Genome-wide association study of pediatric bacteremia and sepsis

Dylan Lawless*1, Flavia Aurelia Hodel 2, 3, Christian W. Thorball 2, 3, Zhi Ming Xu 3, Alessandro Borghesi 3, 4, 5, Eric Giannoni 6, Johannes Trück 7, Martin Stocker 8, Klara M Posfay-Barbe 9, Ulrich Heininger 10, Sara Bernhard-Stirnemann 11, Anita Niederer-Loher 12, Christian R. Kahlert 12, Giancarlo Natalucci 13, Christa Relly 14, Christoph Berger 14, Thomas Riedel 15, Christoph Aebi 15, Philipp Agyeman 15, Jacques Fellay * 2, 3, Luregn J. Schlapbach * 1, on behalf of the Swiss Pediatric Sepsis Study.

1. Department of Intensive Care and Neonatology, University Children's Hospital Zürich, University of Zürich, Switzerland.
2. Precision Medicine Unit, Biomedical Data Science Center, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland.
3. Global Health Institute, School of Life Sciences, École Polytechnique Fédérale de Lausanne, Switzerland.
4. Neonatal Intensive Care Unit, San Matteo Research Hospital, Pavia, Italy.
5. Neonatal Unit, Geneva University Hospitals, Mother, Child and Adolescent Department, Geneva, Switzerland.
6. Clinic of Neonatology, Department Woman-Mother-Child, Lausanne University Hospital and University of Lausanne, Switzerland.
7. Divisions of Allergy and Immunology, University Children's Hospital Zurich and Children's Research Center, University of Zurich (UZH), Zurich, Switzerland.
8. Pediatric Intensive Care and Neonatology, Children's Hospital of Central Switzerland and University of Lucerne, Lucerne, Switzerland.
9. Pediatric Infectious Diseases Unit, Children's Hospital of Geneva, University Hospitals of Geneva, Geneva, Switzerland.
10. Infectious Diseases and Vaccinology, University of Basel Children's Hospital, Basel, Switzerland.
11. Children's Hospital Aarau, Aarau, Switzerland.
12. Division of Infectious Diseases and Infection Control, Children's Hospital of Eastern Switzerland St. Gallen, St. Gallen, Switzerland.
13. Department of Neonatology, University Hospital Zurich, Zurich, Switzerland
14. Division of Infectious Diseases and Hospital Epidemiology, and Children's Research Center, University Children's Hospital Zurich, Switzerland.
15. Division of Infectious Disease, Department of Pediatrics, Inselspital, Bern University Hospital, University of Bern, Switzerland

* These authors contributed equally to the work.

Contacts: dylan.lawless@kispi.uzh.ch , jacques.fellay@epfl.ch , Luregn.Schlapbach@kispi.uzh.ch

Abstract

Background

Sepsis is defined as a dysregulated host response to infection leading to organ dysfunction. It represents a major global health concern, particularly in childhood. The precise underlying pathophysiological mechanisms remain poorly understood.

Methods

Using samples and clinical data from 650 children enrolled in the Swiss Pediatric Sepsis Study, a national multicenter cohort study for culture-proven bacterial sepsis, and from 1395 controls, we explored the human genetic determinants of inter-individual variability in sepsis-related outcomes. We used a case-control study design to search for associations between genome-wide polymorphisms and susceptibility to sepsis, and case-only analysis for specific characteristics of the disease.

Findings

We identified one locus significantly associated with sepsis susceptibility, encompassing the *CTNNAL1* and *ELP1* genes.

Interpretation

Our results suggest a genetic modulator of sepsis susceptibility in children.

Funding

This study was funded by grants from the Swiss National Science Foundation (320030_201060 and 342730 153158/1), the Swiss Society of Intensive Care, the Bangerter Foundation, the Vinetum and Borer Foundation, and the Foundation for the Health of Children and Adolescents.

Keywords

Paediatric; sepsis; bacteremia; genome-wide association study; CTNNAL1; ELP1

Abbreviations

Actl7a and Actl7b (actin-like 7a and 7b), CC (Collaborative Cross), CTNNAL1 (catenin alpha-like 1 gene), CVAD (central venous access device), DALYs (Disability Adjusted Life Years), DEG (differentially expressed gene), ELP1 (elongator complex protein 1), eQTL (expression Quantitative Trait Locus), EUCLIDS (EU Childhood Life-threatening Infectious Disease Study), FAM206A (family with sequence similarity 206, member A), GRM (genetic relationship matrix), GWAS (genome-wide association study), HWE (Hardy-Weinberg Equilibrium), IBS (identify-by-state), ELP1 (elongator acetyltransferase complex subunit 1), IKK (IkB enzyme kinase), LD (linkage disequilibrium), LOCO (leave-one-chromosome-out), MLMA (mixed linear model based asso-ciation), MOF (multiorgan failure), NICU (neonatal intensive care unit), PC (principal component), PCA (principal component analysis), PICU (Pediatric Intensive Care Unit), SNP (single nucleotide polymorphism), sQTL (splicing Quantitative Trait Locus), WGS (whole genome sequencing).

Research in context

Evidence before this study

Blood culture-confirmed bacterial sepsis in children remains an important cause of morbidity and mortality. Several prospective studies from the Swiss Pediatric Sepsis Study (SPSS) have defined its incidence, clinical features, and outcomes in Switzerland. A 2017 nationwide cohort study reported an incidence of 25.1 per 100,000 children annually and a 30-day mortality rate of 7%, with risk of death closely associated with the presence and number of organ dysfunctions. A neonatal-focused analysis identified three distinct clinical subgroups (early-onset, hospital-acquired, and community-acquired late-onset sepsis), each with specific pathogens, risk profiles, and outcomes. Hospital-acquired infections were most frequent and carried the highest case fatality. In a separate cohort of previously healthy children with community-acquired sepsis, rare variants in primary immunodeficiency genes were identified in 20% of cases, although most were of uncertain significance. Another SPSS study assessed organ dysfunction scores in paediatric sepsis and found similar performance in predicting 30-day mortality. Cardiovascular, respiratory, and neurologic dysfunctions were most strongly associated with death, suggesting that simplified assessment may be feasible.

