

# The genomic landscape of Acute Respiratory Distress Syndrome: a meta-analysis by information content of genome-wide studies of the host response.

Jonathan E Millar<sup>1</sup>, Sara Clohisey-Hendry<sup>1</sup>, Megan McMannus<sup>1</sup>, Marie Zechner<sup>1</sup>, Bo Wang<sup>1</sup>, Nick Parkinson<sup>1</sup>, Melissa Jungnickel<sup>1</sup>, Nureen Mohamad Zaki<sup>1</sup>, Erola Pairo-Castineira<sup>1</sup>, Konrad Rawlik<sup>1</sup>, Clark D Russell<sup>2</sup>, Manu Shankar-Hari<sup>2</sup>, Carolyn Calfee<sup>3</sup>, Daniel F McAuley<sup>4</sup>, and J Kenneth Baillie<sup>1</sup>

1. Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom.
2. Centre for Inflammation Research, University of Edinburgh, Edinburgh, United Kingdom.
3. Division of Pulmonary, Critical Care, Allergy & Sleep Medicine, Department of Medicine, University of California San Francisco, San Francisco, USA.
4. Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast, United Kingdom.

## **Abstract**

## Introduction

The acute respiratory distress syndrome (ARDS) is clinically defined as acute hypoxaemic respiratory failure due to non-cardiogenic pulmonary oedema<sup>1</sup>. It occurs following a variety of insults; pulmonary and extra-pulmonary. While this definition has been useful in identifying patients at risk of serious morbidity and death<sup>2</sup>, it overlooks the underlying biology and masks heterogeneity<sup>3</sup>. Arguably, this has contributed to limited success in developing therapeutics<sup>4</sup>. In contrast, a biological definition of ARDS, based on mechanistically distinct sub-phenotypes, may provide the lever necessary for future drug discovery<sup>5</sup>.

Functional genomics technologies enable disease characterisation at unprecedented resolution. The emergence of coronavirus disease 2019 (COVID-19) has provided an opportunity to test their usefulness for drug discovery. A notable success has been the finding that baricitinib, a Janus kinase inhibitor, reduces mortality in patients hospitalised with COVID-19<sup>6</sup>. *A priori* support for baricitinib was greatly enhanced following the discovery of a causal link between elevated tyrosine kinase 2 (*TYK2*) expression and severe COVID-19 in genome-wide association studies (GWAS)<sup>7,8</sup>. The availability of omics data for non-COVID ARDS is limited by comparison, although recent studies have used these techniques to examine signatures of non-COVID ARDS sub-phenotypes<sup>9,10</sup>.

An unresolved challenge is how large omics data can be effectively exploited<sup>11</sup>. Specifically, how can we combine data from heterogeneous sources to derive new insights or recalibrate our current understanding in the light of new data? We have proposed meta-analysis by information content (MAIC) as a data-driven, algorithmic, method for combining gene lists from diverse sources<sup>12</sup>. MAIC is agnostic to the quality or methodology of the sources and combines ranked or un-ranked gene sets by calculating weights for each list and gene, and iteratively updating them to converge on a ranked meta-list. We have successfully applied MAIC to host-genomics studies of Influenza A<sup>12</sup> and SARS-CoV-2<sup>7,13</sup>, and shown that it out-performs existing algorithms when combining ranked and un-ranked lists obtained from heterogeneous sources<sup>14</sup>.

Here we present a living meta-analysis by information content of ARDS host genomics studies as an open-source resource for gene prioritisation, translational genomics, and drug target discovery. A comprehensive, interactive interface is available at <https://baillielab.net/maic/ards>.

## Results

### Systematic review

Our search yielded 8,937 unique citations (Fig. S1). Of these, we retrieved 74 studies for full-text evaluation and ultimately included 40 in our meta-analysis<sup>9,10,15–52</sup>. These studies produced 44 unique gene lists (22 transcriptomic, 13 proteomic, and 9 based on genome-wide association studies (GWAS); see Table 1). Three studies reported results from multiple methodologies<sup>10,34,39</sup>, and several used more than one tissue type<sup>19,22,33</sup>. Excluding GWAS, 14 lists (40%) were from lung or airway samples, and 21 (60%) from blood. We could not retrieve one gene list<sup>27</sup>, and no whole-genome sequencing GWAS were found. Only 36% (n=8) of transcriptomic lists used next-generation sequencing. The earliest study was published in 2004<sup>19</sup>, while almost half (n=19, 47.5%) were published in the last 5 years.

Most studies aimed to identify genes/proteins associated with ARDS susceptibility (n=27, 67.5%). The remainder examined associations with survival (n=6, 15%), sub-phenotype (n=4, 10%), disease progression (n=2, 5%), or severity (n=1, 2.5%). In total, studies included 6,856 ARDS patients. Supplementary Table 1 provides detailed study designs, demographics, and ARDS aetiology.

### Meta-analysis by information content (MAIC)

First, we analysed the 43 available gene lists using MAIC. Lists were categorized by method (GWAS, transcriptomics, proteomics) and technique (e.g., RNA-seq, mass spectrometry; see Table 1). In total, we ranked 7,085 unique genes or SNPs, with a median of 27 genes per list (range 1-4,954). The top 100 ranked genes are summarized in Figure 1. Most genes were found in a single category (n=5,866, 82.8%); only 157 (2.2%) were identified in  $\geq 3$  categories, with a maximum of 5 categories (Figure 1). Similarly, few genes (n=362, 5.1%) were identified by  $> 1$  method, with only *AKR1B10*, *HINT1*, *HSPG2*, *S100A11*, and *SLC18A1* present in transcriptomic, proteomic, and GWAS based lists. To prioritise genes, we used the unit invariant knee method<sup>53</sup> to identify the inflection point in the MAIC score curve. This prioritised 1,306 genes with scores above this point (Figure 1). These genes were more likely to be found in  $\geq 2$  lists or categories and by  $> 1$  method (Figure 1).

