

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

Supplement

Supplementary Methods

Search Strategy

We used the following strategy to search MEDLINE and a direct translation to search Embase.

1 exp Respiratory Distress Syndrome, Adult/

2 “acute lung injury*”.ti,ab,kf,kw

3 1 OR 2

4 “gene*”.mp

5 “genome*”.mp

6 “transcript*”.mp

7 “protein*”.mp

8 4 OR 5 OR 6 OR 7

9 3 AND 8

10 (“COVID-19*” OR “COVID19*” OR “COVID-2019*” OR “covid”).ti,ab,kf,kw

11 (“SARS-CoV-2*” OR “SARSCov-2*” OR “SARSCoV2*” OR “SARS-CoV2”).ti,sh,kf,kw

12 (“2019-nCoV*” OR “2019nCoV*” OR “19- nCoV*” OR “19nCoV*” OR “nCoV2019*” OR “nCoV-2019*” OR “nCoV19*” OR “nCoV- 19*”).ti,ab,kf,kw

13 10 OR 11 OR 12

14 9 NOT 13

15 Letter.pt OR Conference Abstract.pt OR Conference Paper.pt OR Conference Review.pt OR Editorial.pt OR Erratum.pt OR Review.pt OR Note.pt OR Tombstone.pt

16 14 NOT 15

17 exp *adolescence/ or exp *adolescent/ or exp *child/ or exp *childhood disease/ or exp *infant disease/ or (adolescen* or babies or baby or boy? or boyfriend or boyhood or girlfriend or girlhood or child or child* or child*3 or children* or girl? or infan* or juvenil* or juvenile* or kid? or minors or minors* or neonat* or neo-nat* or newborn* or new-born* or paediatric* or paediatric* or pediatric* or perinat* or preschool* or puber* or pubescen* or school* or teen* or toddler? or underage? or under-age? or youth*).ti,kw

18 16 NOT 17

19 ((exp animal/ or nonhuman/) NOT exp human/)

20 18 NOT 19

21 limit 20 to yr=“1967-Current”

Inclusion criteria

Inclusion:

- Human studies: *in-vivo* or *in-vitro*
- Adults (age \geq 18 years)
- Acute Respiratory Distress Syndrome (ARDS)
 - by any contemporaneous definition
- Accepted methodologies:
 - CRISPR screen
 - RNAi screen
 - Protein-protein interaction study
 - Host proteins incorporated into virion or virus-like particle
 - Genome wide association study
 - Transcriptomic study
 - Proteomic study

Exclusion:

- Children (age < 18 years)
- Animal studies
- Meta-analyses, *in-silico* analyses, or re-analysis of previously published data
- Excluded methodologies:
 - *In-vitro* human studies simulating ARDS
 - Candidate *in-vivo* or *in-vitro* transcriptomic or proteomic studies (defined as those investigating < 50 genes)
 - Candidate gene association studies
 - Studies including fewer than 5 individuals in either the control or ARDS arm

Supplementary Results

Supplementary Table 1. Details of included studies.**Supplementary Table 2. Gene list information content and contribution.**

Study	Method	Category	N genes	rIC (%)	rICtb (%)
Sarma ¹	Transcriptomics	RNA-seq	4954	50.8	53.1
Juss ²	Transcriptomics	Microarray	1318	16	15.7
Sarma ¹	Transcriptomics	scRNA-seq	706	9.8	10.3
Nguyen ³	Proteomics	Mass Spec	161	2.2	2.1
Wang ⁴	Transcriptomics	Microarray	137	1.9	1.9
Bhargava ⁵	Proteomics	Mass Spec	233	3.1	1.9
Kovach ⁶	Transcriptomics	Microarray	123	1.8	1.9
Bhargava ⁷	Proteomics	Mass Spec	144	1.9	1.8
Morrell ⁸	Transcriptomics	Microarray	155	1.9	1.7
Christie ⁹	GWAS	Genotyping	143	1.4	1.5
Liao ¹⁰	GWAS	Genotyping	67	0.7	0.8
Sarma ¹	Proteomics	Other	60	0.8	0.7
Jiang ¹¹	Transcriptomics	scRNA-seq	53	0.7	0.6
Batra ¹²	Proteomics	Other	39	0.6	0.6
Bime ¹³	GWAS	Genotyping	51	0.5	0.5
Bos ¹⁴	Transcriptomics	Microarray	53	0.7	0.5
Chang ¹⁵	Proteomics	Mass Spec	37	0.5	0.5
Mirchandani ¹⁶	Transcriptomics	Microarray	41	0.5	0.4
Mirchandani ¹⁶	Proteomics	Mass Spec	29	0.4	0.4
Liao ¹⁰	Transcriptomics	RNA-seq	43	0.4	0.4
Dong ¹⁷	Proteomics	Mass Spec	27	0.4	0.4
Ren ¹⁸	Proteomics	Other	17	0.3	0.3
Tejera ¹⁹	GWAS	Genotyping	19	0.3	0.3
Howrylak ²⁰	Transcriptomics	Microarray	28	0.3	0.2
Xu ²¹	GWAS	WES	16	0.2	0.2
Chen ²²	Proteomics	Mass Spec	16	0.2	0.2
Zhang ²³	Transcriptomics	RNA-seq	20	0.2	0.2
Kangelaris ²⁴	Transcriptomics	Microarray	15	0.2	0.2
Meyer ²⁵	GWAS	Genotyping	10	0.1	0.1
Martucci ²⁶	Transcriptomics	Microarray	13	0.1	0.1
Zhu ²⁷	Transcriptomics	Microarray	14	0.1	0.1
Englert ²⁸	Transcriptomics	RNA-seq	10	0.1	0.1
Lu ²⁹	Transcriptomics	Microarray	12	< 0.1	< 0.1
Scheller ³⁰	Transcriptomics	RNA-seq	9	< 0.1	< 0.1
Nick ³¹	Transcriptomics	Microarray	4	< 0.1	< 0.1
Guillen-Guio ³²	GWAS	Genotyping	6	< 0.1	< 0.1
Meyer ³³	GWAS	Genotyping	4	< 0.1	< 0.1
Dolinay ³⁴	Transcriptomics	Microarray	4	< 0.1	< 0.1

