Genome-wide association study of sepsis-associated acute respiratory distress syndrome in individuals of European ancestry

Beatriz Guillen-Guio, Jose M. Lorenzo-Salazar, Shwu-Fan Ma, Pei-Chi Hou, Tamara Hernandez-Beeftink, Almudena Corrales, M. Isabel García-Laorden, Jonathan Jou, Elena Espinosa, Arturo Muriel, David Domínguez, Leonardo Lorente, María M. Martín, Carlos Rodríguez-Gallego, Jordi Solé-Violán, Alfonso Ambrós, Demetrio Carriedo, Jesús Blanco, José M. Añón, John P. Reilly, Tiffanie K. Jones, Caroline A.G. Ittner, Rui Feng, Franziska Schöneweck, Michael Kiehntopf, Imre Noth, Markus Scholz, Frank M. Brunkhorst, André Scherag, Nuala J. Meyer, Jesús Villar, and Carlos Flores.

Appendix

The GEN-SEP cohort

GEN-SEP is a national, multicenter, observational study conducted in a Spanish network of 11 intensive care units (ICUs) between 2,002 and 2,017. The list of Spanish hospitals involved in this study are: Hospital Universitario de Canarias, Tenerife; Hospital Universitario NS de Candelaria, Tenerife; Hospital Universitario Río Hortega, Valladolid; Hospital Universitario Dr. Negrin, Gran Canaria; Hospital General de Ciudad Real, Ciudad Real; Complejo Hospitalario Universitario de León, León; Hospital Virgen de la Luz, Cuenca; Complejo Hospitalario Universitario de Santiago de Compostela, Santiago de Compostela; Fundació Althaia-Manresa, Barcelona; Hospital Clinic, Barcelona; and Hospital Clínico de Valladolid, Valladolid.

A total of 672 patients with sepsis[1] were included in this stage and diagnosed with ARDS based on Berlin definition criteria:[2] 1) acute onset with PaO2/FiO2 <300 mmHg, 2) bilateral pulmonary infiltrates on frontal chest radiograph, 3) use of invasive mechanical ventilation, and 4) no evidence of cardiac failure. Controls were those sepsis patients that did not develop ARDS during their ICU stay.

Four ml of peripheral blood were withdrawn at the time of inclusion into the study and stored at -20ºC until use. DNA was extracted using the Illustra™ blood genomicPrep Mini Spin Kit (GE Healthcare), quantified with a Qubit 3.0 fluorometer (Thermo Fisher Scientific), and stored at -20ºC until use. Samples with a low concentration of DNA (<10 ng/µl) were cleaned and concentrated using the RealClean & concentrator microspin kit (Real Laboratory).

Genotyping and quality control in discovery and replication stages

For the discovery stage, a total of 587,352 SNPs were genotyped in the National Genotyping Centre (CeGen) using the Axiom Genome-Wide Human CEU 1 Array (Affymetrix). Variant calling was performed in a single batch for all samples using AffyPipe[3] following the authors' recommendations to fine tune the filtering of low quality SNPs and samples. PLINK v1.90[4] and R 3.3.2[5] tools were used to conduct quality controls. Samples with missing clinical information, genotype call rates < 95%, sex mismatches between genotypes and the clinical data, samples with high degree of kinship (PIHAT>0·2), and heterozygosity outliers were removed. Variants with low minor allele frequency (MAF<0·01), genotype call rates (CR) < 95%, or deviated from Hardy Weinberg equilibrium expectations (HWE, $p<1{\cdot}0x10^{-6}$) were excluded. Additionally, a Principal Component (PC) analysis (PCA) was conducted to reduce the effects of population stratification in the analysis. For this purpose, we removed SNPs located at known regions that are in long-distance linkage disequilibrium (LD). We then pruned SNPs in high LD using the function "*indep-pairwise*" of PLINK, setting a r^2 of 0·15 to keep approximately 100,000 independent variants. After excluding eight ancestry outliers, the two first PCs for the discovery sample were plotted overlaid with the HapMap3 populations.[6] The PCA evidenced the similarity between the GEN-SEP samples included in the discovery stage and the European population from HapMap (Supplementary Figure 1).

In the MESSI study, SNPs were genotyped using the Affymetrix Axiom TxArray v.1 (Affymetrix). As described elsewhere,[7] variants were filtered if they were located on sex chromosomes, had a MAF<5%, had a missing genotype rate of >10%, or if deviated from HWE ($p<1{\cdot}0x10^{-3}$). In the SepNet study, HumanOmniExpressExome arrays (Illumina, Inc.) were used for variant genotyping. As described elsewhere,[8] individuals with sex mismatches, missing sex records, CR<98%, implausible heterozygosity

(<20% and >26%), and ancestry outliers based on the PCA were removed. Variants with CR≤95%, MAF<1%, or deviated from HWE (p<1·0x10^{-6}) were also excluded.

Statistical analyses for discovery stage

For variant imputation, phasing was conducted with SHAPEIT v2.r790[9] and the Haplotype Reference Consortium (HRC) version r.1.1 data were used as the reference panel[10] on the Michigan Imputation Server[11]. Logistic regression models assuming an additive inheritance were carried out using EPACTS v3.2.6[12] based on the Wald test. We included sex, age, and the Acute Physiology and Chronic Health Evaluation II (APACHE II) score as covariates. For the variants in the X chromosome, variant imputation and association tests were conducted separately in males and females, and results were subsequently meta-analysed with METASOFT v2.0.1[13]. Fixed-effects or Han and Eskin's Random Effects models were used depending on the significance of the Cochran's Q statistic. Variants with low allele frequency (MAF<1%) or with a low imputation quality (Rsq<0·3) were excluded from the analysis. The genomic inflation factor (λ) of the results was calculated with the GenABEL package v1.8-0.[14]

GCTA-COJO 1.26.0[15] was used for conditional regression analyses to identify independent loci taking into account the underlying LD structure in the study sample. Independent variants showing a p<5·0x10^{-5} were followed up in the replication stage. Regional association plots were generated using LocusZoom[16] based on LD information of European populations from the 1000 Genomes Project (1KGP)[17] and gene information from the UCSC browser data.

Statistical analyses for replication stage

Pre-phasing and variant imputation in MESSI were conducted with MACH v1.0,[18] using the European population from 1KGP Phase 1 v2 as the reference panel. Logistic regression models were performed assuming additive inheritance using R 3.3.2 stats package (glm function, binomial distribution),[5] considering the first two PCs, age, and sex as covariates. All the investigated variants in MESSI had a MAF >1% and Rsq >0·3. As described elsewhere,[8] for the SepNet study SHAPEIT v2.r790[9] was utilized for pre-phasing, and IMPUTE2 v2.3.1[19] was used for variant imputation considering the 1KGP Phase 1 v3 data as the reference panel. Logistic regressions were performed with SNPTEST v2.5,[19] which included the first three PCs, sex, age, and APACHE II as covariates. All the assessed variants in SepNet had MAF >1%, no evidence for deviations from HWE (p>1·0x10^{-10}) and an INFO Score >0·8. To combine the results from MESSI and SepNet, a meta-analysis was assessed using METASOFT v2.0.1.[13]

Meta-analysis of discovery and replication stages

A meta-analysis across discovery and replication stages was performed with METASOFT v2.0.1[13] to estimate the overall effect size of the SNPs reaching nominal significance in replication stage. Fixed-effects or Han and Eskin's Random-effects models were used based on the Cochran's Q test significance. Genome-wide significance was declared with a meta-analysed significance of p<9·26x10^{-8} according to the most recent empirical estimations in European populations.[20] The same approximation was used for the sensitivity analysis of the association of the sentinel variant.

Statistical power

Statistical power was estimated using the Genetic Association Study (GAS) Power Calculator.[21] We assumed a multiplicative model, a GWAS with a sample size of 630 cases and 1,302 controls, a relative risk of 1·5 and a prevalence of 0·1, the study had 80·4% statistical power for detecting associated variants with MAF of 0·30 or greater at significance level of p<9·26x10^{-8}.

FLT1 and *VEGFA* gene expression and functional annotation of genetic variants

Total RNAs from nine lung biopsies of healthy individuals obtained from the Gift of Hope Network Regional Organ Bank of Il (GOH/ROBI) were isolated and subjected to RNA-sequencing. Expression levels of *FLT1* and *VEGFA* were expressed in counts per million (Shwu-Fan Ma and Imre Noth, personal communication). Additionally, the ExAtlas tool[22] was used to explore public gene expression data available (GSE32707) from a peripheral blood transcriptomics analysis in 88 critically-ill adult patients that were evaluated for sepsis and ARDS.[23] For this analysis we used ANOVA followed by pairwise Student's *t*-tests to assess the differences in average intensities of the array probe targeting *FLT1* (ILMN_1752307, which targets exon 30, that is found in the canonical isoform FLT1-201) and *VEGFA* (ILMN_2375879, ILMN_1693060, and ILMN_1803882) between ICU controls (n=34), systemic inflammatory response syndrome (SIRS, n=21), sepsis (n=30), and sepsis-associated ARDS patients (n=18) at study inclusion. We report uncorrected two-sided *p*-values.