To assess the influence of individual lists, we calculated the information content (IC), reflecting the sum of gene scores across all lists (Figure 2), and the information contribution (ICtb), measuring the sum of gene scores contributing to a gene's overall MAIC score. To obtain relative values, we divided the IC/ICtb for each list by the total. This showed that only 10 lists (from 9 studies) contributed  $> 1\%$  of total information by either metric (Tab. S2). Notably, the RNA-seq list from Sarma et al.<sup>10</sup> accounts for  $> 50\%$  of the total IC and ICtb, a function of its length. To account for this, we normalised relative IC/ICtb by the number of genes per list. Along with the proportion of replicated genes, this provides an alternative perspective, with several proteomic lists ranking highly (Figure 2).

### Comparison with existing ARDS sources and COVID-19

To contextualise the results of our meta-analysis, we evaluated the degree of overlap between the genes prioritised by MAIC and those from two established resources: BioLitMine<sup>54</sup>, using an ARDS MeSH search, and the ARDS Database of Genes<sup>55</sup> (Fig. S2). BioLitMine identified 271 ARDS-associated genes, of which 142 (52.4%) were in our analysis. Almost half of the overlapping genes (n = 63, 44.4%) were ranked within our prioritized set (Tab. S3). Of the 239 genes catalogued in the ARDS Database, 177 (74.1%) were also present in our study. However, both sources contain some unsupported gene associations.

Table 1: Summary of studies and gene lists included in the systematic review

Year	Study	Focus	Definition	N <sup>a</sup>	Method	Technique	Tissue	Cell type
2022	Batra <sup>15</sup>	Survival	Berlin	24	Proteomics	Other	Blood	
	Mirchandani <sup>39</sup>	Susceptibility	Berlin	22	Proteomics	Mass Spec	Blood	Monocytes
Sarma <sup>10</sup>	Sub-phenotype	Berlin	41	Transcriptomics	Microarray	Blood	Monocytes	Monocytes
Zhang <sup>51</sup>	AECC	Transcriptomics	RNA-seq	TA	scRNA-Seq	TA	Immune cells	
Liao <sup>34</sup>	Either	GWAS	RNA-Seq	Blood	Blood	Blood	Exosomes	
Martucci <sup>36</sup>	None	Transcriptomics	Microarray	Blood	Blood	Blood	PBMCS	
Xu <sup>49</sup>	Survival	GWAS	WES	Blood	Blood	Blood	Blood	
Zhang <sup>50</sup>	AECC	Transcriptomics	RNA-seq	Blood	Blood	Blood	PBMCS	
Guillen-Guio <sup>28</sup>	AECC	Transcriptomics	Microarray	Blood	Blood	Blood	Blood	
Jiang <sup>30</sup>	AECC	Transcriptomics	RNA-seq	Blood	Blood	Blood	Blood	
Bos <sup>9</sup>	AECC	Transcriptomics	Microarray	Blood	Blood	Blood	Blood	
Englert <sup>26</sup>	AECC	Transcriptomics	RNA-seq	Blood	Blood	Blood	Blood	
Morrell <sup>41</sup>	AECC	Transcriptomics	Microarray	BALF	BALF	AMs	EVs	
Scheller <sup>45</sup>	AECC	Transcriptomics	RNA-seq	Blood	Blood	Blood	Blood	
Bime <sup>18</sup>	AECC	Transcriptomics	Microarray	BALF	BALF	BALF	BALF	
Morrell <sup>40</sup>	AECC	Transcriptomics	Microarray	Blood	Blood	Blood	Blood	
Bhargava <sup>17</sup>	AECC	Transcriptomics	Microarray	Blood	Blood	Blood	Blood	
Lu <sup>35</sup>	AECC	Transcriptomics	Microarray	Blood	Blood	Blood	Blood	
Zhu <sup>52</sup>	AECC	Transcriptomics	Microarray	Blood	Blood	Blood	Blood	

Year	Study	Focus	Definition	N <sup>a</sup>	Method	Technique	Tissue	Cell type
2016	Chen <sup>22</sup> Juss <sup>31</sup>	Severity Susceptibility	AECC Berlin	7 23	Proteomics Transcriptomics	Mass Spec Microarray	BALF/Blood	Neutrophils
	Nick <sup>42</sup>	Sub-phenotype	AECC Berlin	121	Transcriptomics	Microarray	Blood	Neutrophils
	Ren <sup>44</sup>	Susceptibility	AECC Berlin	14	Proteomics	Other	Blood	
2015	Kangellaris <sup>32</sup>	Susceptibility	AECC Berlin	29	Transcriptomics	Microarray	Blood	
	Kovach <sup>33</sup>	Susceptibility	AECC Progression	18	Transcriptomics	Microarray	BALF/Blood	AMs
2014	Bhargava <sup>16</sup> Shortt <sup>46</sup>	Susceptibility	AECC AECC	22	Proteomics	Mass Spec	BALF	
	Chen <sup>21</sup>	Susceptibility	AECC Berlin	213	GWAS	WES	Blood	
	Dong <sup>25</sup>	Progression	None	11	Proteomics	Mass Spec	Blood	
2013	Meyer <sup>38</sup>	Susceptibility	AECC Berlin	14	Proteomics	Mass Spec	BALF	AMs
	Nguyen <sup>43</sup>	Susceptibility	AECC AECC	661	GWAS	Genotyping	Blood	
	Christie <sup>23</sup>	Susceptibility	AECC AECC	30	Proteomics	Mass Spec	BALF	
6	Dolinay <sup>24</sup> Tejera <sup>48</sup>	Susceptibility	AECC AECC	812	GWAS	Genotyping	Blood	
	Frenzel <sup>27</sup>	Susceptibility	AECC AECC	35	Transcriptomics	Microarray	Blood	
	Meyer <sup>37</sup>	Survival	AECC AECC	1400	GWAS	Genotyping	Blood	
2011	Howrylak <sup>29</sup>	Susceptibility	AECC AECC	46	Proteomics	Mass Spec	BALF	
	Chang <sup>20</sup>	Susceptibility	None	1241	GWAS	Genotyping	Blood	
2009	Wang <sup>47</sup>	Susceptibility	AECC	13	Transcriptomics	Microarray	Blood	
2008	Bowler <sup>19</sup>	Susceptibility	AECC	20	Proteomics	Mass Spec	BALF	
				8	Transcriptomics	Microarray	Blood	
2004				16	Proteomics	Mass Spec	BALF/Blood	

a - The number of patients with ARDS included in each study. Abbreviations: AECC - American-European Consensus Conference; AMs - Alveolar macrophages; BALF - Bronchoalveolar lavage fluid; EVs - Extracellular vesicles; GWAS - Genome-wide association study; MS - Mass spectrometry; PBMCs - Peripheral blood mononuclear cells; TA - Tracheal aspirate; WES - Whole-exome sequencing.