Study	Method	Category	N genes	rIC (%)	rICtb (%)
Chen ³⁵	Proteomics	Mass Spec	16	< 0.1	< 0.1
Zhang ³⁶	Transcriptomics	RNA-seq	5	< 0.1	< 0.1
Shortt ³⁷	GWAS	WES	3	< 0.1	< 0.1
Bowler ³⁸	Proteomics	Mass Spec	18	< 0.1	< 0.1
Morrell ³⁹	Transcriptomics	Microarray	1	< 0.1	< 0.1

Abbreviations: GWAS - Genome-wide association study; Mass Spec - Mass spectrometry; rIC - Relative information content; rICtb - Relative information contribution; WES - Whole-exome sequencing.

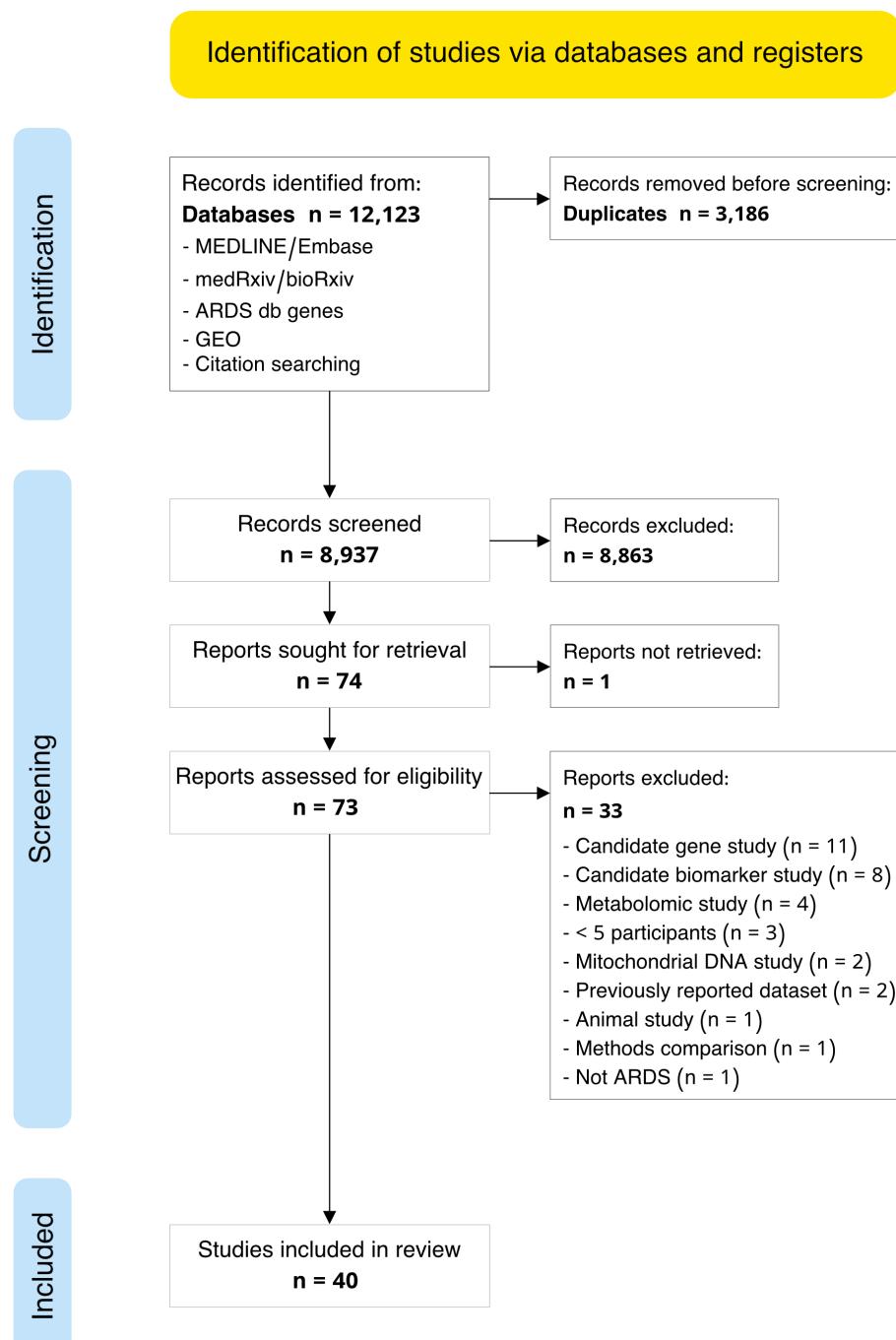


Figure 1: **Systematic review inclusion diagram.** Abbreviations: db - data base; GEO - NCBI Gene Expression Omnibus.

