻¹³
177*	124	271	1.03x10 ⁻⁶	Toll-like receptor cascades	1.42x10 ⁻¹⁸
46	88	269	1.31x10 ⁻⁶	mRNA splicing	9.94x10 ⁻²⁵
204*	121	271	1.61x10 ⁻⁶	mRNA splicing	3.51x10 ⁻³²

156*	91	268	1.72×10^{-6}	Integrin cell surface interactions	6.30×10^{-20}
24*	85	270	2.39×10^{-6}	Integrin cell surface interactions	4.97×10^{-16}
181	121	263	3.93×10^{-6}	Krebs cycle and the respiratory electron transport	7.65×10^{-69}
126*	93	246	4.92×10^{-6}	Krebs cycle and the respiratory electron transport	5.81×10^{-53}
171*	109	268	5.68×10^{-6}	Programmed cell death	4.82×10^{-16}

*Clusters not depicted in supplementary figures because they show >20% gene-level similarity with any of the represented clusters.

ARTICLE IN PRESS

Table 2 Association of qualifying variant carriers (CADD ≥ 15) with sepsis-associated ARDS risk in the 10 significant gene clusters with the lowest pairwise similarities.

Cluster ID	Top enriched pathway	Genes	Odds Ratio (95% CI)	<i>p</i> -value
12	Amino acids regulate mTORC1	110	6.57 (2.75 - 15.68)	2.0×10^{-5}
170	Toll-like receptor cascades	114	3.72 (2.05 - 6.71)	1.0×10^{-5}
169	Regulation of white adipocyte differentiation	99	3.05 (1.61 - 5.80)	6.3×10^{-4}
186	Integrin cell surface interactions	110	6.10 (2.15 - 17.29)	6.6×10^{-4}
153	Interferon signaling	88	2.61 (1.48 - 4.61)	9.2×10^{-4}
97	DNA repair	91	2.78 (1.46 - 5.29)	1.8×10^{-3}
19	Protein-protein interactions at synapses	80	3.88 (1.86 - 8.07)	2.8×10^{-4}
182	mTOR signaling	102	2.63 (1.30 - 5.34)	7.2×10^{-3}
46	mRNA splicing	88	3.42 (1.65 - 7.09)	9.2×10^{-4}
181	Krebs cycle and respiratory electron transport	121	3.04 (1.75 - 5.28)	7.0×10^{-5}

CI: confidence interval. † Models not converging, attributable to disproportionate distribution of carriers and non-carriers and leading to unreliable effect size estimates.

Table 3 Results of the 28-day survival analyses for the 10 significant gene clusters with the lowest pairwise similarities (QVs with CADD \geq 15).

Cluster ID	Top enriched pathway	Genes	Model 1: sex, age, APACHE II	Model 2: sex, age, APACHE II, ARDS status
			Hazard Ratio (95% CI); <i>p</i> -value	Hazard Ratio (95% CI); <i>p</i> -value
12	Amino acids regulate mTORC1	110	1.14 (0.64 - 2.01); <i>p</i> =0.645	0.98 (0.55 - 1.76); <i>p</i> =0.964
170	Toll-like receptor cascades	114	1.26 (0.76 - 2.10); <i>p</i> =0.359	1.14 (0.68 - 1.91); <i>p</i> =0.611
169	Regulation of white adipocyte differentiation	99	0.60 (0.37 - 0.97); <i>p</i> =0.038	0.54 (0.33 - 0.88); <i>p</i> =0.013
186	Integrin cell surface interactions	110	1.30 (0.61 - 2.79); <i>p</i> =0.487	1.14 (0.53 - 2.47); <i>p</i> =0.723
153	Interferon signaling	88	0.95 (0.59 - 1.50); <i>p</i> =0.825	0.89 (0.56 - 1.42); <i>p</i> =0.635
97	DNA repair	91	1.07 (0.60 - 1.91); <i>p</i> =0.793	0.97 (0.54 - 1.73); <i>p</i> =0.924
19	Protein-protein interactions at synapses	80	3.80 (1.40 - 10.26); <i>p</i> =0.008	3.44 (1.26 - 9.35); <i>p</i> =0.015
182 [#]	mTOR signaling	102	0.97 (0.54 - 1.71); <i>p</i> =0.918	0.90 (0.51 - 1.60); <i>p</i> =0.729
46	mRNA splicing	88	1.29 (0.65 - 2.53); <i>p</i> =0.458	1.17 (0.59 - 2.32); <i>p</i> =0.643
181	Krebs cycle and respiratory electron transport	121	1.23 (0.74 - 2.04); <i>p</i> =0.420	1.10 (0.66 - 1.85); <i>p</i> =0.703

APACHE II: Acute Physiology and Chronic Health Disease Classification System II. CI: confidence interval. Analysis performed in 665 of the patients with available 28-day mortality data. [#]Used the weighted Cox regression model, not assuming the proportionality of hazards.