For the BioLitMine search, we identified 4 such genes not initially found in the ARDS MAIC set after correcting historical gene symbol aliases. A further 104 were supported by a single publication. For the remaining 21, we obtained their 100 most co-expressed genes using ARCHS4<sup>56</sup> (returning data for 18) and assessed the overlap with ARDS MAIC (Fig. S2). Two-thirds exhibited <50% overlap. Finally, we compared the overlap between the genes ranked by MAIC for ARDS and by a previous MAIC of the host response to COVID-19<sup>13</sup> (Fig. S2). In total, 2,606 ARDS genes (36.8%) were also found in COVID-19, of which 143 were prioritized by both analyses (Fig. S2).

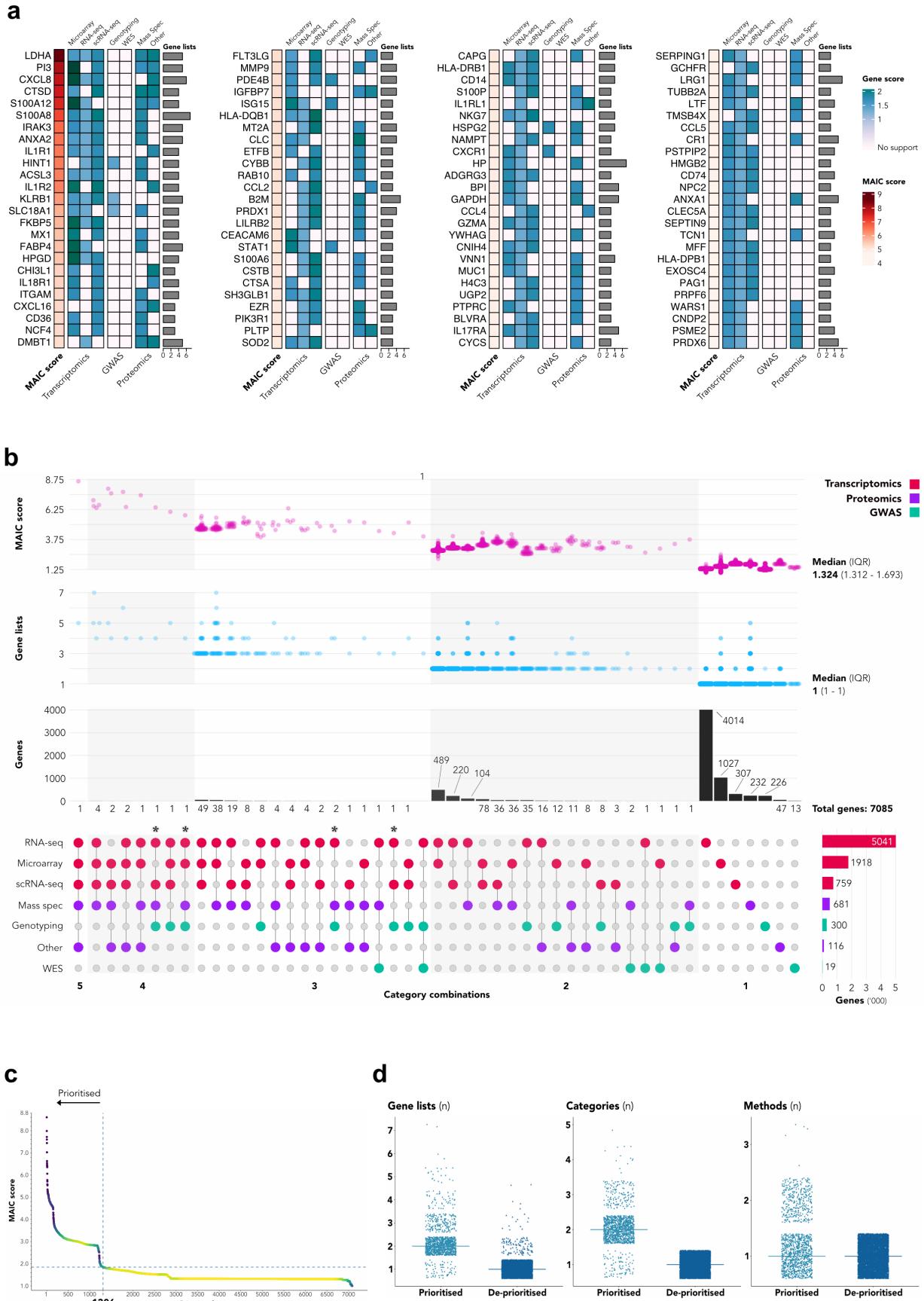
### **Tissue and cell-specific expression**

Despite the dominance of blood sampling, the majority of genes included in the meta-analysis were identified in airways samples (n=5,847, 82.5%) (Fig. S3). This was true for the prioritised set of genes, however, here most were also identified in blood (n=818, 62.6%) (Fig. S3). For the genes solely found in blood sampling lists (n=1,238), almost three-quarters are known to be expressed in the lung using scRNA-seq data ( $\geq 5$  normalised transcripts per million (nTPM)), with a quarter highly-expressed ( $\geq 100$  nTPM) (Fig S2).