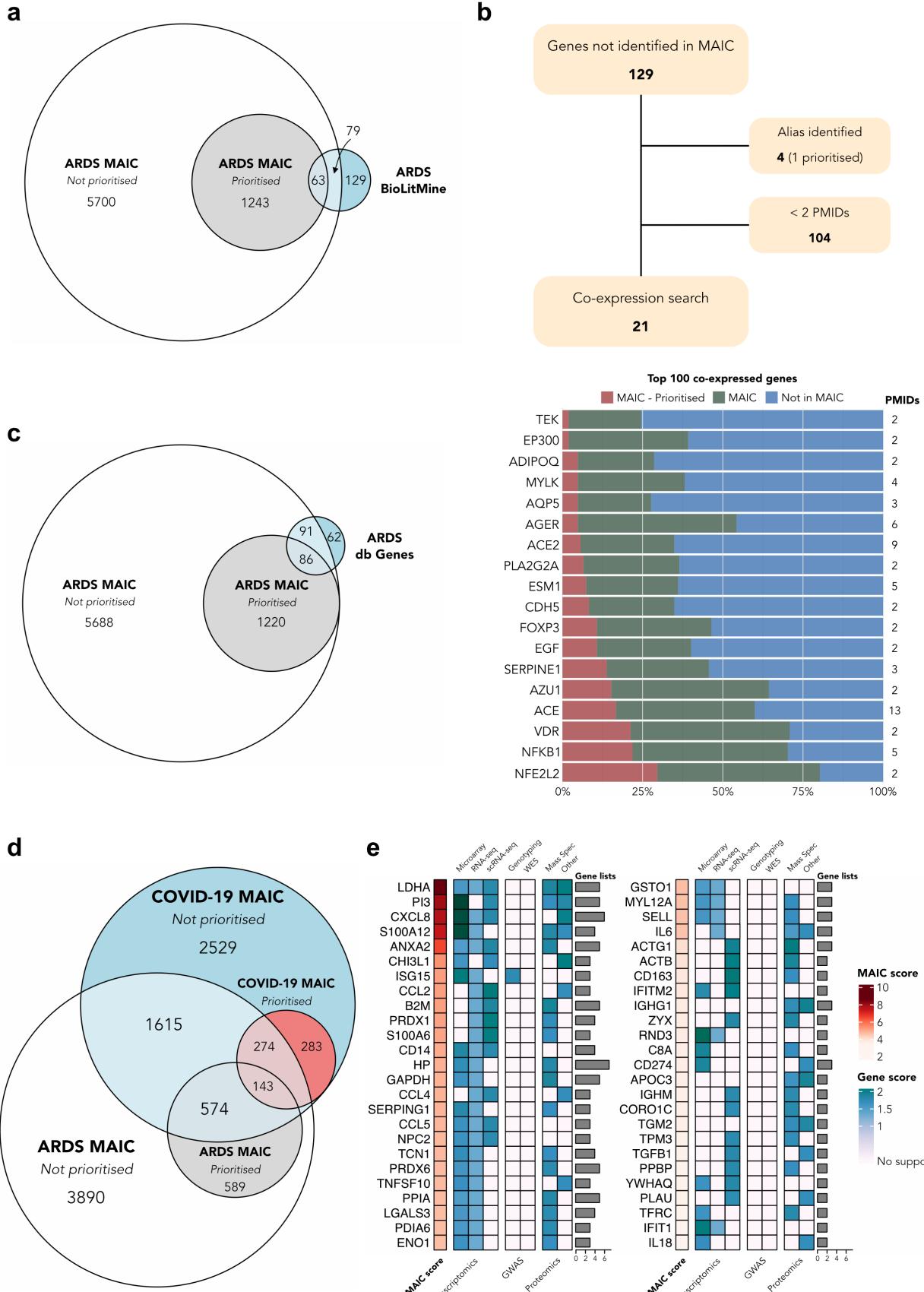


Figure 2: Overlap between ARDS MAIC and ARDS-associated genes and ARDS MAIC and COVID-19 MAIC.

(a) Euler diagram of gene overlap between ARDS MAIC and a BioLitMine search using the ARDS MeSH term. (b) Schematic overview of a co-expression search for genes identified in the BioLitMine search but not present in ARDS MAIC and a stacked bar plot of the proportion of the 100 most co-expressed genes of this group and ARDS MAIC. (c) Euler diagram of gene overlap between ARDS MAIC and the ARDS Database of Genes. (d) Euler diagram of gene overlap between ARDS MAIC and a MAIC of COVID-19 host-response studies. (e) Heatmap of the 50 top ranked ARDS MAIC genes also prioritised by the COVID-19 MAIC, displaying the ARDS MAIC score for each gene, highest gene score in each category, and the number of supporting gene lists.