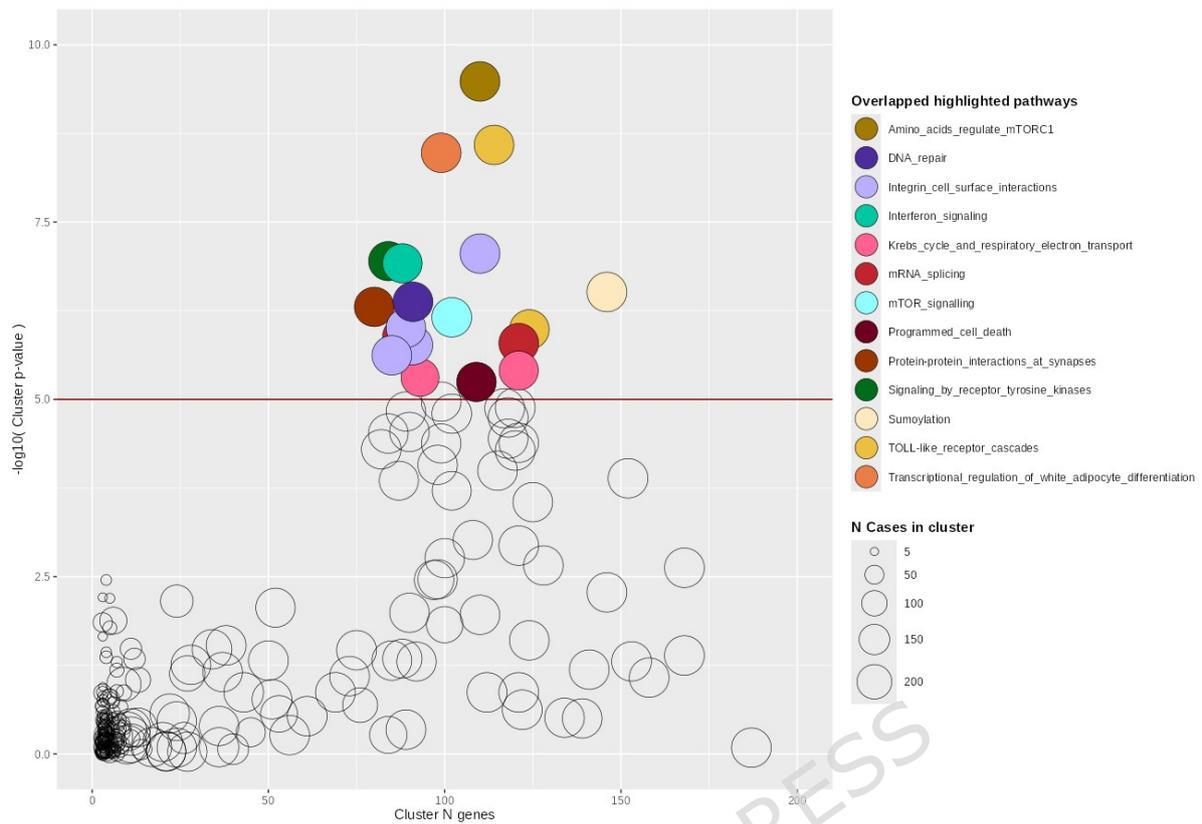


Fig. 1 Bubble plot of gene clusters identified by the Network-based Heterogeneity Clustering analysis of ARDS cases compared to sepsis controls. Gene clusters are represented by bubbles, where the size of each is proportional to the number of ARDS cases assigned. Colors denote significant clusters ($p \leq 1 \times 10^{-5}$) that share the top significant pathway. The x-axis represents the number of genes included in each cluster, and the y-axis represents the significance ($-\log_{10}[p\text{-value}]$).