Next, we used a combination of tools and datasets to evaluate the regulatory potential of the associated variants in gene expression (through epigenetic mechanisms, long-distance physical interactions, and tissue-specific cis-eQTLs), and the likelihood of deleteriousness. These included Capture Hi-C Plotter (CHiCP),[24] DeepSea,[25] DSNetwork,[26] GTEx Analysis Release V7,[27] HaploReg v4.1,[28] Open Targets Genetics,[29] RegulomeDB,[30] SNPdelScore,[31] TIVAN,[32] and Variant Effect Predictor (VEP).[33]

CHiCP allows for the determination of empirically-observed physical interactions between distal DNA regulatory elements and gene promoters in multiple tissues. DeepSea predicts the epigenetics state of a sequence and prioritize regulatory variants by calculating functional significance scores, while DSNetwork allows for the selection of the most probable functional SNP from a list of variants according to nearly sixty prediction approaches. The GTEx Portal allows for the study of Single-Tissue expression quantitative trait loci (eQTL) and tissue-specific gene expression and regulation. HaploReg, Open Targets Genetics, and RegulomeDB explore annotations of coding and non-coding variants integrating data from chromatin states, regulatory motifs, eQTLs, pQTLs, DNase I hypersensitive sites, enhancer-transcription start sites, and promoter capture Hi-C experiments from different cell lines. Open Targets Genetics puts functional information in the context of the UK Biobank association evidence, allowing one to link each variant to its proximal and distal target gene(s), using a single evidence score. SNPdelScore combines different methods to address deleterious effects of noncoding variants, including the CellulAr dePendent dEactivating (CAPE) mutations predictor. Finally, TIVAN allows for the prediction of tissue-specific cis-eQTL single nucleotide variants, and VEP determines the effect of the variants analysed on genes, transcripts, protein sequence, and regulatory regions.

Constructs, transient transfections, and dual-luciferase reporter assays

A Dual-Luciferase Reporter Assay System® (Promega, Madison, WI) was used to evaluate the potential regulatory effect of the ARDS-associated variant on promoter activity. The reporter construct was generated by synthesizing (GenScript Inc, Piscataway, NJ) a fragment of 1,032 bp of the *FLT1* promoter (Ensembl ID: ENSR00000060438) plus 284 bp of the 5' UTR of exon one and 784 bp of the upstream sequence (chr13: 29,068,982-29,070,013; GRCh37/hg19 coordinates), and inserting it into a promoterless pGL4.10 [luc2] luciferase reporter vector (Promega). This *FLT1* promoter region was chosen for having the highest activity *in vitro* in a previous characterization of the gene promoter.[34] In addition, two regulatory constructs were generated by synthesizing (GenScript Inc, Piscataway, NJ) a 1·9 kb intron 10 fragment containing either the reference or alternative alleles of the most significantly associated variant within *FLT1* and its perfect LD proxies (i.e. rs9508032, rs9508034, rs722503, rs8002446, rs9513111, r^2=1·0) in Europeans (chr13: 28,995,800-28,997,700; GRCh37/hg19 coordinates) and inserting them into the reporter construct.

The constructed plasmids (50 ng DNA each) and the control plasmid pGL4.74 [hRluc/TK] (10 ng DNA) were transiently co-transfected into human lung epithelial (A549) or peripheral blood monocyte (THP-1) cell lines using the TransIT®-LT1 Transfection Reagent (Mirus Bio LLC, Madison, WI) following manufacturer's protocol. A549 and THP-1 cells were separately grown in 10% DMEM or 10% RPM 1640 media, respectively, and were plated into white 96-well plates until confluency. Twenty-four hours after transfection, cells were collected and luminescence was measured by Dual-Luciferase Reporter Assay System according to manufacturer's instructions using a Cytation5 plate reader (BioTek, Winooski, VT). Luminescence experiments were performed four to eight times, with each transfection in triplicate. Following manufacturer's instructions,[35] to reduce variability, simplify comparisons and improve significance, promoter activities were expressed as the relative response ratio of *Firefly* luciferase/Renilla luciferase luminescence according to the formula:

$$Relative\ response\ ratio = \frac{(experimental\ sample\ ratio) - (negative\ control\ ratio)}{(positive\ control\ ratio) - (negative\ control\ ratio)}$$

We considered the construct including only the *FLT1* promoter as the positive control and the promoterless construct as the negative control (see figure 4A). Mean differences among the independent experimental groups were assessed by non-parametric Wilcoxon signed-rank test. Again, we report uncorrected two-sided *p*-values.

Literature mining of previously reported ARDS-associated genes

In order to evaluate genes that were previously associated with ARDS, we conducted a bibliographic search on PubMed for all studies reporting genes which were significantly associated with ARDS from December 2015 to November 2018. This updated result was merged with a list of all published studies we collected up to December 2015 available elsewhere.[36–38] For that search, we used combinations of the terms "acute respiratory distress syndrome", "ARDS" OR "acute lung injury" with "polymorphism" OR "genetic variant" and retrieved seven publications reporting five additional candidate ARDS genes in adults. Association results in the discovery stage were extracted and an effective number of independent signals per gene was measured using the Genetic Type I Error Calculator[39] in order to adjust for multiple testing. Significant association was declared if any of the individual variants surpassed one of two Bonferroni-corrected significant levels. We considered a study-wise adjustment accounting for all the independent tests across all genes, and a gene-wise adjustment just accounting - i.e. adjusting for the independent variants mapping at individual genes.

Evaluation of *FLT1* variants in a trauma-associated ARDS cohort

We evaluated if the *FLT1* sentinel variant and perfect proxies also associated with non-sepsis ARDS. For that, we accessed the table S2 of the only GWAS of trauma-associated ARDS published to date,[40] containing publicly available (but incomplete) summary data. We found that none of the *FLT1* variants that achieved genome-wide significance in sepsis-associated ARDS were present in that study because of the reference panel used for variant imputation. Despite this, it was reassuring to find that out of 13 *FLT1* SNPs listed (all within a region of 31 kb and showing nominally significant associations with ARDS after trauma), six were also located in intron 10 (*p*-values in the range of $9{\cdot}15x10^{-4}$ to $2{\cdot}44x10^{-3}$). However, their LD with the sentinel variant of our study was weak in Europeans (r^2=0·13).

References

1 Singer M, Deutschman CS, Seymour CW, *et al.* The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016; **315**: 801–10.

2 ARDS Definition Task Force, Ranieri VM, Rubenfeld GD, *et al.* Acute respiratory distress syndrome: the Berlin Definition. *JAMA* 2012; **307**: 2526–33.

3 Nicolazzi EL, Iamartino D, Williams JL. AffyPipe: an open-source pipeline for Affymetrix Axiom genotyping workflow. *Bioinformatics* 2014; **30**: 3118–9.

4 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; **4**: 7.

5 R Core Team. R: A language and environment for statistical computing. *R Found Stat Comput Vienna, Austria* 2013. http://www.r-project.org/. Accessed September 2017.

6 International HapMap 3 Consortium, Altshuler DM, Gibbs RA, *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* 2010; **467**: 52–8.

7 Reilly JP, Wang F, Jones TK, *et al.* Plasma angiopoietin-2 as a potential causal marker in sepsis-associated ARDS development: evidence from Mendelian randomization and mediation analysis. *Intensive Care Med* 2018; **44**: 1849–58.

8 Scherag A, Schöneweck F, Kesselmeier M, *et al.* Genetic Factors of the Disease Course after Sepsis: A Genome-Wide Study for 28Day Mortality. *EBioMedicine* 2016; **12**: 239–46.

9 Delaneau O, Howie B, Cox AJ, Zagury J-F, Marchini J. Haplotype estimation using sequencing reads. *Am J Hum Genet* 2013; **93**: 687–96.

10 McCarthy S, Das S, Kretzschmar W, *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016; **48**: 1279–83.

11 Das S, Forer L, Schönherr S, *et al.* Next-generation genotype imputation service and methods. *Nat Genet* 2016; **48**: 1284–7.

12 Kang H. Efficient and parallelisable association container toolbox (EPACTS). http://genome.sph.umich.edu/wiki/EPACTS 2014. Accessed October 2017.

13 Han B, Eskin E. Random-Effects Model Aimed at Discovering Associations in Meta-Analysis of Genome-wide Association Studies. *Am J Hum Genet* 2011; **88**: 586–98.

14 Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 2007; **23**: 1294–6.

15 Zhu Z, Zheng Z, Zhang F, *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* 2018; **9**: 224.

16 Pruim RJ, Welch RP, Sanna S, *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010; **26**: 2336–7.

17 1000 Genomes Project Consortium RA, Auton A, Brooks LD, *et al.* A global reference for human genetic variation. *Nature* 2015; **526**: 68–74.

18 Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 2010; **34**: 816–34.

19 Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; **39**: 906–13.

20 Kanai M, Tanaka T, Okada Y. Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. *J Hum Genet* 2016; **61**: 861–6.

21 Johnson JL, Abecasis GR. GAS Power Calculator: web-based power calculator for genetic

association studies. *bioRxiv* 2017. Doi: https://doi.org/10.1101/164343.

22 Sharov AA, Schlessinger D, Ko MSH. ExAtlas: An interactive online tool for meta-analysis of gene expression data. *J Bioinform Comput Biol* 2015; **13**: 1550019.

23 Dolinay T, Kim YS, Howrylak J, *et al.* Inflammasome-regulated cytokines are critical mediators of acute lung injury. *Am J Respir Crit Care Med* 2012; **185**: 1225–34.

24 Schofield EC, Carver T, Achuthan P, *et al.* CHiCP: a web-based tool for the integrative and interactive visualization of promoter capture Hi-C datasets. *Bioinformatics* 2016; **32**: 2511–3.

25 Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat Methods* 2015; **12**: 931–4.

26 Lemaçon A, Scott-Boyer M-P, Soucy P, Ongaro-Carcy R, Simard J, Droit A. DSNetwork: An integrative approach to visualize predictions of variants' deleteriousness. *bioRxiv 526335* 2019. Doi:https://doi.org/10.1101/526335.

27 GTEx Consortium J, Thomas J, Salvatore M, *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013; **45**: 580–5.

28 Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012; **40**: D930-4.