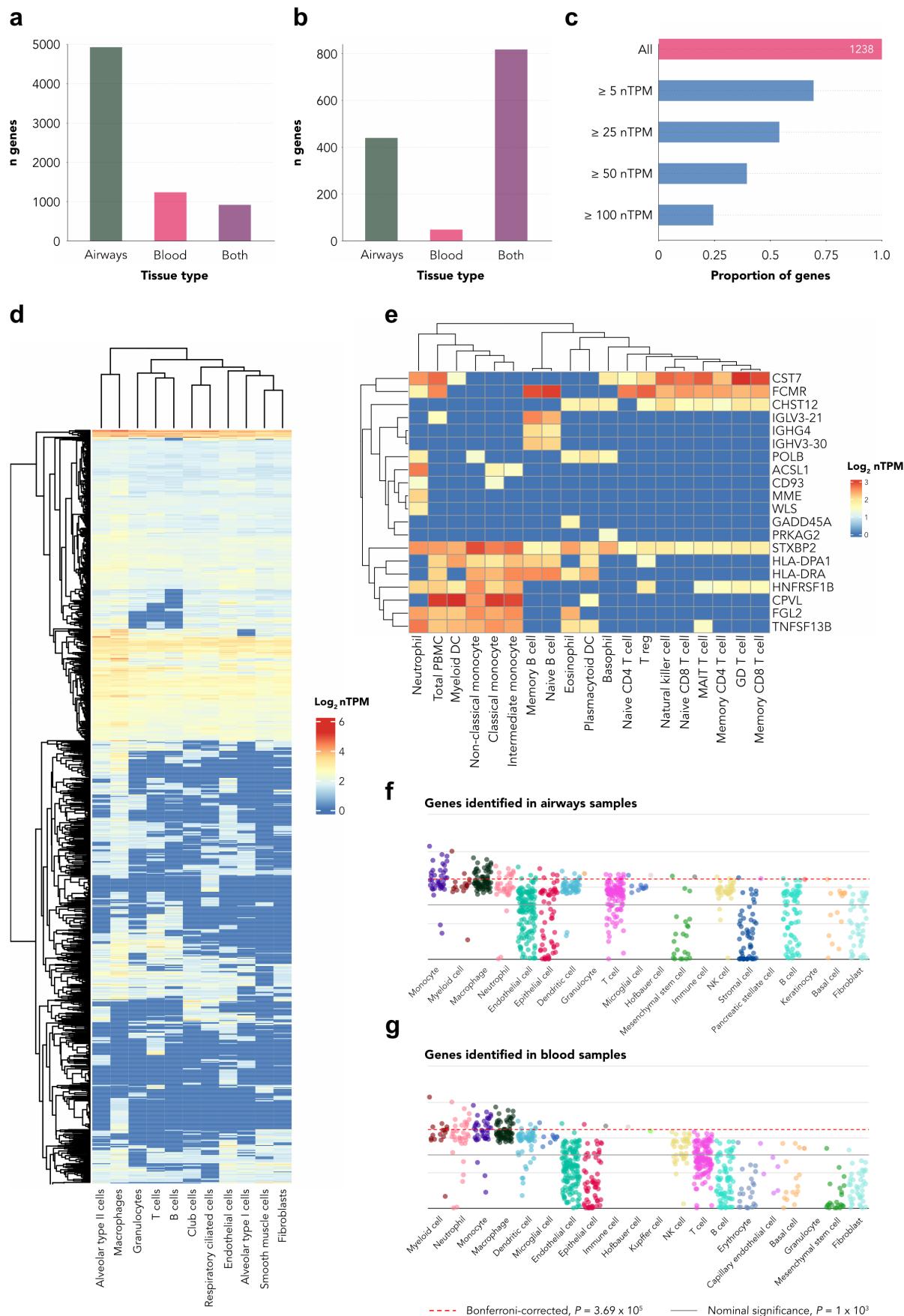


Figure 3: Tissue and cell-specific expression. (a) Bar plot of the tissue type in which genes are identified - all genes (n=7,085). (b) Bar plot of the tissue type in which genes are identified - prioritised genes (n=1,306). (c) Bar plot of the proportion of genes identified solely in blood meeting mRNA expression thresholds in bulk lung tissue. nTPM - normalised transcripts per million. (d) Heatmap of mRNA expression in lung cell-types for genes identified in studies based on airways sampling. (e) Heatmap of mRNA expression in blood cell-types for genes identified solely in studies based on blood sampling. (f) Manhattan plot of the top 20 cell types overenriched for expression of genes identified by studies based on airways sampling. (g) Manhattan plot of the top 20 cell types overenriched for expression of genes identified by studies based on blood sampling.