Added value of this study

This is the first genome-wide association study of paediatric bacterial sepsis based on a national cohort with blood culture-confirmed diagnoses. The cohort included 650 sepsis cases and 1395 population-based controls. A genome-wide significant susceptibility locus was identified on chromosome 9 (CTNNAL1/ELP1). Some biological relevance of this locus has previously been demonstrated independently. The study expands previous SPSS research by integrating epidemiological data with genome-wide genetics and by applying stringent ancestry correction and imputation within the clinical phenotype. The findings offer new avenues for investigating genetic contributions to paediatric sepsis susceptibility.

Implications of all the available evidence

Combined with earlier SPSS studies, these findings support a model in which both genetic and clinical factors contribute to sepsis susceptibility and outcome in children. The identification of a susceptibility locus near CTNNAL1 and ELP1 provides a new basis for mechanistic studies and implicates transcriptional regulation and NF-κB signalling pathways. Use of a microbiologically confirmed phenotype enhances interpretability and relevance for translational research. These results provide a foundation for future work on host-pathogen interactions, genetic risk stratification, and targeted interventions in paediatric sepsis.

Introduction

Sepsis represents a leading cause of death and disability in adults and children [1], [2], [3]. Globally, pediatric sepsis results in over one million annual deaths [4], [5]. Even in countries with low childhood mortality, sepsis contributes to 10 to 20% of childhood deaths, and to one out of four deaths in pediatric intensive care units (PICU) [6], [7]. Sepsis is the result of a dysregulated host response to infection leading to organ dysfunction, yet the biomolecular mechanisms and host-pathogen interactions involved remain poorly understood [8], [9], [10]. A key feature of sepsis is that it usually develops after an insidious onset of non-specific signs common to many mild infections before progressing to organ dysfunction. In clinical practice, recognizing this transition from an infected child who may solely require antibiotics to a septic child progressing to multiorgan failure (MOF) and death, is extremely challenging.

The Swiss Pediatric Sepsis Study (SPSS) is a national observational multicenter cohort study investigating blood culture-proven bacterial sepsis in children in Switzerland. Clinical and laboratory data have been have been previously reported [2], [11], [12]. We have also previously reported population-based incidence estimates and findings on the epidemiology of sepsis [1]. Based on our data, 4.4% (95%-CI 3.5-5.4) of all childhood deaths in Switzerland were associated with blood culture-proven bacterial sepsis, ranking 5th as the cause of all childhood deaths. The age-standardized incidence of sepsis was 25.1 per 100'000 children per year with an average 30-day in-hospital mortality of 6.9% (95%-CI 5.6-8.6), with highest rates observed in children under five years of age.

The incidence of sepsis is highest during early life [13], a vulnerable period of adaptive immunity development [14], [15], [16]. The mechanisms underlying susceptibility and resistance to severe infectious diseases in otherwise healthy patients remains largely unknown, however epidemiological studies suggest a strong genetic contribution [17], [18]. Previous studies have used mostly candidate gene and linkage analyses [19]. Genome-wide association studies (GWAS), which allow a comprehensive interrogation of potential associations throughout the genome, have identified common genetic variants that are associated with many infectious diseases [20]. Understanding why a minority of children become very ill during the course of an infection remains one of the unsolved riddles in host genetic research, and may harbor clues for future targeted interventions. We here use a GWAS approach to elucidate host genetic determinants of susceptibility to pediatric sepsis and disease characteristics.

Methods

Sample collection

The SPSS ran prospectively from September 1, 2011, to December 31, 2015. During the recruitment period, participating centers accounted for 77.6% of all hospital admissions and 97.8% of all pediatric intensive care unit admissions with an ICD-10 code for pathogen-specific sepsis in children below the age of 17 years in Switzerland. A total of 1269 blood culture-proven sepsis episodes were recorded in 1164 children. Detailed clinical and laboratory data were collected to assess predictors including pediatric organ dysfunction scores [2], [3], [4].

Genotype quality control and imputation

Genomic DNA was extracted from whole blood of 704 participants. From these, 670 were patients with clinical features and genomic data suitable for our analysis. Samples were genotyped using Illumina OmniExpressExome-8 v1.4 genotyping array and genotypes were called using Illumina GenomeStudio. An in-house control cohort of 1395 Swiss population samples was included. Study participants were excluded based on a missing genotype call rate of 10%. Subject independence was assessed using KING; nine samples were removed due to a high degree of kinship or duplication (pairwise identify-by-state (IBS) estimated kinship coefficient > 0.18) [21]. From the distribution of genetic distance, no exclusion of outliers was necessary. One sample was removed based on retraction of genetic consent.

Variants were removed for minor allele frequencies < 0.05, missingness > 0.1, and additionally for controls, Hardy-Weinberg Equilibrium (HWE) P < 1E-6. Reported and estimated sex was examined for discrepancy. We compared the genetic ancestry in cases to self-reported ethnicity to check for mislabeling. After quality control (QC), 650 sepsis case samples and 1395 controls remained. Genotyping data was phased (SHAPEIT2) and imputed (IMPUTE2) using the 1000 Genomes Project phase 3 reference panel. The reference genome build and LD population used was hg19/1000G Nov2014 EUR. Imputation quality was assessed and SNPs with an information score of < 0.8 or minor allele frequency < 0.05 were removed (**Figure S2**).