29 Carvalho-Silva D, Pierleoni A, Pignatelli M, *et al.* Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res* 2019; **47**: D1056–65.

30 Boyle AP, Hong EL, Hariharan M, *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012; **22**: 1790–7.

31 Vera Alvarez R, Li S, Landsman D, Ovcharenko I. SNPDelScore: combining multiple methods to score deleterious effects of noncoding mutations in the human genome. *Bioinformatics* 2017; **34**: 289–91.

32 Chen L, Wang Y, Yao B, Mitra A, Wang X, Qin X. TIVAN: Tissue-specific cis-eQTL single nucleotide variant annotation and prediction. *Bioinformatics* 2018; published online Oct 10. Doi:10.1093/bioinformatics/bty872.

33 McLaren W, Gil L, Hunt SE, *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* 2016; **17**: 122.

34 Morishita K, Johnson DE, Williams LT. A novel promoter for vascular endothelial growth factor receptor (flt-1) that confers endothelial-specific gene expression. *J Biol Chem* 1995; **270**: 27948–53.

35 Eggers C, Hook B, Lewis S, Strayer C, Landreman A. Designing a Bioluminescent Reporter Assay: Normalization. *Promega Corp. Web site.* http://www.promega.es/resources/pubhub/designing-a-bioluminescent-reporter-assay-normalization/ 2016. Accessed April 23, 2019.

36 Flores C, Pino-Yanes M del M, Villar J. A quality assessment of genetic association studies supporting susceptibility and outcome in acute lung injury. *Crit Care* 2008; **12**: R130.

37 Acosta-Herrera M, Pino-Yanes M, Perez-Mendez L, Villar J, Flores C. Assessing the quality of studies supporting genetic susceptibility and outcomes of ARDS. *Front Genet* 2014; **5**: 20.

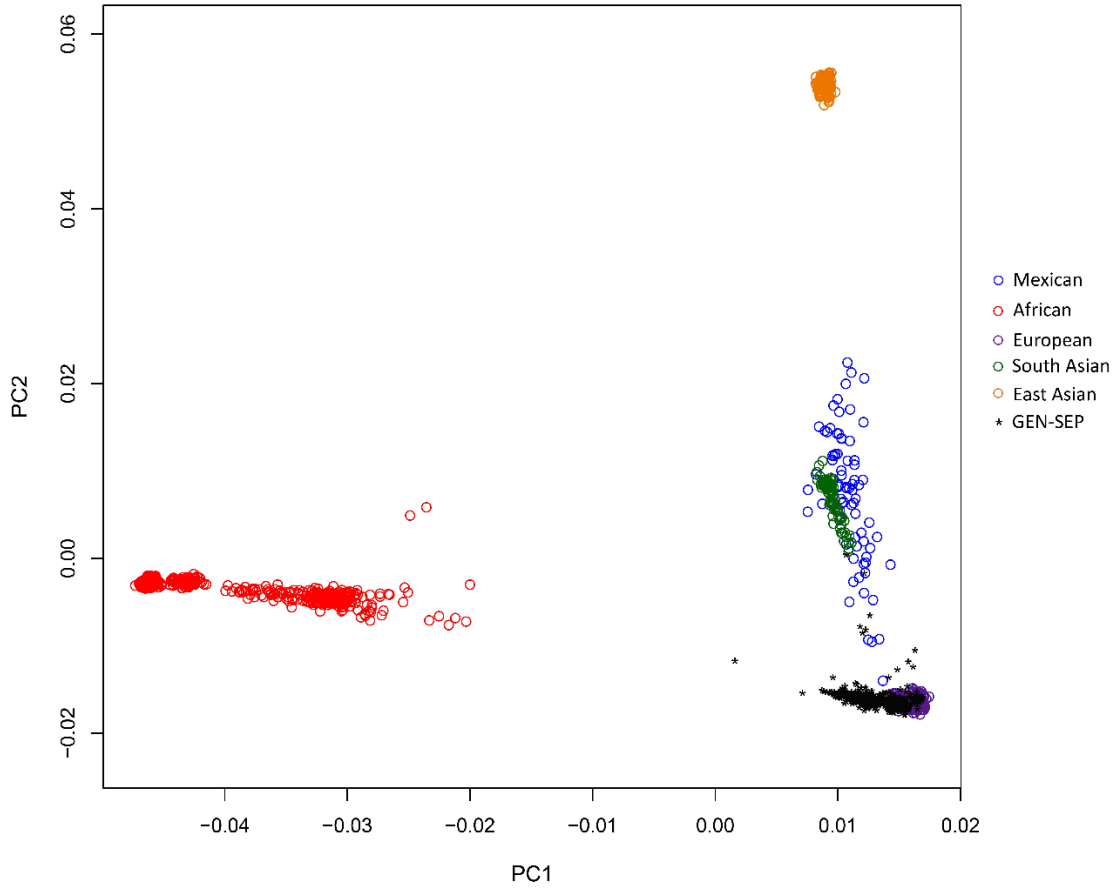
38 Guillén-Guío B, Acosta-Herrera M, Villar J, Flores C. Genetics of Acute Respiratory Distress Syndrome. *eLS. John Wiley Sons* 2016. DOI: 10.4046/trd.2001.51.1.5

39 Li M-X, Yeung JMY, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum Genet* 2012; **131**: 747–56.

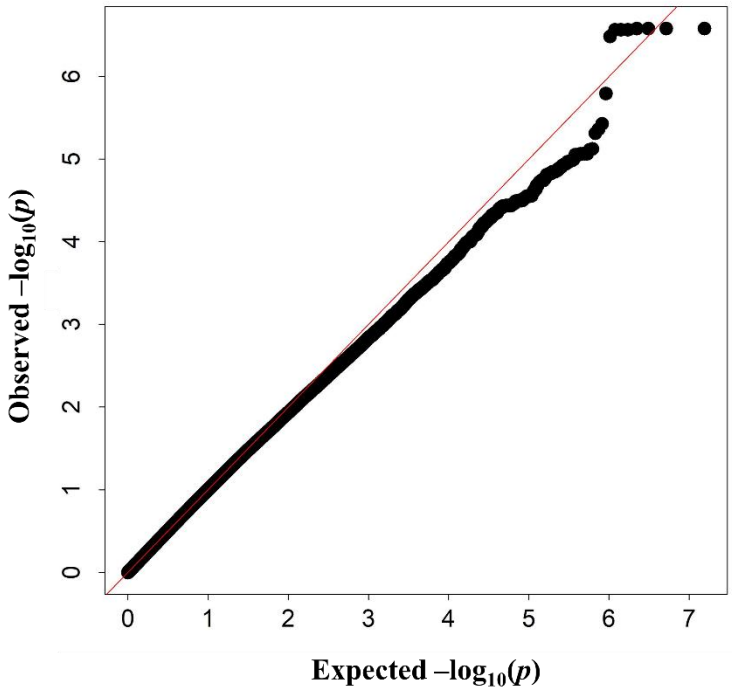
40 Christie JD, Wurfel MM, Feng R, et al. Genome wide association identifies PPFIA1 as a candidate gene for acute lung injury risk following major trauma. PLoS One 2012; **7**: e28268.

Supplementary Figures

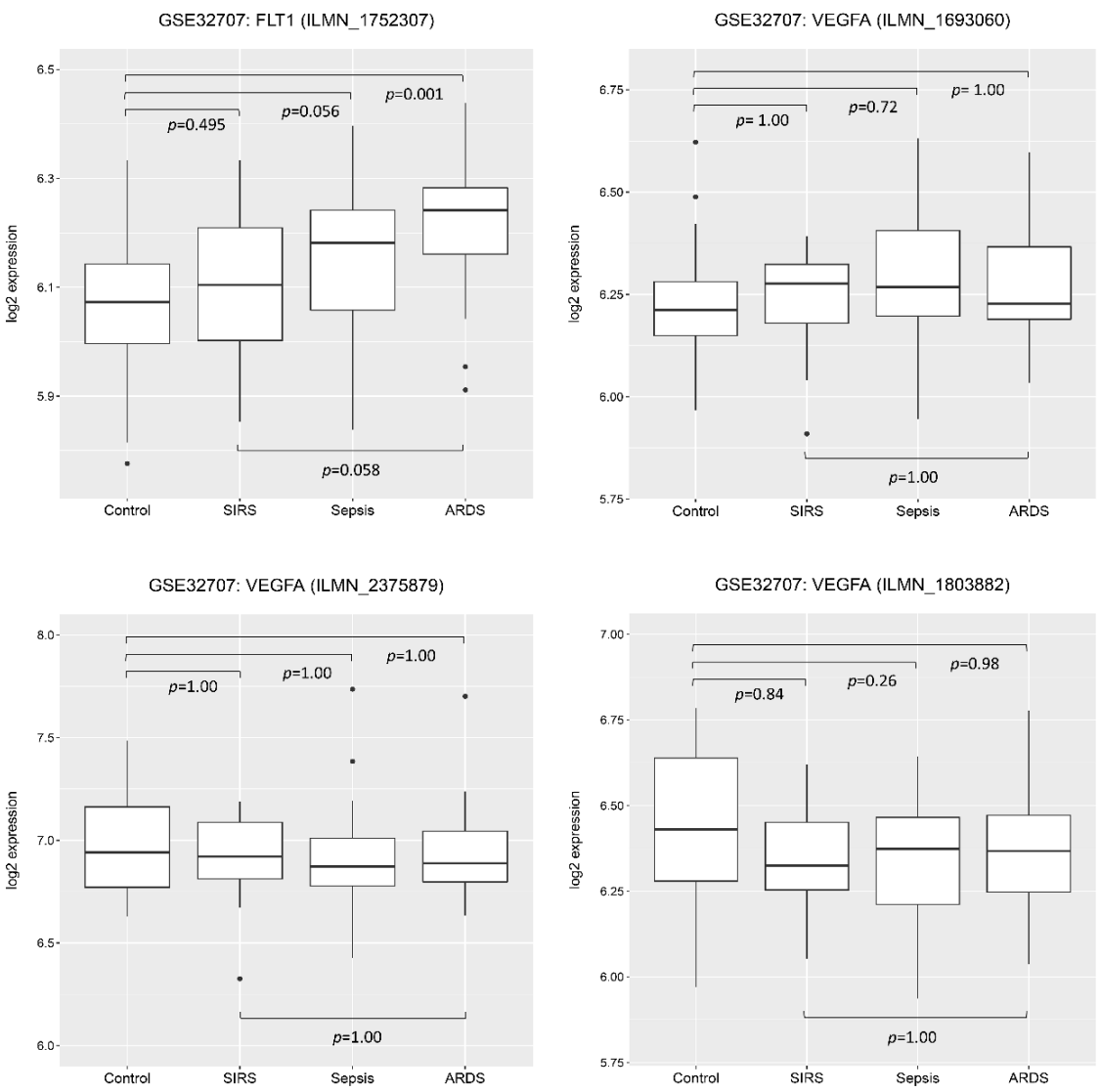
Supplementary Figure 1. Principal component analysis. Plot of the first two principal components (PC) of individuals analyzed in the discovery stage were projected on the HapMap3 reference dataset**.**