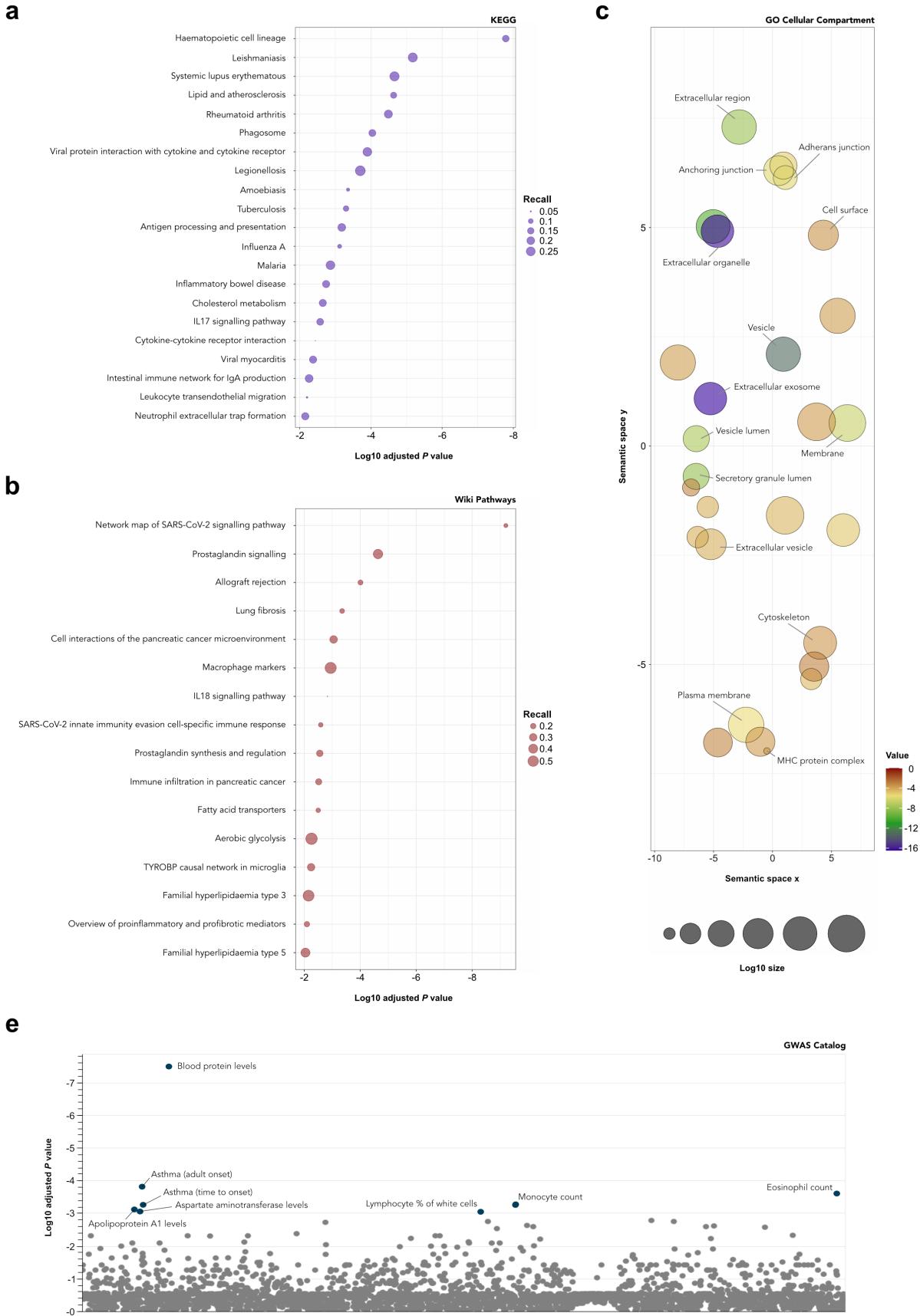


Figure 4: Functional enrichment. (a) Significantly enriched KEGG terms ($P < 0.01$) for prioritised genes. Terms size proportional to recall. (b) Significantly enriched WikiPathways terms ($P < 0.01$) for prioritised genes. Terms size proportional to recall. (c) Scatter plot of the semantic similarity between significantly enriched GO cellular component terms for prioritised genes (d) Manhattan plot of the overenrichment of prioritised genes against the GWAS catalog.

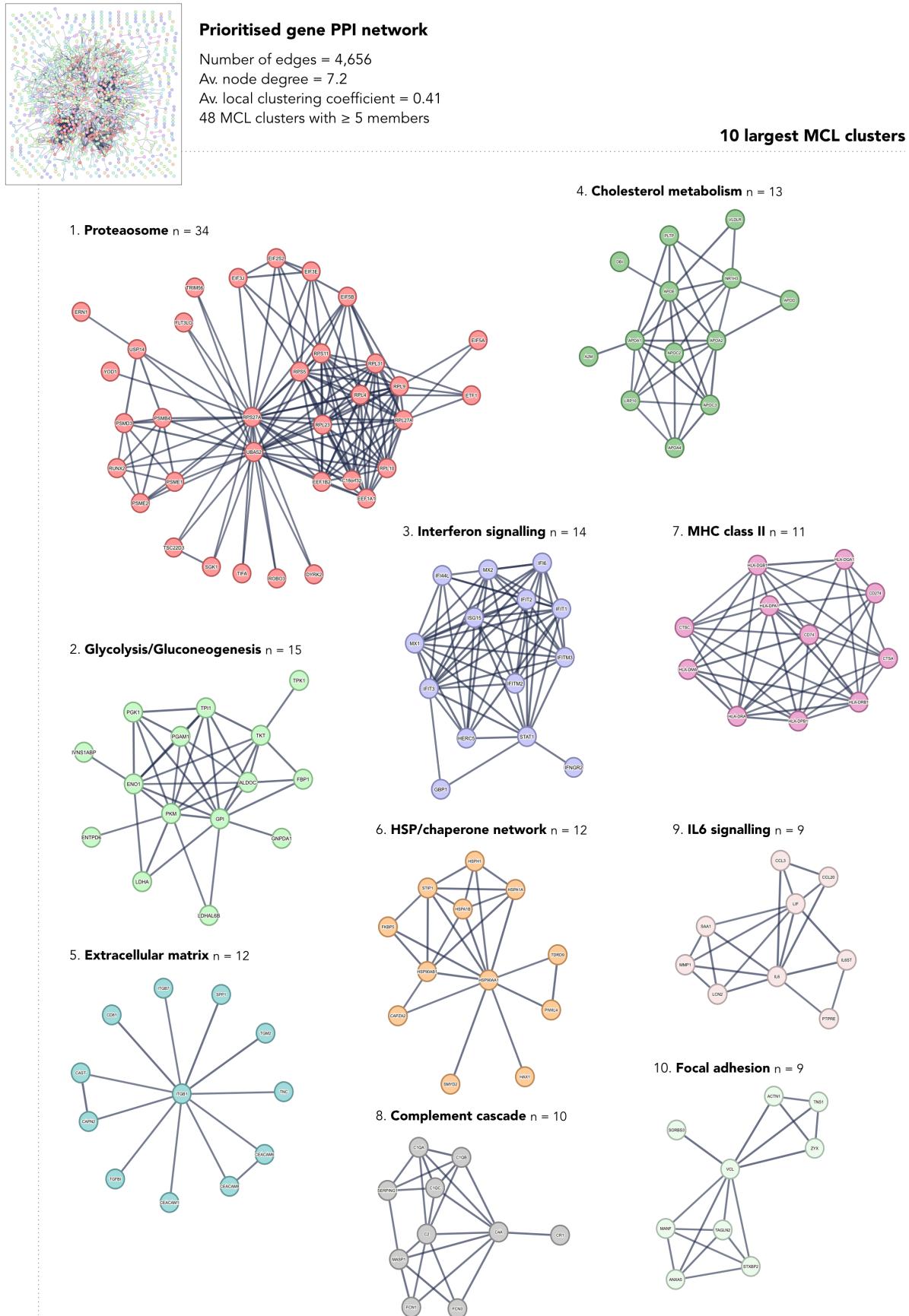


Figure 5: PPI clusters. A protein-protein interaction network of prioritised genes and the 10 largest graph-based clusters. Functional annotation by hand based on a concensus of enriched Reactome, KEGG, WikiPathways, and GO Biological Process terms.