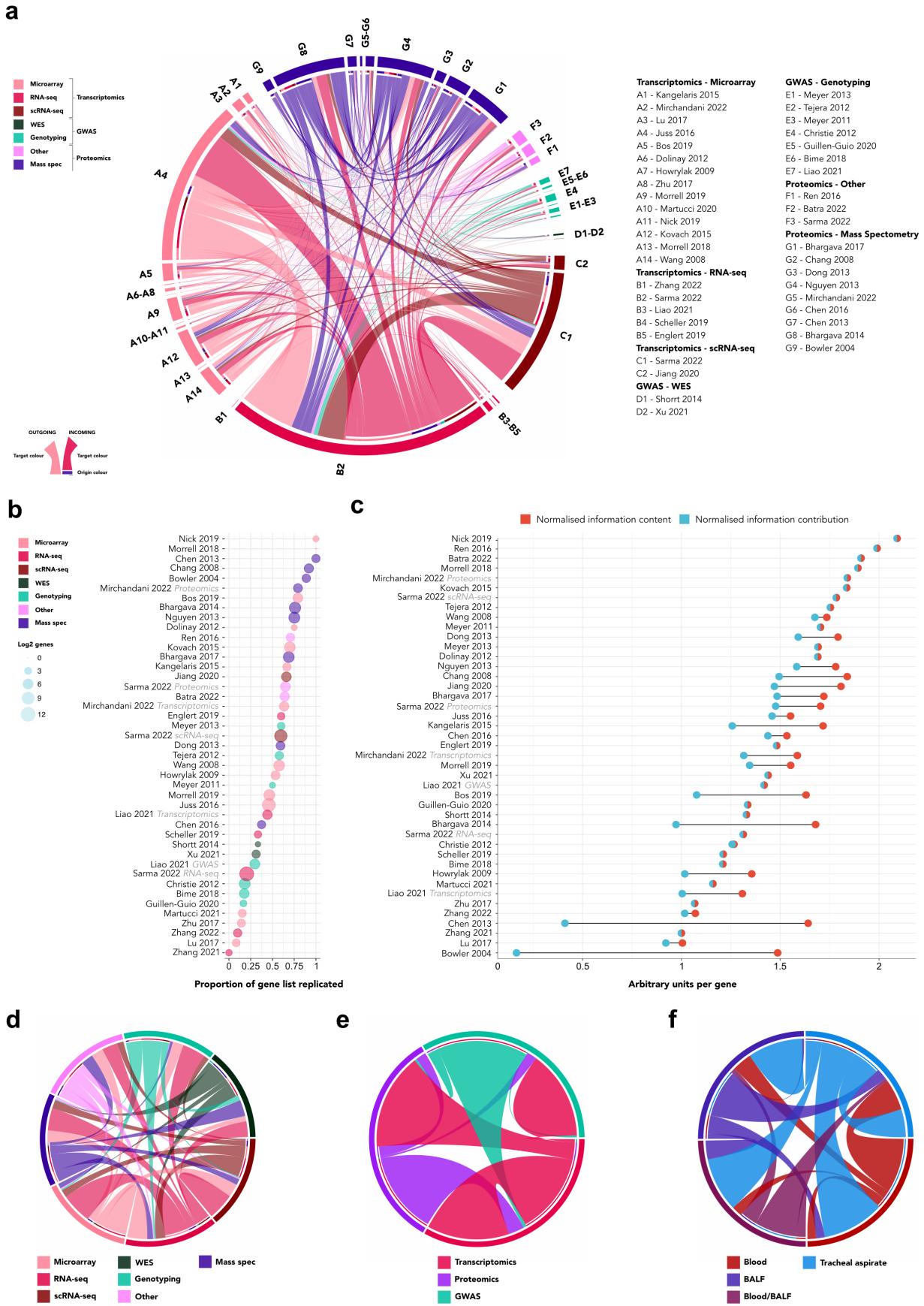
### **Functional enrichment**

#### **In-silico perturbation**

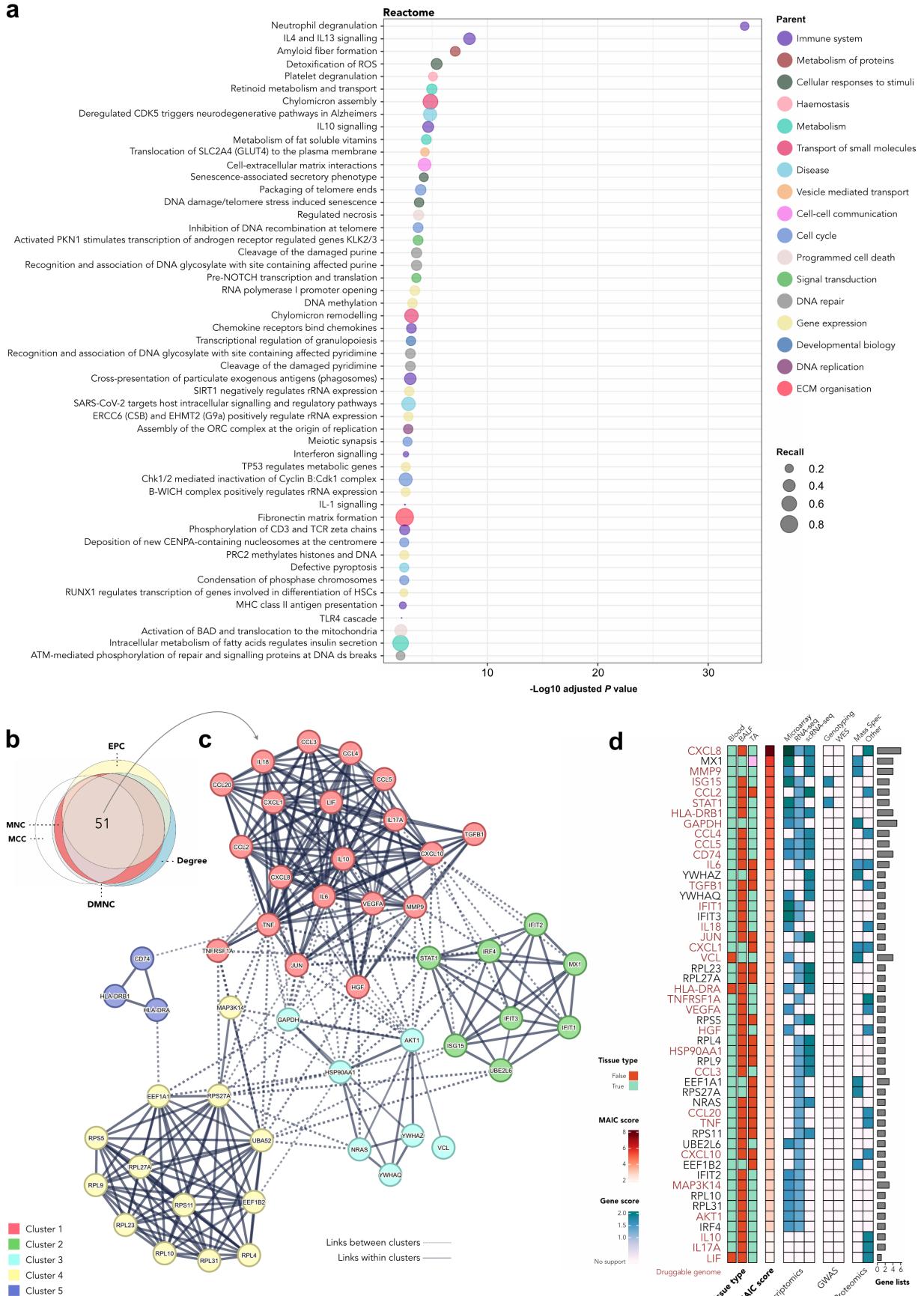
#### **Sub-groups**



**Figure 1: Meta-analysis by information content.** (a) Heatmap of top 100 ranked genes showing MAIC score, highest score per category, and number of supporting lists. (b) UpSet plot of MAIC genes showing total numbers for each category combination, MAIC score distribution, and supporting lists. (b) Gene prioritization using the Unit Invariant Knee method. Intersection of lines identifies elbow point of best-fit curve. 1,306 genes in upper left quadrant were prioritized. (c) Strip plots comparing number of lists and categories/methods per gene between prioritized and de-prioritized sets.



**Figure 2: Attributing information in MAIC.** (a) Shared information content (IC) between gene lists. Links indicate absolute IC (sum of common gene scores) between studies. (b) Proportion of replicated genes. Circle diameter is logarithm (base 2) of gene number per list. (c) IC normalized by number of genes. Overlapping circles denote equal normalized IC and contribution (IC<sub>tb</sub> - sum of common gene scores contributing to MAIC), indicating all gene scores contributed to MAIC. (d) Shared IC between categories, scaled so links show fraction of total IC. (e) Shared IC between methods, scaled. (f) Shared IC between tissue types, scaled.



**Figure 3: Functional enrichment of prioritised genes.** (a) Significantly enriched Reactome terms ( $P < 0.01$ ). Terms colored by parent class and size proportional to recall. (b) Euler diagram of the overlap of hub genes identified by five methods. MNC - Maximum Neighbourhood Component, MCC - Maximal Clique Centrality, DMNC - Density of MNC, EPC - Edge Percolated Component. (c) Protein-protein interaction (PPI) network of hub genes, clustered using the Markov Chain Algorithm. (d) Heatmap of common hub genes displaying tissue type(s), MAIC score, highest category score, supporting lists, and presence in the druggable genome.

## **Discussion**

### **Methods**

The systematic review and meta-analysis protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO; CRD42022306270). The review is reported in compliance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines<sup>57</sup>.