Supplementary Figure 2. Quantile-Quantile (Q-Q) plot. Observed versus expected -$\log_{10}$ *p*-values for the GWAS results of the discovery stage.



Supplementary Figure 3. *FLT1* and *VEGFA* gene expression in critical care patients. Probe intensities of expression arrays obtained from peripheral blood samples from ICU controls (n=34), systemic inflammatory response syndrome (SIRS, n=21), sepsis (n=30) and sepsis-associated ARDS patients (n=18) at study inclusion. Note that the probe ILMN_1752307 targets *FLT1* exon 30, which is found in the canonical isoform (FLT1-201), one of the most highly expressed isoforms of the gene. Differences in average intensities were assessed using ANOVA followed by t-tests. GEO accession: GSE32707.[23]



Supplementary Tables

Supplementary Table 1. Top 53 independent SNPs associated with ARDS in the discovery stage ($p<5{\cdot}0x10^{-5}$).

Variant	Chr	Position (hg19)	Gene	A1/A2	MAF	OR [95% CI]	*p*-value
rs598782	1	202572596	*SYT2*	T/C	0·173	0·49 [0·35, 0·69]	$3{\cdot}16x10^{-5}$
rs10917581	1	162624504	*DDR2*	G/A	0·262	0·56 [0·42, 0·74]	$4{\cdot}37x10^{-5}$
rs56865040	2	30907832	*LCLAT1-CAPN13*	G/A	0·066	3·35 [1·95, 5·77]	$1{\cdot}24x10^{-5}$
rs58982889	3	85080936	*CADM2*	C/G	0·483	0·57 [0·44, 0·73]	$1{\cdot}37x10^{-5}$
rs12494792	3	54631523	*CACNA2D3*	A/G	0·251	1·82 [1·38, 2·40]	$2{\cdot}52x10^{-5}$
rs71331755	3	134040335	*RYK-AMOTL2*	C/G	0·237	1·83 [1·37, 2·43]	$3{\cdot}20x10^{-5}$
rs76763432	4	20933002	*KCNIP4*	T/C	0·114	2·36 [1·60, 3·48]	$1{\cdot}38x10^{-5}$
rs12513121	4	126763999	*FAT4-INTU*	A/C	0·302	0·55 [0·43, 0·72]	$1{\cdot}48x10^{-5}$
rs11097547	4	77763070	*SHROOM3-SOWAHB*	G/A	0·200	1·95 [1·44, 2·64]	$1{\cdot}51x10^{-5}$
rs78119818	4	78068598	*CCNI-CCNG2*	A/T	0·048	3·79 [2·03, 7·06]	$2{\cdot}81x10^{-5}$
rs10518480	4	126898260	*FAT4-INTU*	G/A	0·202	1·93 [1·41, 2·64]	$3{\cdot}46x10^{-5}$
rs66691935	4	184540486	*ING2-RWDD4*	T/C	0·175	1·97 [1·43, 2·71]	$3{\cdot}67x10^{-5}$
rs62300402	4	66422737	*EPHA5*	G/A	0·276	0·56 [0·42, 0·74]	$4{\cdot}46x10^{-5}$
rs66486976	5	177602232	*NHP2-GMCL2*	T/C	0·313	0·55 [0·42, 0·71]	$1{\cdot}00x10^{-5}$
rs58681704	5	133268913	*FSTL4-C5orf15*	A/G	0·202	0·51 [0·37, 0·70]	$3{\cdot}18x10^{-5}$
rs62390494	5	177493565	*FAM153C-N4BP3*	T/C	0·199	1·90 [1·40, 2·60]	$4{\cdot}53x10^{-5}$
rs9453845	6	67330152	*EYS-ADGRB3*	T/G	0·107	0·41 [0·27, 0·62]	$2{\cdot}81x10^{-5}$
rs3003179	6	74677167	*CD109-COL12A1*	A/G	0·279	1·76 [1·35, 2·30]	$3{\cdot}20x10^{-5}$
rs58277258	6	129194762	*LAMA2*	C/T	0·084	2·69 [1·68, 4·30]	$3{\cdot}88x10^{-5}$
rs12197618	6	85969855	*TBX18-NT5E*	A/G	0·053	3·58 [1·95, 6·57]	$4{\cdot}01x10^{-5}$
rs9367172	6	43709993	*MRPS18A-VEGFA*	A/G	0·237	0·55 [0·41, 0·73]	$4{\cdot}69x10^{-5}$
rs72611587	7	146905995	*CNTNAP2*	T/C	0·140	2·16 [1·51, 3·09]	$2{\cdot}78x10^{-5}$
rs7777943	7	150483237	*GIMAP5-TMEM176B*	G/A	0·277	0·57 [0·44, 0·74]	$3{\cdot}18x10^{-5}$
rs12678166	8	8520530	*PRAG1-CLDN23*	T/C	0·152	2·05 [1·46, 2·89]	$3{\cdot}80x10^{-5}$
rs796455145	9	5487547	*PLGRKT*	C/T	0·411	1·74 [1·37, 2·22]	$7{\cdot}49x10^{-6}$
rs4740791	9	4611901	*SPATA6L*	T/C	0·087	2·62 [1·66, 4·12]	$3{\cdot}14x10^{-5}$
rs2734600	9	33753355	*PRSS3*	T/C	0·152	0·48 [0·33, 0·68]	$3{\cdot}76x10^{-5}$
rs1751276	10	4477665	*KLF6-AKR1E2*	A/G	0·123	2·53 [1·70, 3·77]	$4{\cdot}85x10^{-6}$
rs1867966	10	71187839	*TACR2-TSPAN15*	G/A	0·430	0·58 [0·45, 0·74]	$1{\cdot}32x10^{-5}$
rs11195238	10	112388857	*SMC3-RBM20*	T/C	0·151	0·47 [0·33, 0·67]	$2{\cdot}97x10^{-5}$
rs10795549	10	7582855	*SFMBT2-ITIH5*	C/A	0·478	0·60 [0·47, 0·76]	$3{\cdot}02x10^{-5}$
rs10736526	11	122589092	*UBASH3B*	C/T	0·207	0·50 [0·37, 0·67]	$3{\cdot}73x10^{-6}$
rs61710829	11	126566557	*KIRREL3*	G/C	0·378	1·71 [1·33, 2·20]	$3{\cdot}07x10^{-5}$
rs602124	11	69388853	*MYEOV-CCND1*	C/G	0·297	1·77 [1·35, 2·31]	$3{\cdot}15x10^{-5}$
rs76921243	12	26606699	*ITPR2*	A/G	0·056	0·25 [0·13, 0·47]	$1{\cdot}81x10^{-5}$
rs1861180	12	12958559	*DDX47*	T/C	0·088	0·40 [0·25, 0·62]	$4{\cdot}07x10^{-5}$
rs1904566	12	68125847	*DYRK2-IFNG*	C/A	0·420	1·69 [1·31, 2·17]	$4{\cdot}49x10^{-5}$
rs9508032	13	28995940	*FLT1*	T/C	0·288	0·49 [0·38, 0·65]	$2{\cdot}62x10^{-7}$
rs8001184	13	90603540	*SLITRK5-GPC5*	A/C	0·483	1·65 [1·30, 2·10]	$4{\cdot}48x10^{-5}$
rs946626	14	49140883	*MDGA2-RPS29*	A/G	0·428	1·73 [1·36, 2·20]	$8{\cdot}59x10^{-6}$
rs7161717	14	26389695	*STXBP6-NOVA1*	C/T	0·132	0·44 [0·30, 0·64]	$2{\cdot}54x10^{-5}$
rs4887263	15	86626153	*AGBL1*	A/C	0·096	2·73 [1·74, 4·28]	$1{\cdot}19x10^{-5}$
rs12902176	15	65518664	*CILP-PARP16*	G/A	0·268	1·78 [1·35, 2·35]	$4{\cdot}04x10^{-5}$
rs11647343	16	84454267	*ATP2C2*	C/A	0·384	1·75 [1·36, 2·24]	$1{\cdot}05x10^{-5}$
rs244783	16	84360055	*WFDC1*	T/G	0·212	1·97 [1·45, 2·69]	$1{\cdot}77x10^{-5}$
rs4791367	17	9724374	*GLP2R*	G/A	0·092	0·35 [0·22, 0·56]	$1{\cdot}25x10^{-5}$
rs9675656	18	2947220	*LPIN2*	C/G	0·109	2·44 [1·63, 3·66]	$1{\cdot}65x10^{-5}$
rs397195	19	6619504	*CD70-TNFSF14*	G/C	0·354	1·71 [1·32, 2·20]	$3{\cdot}44x10^{-5}$

rs285251	19	16415993	*AP1M1-KLF2*	C/T	0·286	0·56 [0·42, 0·74]	$4{\cdot}00x10^{-5}$
rs6040856	20	11702045	*JAG1-BTBD3*	G/C	0·354	1·68 [1·32, 2·15]	$3{\cdot}16x10^{-5}$
rs2831537	21	29516376	*ADAMTS5-N6AMT1*	T/C	0·462	0·57 [0·45, 0·73]	$7{\cdot}65x10^{-6}$
rs4817154	21	28477085	*ADAMTS5-N6AMT1*	A/G	0·082	2·59 [1·64, 4·10]	$4{\cdot}57x10^{-5}$
rs1155955	X	67297091	*OPHN1*	G/A	0·120	3·12 [1·83, 5·31]	$2{\cdot}87x10^{-5}$

A1, Effect allele; A2, Non-effect allele; CI, Confidence Interval; MAF, Minor Allele Frequency; OR, Odd Ratio.