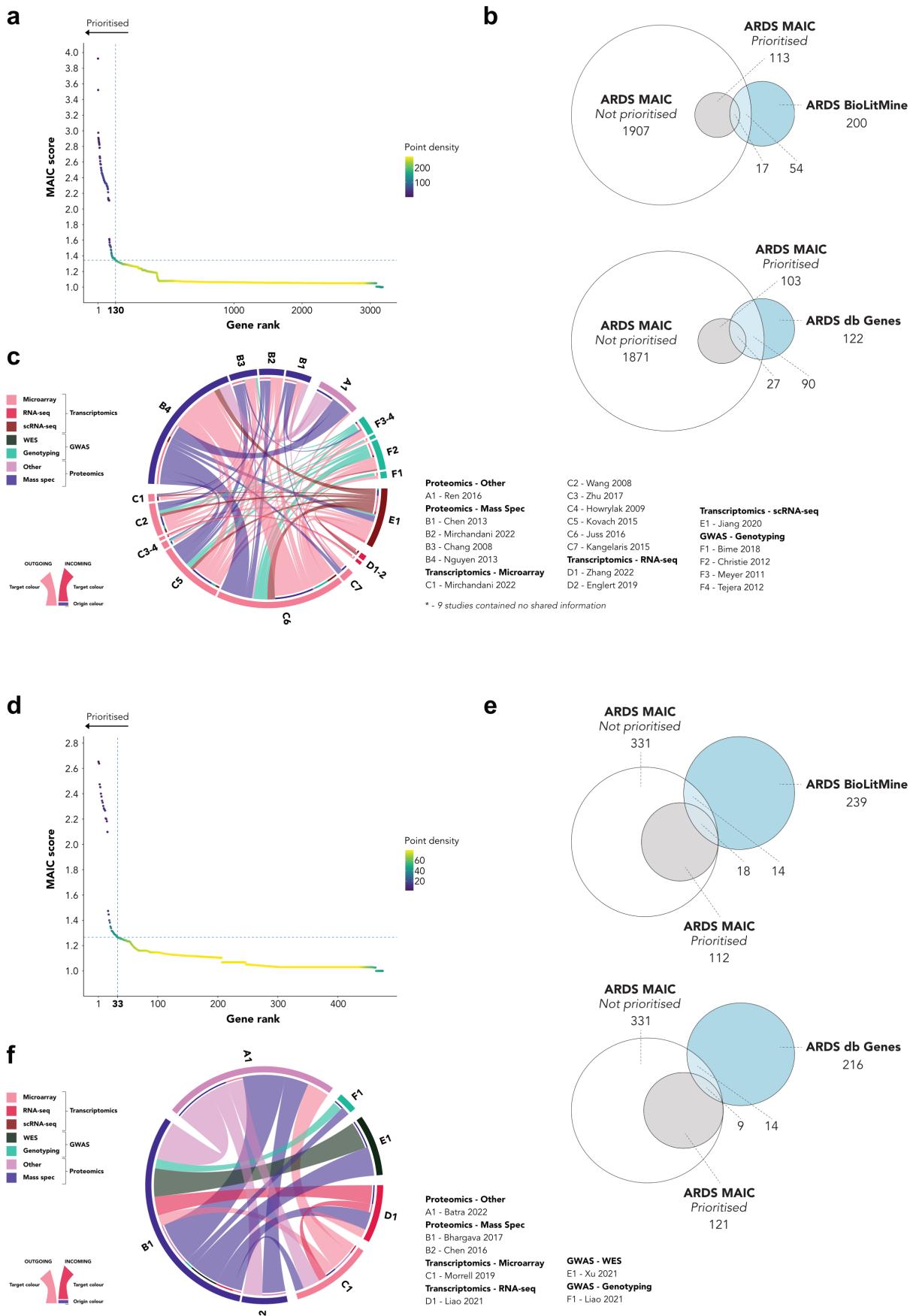


Figure 6: Details of MAIC on sub-groups. (a) Gene prioritization for the ARDS MAIC susceptibility sub-group using the Unit Invariant Knee method. Intersection of lines identifies elbow point of best-fit curve. 130 genes in upper left quadrant were prioritized. (b) Euler diagrams of gene overlap between the ARDS MAIC susceptibility sub-group and a BioLitMine search using the ARDS MeSH term and the ARDS Database of Genes. (c) Shared information content (IC) between susceptibility gene lists. Links indicate absolute IC (sum of common gene scores) between studies. (d) Gene prioritization for the ARDS MAIC survival sub-group using the Unit Invariant Knee method. Intersection of lines identifies elbow point of best-fit curve. 33 genes in upper left quadrant were prioritized. (e) Euler diagrams of gene overlap between the ARDS MAIC survival sub-group and a BioLitMine search using the ARDS MeSH term and the ARDS Database of Genes. (f) Shared information content (IC) between survival gene lists. Links indicate absolute IC (sum of common gene scores) between studies.