Association testing

GCTA was used to calculate the genetic relationship matrix (GRM) and to perform principal component analysis (PCA) to quantify population structure [22]. For case-control analysis of sepsis susceptibility, population structure/stratification was alleviated by sub-setting the dataset to remove outlying clusters as evident from PCA (**Figure S3**). This removed 140 cases and 401 controls, leaving a case-control subset of 510 cases and 994 controls of European ancestry. Datasets were merged using PLINK v1.9. SNP positions and identifiers were updated according to dbNSFP4.0a (hg19) [23]. QC was repeated after merging cases

and controls for combined cohort-specific frequencies and GCTA mixed linear model-based association (MLMA) with leaving-one-chromosome-out (LOCO) was used for analysis. SNP QC filters and association model summaries are listed in **Tables S1 and S2**, respectively. GCTA MLMA-LOCO was used for within-cohort (case only) analyses of several phenotypes as listed in **Table 1 and S2** [24]. Population structure was controlled by GRM eigenvectors and analysis covariates consisted of sex, age, and study site.

Testing for overlap with an independent mouse model study

We identified *ELP1* and *CTNNAL1* as candidate genes for susceptibility to paediatric sepsis. These genes (*Elp1* and *Ctnnal1* in mice) were independently reported by Vered et al. [26] as part of the Collaborative Cross (CC), a systems genetics resource for host-pathogen interaction studies using a multi-parental mouse genetic reference population, as reviewed by Noll et al. [25]. Vered et al. [26] challenged 73 pre-CC mouse lines with *Klebsiella pneumoniae* and identified 2 and 6 significantly associated intronic SNPs in Elp1 and Ctnnal1, respectively. To test whether the overlap of these genes across the two studies exceeded that expected by chance, we applied a weighted Z-test and Fisher's combined probability test using the reported study sizes (n=1504 for SPSS and n=328 for CC).

Ethics statement

The study was approved by the respective ethics committees of all participating centers (Cantonal Ethics Committee Bern, approval number KEK-029/11) and the study was conducted in accordance with the Declaration of Helsinki. Participants or their legally acceptable representatives provided their written consent to participate in the study after having been informed about the nature and purpose of the study, participation/termination conditions, and risks and benefits of treatment.

Role of funders

The funders were not involved in study design, data collection and analysis, interpretation, and writing of the manuscript.

Results

The demographic and clinical characteristics of the 650 study participants for whom within-cohort analysis was performed are shown in **Table 1** and **Figure S1**. Within-cohort (case only) association testing was performed on all 650 case samples. We found no association with specific sepsis characteristics, including site or type of infection, cardiovascular failure, comorbidities, pathogens identified, case fatality, PICU admission, hospital-acquired sepsis, length of PICU stay, time to death from sepsis onset, number of organ dysfunctions, length of

stay after sepsis onset, invasive ventilation, and age groups (**Table S2**). The sub-group summaries are listed with demographic and clinical characteristics in **Table 1** and **Figure S1**.

A subset of 510 cases and 994 Swiss population-matched controls were included in a case-control GWAS for sepsis susceptibiliy as shown in **Figure 1.** This analysis highlighted one significantly associated region of chromosome 9 with 13 genome-wide significant SNPs and rs28361152 as the top-associated variant (**Table 2** and **Figure 2A**). Conditional analysis with GCTA-COJO showed that all SNPs reflect the same association signal. All 13 SNPs were intronic and did not initially appear to affect protein structure, but deeper analyses revealed their functional effects as expression Quantitative Trait Loci (eQTL). The QQ-plot illustrated in **Figure 2B** indicates that genomic inflation was controlled.

The lead SNP, associated variants and gene positions in this region are shown in a LocusZoom plot in **Figure 3**. The strength and extent of the association signal relative to genomic position, local LD and recombination patterns indicate five genes of potential interest: *ELP1* (elongator acetyltransferase complex subunit 1), *FAM206A* (family with sequence similarity 206, member A), *CTNNAL1* (catenin alpha-like 1 gene), *TMEM245* (Transmembrane protein 245), and *FRRS1L* (DOMON domain-containing protein FRRS1L). Using GTEx v.8 data (**Figure S4**), we observed that the top associated variant is a significant cis-eQTL for all five genes in the LD block (**Figure 3**).

The details of all coding and non-coding variants in LD are provided in **Table S3** and **Table S4**, respectively. **Table S5** shows all significant eQTLs for the 13 SNPs associated with susceptibility to sepsis in our study. The strongest association between rs28361152 and mRNA expression differences was found for *FRRS1L*, approximately 165kb downstream of rs28361152 with increased expression in presence of the minor allele. rs28361152 was also found to act as a splicing Quantitative Trait Locus (sQTL) in *ELP1* for intron ID 108906485:108908305:clu_63656 (GRCh38) affecting exon 14 ENSE00000983463 and intron 14-15 (**Table S6**). All genes in the region are highly expressed in most tissues, with the exception of *FRRS1L*, which has low expression globally (**Figure S4**).

ELP1 and *CTNNAL1* (and murine *Elp1* and *Ctnnal1*) were identified as candidate causal genes in both our paediatric sepsis susceptibility scan and in a previously reported independent scan of 73 mouse lines challenged with *K. pneumoniae* by Vered et al. [26], in the CC study [25]. *K. pneumoniae* causes severe pneumonia and sepsis in immunocompromised individuals. We tested the probability of identifying the same two genes in the two independent scans and found it highly unlikely to be due to chance: P-

value=1.28E-13 using weighted *Z*-test (study sizes 1504 and 328 for SPSS and CC, respectively), and 1.53E-17 using Fisher's combined probability test. Infection with *Klebsiella* spp. was reported in 28 of our 510 sepsis cases. The direction and strength of the association with rs28361152 was comparable between sepsis due to *Klebsiella* spp. (OR 0.68, 95% CI 0.3-1.56) and sepsis due to other pathogens (OR 0.46, 95% CI 0.37-0.58).

Discussion

As part of the Swiss Pediatric Sepsis Study, we explored culture-proven bacteremia and sepsis-related outcomes for 650 cases and performed case-control analysis for 510 cases and 994 controls of European ancestry. Our search for human genetic determinants of susceptibility to pediatric sepsis identified a locus of interest on chromosome 9 with 13 genome-wide significant SNPs. The top associated variant, rs28361152, is an eQTL for five genes in the immediate vicinity - *ELP1, FAM206A*, *CTNNAL1*, TMEM245, and *FRRS1L*. It is also an sQTL for *ELP1*. Of note, *elp1* and *ctnnal1* have been previously reported as top candidate genes in a murine model of sepsis [26].