Supplementary Table 2. Main demographic and clinical features of individuals included in the study.

	GEN-SEP			MESSI			SepNet			Comparison between studies	
	Controls (n=316)	**Cases (n=274)**	***p*-value***	**Controls (n=337)**	**Cases (n=268)**	***p*-value***	**Controls (n=649)**	**Cases (n=91)**	***p*-value***	***Controls* *p*-value***	***Cases* *p*-value***
Sex (n males/N)	197/316 (62·3%)	194/274 (70·8%)	0·04	200/337 (59·3%)	163/268 (60·8%)	0·74	379/649 (58·4%)	59/91 (64·8%)	0·50	0·50	0·05
Mean age (years)	63·0 ± 15·0	62·5 ± 14·1	0·47	61·5 ± 13·8	58·8 ± 14·1	0·02	65·4 ± 14.1	62·1 ± 13·6	0·02	0·04	0·88
Pneumonia (n/N)	83/267 (31·1%)	128/252 (50·8%)	$7{\cdot}50x10^{-6}$	134/337 (39.8%)	166/268 (61·9%)	$1{\cdot}2x10^{-7}$	249/646 (38·5%)	49/91 (53·8%)	$5{\cdot}0x10^{-3}$	0·05	0·03
APACHE (median) (P_{25}–P_{75})†	20 (15-24)	22 (18-27)	$2{\cdot}22x10^{-5}$	71 (57-88)	85 (67-104)	$1{\cdot}4x10^{-7}$	20 (16-24)	19 (16-23)	0·47	0·42	$1{\cdot}6x10^{-3}$
Mortality (n/N)‡	79/310 (25.5%)	115/268 (42.9%)	$1{\cdot}5x10^{-5}$	127/337 (37·7%)	182/268 (67·9%)	$2{\cdot}2x10^{-13}$	137/649 (21·1%)	12/91 (13·2%)	0·08	-	-

n=number of individuals with data available, N=group size. All individuals have age and APACHE data, except for MESSI cohort, where only 229 cases and 269 controls have APACHE data. *Categorical data compared by chi square test, continuous data compared by Wilcoxon test (two-sample comparison) or Kruskal-Wallis test (three-sample comparison); †APACHE III was measured for the MESSI Cohort and APACHE II for GENSEP and SepNet Cohorts. Therefore, only APACHE II scores were considered in the comparison between studies; ‡Patient mortality was not compared between studies since ICU mortality was considered for the GENSEP cohort, 30-day mortality was considered for the MESSI Cohort, and 28-day mortality for the SepNet cohort. APACHE, Acute Physiology and Chronic Health Evaluation; P25, percentile 25; P75, percentile 75.

Supplementary Table 3. Sensitivity analysis for the rs9508032 in the three cohorts together.

	OR [95% CI]	*p*-value
Unadjusted[*]	0·62 [0·43, 0·90]	$1{\cdot}07x10^{-7}$
Sex	0·62 [0·43, 0·90]	$1{\cdot}11x10^{-7}$
Age	0·62 [0·43, 0·90]	$1{\cdot}30x10^{-7}$
APACHE[†]	0·61 [0·40, 0·93]	$1{\cdot}81x10^{-8}$
Smokers[‡]	0·58 [0·42, 0·80]	$9{\cdot}72x10^{-4}$[††]
Previous surgery[§]	0·64 [0·50, 0·83]	$7{\cdot}35x10^{-4}$[††]
Ischemic cardiac disease[§]	0·56 [0·45, 0·70]	$3{\cdot}05x10^{-7}$[‡‡]
Pulmonary sepsis	0·61 [0·41, 0·91]	$9{\cdot}12x10^{-8}$[‡‡]
Mortality[¶]	0·62 [0·41, 0·94]	$2{\cdot}51x10^{-7}$
Pathogen[‡]	0·48 [0·34, 0·68]	$2{\cdot}34x10^{-5}$[††]
Multi organ dysfunction[‖]	0·61 [0·41, 0·91]	$1{\cdot}03x10^{-7}$
Comorbidities[**]	0·62 [0·40, 0·94]	$1{\cdot}56x10^{-7}$[‡‡]

[*]Unadjusted data for GEN-SEP, adjustment for 2 PC for MESSI, and adjustment for 3 PC for SepNet; [†]APACHE III was measured for the MESSI Cohort and APACHE II for GEN-SEP and SepNet Cohorts; [‡]There was only information available for GEN-SEP study; [§]There was not information available for MESSI study; [¶]ICU mortality was considered for the GENSEP cohort, 30-day mortality was considered for the MESSI Cohort, and 28-day mortality for the SepNet cohort; [‖]Two or more affected organs; [**]For the GEN-SEP and SepNet studies, comorbidities considered are autoimmune diseases, cancer, chronic diseases, diabetes, hepatopathies, immunosuppression, kidney diseases, morbid obesity, pregnancy, severe infections, severe brain damage, and valvulopathies. For MESSI, comorbidities are defined as immunocompromise (cancer receiving treatment, hematologic malignancy, AIDS, metastatic cancer, or receiving immunosuppressive medication), cirrhosis, congestive heart failure, or chronic renal insufficiency including dialysis; [††]Missing data for more than 35% of individuals from the GEN-SEP study; [‡‡]Missing data for the 15-20% of individuals from the GEN-SEP study. APACHE, Acute Physiology and Chronic Health Evaluation; ICU, intensive care unit; OR, Odd Ratio; CI, Confidence Interval.

Supplementary Table 4. Association results for all SNPs within *FLT1* that have data available in all studies.