#### **Search strategy and selection criteria**

A detailed description of our search strategy and eligibility criteria is provided in the Supplementary Methods. Briefly, we searched MEDLINE, Embase, bioRxiv, medRxiv, the ARDS Database of Genes<sup>55</sup>, and the NCBI Gene Expression Omnibus from inception to December 1<sup>st</sup>, 2022 without language restrictions. We also performed single-level backwards and forwards citation searches using SpiderCite<sup>58</sup> and hand-searched recent review articles<sup>59–62</sup>.

We included human genome-wide studies reporting associations between genes, transcripts, or proteins and ARDS susceptibility, severity, survival, or phenotype, accepting any contemporaneous ARDS definition. We excluded paediatric studies (age < 18 years), animal studies, *in-vitro* human ARDS models, candidate *in-vivo* or *in-vitro* studies (< 50 genes/proteins), candidate gene associations, and studies with < 5 patients per arm (except scRNA-seq).

### **Outcomes**

We retrieved ranked lists of genes associated with the ARDS host response, preferring measures of significance and adjusted *P* values over raw *P* values when multiple ranking measures were used. We obtained both summary lists (all implicated genes) and author-defined subgroup lists. To combine subgroup lists into summary lists, we took the minimum *P* value or maximum effect size. We excluded genes below the author-defined threshold for significance/effect magnitude. If unavailable, we excluded genes with *P* > 0.05, z-score < 1.96, or log fold change < 1.5.

#### **Study selection and data extraction**

Article titles and abstracts from our search were stored in Zotero v6.0-beta (Corporation for Digital Scholarship, United States). Titles were initially screened by one author using Screenatron<sup>58</sup>. Two authors then independently screened abstracts against eligibility criteria, with a third resolving inconsistencies. Full texts and supplements of eligible studies were retrieved and inclusion adjudicated by consensus.

Data were extracted by one author and cross-checked by a second. Gene, transcript, or protein identifiers were mapped to HGNC symbols or Ensembl/RefSeq equivalents if no HGNC symbol was available. Unannotated SNPs were searched in NCBI dbSNP. miRBase (University of Manchester, United Kingdom) provided miRNA symbols. For microarray probes without symbols, we used the DAVID Gene Accession Conversion tool (Laboratory of Human Retrovirology and Immunoinformatics, Frederick National Laboratory for Cancer Research, United States) to map them to HGNC symbols. We extracted information relating to study design, methodology, tissue/cell type, demographics, ARDS aetiology, risk factors, severity, and outcomes.