CTNNAL1 encodes alpha-catulin, a regulatory protein that is associated with inflammatory and immune pathways through modulation of the NF-kB pathway activity [27]. Murine Ctnnal1 is also important to cell adhesion, in response to bacterial triggers, to eliminate infection [28], [29]. *ELP1* encodes the elongator complex protein 1, a component of the RNA polymerase II elongator complex, which is involved in transcriptional elongation. This complex catalyzes the formation of carboxymethyluridine in the wobble base at position 34 in tRNAs [30]. Loss of mouse *Elp1* is embryonic lethal and can be rescued by human *ELP1* [31].

The Swiss Pediatric Sepsis Study addressed the pressing need for comprehensive, longitudinal, population-based cohorts in paediatric sepsis - a condition that remained a leading cause of death and disability globally [5]. The relatively small sample size limited the power of some analyses. In particular, the case-only analyses were hindered by insufficient sample numbers, and the case-control analysis was constrained by the limited availability of well-matched controls for non-European genetic backgrounds.

Differences in allele frequencies further highlighted the challenges in this research area. The in-house Swiss population control cohort revealed a minor allele frequency of approximately 0.3 for the lead variant, consistent with data from gnomAD and FinnGen. In contrast, African and South Asian populations exhibited much lower frequencies, 0.05 and 0.0, respectively, suggesting that sepsis susceptibility might have varied significantly with ancestry. Increasing

sample sizes and increasing the controls in future studies could enhance the ability to detect additional genetic associations more effectively.

Although GWAS provided valuable statistical associations, they fall short of pinpointing causal variants, especially in non-coding regions where genomic complexity is a major hurdle. Replication studies with independent cohorts, along with functional assays and fine mapping, are preferable to validate these associations and to provide a definitive explanation of the underlying mechanisms. Large-scale genomic approaches, such as targeted exome and whole genome sequencing (WGS), offer promising avenues for uncovering novel candidate genes and clarifying the biological mechanisms of host susceptibility and host-pathogen interactions. These insights may ultimately pave the way for targeted interventions, even though current clinical guidelines do not yet incorporate these genomic findings [32], [33]. With increased power in future studies, GWAS or WGS approaches could be employed for host-pathogen interaction analyses that reveal avenues for targeted interventions.

Through genomic analysis as part of a national multicenter cohort study for culture-proven bacterial sepsis in children in Switzerland, we identified a genomic region significantly associated with individual susceptibility to sepsis. This offers a potential route for improving the prognosis for those affected and uncovering pathogenic mechanisms involved in immune responses and host-pathogen interactions.

Conclusion

This genome-wide analysis in the Swiss Pediatric Sepsis Study identified a significantly associated locus on chromosome 9, centred around *CTNNAL1* and *ELP1*, considered to modulate paediatric sepsis susceptibility. These findings provide a potential avenue for improving prognosis and shed light on the underlying host-pathogen interactions, offering a foundation for future targeted interventions.

Weblinks

FinnGen https://r4.finngen.fi

FUMA https://fuma.ctglab.nl

GCTA https://cnsgenomics.com/software/gcta/

GTEx https://www.gtexportal.org/home/

HaploReg https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php

IMPUTE2 https://mathgen.stats.ox.ac.uk/impute/impute_v2.html

KING https://people.virginia.edu/~wc9c/KING/

LDLink https://ldlink.nci.nih.gov

LocusZoom http://locuszoom.org
PLINK http://zzz.bwh.harvard.edu/plink/
SHAPEIT2 https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html

Contributors

Dylan Lawless performed the genomic analysis and wrote the manuscript. Flavia Aurelia Hodel, Zhi Ming Xu, Christian W. Thorball, Alessandro Borghesi, Eric Giannoni, Johannes Trück, Martin Stocker, Klara M Posfay-Barbe, Ulrich Heininger, Sara Bernhard-Stirnemann, Anita Niederer-Loher, Christian R. Kahlert, Giancarlo Natalucci, Christa Relly, Christoph Berger, Thomas Riedel, Christoph Aebi, and Philipp Agyeman contributed to data collection and manuscript drafting. Jacques Fellay and Luregn J. Schlapbach led the project and supervised the study.

Declaration of Interest

The authors declare no conflict of interest.

Data Sharing Statement

Summary statistic data is available from the public repository. Access to individual-level genetic data is restricted due to research use agreements. Researchers may contact the corresponding authors (L.J.S. and J.F.) to request access, subject to institutional and ethical approvals.

Acknowledgements

We gratefully acknowledge the participation of all volunteers, and thank the medical centres and staff for their contribution.

References

[1] P. K. A. Agyeman *et al.*, “Epidemiology of blood culture-proven bacterial sepsis in children in Switzerland: a population-based cohort study,” *The Lancet Child & Adolescent Health*, vol. 1, no. 2. pp. 124–133, 2017. doi: 10.1016/s2352-4642(17)30010-x.

[2] L. J. Schlapbach *et al.*, “Mortality related to invasive infections, sepsis, and septic shock in critically ill children in Australia and New Zealand, 2002-13: a multicentre retrospective cohort study,” *Lancet Infect Dis*, vol. 15, no. 1. pp. 46–54, 2015. doi: 10.1016/S1473-3099(14)71003-5.

[3] R. P. Dellinger *et al.*, “Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012,” *Intensive Care Medicine*, vol. 39, no. 2. pp. 165–228, 2013. doi: 10.1007/s00134-012-2769-8.