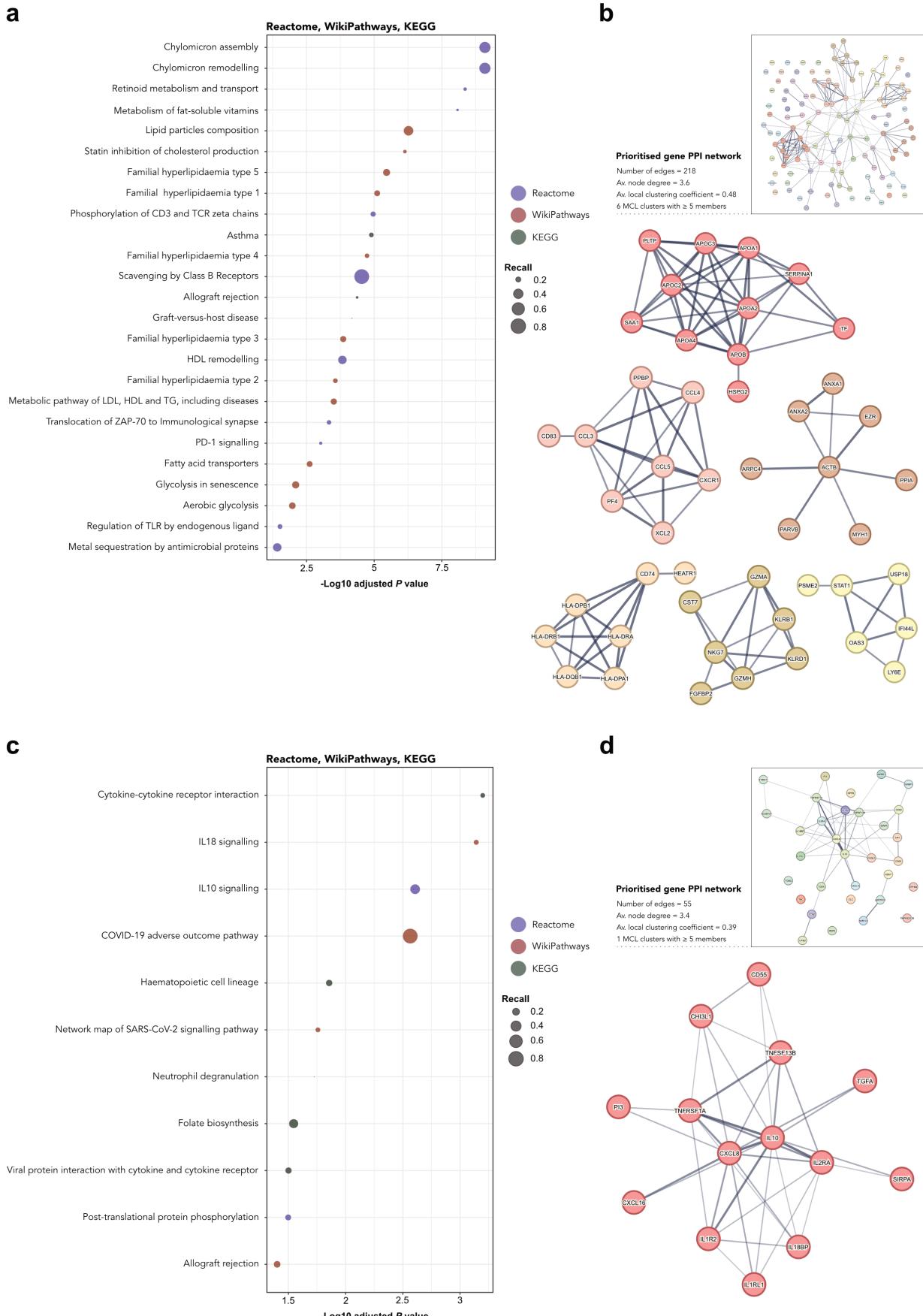


Figure 7: Sub-group functional enrichment. (a) Significantly enriched Reactome, WikiPathways, and KEGG terms ($P < 0.01$) for prioritised genes in the susceptibility sub-group. Terms are coloured by pathway and size is proportional to recall. (b) A protein-protein interaction network of prioritised genes in the susceptibility cohort and graph-based clusters with ≥ 5 members. (c) Significantly enriched Reactome, WikiPathways, and KEGG terms ($P < 0.01$) for prioritised genes in the survival sub-group. Terms are coloured by pathway and size is proportional to recall. (d) A protein-protein interaction network of prioritised genes in the survival cohort and graph-based clusters with ≥ 5 members.

Supplementary Table 3. ARDS MAIC prioritised genes found in common by BioLitMine with >= 2 associated publications.

Publication			MAIC
Gene	count	PubMed IDs	rank
TGFB1	8	30395619, 29083412, 28188225, 27309347, 22034170, 20142324, 16100012, 12654639	225
VEGFA	8	24356493, 23542734, 21797753, 19543148, 19349383, 17289863, 15920019, 15741444	320
IL10	8	32217834, 31936183, 30280795, 28432351, 22033829, 21138342, 18242340, 16585075	1268
SFTPB	6	21128671, 18679120, 16100012, 15190959, 14718442, 12490037	177
IL17A	6	34239039, 32795834, 32651218, 30655311, 26709006, 26002979	1294
PI3	5	28187039, 24617927, 19251943, 19197381, 18203972	2
CXCL8	5	22897124, 22080750, 21348591, 17498967, 14729508	3
IL6	5	34757857, 33250487, 32826331, 31261506, 18593632	144
TNF	5	31261506, 22507624, 21784970, 17034639, 16135717	651
NAMPT	4	24821571, 24053186, 18486613, 17392604	58
IL1RN	4	30095747, 29943912, 23449693, 18838927	175
SCGB1A1	4	32787812, 28548310, 18521628, 16215398	187
NPPB	4	28322314, 26359292, 21696613, 19830720	1239
HGF	3	18065658, 17702746, 11943656	343
IL33	3	33936076, 31147742, 23000728	385
CXCL10	3	31651197, 23542734, 23144331	671
S100A12	2	26274928, 24887223	5
MUC1	2	21418654, 17565019	69
PLAU	2	23064953, 17994220	244
EPAS1	2	28613249, 25574837	425
FASLG	2	30385692, 12414525	503
EDN1	2	27765761, 17875064	643
AKT1	2	27607575, 15961723	950
MMP8	2	24651234, 15187163	1223

Supplementary Table 4. ARDS susceptibility gene list information content and contribution.