#### **Meta-analysis by information content (MAIC)**

The MAIC algorithm has been described in detail<sup>7,12–14</sup>. Full documentation and the source code are available at <https://github.com/baillielab/maic>. Briefly, MAIC combines ranked and unranked lists of related named entities, such as genes, from heterogeneous experimental categories, without prior regard to the quality of each source. The algorithm

makes four key assumptions; (1) genes associated with ARDS exist as true positives, (2) a gene is more likely to be a true positive if it is found in more than one source, (3) the probability of being a true positive is enhanced if the gene appears in a list that contains a higher proportion of replicated genes, and (4) the probability is further enhanced if it is found in more than one category of experiment. Based on these assumptions, MAIC compares lists with each other, forming a weighting for each source based on its information content, which is then used to calculate a score for each gene. The output is a ranked list summarizing the total information supporting each gene's association with ARDS. We have shown MAIC outperforms available algorithms, especially with ranked and unranked heterogeneous data<sup>14</sup>.

As our primary analysis, we performed MAIC on all summary gene lists, regardless of study focus. Lists were assigned categories based on their methodology and experimental technique: genome-wide association study (GWAS) - genotyping, GWAS - whole exome sequencing, transcriptomics - microarray, transcriptomics - RNA-sequencing (RNA-seq), transcriptomics - single cell RNA-seq (scRNA-seq), proteomics - mass spectrometry, and proteomics - other. For secondary analyses, we performed MAIC on subsets of lists based on study focus (i.e., susceptibility to ARDS or survival/severity).

For each MAIC iteration, we prioritised genes with sufficient evidentiary support for further study (i.e., the gene set before which information content diminished such that there was little/no corroboration for the remainder's ARDS association). We used the unit invariant knee method<sup>53,63</sup> to identify the elbow point in the best-fit curve of MAIC scores. Genes with values above this point were prioritized for downstream analyses.

### **ARDS literature and SARS-CoV-2 associations**

We used BioLitMine<sup>54</sup> to query the NCBI Gene database for genes associated with the Medical Subject Heading (MeSH) term "Respiratory Distress Syndrome, Acute", generating a list of genes and publications. We descriptively compared the overlap between this list and the MAIC-ranked gene list. Similar comparisons were made between the ARDS MAIC results and the gene set in the ARDS Database of Genes<sup>55</sup> and a prior MAIC of SARS-CoV-2 host genomics<sup>13</sup>.

### **Tissue expression and enrichment**

Transcript and protein expression data for prioritised genes were retrieved from the Human Protein Atlas (HPA, version 21.0)<sup>64</sup>. We investigated transcript expression in a consensus RNA-seq dataset of 55 tissues, combining data from GTEx<sup>65,66</sup> (RSEMv1.3.0 v8) and the HPA<sup>67</sup>, and in the HPA RNA-seq blood dataset<sup>68</sup>, containing expression levels in 18 immune cell types and total peripheral blood mononuclear cells. We calculated tissue enrichment scores, using the R package *TissueEnrich*<sup>69</sup>, for genes in comparison to all genes present in GTEx<sup>65,66</sup> (RSEMv1.3.0 v8) and corrected for multiple comparisons using the Benjamini-Hochberg method<sup>69</sup>. To investigate protein expression, we retrieved tissue-specific expression scores from the HPA<sup>67</sup>.

### **Functional enrichment**

We performed functional enrichment of prioritised genes against the universe of all annotated genes using g:Profiler<sup>70</sup>. The following data sources were used; Gene Ontology (GO) Biological Process<sup>71</sup>, Kyoto Encyclopaedia of Genes and Genomes (KEGG)<sup>72</sup>, Reactome<sup>73</sup>, and WikiPathways<sup>74</sup>. Multiple testing was corrected for using the g:SCS algorithm<sup>70</sup>, with a threshold of  $P < 0.01$ . The input list was ordered by MAIC score. To address pathway redundancy, we used the output of the functional enrichment analysis as an input to Enrichment Map<sup>75</sup>. We created networks based on a Jaccard similarity cutoff of 0.25 for each of; GO:Molecular Function, GO:Cellular Component, GO:Biological

Process, Reactome, and Wiki Pathways. We then repeated this for all combined. Protein-protein interaction enrichment was performed using STRING v11<sup>76</sup>. We included all possible interaction sources but specified a minimum interaction score of 0.9. We used the the whole annotated genome as the statistical background. Markov Clustering Analysis (MCL) was applied to the resulting network with an inflation parameter of 3. Clusters were annotated with the Reactome pathway with the smallest False Discovery Rate.

### ***In-silico perturbation***

### **Software and code availability**

MAIC is implemented in Python v3.9.7 (Python Software Foundation, Wilmington, United States). All other analyses were performed with R v4.2.2 (R Core Team, R Foundation for Statistical Computing, Vienna, Austria). All code required to reproduce the analyses is available at [https://github.com/JonathanEMillar/ards\\_maic\\_analysis](https://github.com/JonathanEMillar/ards_maic_analysis).