[4] C. Fleischmann-Struzek, D. M. Goldfarb, P. Schlattmann, L. J. Schlapbach, K. Reinhart, and N. Kissoon, “The global burden of paediatric and neonatal sepsis: a systematic review,” *Lancet Respir Med*, vol. 6, no. 3. pp. 223–230, 2018. doi: 10.1016/S2213-2600(18)30063-8.

[5] K. Reinhart, R. Daniels, N. Kissoon, F. R. Machado, R. D. Schachter, and S. Finfer, “Recognizing Sepsis as a Global Health Priority - A WHO Resolution,” *N Engl J Med*. 2017. doi: 10.1056/NEJMp1707170.

[6] S. Ladhani, R. G. Pebody, M. E. Ramsay, T. L. Lamagni, A. P. Johnson, and M. Sharland, “Continuing impact of infectious diseases on childhood deaths in England and Wales, 2003-2005.,” *Pediatr Infect Dis J*, vol. 29, no. 4. pp. 310–3, 2010. doi: 10.1097/INF.0b013e3181d73322.

[7] M. Lantto, M. Renko, and M. Uhari, “Changes in infectious disease mortality in children during the past three decades.,” *Pediatr Infect Dis J*, vol. 32, no. 9. pp. e355-9, 2013. doi: 10.1097/INF.0b013e3182930694.

[8] M. Singer *et al.*, “The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3),” *JAMA*, vol. 315, no. 8. pp. 801–10, 2016. doi: 10.1001/jama.2016.0287.

[9] L. J. Schlapbach and N. Kissoon, “Defining Pediatric Sepsis,” *JAMA Pediatr*, vol. 172, no. 4. pp. 312–314, 2018. doi: 10.1001/jamapediatrics.2017.5208.

[10] J. Cohen *et al.*, “Sepsis: a roadmap for future research,” *Lancet Infect Dis*, vol. 15, no. 5. pp. 581–614, 2015. doi: 10.1016/S1473-3099(15)70112-X.

[11] S. Leteurtre *et al.*, “PELOD-2: an update of the PEdiatric logistic organ dysfunction score,” *Crit Care Med*, vol. 41, no. 7. pp. 1761–73, 2013. doi: 10.1097/CCM.0b013e31828a2bbd.

[12]T. J. Matics and L. N. Sanchez-Pinto, “Adaptation and Validation of a Pediatric Sequential Organ Failure Assessment Score and Evaluation of the Sepsis-3 Definitions in Critically Ill Children,” *JAMA Pediatr*, vol. 171, no. 10. p. e172352, 2017. doi: 10.1001/jamapediatrics.2017.2352.

[13]R. S. Watson, J. A. Carcillo, W. T. Linde-Zwirble, G. Clermont, J. Lidicker, and D. C. Angus, “The epidemiology of severe sepsis in children in the United States,” *Am J Respir Crit Care Med*, vol. 167, no. 5. pp. 695–701, 2003. doi: 10.1164/rccm.200207-682OC 200207-682OC [pii].

[14]D. C. Angus and T. van der Poll, “Severe sepsis and septic shock,” *The New England journal of medicine*, vol. 369, no. 9. pp. 840–51, 2013. doi: 10.1056/NEJMra1208623.

[15]R. L. Schelonka and A. J. Infante, “Neonatal immunology,” *Semin Perinatol*, vol. 22, no. 1. pp. 2–14, 1998.

[16]M. PrabhuDas *et al.*, “Challenges in infant immunity: implications for responses to infection and vaccines,” *Nature Immunology*, vol. 12, no. 3. pp. 189–94, 2011. doi: 10.1038/ni0311-189.

[17]N. Obel, K. Christensen, I. Petersen, T. I. A. Sorensen, and A. Skytthe, “Genetic and Environmental Influences on Risk of Death due to Infections Assessed in Danish Twins, 1943-2001,” *American Journal of Epidemiology*, vol. 171, no. 9. pp. 1007–1013, 2010. doi: 10.1093/aje/kwq037.

[18]D. Burgner and M. Levin, “Genetic susceptibility to infectious diseases,” *Pediatr Infect Dis J*, vol. 22, no. 1. pp. 1–6, 2003. doi: 10.1097/01.inf.0000043008.07700.6a.

[19]A. M. Sutherland and K. R. Walley, “Bench-to-bedside review: Association of genetic variation with sepsis,” *Crit Care*, vol. 13, no. 2. p. 210, 2009. doi: cc7702 [pii] 10.1186/cc7702.

[20]J. S. Bourgeois, C. M. Smith, and D. C. Ko, “These Are the Genes You’re Looking For: Finding Host Resistance Genes,” *Trends in Microbiology*, 2020, doi: 10.1016/j.tim.2020.09.006.

[21]A. Manichaikul, J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, and W.-M. Chen, “Robust relationship inference in genome-wide association studies,” *Bioinformatics*, vol. 26, no. 22, pp. 2867–2873, 2010.

[22]J. Yang, S. H. Lee, M. E. Goddard, and P. M. Visscher, “GCTA: a tool for genome-wide complex trait analysis,” *The American Journal of Human Genetics*, vol. 88, no. 1, pp. 76–82, 2011.

[23]X. Liu, C. Wu, C. Li, and E. Boerwinkle, “dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs,” *Human Mutation*, vol. 37, no. 3, pp. 235–241, 2016, doi: 10.1002/humu.22932.

[24]J. Yang, N. A. Zaitlen, M. E. Goddard, P. M. Visscher, and A. L. Price, “Advantages and pitfalls in the application of mixed-model association methods,” *Nature genetics*, vol. 46, no. 2, p. 100, 2014.

[25]K. E. Noll, M. T. Ferris, and M. T. Heise, “The Collaborative Cross: a systems genetics resource for studying host-pathogen interactions,” *Cell host & microbe*, vol. 25, no. 4, pp. 484–498, 2019.

[26]K. Vered, C. Durrant, R. Mott, and F. A. Iraqi, “Susceptibility to klebsiella pneumonaie infection in collaborative cross mice is a complex trait controlled by at least three loci acting at different time points,” *BMC genomics*, vol. 15, no. 1, p. 865, 2014.