Study	Method	Category	N genes	rIC (%)	rICtb (%)
Juss ²	Transcriptomics	Microarray	1318	54.7	54.7
Nguyen ³	Proteomics	Mass Spec	161	8.1	7.7
Christie ⁹	GWAS	Genotyping	143	6	6.3
Kovach ⁶	Transcriptomics	Microarray	123	5.8	6.1
Wang ⁴	Transcriptomics	Microarray	137	5.8	6
Jiang ¹¹	Transcriptomics	scRNA-seq	53	2.9	3
Bime ¹³	GWAS	Genotyping	51	2.2	2.3
Mirchandani ¹⁶	Transcriptomics	Microarray	41	1.7	1.6
Chang ¹⁵	Proteomics	Mass Spec	37	1.9	1.5
Mirchandani ¹⁶	Proteomics	Mass Spec	29	1.4	1.3
Howrylak ²⁰	Transcriptomics	Microarray	28	1.2	1.3
Ren ¹⁸	Proteomics	Other	17	1	1.1
Tejera ¹⁹	GWAS	Genotyping	19	0.9	1
Chen ³⁵	Proteomics	Mass Spec	16	0.9	0.9
Zhang ²³	Transcriptomics	RNA-seq	20	0.8	0.9
Zhu ²⁷	Transcriptomics	Microarray	14	0.6	0.6
Kangelaris ²⁴	Transcriptomics	Microarray	15	0.7	0.6
Englert ²⁸	Transcriptomics	RNA-seq	10	0.6	0.6
Lu ²⁹	Transcriptomics	Microarray	12	0.5	0.5
Meyer ²⁵	GWAS	Genotyping	10	0.4	0.4
Bowler ³⁸	Proteomics	Mass Spec	18	0.9	0.4
Scheller ³⁰	Transcriptomics	RNA-seq	9	0.4	0.3
Guillen-Guio ³²	GWAS	Genotyping	6	0.2	0.2
Zhang ³⁶	Transcriptomics	RNA-seq	5	0.2	0.2
Dolinay ³⁴	Transcriptomics	Microarray	4	0.2	0.2
Shortt ³⁷	GWAS	WES	3	0.1	0.1
Meyer ³³	GWAS	Genotyping	4	< 0.1	0.1
Morrell ³⁹	Transcriptomics	Microarray	1	< 0.1	< 0.1

Abbreviations: GWAS - Genome-wide association study; Mass Spec - Mass spectrometry; rIC - Relative information content; rICtb - Relative information contribution;

WES - Whole-exome sequencing.

Supplementary Table 5. ARDS survival/severity gene list information content and contribution.

Study	Method	Category	N genes	rIC (%)	rICtb (%)
Bhargava ⁷	Proteomics	Mass Spec	144	30.4	30.3
Morrell ⁸	Transcriptomics	Microarray	155	29.7	29.7
Liao ¹⁰	GWAS	Genotyping	67	12.9	13
Batra ¹²	Proteomics	Other	39	9.4	9.4
Liao ¹⁰	Transcriptomics	RNA-seq	43	8.5	8.5
Xu ²¹	GWAS	WES	16	3.5	3.5
Chen ²²	Proteomics	Mass Spec	16	3.4	3.4
Lu ²⁹	Transcriptomics	Microarray	12	2.2	2.2

Abbreviations: GWAS - Genome-wide association study; Mass Spec - Mass spectrometry; rIC - Relative information content; rICtb - Relative information contribution; WES - Whole-exome sequencing.

References

1. 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).
2. 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).
3. Nguyen, E. V. *et al.* Proteomic profiling of bronchoalveolar lavage fluid in critically ill patients with ventilator-associated pneumonia. *PLoS One* **8**, e58782 (2013).
4. 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).
5. Bhargava, M. *et al.* Proteomic profiles in acute respiratory distress syndrome differentiates survivors from non-survivors. *PLoS One* **9**, e109713 (2014).
6. 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).
7. 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).
8. 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).
9. 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).
10. Liao, S. Y. *et al.* Identification of early and intermediate biomarkers for ARDS mortality by multi-omic approaches. *Sci. Rep.* **11**, 18874 (2021).
11. Jiang, Y. *et al.* Single cell RNA sequencing identifies an early monocyte gene signature in acute respiratory distress syndrome. *JCI Insight* **5**, (2020).
12. Batra, R. *et al.* Multi-omic comparative analysis of COVID-19 and bacterial sepsis-induced ARDS. *PLoS Pathog.* **18**, e1010819 (2022).
13. 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).
14. 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).
15. 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).
16. 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).
17. 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).
18. 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).
19. 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).

20. Howrylak, J. A. *et al.* Discovery of the gene signature for acute lung injury in patients with sepsis. *Physiol. Genomics* **37**, 133–139 (2009).
21. Xu, J.-Y. *et al.* Nucleotide polymorphism in ARDS outcome: A whole exome sequencing association study. *Ann. Transl. Med.* **9**, 780 (2021).
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. 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).
24. 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).
25. 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).
26. Martucci, G. *et al.* Identification of a circulating miRNA signature to stratify acute respiratory distress syndrome patients. *J. Pers. Med.* **11**, 15 (2020).
27. Zhu, Z. *et al.* Whole blood microRNA markers are associated with acute respiratory distress syndrome. *Intensive Care Med. Exp.* **5**, 38 (2017).
28. 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).
29. 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).
30. 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).
31. 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).
32. 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).
33. 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).
34. Dolinay, T. *et al.* Inflammasome-regulated cytokines are critical mediators of acute lung injury. *Am. J. Respir. Crit. Care Med.* **185**, 1225–1234 (2012).
35. 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).
36. Zhang, S. *et al.* miR-584 and miR-146 are candidate biomarkers for acute respiratory distress syndrome. *Exp. Ther. Med.* **21**, 445 (2021).
37. Shortt, K. *et al.* Identification of novel single nucleotide polymorphisms associated with acute respiratory distress syndrome by exome-seq. *PLoS One* **9**, e111953 (2014).

38. 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).
39. Morrell, E. D. *et al.* Cytometry TOF identifies alveolar macrophage subtypes in acute respiratory distress syndrome. *JCI Insight* **3**, (2018).