							Discovery (274:316)*		Replication (359:986)*		Meta-analysis (633:1302)*	
rs ID	**Chr**	**Position**	**Location†**	**A1/A2**	**MAF**	**r^2**	**OR [95% CI]**	***p*-value**	**OR [95% CI]**	***p*-value**	**OR [95% CI]**	***p*-value**
rs9508032	**13**	**28995940**	**Intron 10**	**T/C**	**0·290**	**1·00**	**0·49 [0·38, 0·65]**	**$2{\cdot}62x10^{-7}$**	**0·74 [0·60, 0·92]**	**$5{\cdot}98x10^{-3}$**	**0·61 [0·41, 0·91]**	**$5{\cdot}18x10^{-8}$**
rs8002446	13	28997400	Intron 10	A/G	0·288	1·00	0·50 [0·38, 0·65]	$2{\cdot}71x10^{-7}$	0·75 [0·61, 0·92]	$6{\cdot}82x10^{-3}$	0·61 [0·41, 0·92]	$6{\cdot}03x10^{-8}$
rs9513111	13	28997563	Intron 10	T/C	0·288	1·00	0·50 [0·38, 0·65]	$2{\cdot}71x10^{-7}$	0·75 [0·61, 0·92]	$6{\cdot}82x10^{-3}$	0·61 [0·41, 0·92]	$6{\cdot}03x10^{-8}$
rs722503	13	28997052	Intron 10	T/C	0·288	1·00	0·50 [0·38, 0·65]	$2{\cdot}71x10^{-7}$	0·75 [0·61, 0·92]	$6{\cdot}94x10^{-3}$	0·61 [0·41, 0·92]	$6{\cdot}11x10^{-8}$
rs9508034	13	28996604	Intron 10	C/A	0·290	1·00	0·49 [0·38, 0·65]	$2{\cdot}62x10^{-7}$	0·75 [0·61, 0·93]	$7{\cdot}58x10^{-3}$	0·61 [0·41, 0·92]	$6{\cdot}47x10^{-8}$
rs9508033	13	28996568	Intron 10	C/T	0·290	0·99	0·49 [0·38, 0·65]	$2{\cdot}62x10^{-7}$	0·75 [0·61, 0·92]	$7{\cdot}19x10^{-3}$	0·61 [0·41, 0·92]	$6{\cdot}18x10^{-8}$
rs6491284	13	28995390	Intron 10	T/C	0·288	0·98	0·50 [0·38, 0·65]	$3{\cdot}26x10^{-7}$	0·76 [0·61, 0·94]	$9{\cdot}91x10^{-3}$	0·62 [0·41, 0·93]	$1{\cdot}05x10^{-7}$
rs2281827	13	29001721	Intron 9	T/C	0·265	0·86	1·97 [1·49, 2·60]	$1{\cdot}60x10^{-6}$	1·37 [1·10, 1·70]	$4{\cdot}20x10^{-3}$	1·63 [1·14, 2·32]	$1{\cdot}69x10^{-7}$
rs7324510	13	29007035	Intron 6	A/C	0·242	0·72	0·54 [0·40, 0·72]	$3{\cdot}17x10^{-5}$	0·69 [0·55, 0·86]	$1{\cdot}01x10^{-3}$	0·63 [0·53, 0·75]	$2{\cdot}81x10^{-7}$
rs9513115	13	29011570	Intron 4	A/C	0·246	0·72	0·55 [0·41, 0·73]	$4{\cdot}17x10^{-5}$	0·69 [0·55, 0·86]	$1{\cdot}12x10^{-3}$	0·63 [0·53, 0·75]	$3{\cdot}69x10^{-7}$
rs9513114	13	29009059	Intron 4	T/C	0·245	0·71	0·52 [0·39, 0·70]	$1{\cdot}15x10^{-5}$	0·70 [0·56, 0·88]	$1{\cdot}76x10^{-3}$	0·63 [0·53, 0·75]	$2{\cdot}62x10^{-7}$
rs8001784	13	29009213	Intron 4	G/A	0·242	0·71	0·54 [0·40, 0·72]	$3{\cdot}17x10^{-5}$	0·69 [0·56, 0·87]	$1{\cdot}28x10^{-3}$	0·63 [0·53, 0·75]	$3{\cdot}68x10^{-7}$
rs4771249	13	29013414	Intron 3	C/G	0·241	0·71	0·54 [0·41, 0·73]	$3{\cdot}84x10^{-5}$	0·70 [0·56, 0·87]	$1{\cdot}36x10^{-3}$	0·63 [0·53, 0·76]	$4{\cdot}53x10^{-7}$
rs9508035	13	29009099	Intron 4	A/C	0·242	0·71	0·54 [0·40, 0·72]	$3{\cdot}17x10^{-5}$	0·70 [0·56, 0·88]	$1{\cdot}76x10^{-3}$	0·64 [0·53, 0·76]	$5{\cdot}37x10^{-7}$
rs3794405	13	29006847	Intron 6	T/C	0·243	0·71	0·54 [0·41, 0·73]	$4{\cdot}16x10^{-5}$	0·70 [0·56, 0·87]	$1{\cdot}29x10^{-3}$	0·64 [0·53, 0·76]	$4{\cdot}56x10^{-7}$
rs9505994	13	29024346	Intron 3	G/A	0·239	0·71	0·53 [0·40, 0·71]	$2{\cdot}41x10^{-5}$	0·69 [0·55, 0·86]	$1{\cdot}10x10^{-3}$	0·63 [0·53, 0·75]	$2{\cdot}53x10^{-7}$
rs34961350	13	28991902	Intron 10	G/C	0·220	0·57	1·89 [1·41, 2·54]	$2{\cdot}38x10^{-5}$	1·27 [1·00, 1·60]	$5{\cdot}02x10^{-2}$	1·53 [1·03, 2·27]	$2{\cdot}55x10^{-5}$

*Cases:Controls; †According to the principal *FLT1* transcript (ENST00000282397.8) ; A1, Effect allele; A2, Non-effect allele; CI, Confidence Interval; MAF, Minor Allele Frequency (GEN-SEP cohort) ; OR, Odd Ratio; r^2, squared coefficient of correlation (with respect to rs9508032). Alleles referred to the positive strand of hg19. The top hit is indicated in bold.

Supplementary Table 5. Prediction of ICU survival for the sentinel variant at *FLT1* in GEN-SEP

	OR [95% CI]	*p*-value
Sepsis	1·00 [0·81, 1·24]	0·974
ARDS	1·16 [0·87, 1·54]	0·307

Data were obtained using Cox regression models adjusted for age, sex, and APACHE II scores. APACHE II, Acute Physiology and Chronic Health Evaluation II; ICU, intensive care unit; OR, Odd Ratio; CI, Confidence Interval.

Supplementary Table 6. Expression of *FLT1* and *VEGFA* isoforms in lung tissue.

Name	Average	SD
FLT1-201	5,928	2,803
FLT1-202	5	5
FLT1-203	6	4
FLT1-204	20	12
FLT1-205	6	7
FLT1-206	12	12
FLT1-207	3,998	2,492
FLT1-208	2	2
VEGFA-203	0	0
VEGFA-226	11	28
VEGFA-209	0	0
VEGFA-207	12	14
VEGFA-205	3,804	3,543
VEGFA-213	0	0
VEGFA-228	5	12
VEGFA-229	2,748	2,824
VEGFA-208	0	0
VEGFA-204	0	0
VEGFA-227	1,804	3,203
VEGFA-220	5	7
VEGFA-202	0	0
VEGFA-224	0	0
VEGFA-222	1,536	1,704
VEGFA-218	94	81
VEGFA-225	9	27
VEGFA-219	0	1
VEGFA-223	1	4
VEGFA-201	69	75
VEGFA-206	1,232	1,325
VEGFA-210	1	4
VEGFA-221	1	1
VEGFA-212	3,393	2,676
VEGFA-215	4,193	3,631
VEGFA-211	52	41
VEGFA-214	83	105
VEGFA-217	8	15
VEGFA-216	161	158

No data was available for the isoforms FLT1-209, VEGFA-230, VEGFA-231, or VEGFA-232. SD, standard deviation. Values are given in counts per million.

Supplementary Table 7. Functional annotation of the *FLT1* top hit (rs9508032) and the most promising two proxies.

	rs9508032	rs722503	rs8002446
Functional significance score predicted with DeepSEA	0·13	<0·05	0·10
regulomedB Score	(5) TF binding or Dnase peak	(3a) TF binding + any motif + Dnase peak	(4) TF binding + Dnase peak
Enhancer histone marks	H3K4me1[*], H3K27ac[†]	H3K4me1[‡], H3K27ac[§]	H3K4me1[¶], H3K27ac[‖]
Promoter histone marks	H3K4me3[**], H3K9ac[††]	H3K4me3[‡‡], H3K9ac[§§]	H3K4me3[¶¶], H3K9ac[§§]
DNAse	HSC & B-cell, Monocytes-CD14+ RO01746 Primary Cells	HSC & B-cell, ES-deriv	IMR90, iPSC, Blood & T-cell, HSC & B-cell, Epithelial, Thymus, Muscle, Fetal Kidney, Fetal Lung, Ovary, Placenta, GM12878 Lymphoblastoid Cells, Monocytes-CD14+ RO01746 Primary Cells
Altered regulatory motifs	Cdc5, Gfi-1, HNF1, Mef2	CCNT2, MAZR, NF-kappaB, Spz1	None
Proteins bound	None	POL2, NFKB	BCL11A, EBF1, EBF1, ELF1, PAX5C20, PAX5N19, PU1, SP1, PU1, POL2, CMYC, MAX
CHICP	CD34: *POMP* (11·36); GM12878: *POMP* (9·71), *FLT1* (10·97), *SLC46A3/RNU6-53P* (9·91), *PAN3* (9·13), *SLC46A3/CYP51A1P2* (8·63)		
Open Targets Genetics	None	None	*FLT3* (top ranked), *POMP*, *PAN3*
eQTLs	None	None	None
Score CAPE dsQTL >0·5	HUVEC, A549 EtOH 0.02pct Lung Carcinoma Cell Line	None	None
Score CAPE eQTL >0·5	None	HUVEC, NHLF Lung Fibroflast Primary Cells, NHDF-Adult Dermal Fibroflast Primary Cells, Monocytes-CD14+ RO01746 Primary Cells, A549 EtOH 0·02pct Lung Carcinoma Cell Line, Foreskin Fibroblast Primary Cells skin01, IMR90 fetal lung fibroblasts Cell Line	A549 EtOH 0·02pct Lung Carcinoma Cell Line

CAPE, Cellular dependent deactivating mutations; CD34, human hematopoietic progenitor cell line; GM12878, lymphoblastoid cell line; HUVEC, Human umbilical vein endothelial cell; IMR90, Human foetal lung cells; NHDF, Normal Human Dermal Fibroblasts; NHLF, Normal human lung fibroblasts. [*]IMR90, ESC, iPSC, ES-deriv, Blood & T-cell, HSC & B-cell, Epithelial, Brain, Adipose, Muscle, Heart, Fetal Lung, Fetal Adrenal Gland, Liver, Spleen, GM12878 Lymphoblastoid Cells, HUVEC Umbilical Vein Endothelial Primary Cells, Monocytes-CD14+ RO01746 Primary Cells. [†]iPSC, ES-deriv, HSC & B-cell, Epithelial, Adipose, Spleen, GM12878 Lymphoblastoid Cells, HUVEC Umbilical Vein Endothelial Primary Cells. [‡]IMR90, ESC, iPSC, ES-deriv, Blood & T-cell, HSC & B-cell, Epithelial, Thymus, Brain, Adipose, Muscle, Heart, Digestive, Fetal Lung, Fetal Adrenal Gland, Placenta, Liver, Lung, Spleen, Dnd41 TCell Leukemia Cell Line, GM12878 Lymphoblastoid Cells, HUVEC Umbilical Vein Endothelial Primary Cells, K562 Leukemia Cells, Monocytes-CD14+ RO01746 Primary Cells. [§]iPSC, HSC & B-cell, Brain, Adipose, Heart, Digestive, Liver, Dnd41 TCell Leukemia Cell Line, GM12878 Lymphoblastoid Cells, HUVEC Umbilical Vein Endothelial Primary Cells, Monocytes-CD14+ RO01746 Primary Cells. [¶]ESC, iPSC, ES-deriv, Blood & T-cell, HSC & B-cell, Epithelial, Thymus, Brain, Adipose, Muscle, Heart, Digestive, Fetal Lung, Fetal Adrenal Gland, Placenta, Liver, Lung, Spleen, Dnd41 TCell Leukemia Cell Line, GM12878 Lymphoblastoid Cells, HUVEC Umbilical Vein Endothelial Primary Cells, Monocytes-CD14+ RO01746 Primary Cells. [‖]iPSC, HSC & B-cell, Epithelial, Brain, Adipose, Heart, Digestive, Ovary, Liver, Dnd41 TCell Leukemia Cell Line, GM12878 Lymphoblastoid Cells, HUVEC Umbilical Vein Endothelial Primary Cells, Monocytes-CD14+ RO01746 Primary Cells. [**]HSC & B-cell, Monocytes-CD14+ RO01746 Primary Cells. [††]iPSC, Adipose, Digestive. [‡‡]Blood & T-cell, HSC & B-cell, Brain, Adipose, Heart, Digestive, Liver, Dnd41 TCell Leukemia Cell Line, GM12878 Lymphoblastoid Cells, Monocytes-CD14+ RO01746 Primary Cells. [§§]Blood & T-cell, Adipose, Sm. Muscle, Dnd41 TCell Leukemia Cell Line, GM12878 Lymphoblastoid Cells, Monocytes-CD14+ RO01746 Primary Cells. [¶¶]Blood & T-cell, HSC & B-cell, Brain, Dnd41 TCell Leukemia Cell Line, GM12878 Lymphoblastoid Cells, Monocytes-CD14+ RO01746 Primary Cells