[27]C. Wiesner *et al.*, “α-Catulin, a Rho signalling component, can regulate NF-κB through binding to IKK-β, and confers resistance to apoptosis,” *Oncogene*, vol. 27, no. 15, pp. 2159–2169, 2008.

[28]M. Ko *et al.*, “Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development,” *Development*, vol. 127, no. 8, pp. 1737–1749, 2000.

[29]W. C. Skarnes *et al.*, “A conditional knockout resource for the genome-wide study of mouse gene function,” *Nature*, vol. 474, no. 7351, pp. 337–342, 2011.

[30]U. Dalwadi and C. K. Yip, “Structural insights into the function of Elongator,” *Cellular and Molecular Life Sciences*, vol. 75, no. 9, pp. 1613–1622, 2018.

[31]Y.-T. Chen *et al.*, “Loss of mouse Ikbkap, a subunit of elongator, leads to transcriptional deficits and embryonic lethality that can be rescued by human IKBKAP,” *Molecular and cellular biology*, vol. 29, no. 3, pp. 736–744, 2009.

[32]T. Y. Tan *et al.*, “Diagnostic Impact and Cost-effectiveness of Whole-Exome Sequencing for Ambulant Children With Suspected Monogenic Conditions,” *JAMA Pediatr*, vol. 171, no. 9. pp. 855–862, 2017. doi: 10.1001/jamapediatrics.2017.1755.

[33]A. G. Randolph and R. J. McCulloh, “Pediatric sepsis: important considerations for diagnosing and managing severe infections in infants, children, and adolescents,” *Virulence*, vol. 5, no. 1. pp. 179–89, 2014. doi: 10.4161/viru.27045.

Table and Figure Legends

Table 1: Demographic and clinical characteristics of 650 Swiss Pediatric Sepsis Study GWAS participants.

Categorical variables are presented as frequencies (%) and continuous variables as median (IQR). Column percentages are presented based on available data for each variable. [a]Missing in 19, [b]Missing in 1, [c]Missing in 1, [d]Missing in 3, [e]Missing in 9, [f]In 3 cases laboratory and physiological parameters needed to define organ dysfunctions according to the 2005 consensus definition were missing, but according to the initial investigator-based classification these patients did not suffer from any organ dysfunction. CVAD, central venous access device; PICU, paediatric intensive care unit; NICU, neonatal intensive care unit.

Table 2. Significant associations identified for case and control susceptibility.

Lead SNP rs28361152 is a non-coding variant that falls within *CTNNAL1*. The remaining associated SNPs in LD ($r^2 \geq 0.8$) span *ELP1* and *FAM206A*.

Figure 1: Cohort analysis summary.

Swiss Pediatric Sepsis cohort episodes, genotyping, and quality control. Within-cohort analysis are listed in Table S2. Case-control analysis subset to suitably match cohorts.

Figure 2: GWAS Manhattan plot and QQ plot.

(A) Manhattan plot of association shows significant association signals in chromosome 9 (significant threshold P-value = $5e^{-08}$). (B) QQ plot demonstrates that the observed distribution of P-values corresponds to the expected distribution under the null hypothesis, indicating that potential confounders are well controlled, λ=1.019.

Figure 3: A LocusZoom plot for SNPs in the region flanking 200kb on either side of the selected SNP rs28361152 on chromosome 9.

P-values in –log10 scale are shown on the left vertical axis, the recombination rates are on the right vertical axis, and the chromosomal positions are on the horizontal axis. LD is shown by r^2 and represented by color. The lead SNP is in strong LD with nearest gene *CTNNAL1*, *FAM206A,* and *ELP1.* The LD block extends to *TMEM245* and *FRRS1L,* but no associated with disease is evident within the SPSS cohort.

Supplemental

Table S1: SNP quality control filters.
Samples were genotyped using Illumina OmniExpressExome-8 v1.4 genotyping array, phased with (SHAPEIT2) and imputed (IMPUTE2) using the 1000 Genomes Project phase 3 reference panel. Filtering included missing genotype call rate 10%, minor allele frequencies < 0.05, control HWE $1e^{-6}$, and subject independence by identify-by-state (IBS). Geno, genotype call rate; HWE, Hardy-Weinberg Equilibrium; MAF, minor allele frequency; LD, linkage disequilibrium.

Table S2: Association model summary.
Covariates were included to account for age, sex, study sites. Ancestry was accounted for by GRM. Case-only analysis was also performed for key criteria of clinical characteristics that may be used to assess sepsis.

Table S3. List of protein-altering variants in LD with lead SNP
Genomic coordinates are based on the GRCh 37 build. LD estimates are based on the HRC reference panel. Haploreg v4 was used to assess the protein-altering functional annotation of the proxy SNPs.

Table S4. Extended list of all variants in LD with lead SNP
The extended list of variants in LD with the lead SNP contain 1519 variants, with ranges in their MAF 0.1-0.5; D' 0.1-1.0; and R^2 0.0-1.0.

Table S5. eQTL associations for associated SNPs in GTEx v8
Genome wide significant SNPs in sepsis susceptibility case-control analysis were used to query for eQTL associations in GTEx v8. Results for the lead SNP rs28361152 are shown. P-values included are below the gene-level significant threshold of variant-gene pairs. Genomic coordinates are based on the GRCh 38 build. NES, normalized effect size; sQTL, splicing Quantitative Trait Loci.

Table S6. sQTL associations for associated SNPs in GTEx v8
Genome wide significant SNPs in sepsis susceptibility case-control analysis were used to query for eQTL associations in GTEx v8. Results for the lead SNP rs28361152 are shown. P-values included are below the gene-level significant threshold of variant-gene pairs. Genomic coordinates are based on the GRCh 38 build. eQTL, expression Quantitative Trait Loci; NES, normalized effect size.

Figure S1: Patient cohort characteristics.