## References

1. ARDS Definition Task Force *et al.* Acute respiratory distress syndrome: The Berlin definition. *JAMA* **307**, 2526–2533 (2012).
2. Bellani, G. *et al.* Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries. *JAMA* **315**, 788–800 (2016).
3. Wilson, J. G. & Calfee, C. S. ARDS subphenotypes: Understanding a heterogeneous syndrome. *Crit. Care* **24**, 102 (2020).
4. Laffey, J. G. & Kavanagh, B. P. Negative trials in critical care: Why most research is probably wrong. *Lancet Respir. Med.* **6**, 659–660 (2018).
5. Bos, L. D. J. *et al.* Towards a biological definition of ARDS: Are treatable traits the solution? *Intensive Care Med. Exp.* **10**, 8 (2022).
6. Peter W Horby, and *et al.* Baricitinib in patients admitted to hospital with COVID-19 (RECOVERY): A randomised, controlled, open-label, platform trial and updated meta-analysis. (2022) doi:[10.1101/2022.03.02.22271623](https://doi.org/10.1101/2022.03.02.22271623).
7. Pairo-Castineira, E. *et al.* Genetic mechanisms of critical illness in COVID-19. *Nature* **591**, 92–98 (2021).
8. Kousathanas, A. *et al.* Whole-genome sequencing reveals host factors underlying critical COVID-19. *Nature* **607**, 97–103 (2022).
9. Bos, L. D. J. *et al.* Understanding heterogeneity in biologic phenotypes of acute respiratory distress syndrome by leukocyte expression profiles. *Am. J. Respir. Crit. Care Med.* **200**, 42–50 (2019).
10. Sarma, A. *et al.* Hyperinflammatory ARDS is characterized by interferon-stimulated gene expression, t-cell activation, and an altered metatranscriptome in tracheal aspirates. *bioRxiv* (2022).
11. Gomez-Cabrero, D. *et al.* Data integration in the era of omics: Current and future challenges. *BMC Syst. Biol.* **8 Suppl 2**, I1 (2014).
12. Li, B. *et al.* Genome-wide CRISPR screen identifies host dependency factors for influenza a virus infection. *Nat. Commun.* **11**, 164 (2020).
13. Parkinson, N. *et al.* Dynamic data-driven meta-analysis for prioritisation of host genes implicated in COVID-19. *Sci. Rep.* **10**, 22303 (2020).
14. Wang, B. *et al.* Systematic comparison of ranking aggregation methods for gene lists in experimental results. *bioRxiv* (2022).
15. Batra, R. *et al.* Multi-omic comparative analysis of COVID-19 and bacterial sepsis-induced ARDS. *PLoS Pathog.* **18**, e1010819 (2022).
16. Bhargava, M. *et al.* Proteomic profiles in acute respiratory distress syndrome differentiates survivors from non-survivors. *PLoS One* **9**, e109713 (2014).
17. Bhargava, M. *et al.* Bronchoalveolar lavage fluid protein expression in acute respiratory distress syndrome provides insights into pathways activated in subjects with different outcomes. *Sci. Rep.* **7**, 7464 (2017).
18. Bime, C. *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.* **197**, 1421–1432 (2018).

19. Bowler, R. P. *et al.* Proteomic analysis of pulmonary edema fluid and plasma in patients with acute lung injury. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **286**, L1095–104 (2004).
20. Chang, D. W. *et al.* Proteomic and computational analysis of bronchoalveolar proteins during the course of the acute respiratory distress syndrome. *Am. J. Respir. Crit. Care Med.* **178**, 701–709 (2008).
21. Chen, X., Shan, Q., Jiang, L., Zhu, B. & Xi, X. Quantitative proteomic analysis by iTRAQ for identification of candidate biomarkers in plasma from acute respiratory distress syndrome patients. *Biochem. Biophys. Res. Commun.* **441**, 1–6 (2013).
22. Chen, C., Shi, L., Li, Y., Wang, X. & Yang, S. Disease-specific dynamic biomarkers selected by integrating inflammatory mediators with clinical informatics in ARDS patients with severe pneumonia. *Cell Biol. Toxicol.* **32**, 169–184 (2016).
23. Christie, J. D. *et al.* Genome wide association identifies PPFIA1 as a candidate gene for acute lung injury risk following major trauma. *PLoS One* **7**, e28268 (2012).
24. Dolinay, T. *et al.* Inflammasome-regulated cytokines are critical mediators of acute lung injury. *Am. J. Respir. Crit. Care Med.* **185**, 1225–1234 (2012).
25. Dong, H. *et al.* Comparative analysis of the alveolar macrophage proteome in ALI/ARDS patients between the exudative phase and recovery phase. *BMC Immunol.* **14**, 25 (2013).
26. Englert, J. A. *et al.* Whole blood RNA sequencing reveals a unique transcriptomic profile in patients with ARDS following hematopoietic stem cell transplantation. *Respir. Res.* **20**, 15 (2019).
27. Frenzel, J. *et al.* Outcome prediction in pneumonia induced ALI/ARDS by clinical features and peptide patterns of BALF determined by mass spectrometry. *PLoS One* **6**, e25544 (2011).
28. Guillen-Guió, B. *et al.* Sepsis-associated acute respiratory distress syndrome in individuals of european ancestry: A genome-wide association study. *Lancet Respir. Med.* **8**, 258–266 (2020).
29. Howrylak, J. A. *et al.* Discovery of the gene signature for acute lung injury in patients with sepsis. *Physiol. Genomics* **37**, 133–139 (2009).
30. Jiang, Y. *et al.* Single cell RNA sequencing identifies an early monocyte gene signature in acute respiratory distress syndrome. *JCI Insight* **5**, (2020).
31. Juss, J. K. *et al.* Acute respiratory distress syndrome neutrophils have a distinct phenotype and are resistant to phosphoinositide 3-kinase inhibition. *Am. J. Respir. Crit. Care Med.* **194**, 961–973 (2016).
32. Kangelaris, K. N. *et al.* Increased expression of neutrophil-related genes in patients with early sepsis-induced ARDS. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **308**, L1102–13 (2015).
33. Kovach, M. A. *et al.* Microarray analysis identifies IL-1 receptor type 2 as a novel candidate biomarker in patients with acute respiratory distress syndrome. *Respir. Res.* **16**, 29 (2015).
34. Liao, S. Y. *et al.* Identification of early and intermediate biomarkers for ARDS mortality by multi-omic approaches. *Sci. Rep.* **11**, 18874 (2021).
35. Lu, X.-G. *et al.* Circulating miRNAs as biomarkers for severe acute pancreatitis associated with acute lung injury. *World J. Gastroenterol.* **23**, 7440–7449 (2017).
36. Martucci, G. *et al.* Identification of a circulating miRNA signature to stratify acute respiratory distress syndrome patients. *J. Pers. Med.* **11**, 15 (2020).