Supplementary Table 8. Signals of replication at gene level in the GEN-SEP dataset within 100 kb of previously reported candidate genes.

Gene	Independent signals	Gene-wise Bonferroni *p*-value threshold	SNP min *p*-value	A1/A2	OR [95% CI]	*p*-value
ABCC1	275·18	$1{\cdot}82\times10^{-4}$	rs246233	G/T	1·98 [1·21, 3·25]	$6{\cdot}79\times10^{-3}$
ACE	72·94	$6{\cdot}85\times10^{-4}$	rs9857615	T/C	0·70 [0·49, 0·98]	$3{\cdot}89\times10^{-2}$
ADA	137·74	$3{\cdot}63\times10^{-4}$	rs17687734	G/A	2·60 [1·31, 5·16]	$6{\cdot}31\times10^{-3}$
ADGRV1	640·97	$7{\cdot}80\times10^{-5}$	rs6094023	A/G	0·48 [0·29, 0·79]	$3{\cdot}65\times10^{-3}$
ADIPOQ	288·87	$1{\cdot}73\times10^{-4}$	rs114210898	A/G	1·34 [1·05, 1·72]	$2{\cdot}06\times10^{-2}$
ADRBK2	318·26	$1{\cdot}57\times10^{-4}$	rs1467387	C/T	0·67 [0·50, 0·89]	$5{\cdot}96\times10^{-3}$
*AGER**	363·32	$1{\cdot}38\times10^{-4}$	rs61746206	T/C	0·32 [0·13, 0·83]	$1{\cdot}83\times10^{-2}$
AGT	303·17	$1{\cdot}65\times10^{-4}$	rs1078499	G/A	0·69 [0·53, 0·91]	$8{\cdot}68\times10^{-3}$
AGTR1	281·4	$1{\cdot}78\times10^{-4}$	rs275643	A/G	1·86 [1·24, 2·80]	$2{\cdot}70\times10^{-3}$
AHR	201·39	$2{\cdot}48\times10^{-4}$	rs140084506	T/C	4·69 [1·26, 17·4]	$2{\cdot}10\times10^{-2}$
ANGPT2	456·67	$1{\cdot}09\times10^{-4}$	rs2442570	A/G	0·43 [0·24, 0·76]	$3{\cdot}76\times10^{-3}$
APOA1	236·49	$2{\cdot}11\times10^{-4}$	rs2513094	C/T	1·36 [1·03, 1·79]	$2{\cdot}99\times10^{-2}$
ARSD	159·08	$3{\cdot}14\times10^{-4}$	rs1698814	C/T	1·45 [1·08, 1·96]	$1{\cdot}46\times10^{-2}$
BCL11A	262·4	$1{\cdot}91\times10^{-4}$	rs76064527	A/C	1·90 [1·08, 3·37]	$2{\cdot}69\times10^{-2}$
CBS	160·35	$3{\cdot}12\times10^{-4}$	rs2401154	T/C	1·49 [1·06, 2·08]	$2{\cdot}03\times10^{-2}$
CELF2	669·82	$7{\cdot}46\times10^{-5}$	rs76209150	T/G	2·65 [1·33, 5·26]	$5{\cdot}50\times10^{-3}$
CHIT1	269·23	$1{\cdot}86\times10^{-4}$	rs1845466	T/G	0·64 [0·49, 0·83]	$8{\cdot}63\times10^{-4}$
CLASRP	113·15	$4{\cdot}42\times10^{-4}$	rs10405859	C/T	0·71 [0·56, 0·91]	$6{\cdot}50\times10^{-3}$
CXCL2	96·08	$5{\cdot}20\times10^{-4}$	rs28574621	C/G	0·30 [0·09, 0·94]	$3{\cdot}85\times10^{-2}$
CXCR2	98·44	$5{\cdot}08\times10^{-4}$	rs12989315	A/G	2·26 [1·26, 4·07]	$6{\cdot}36\times10^{-3}$
CYP1A1	82·88	$6{\cdot}03\times10^{-4}$	rs17861120	G/A	0·51 [0·32, 0·83]	$6{\cdot}20\times10^{-3}$
DARC	202·96	$2{\cdot}46\times10^{-4}$	rs55833893	T/C	0·31 [0·14, 0·68]	$3{\cdot}66\times10^{-3}$
DIO2	158·36	$3{\cdot}16\times10^{-4}$	rs17176215	G/A	0·24 [0·07, 0·87]	$2{\cdot}96\times10^{-2}$
EGF	240·96	$2{\cdot}08\times10^{-4}$	rs146141236	C/T	0·22 [0·09, 0·55]	$1{\cdot}25\times10^{-3}$
*EGLN1**	144·18	$3{\cdot}47\times10^{-4}$	rs141921538	T/C	4·56 [1·44, 14·39]	$9{\cdot}75\times10^{-3}$
EPAS1	417·3	$1{\cdot}20\times10^{-4}$	rs11888926	G/C	1·65 [1·19, 2·28]	$2{\cdot}63\times10^{-3}$
F5	269·74	$1{\cdot}85\times10^{-4}$	rs144628673	A/G	5·37 [1·51, 19·05]	$9{\cdot}37\times10^{-3}$
FAAH	187·13	$2{\cdot}67\times10^{-4}$	rs78918625	G/C	0·17 [0·04, 0·78]	$2{\cdot}30\times10^{-2}$
FAS	239·16	$2{\cdot}09\times10^{-4}$	rs61852572	G/A	0·60 [0·41, 0·89]	$1{\cdot}11\times10^{-2}$
*FER**	644·44	$7{\cdot}76\times10^{-5}$	rs10515395	C/T	1·82 [1·22, 2·71]	$3{\cdot}28\times10^{-3}$
FTL	143·23	$3{\cdot}49\times10^{-4}$	rs140747916	T/A	1·70 [1·08, 2·66]	$2{\cdot}12\times10^{-2}$
FZD2	125·26	$3{\cdot}99\times10^{-4}$	rs9900767	T/C	0·50 [0·28, 0·90]	$2{\cdot}09\times10^{-2}$
GADD45A	184·53	$2{\cdot}71\times10^{-4}$	rs344923	G/A	1·33 [1·05, 1·70]	$1{\cdot}92\times10^{-2}$
GHR	322·68	$1{\cdot}55\times10^{-4}$	rs41271073	A/G	0·34 [0·15, 0·74]	$6{\cdot}48\times10^{-3}$
GP5	210·56	$2{\cdot}37\times10^{-4}$	rs7611390	T/C	0·57 [0·41, 0·81]	$1{\cdot}63\times10^{-3}$
GRM3	285·81	$1{\cdot}75\times10^{-4}$	rs6974073	A/C	1·59 [1·02, 2·47]	$4{\cdot}01\times10^{-2}$
HAS1	225·37	$2{\cdot}22\times10^{-4}$	rs113174648	G/C	1·51 [1·12, 2·03]	$7{\cdot}03\times10^{-3}$
HECTD2	174·41	$2{\cdot}87\times10^{-4}$	rs11186608	T/G	0·71 [0·56, 0·89]	$3{\cdot}23\times10^{-3}$
HMOX1	129·66	$3{\cdot}86\times10^{-4}$	rs4645773	T/C	0·35 [0·16, 0·75]	$6{\cdot}49\times10^{-3}$
HMOX2	119·94	$4{\cdot}17\times10^{-4}$	rs190300249	T/C	0·34 [0·12, 0·93]	$3{\cdot}51\times10^{-2}$
HSPG2	351·59	$1{\cdot}42\times10^{-4}$	rs72662414	C/A	3·05 [1·03, 9·07]	$4{\cdot}50\times10^{-2}$
HTR2A	286·1	$1{\cdot}75\times10^{-4}$	rs1923886	T/C	1·54 [1·21, 1·96]	$5{\cdot}36\times10^{-4}$
IL10	185·14	$2{\cdot}70\times10^{-4}$	rs79474100	A/T	0·27 [0·10, 0·73]	$9{\cdot}99\times10^{-3}$
IL13	86·87	$5{\cdot}76\times10^{-4}$	rs60153262	T/C	2·94 [1·50, 5·77]	$1{\cdot}74\times10^{-3}$
IL18	143·25	$3{\cdot}49\times10^{-4}$	rs360723	T/A	0·69 [0·49, 0·96]	$2{\cdot}87\times10^{-2}$
IL1RN	281·1	$1{\cdot}78\times10^{-4}$	rs6746416	G/A	1·37 [1·07, 1·76]	$1{\cdot}28\times10^{-2}$
IL32	70·73	$7{\cdot}07\times10^{-4}$	rs12598558	G/T	0·58 [0·38, 0·89]	$1{\cdot}18\times10^{-2}$
IL4	91·11	$5{\cdot}49\times10^{-4}$	rs60153262	T/C	2·94 [1·50, 5·77]	$1{\cdot}74\times10^{-3}$
IL6	236·93	$2{\cdot}11\times10^{-4}$	rs75897827	A/G	2·70 [1·18, 6·21]	$1{\cdot}91\times10^{-2}$