(A) categorical and (B) continuous variables for patient cohort. Our cohort consisted of 650 cases after QC (quality control). Color scale low (blue) to high (yellow).

Figure S2: Imputation quality.

The information score is plotted for illustration of a randomly selected region of $5e^6$ SNPs. High quality imputation is present for variants with scores above 0.7-0.8 as shown for (A) total count and (B) cumulative sum. INFO score >0.8 was used for analysis.

Figure S3: Principal component analysis.

(A) The Swiss Pediatric Sepsis Study cohort is predominantly made up of individuals of European ancestry (83%, **Table 1**). Within-cohort association was tested for all samples with eight eigenvalues included as quantitative covariates to control for population structure in a mixed model. (B) For case-control analysis population structure required pruning to European only ancestry. The first twenty PCs are shown in (C).

Figure S4: Gene expression in candidate gene locus

Gene expression heatmap from 5 genes of interest on Chr 9 in association with sepsis. Annotated for biological context based on Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA). Significantly associated SNPs overlap *ELP1, FAM206A*, and *CTNNAL1* and all five genes were identified as expression Quantitative Trait Loci.

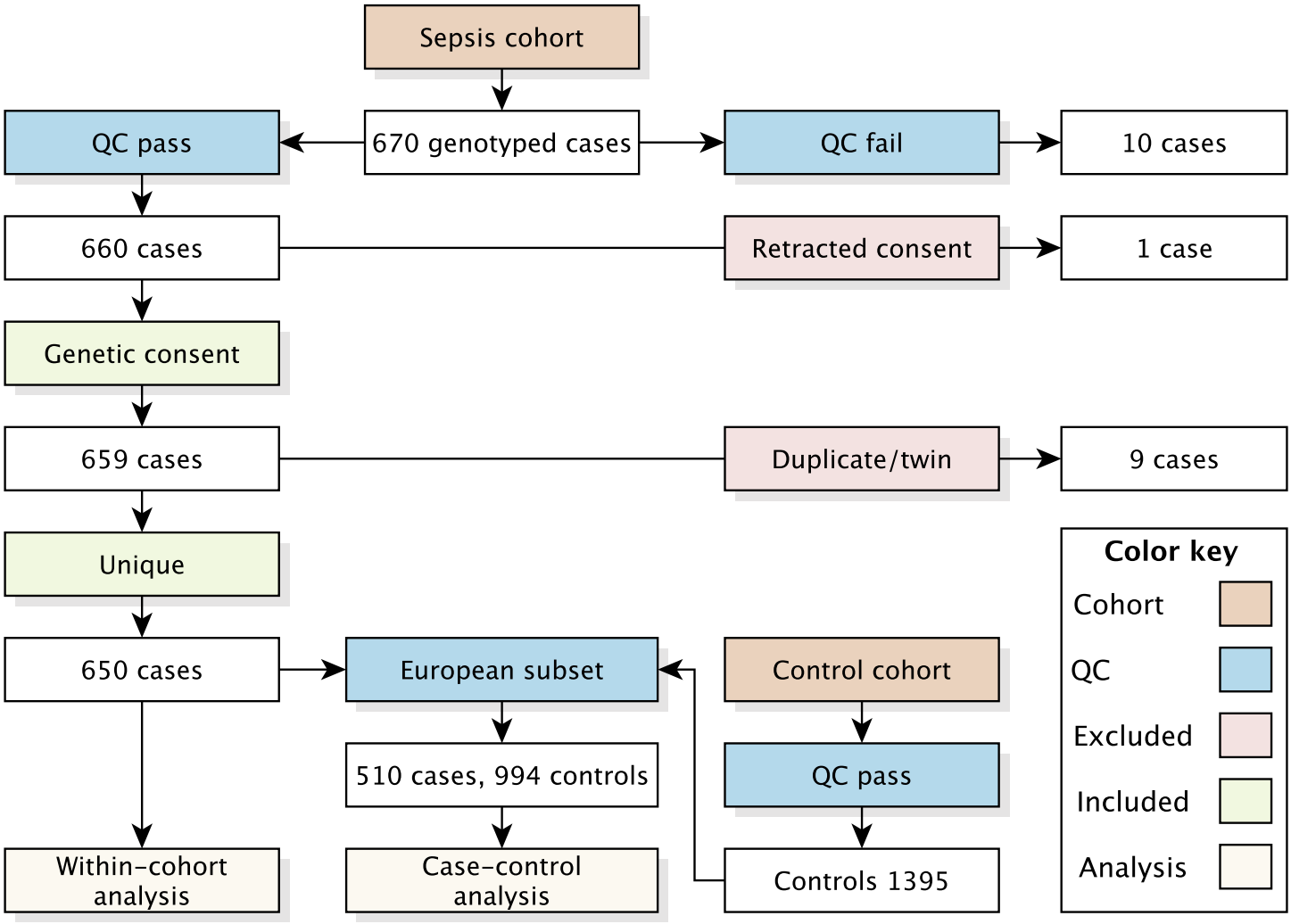
Sepsis cohort
670 genotyped cases
QC pass
QC fail
10 cases
660 cases
Retracted consent
1 case
Genetic consent
659 cases
Duplicate/twin
9 cases
Unique
650 cases
European subset
Control cohort
510 cases, 994 controls
QC pass
Within-cohort analysis
Case-control analysis
Controls 1395
Color key
Cohort
QC
Excluded
Included
Analysis