37. Meyer, N. J. *et al.* ANGPT2 genetic variant is associated with trauma-associated acute lung injury and altered plasma angiopoietin-2 isoform ratio. *Am. J. Respir. Crit. Care Med.* **183**, 1344–1353 (2011).
38. Meyer, N. J. *et al.* IL1RN coding variant is associated with lower risk of acute respiratory distress syndrome and increased plasma IL-1 receptor antagonist. *Am. J. Respir. Crit. Care Med.* **187**, 950–959 (2013).
39. Mirchandani, A. S. *et al.* Hypoxia shapes the immune landscape in lung injury and promotes the persistence of inflammation. *Nat. Immunol.* **23**, 927–939 (2022).
40. Morrell, E. D. *et al.* Cytometry TOF identifies alveolar macrophage subtypes in acute respiratory distress syndrome. *JCI Insight* **3**, (2018).
41. Morrell, E. D. *et al.* Alveolar macrophage transcriptional programs are associated with outcomes in acute respiratory distress syndrome. *Am. J. Respir. Crit. Care Med.* **200**, 732–741 (2019).
42. Nick, J. A. *et al.* Extremes of interferon-stimulated gene expression associate with worse outcomes in the acute respiratory distress syndrome. *PLoS One* **11**, e0162490 (2016).
43. Nguyen, E. V. *et al.* Proteomic profiling of bronchoalveolar lavage fluid in critically ill patients with ventilator-associated pneumonia. *PLoS One* **8**, e58782 (2013).
44. Ren, S. *et al.* Deleted in malignant brain tumors 1 protein is a potential biomarker of acute respiratory distress syndrome induced by pneumonia. *Biochem. Biophys. Res. Commun.* **478**, 1344–1349 (2016).
45. Scheller, N. *et al.* Proviral MicroRNAs detected in extracellular vesicles from bronchoalveolar lavage fluid of patients with influenza virus-induced acute respiratory distress syndrome. *J. Infect. Dis.* **219**, 540–543 (2019).
46. Shortt, K. *et al.* Identification of novel single nucleotide polymorphisms associated with acute respiratory distress syndrome by exome-seq. *PLoS One* **9**, e111953 (2014).
47. Wang, Z., Beach, D., Su, L., Zhai, R. & Christiani, D. C. A genome-wide expression analysis in blood identifies pre-elafin as a biomarker in ARDS. *Am. J. Respir. Cell Mol. Biol.* **38**, 724–732 (2008).
48. Tejera, P. *et al.* Distinct and replicable genetic risk factors for acute respiratory distress syndrome of pulmonary or extrapulmonary origin. *J. Med. Genet.* **49**, 671–680 (2012).
49. Xu, J.-Y. *et al.* Nucleotide polymorphism in ARDS outcome: A whole exome sequencing association study. *Ann. Transl. Med.* **9**, 780 (2021).
50. Zhang, S. *et al.* miR-584 and miR-146 are candidate biomarkers for acute respiratory distress syndrome. *Exp. Ther. Med.* **21**, 445 (2021).
51. Zhang, C. *et al.* Differential expression profile of plasma exosomal microRNAs in acute type a aortic dissection with acute lung injury. *Sci. Rep.* **12**, 11667 (2022).
52. Zhu, Z. *et al.* Whole blood microRNA markers are associated with acute respiratory distress syndrome. *Intensive Care Med. Exp.* **5**, 38 (2017).
53. Christopoulos, D. Introducing unit invariant knee (UIK) as an objective choice for elbow point in multivariate data analysis techniques. *SSRN Electron. J.* (2016).
54. Hu, Y. *et al.* BioLitMine: Advanced mining of biomedical and biological literature about human genes and genes from major model organisms. *G3 (Bethesda)* **10**, 4531–4539 (2020).
55. Quintanilla, E., Diwa, K., Nguyen, A., Vu, L. & Toby, I. T. A data report on the curation and development of a database of genes for acute respiratory distress syndrome. *Front. Genet.* **12**, 750568 (2021).

56. Lachmann, A. *et al.* Massive mining of publicly available RNA-seq data from human and mouse. *Nat. Commun.* **9**, 1366 (2018).
57. Page, M. J. *et al.* The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **372**, n71 (2021).
58. Clark, J. *et al.* A full systematic review was completed in 2 weeks using automation tools: A case study. *J. Clin. Epidemiol.* **121**, 81–90 (2020).
59. Battaglini, D. *et al.* Personalized medicine using omics approaches in acute respiratory distress syndrome to identify biological phenotypes. *Respir. Res.* **23**, 318 (2022).
60. Hernández-Beeftink, T., Guillen-Guió, B., Villar, J. & Flores, C. Genomics and the acute respiratory distress syndrome: Current and future directions. *Int. J. Mol. Sci.* **20**, 4004 (2019).
61. Reilly, J. P., Christie, J. D. & Meyer, N. J. Fifty years of research in ARDS. Genomic contributions and opportunities. *Am. J. Respir. Crit. Care Med.* **196**, 1113–1121 (2017).
62. Zheng, F. *et al.* Novel biomarkers for acute respiratory distress syndrome: Genetics, epigenetics and transcriptomics. *Biomark. Med.* **16**, 217–231 (2022).
63. Christopoulos, D. T. *Inflection: Finds the inflection point of a curve*. (2019).
64. Uhlen, M. *et al.* Towards a knowledge-based human protein atlas. *Nat. Biotechnol.* **28**, 1248–1250 (2010).
65. GTEx Consortium. The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
66. Carithers, L. J. *et al.* A novel approach to high-quality postmortem tissue procurement: The GTEx project. *Biopreserv. Biobank.* **13**, 311–319 (2015).
67. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
68. Uhlen, M. *et al.* A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. *Science* **366**, eaax9198 (2019).
69. Jain, A. & Tuteja, G. TissueEnrich: Tissue-specific gene enrichment analysis. *Bioinformatics* **35**, 1966–1967 (2019).
70. Raudvere, U. *et al.* G:profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).
71. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).
72. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
73. Gillespie, M. *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* **50**, D687–D692 (2022).
74. Martens, M. *et al.* WikiPathways: Connecting communities. *Nucleic Acids Res.* **49**, D613–D621 (2021).

75. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: A network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5**, e13984 (2010).
76. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).