IL8	154·71	$3{\cdot}23x10^{-4}$	rs7686667	G/A	2·43 [1·06, 5·61]	$3{\cdot}69x10^{-2}$
IRAK3	186·89	$2{\cdot}68x10^{-4}$	rs569436368	A/G	0·26 [0·09, 0·81]	$2{\cdot}05x10^{-2}$
ISG15	8·12	$6{\cdot}16x10^{-3}$	rs12093451	A/C	0·50 [0·29, 0·88]	$1{\cdot}55x10^{-2}$
KLK2	201·25	$2{\cdot}48x10^{-4}$	rs1701934	T/C	5·03 [1·65, 15·4]	$4{\cdot}61x10^{-3}$
LRRC16A	781·8	$6{\cdot}40x10^{-5}$	rs2690123	G/A	1·41 [1·11, 1·8]	$5{\cdot}64x10^{-3}$
LTA	459·6	$1{\cdot}09x10^{-4}$	rs45552734	T/C	0·58 [0·41, 0·82]	$2{\cdot}40x10^{-3}$
*MAP3K1**	262·39	$1{\cdot}91x10^{-4}$	rs1910019	T/C	1·77 [1·25, 2·50]	$1{\cdot}30x10^{-3}$
MAP3K6	86·03	$5{\cdot}81x10^{-4}$	rs12742921	C/T	0·72 [0·52, 0·99]	$4{\cdot}53x10^{-2}$
MBL2	250·05	$2{\cdot}00x10^{-4}$	rs34546527	C/A	0·44 [0·28, 0·68]	$2{\cdot}50x10^{-4}$
MIF	141·4	$3{\cdot}54x10^{-4}$	rs75761219	T/C	0·22 [0·06, 0·76]	$1{\cdot}72x10^{-2}$
*MUC5B**	71·73	$6{\cdot}97x10^{-4}$	rs2071175	T/C	2·10 [1·18, 3·74]	$1{\cdot}16x10^{-2}$
MYLK	307·19	$1{\cdot}63x10^{-4}$	rs16834826	G/A	1·48 [1·15, 1·91]	$2{\cdot}45x10^{-3}$
NAMPT	238·31	$2{\cdot}10x10^{-4}$	rs56844330	G/A	0·64 [0·47, 0·87]	$5{\cdot}02x10^{-3}$
NFE2L2	196·18	$2{\cdot}55x10^{-4}$	rs2588866	T/C	0·51 [0·30, 0·87]	$1{\cdot}25x10^{-2}$
NFKB1	210·58	$2{\cdot}37x10^{-4}$	rs76615823	G/A	3·21 [1·13, 9·14]	$2{\cdot}88x10^{-2}$
NFKBIA	254·35	$1{\cdot}97x10^{-4}$	rs75208350	T/C	2·44 [1·15, 5·19]	$2{\cdot}01x10^{-2}$
NOS3	195·41	$2{\cdot}56x10^{-4}$	rs41307316	A/G	0·22 [0·10, 0·51]	$4{\cdot}03x10^{-4}$
NQO1	143·23	$3{\cdot}49x10^{-4}$	rs116423606	A/G	0·30 [0·09, 0·95]	$3{\cdot}98x10^{-2}$
PDE4B	530·28	$9{\cdot}43x10^{-5}$	rs6664875	C/A	0·64 [0·48, 0·85]	$1{\cdot}96x10^{-3}$
PI3	142·59	$3{\cdot}51x10^{-4}$	rs877608	A/T	1·98 [1·08, 3·63]	$2{\cdot}71x10^{-2}$
PLAU	116·57	$4{\cdot}29x10^{-4}$	rs72816344	A/G	3·50 [1·08, 11·36]	$3{\cdot}73x10^{-2}$
POPDC3	181·91	$2{\cdot}75x10^{-4}$	rs1051484	T/C	0·62 [0·46, 0·82]	$1{\cdot}12x10^{-3}$
PPARGC1A	978·39	$5{\cdot}11x10^{-5}$	rs6847465	T/C	3·92 [1·51, 10·15]	$4{\cdot}96x10^{-3}$
PPFIA1-SHANK2	881·08	$5{\cdot}67x10^{-5}$	rs11602848	C/T	1·71 [1·20, 2·43]	$3{\cdot}01x10^{-3}$
PRKAG2	648·11	$7{\cdot}71x10^{-5}$	rs10231047	C/T	0·71 [0·56, 0·89]	$2{\cdot}91x10^{-3}$
S1PR3	243·08	$2{\cdot}06x10^{-4}$	rs150901384	G/T	4·57 [1·44, 14·5]	$9{\cdot}79x10^{-3}$
SELPLG	176·88	$2{\cdot}83x10^{-4}$	rs8179106	A/G	1·77 [1·17, 2·66]	$6{\cdot}44x10^{-3}$
SERPINE1	207·41	$2{\cdot}41x10^{-4}$	rs73168394	A/G	0·33 [0·17, 0·67]	$2{\cdot}22x10^{-3}$
SFTPA1	143·17	$3{\cdot}49x10^{-4}$	rs17886197	G/T	1·60 [1·14, 2·26]	$7{\cdot}27x10^{-3}$
SFTPA2	166·37	$3{\cdot}01x10^{-4}$	rs17886197	G/T	1·60 [1·14, 2·26]	$7{\cdot}27x10^{-3}$
SFTPB	165·98	$3{\cdot}01x10^{-4}$	rs75830997	T/G	3·64 [1·51, 8·73]	$3{\cdot}86x10^{-3}$
SFTPD	222·05	$2{\cdot}25x10^{-4}$	rs7082484	C/A	2·27 [1·22, 4·24]	$9{\cdot}55x10^{-3}$
SOD3	175·95	$2{\cdot}84x10^{-4}$	rs2361079	C/T	0·59 [0·40, 0·85]	$4{\cdot}55x10^{-3}$
STAT1	161·06	$3{\cdot}10x10^{-4}$	rs4853453	A/G	1·58 [1·12, 2·24]	$9{\cdot}30x10^{-3}$
TGFB2	244·72	$2{\cdot}04x10^{-4}$	rs75854892	C/T	5·78 [1·60, 20·9]	$7{\cdot}43x10^{-3}$
TIA1	104·96	$4{\cdot}76x10^{-4}$	rs11694045	G/T	1·42 [1·11, 1·82]	$5{\cdot}48x10^{-3}$
TIRAP	187·4	$2{\cdot}67x10^{-4}$	rs12283024	A/G	0·34 [0·15, 0·78]	$1{\cdot}08x10^{-2}$
TLR1	227·88	$2{\cdot}19x10^{-4}$	rs193202734	C/T	5·63 [1·86, 17·01]	$2{\cdot}20x10^{-3}$
TNF	448·83	$1{\cdot}11x10^{-4}$	rs45552734	T/C	0·58 [0·41, 0·82]	$2{\cdot}40x10^{-3}$
TNFRSF11A	246·87	$2{\cdot}03x10^{-4}$	rs7235828	A/G	0·65 [0·47, 0·89]	$7{\cdot}73x10^{-3}$
TRAF6	185·4	$2{\cdot}70x10^{-4}$	rs2458928	A/G	0·68 [0·52, 0·89]	$5{\cdot}39x10^{-3}$
UGT2B7	186·83	$2{\cdot}68x10^{-4}$	rs139914109	C/T	7·69 [1·65, 35·87]	$9{\cdot}42x10^{-3}$
VEGFA	**262·98**	$\mathbf{1{\cdot}90x10^{-4}}$	**rs9367172**	**A/G**	**0·55 [0·41, 0·73]**	$\mathbf{4{\cdot}69x10^{-5}}$
VLDLR	375·76	$1{\cdot}33x10^{-4}$	rs10491716	C/A	1·55 [1·19, 2·02]	$1{\cdot}20x10^{-3}$
VWF	403·43	$1{\cdot}24x10^{-4}$	rs2239160	G/A	0·42 [0·26, 0·68]	$3{\cdot}28x10^{-4}$
XKR3	97·09	$5{\cdot}15x10^{-4}$	rs5994042	A/T	3·23 [1·20, 8·71]	$2{\cdot}03x10^{-2}$
ZNF335	141·92	$3{\cdot}52x10^{-4}$	rs1736493	G/A	0·53 [0·31, 0·89]	$1{\cdot}72x10^{-2}$

A1, Effect allele; A2, Non-effect allele; CI, Confidence Interval; OR, Odds ratio for the effect alleles. In bold, genes harboring variants reaching the bonferroni threshold. *Genes identified for this study (December 2015 to November 2018) based on the search of terms “acute respiratory distress syndrome”, “ARDS” OR “acute lung injury” with “polymorphism” OR “genetic variant”.