Figure 1

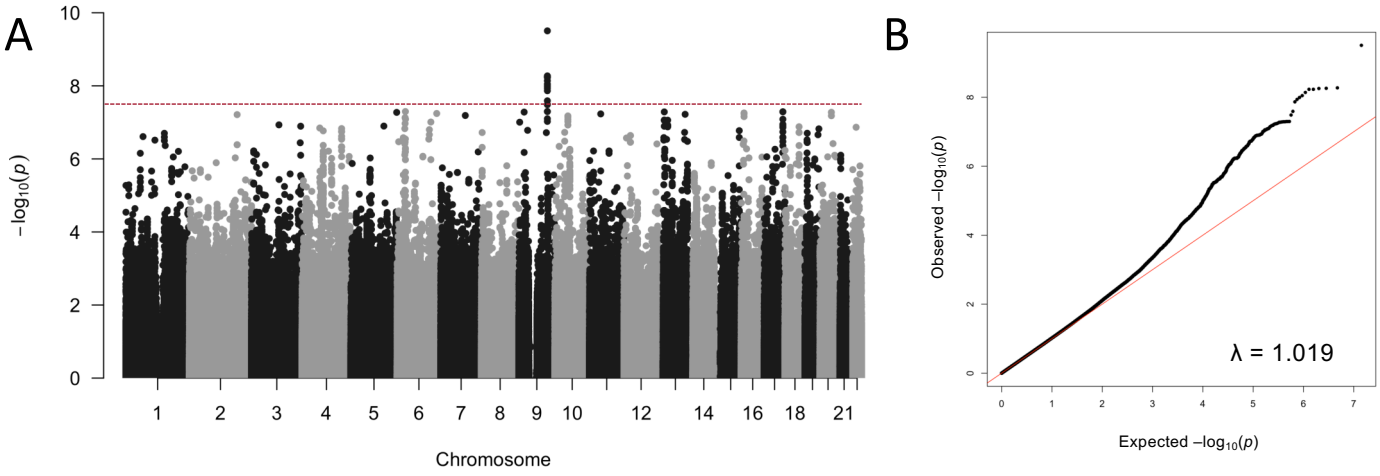
A
10
8
6
4
2
0
$-\log_{10}(p)$
1
2
3
4
5
6
7
8
9
10
12
14
16
18
21
Chromosome
B
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
λ = 1.019

Figure 2

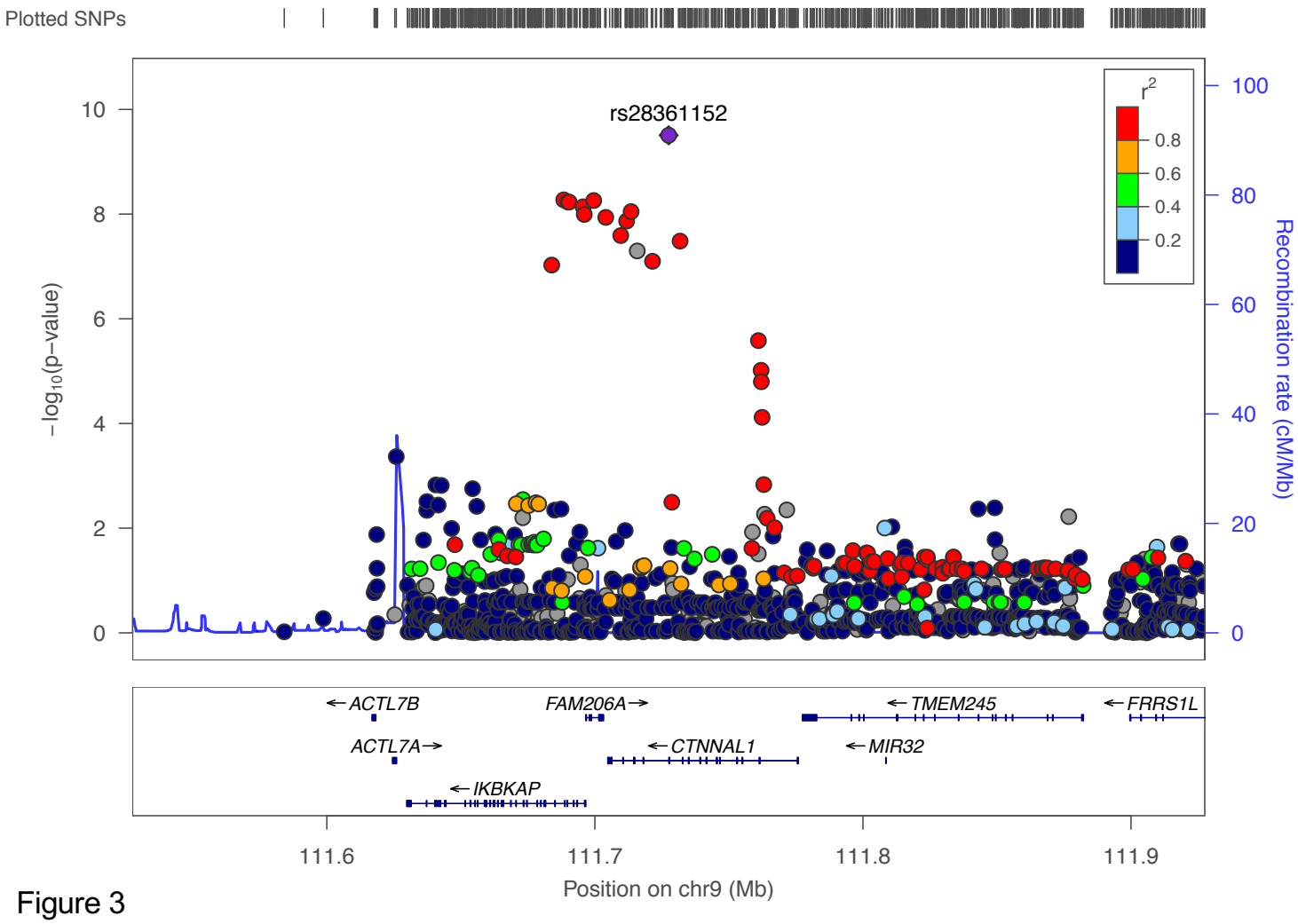
Plotted SNPs
rs28361152
r2
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
ACTL7B
FAM206A
TMEM245
FRRS1L
ACTL7A
CTNNAL1
MIR32
IKBKAP
111.6
111.7
111.8
111.9
Position on chr9 (Mb)

Figure 3

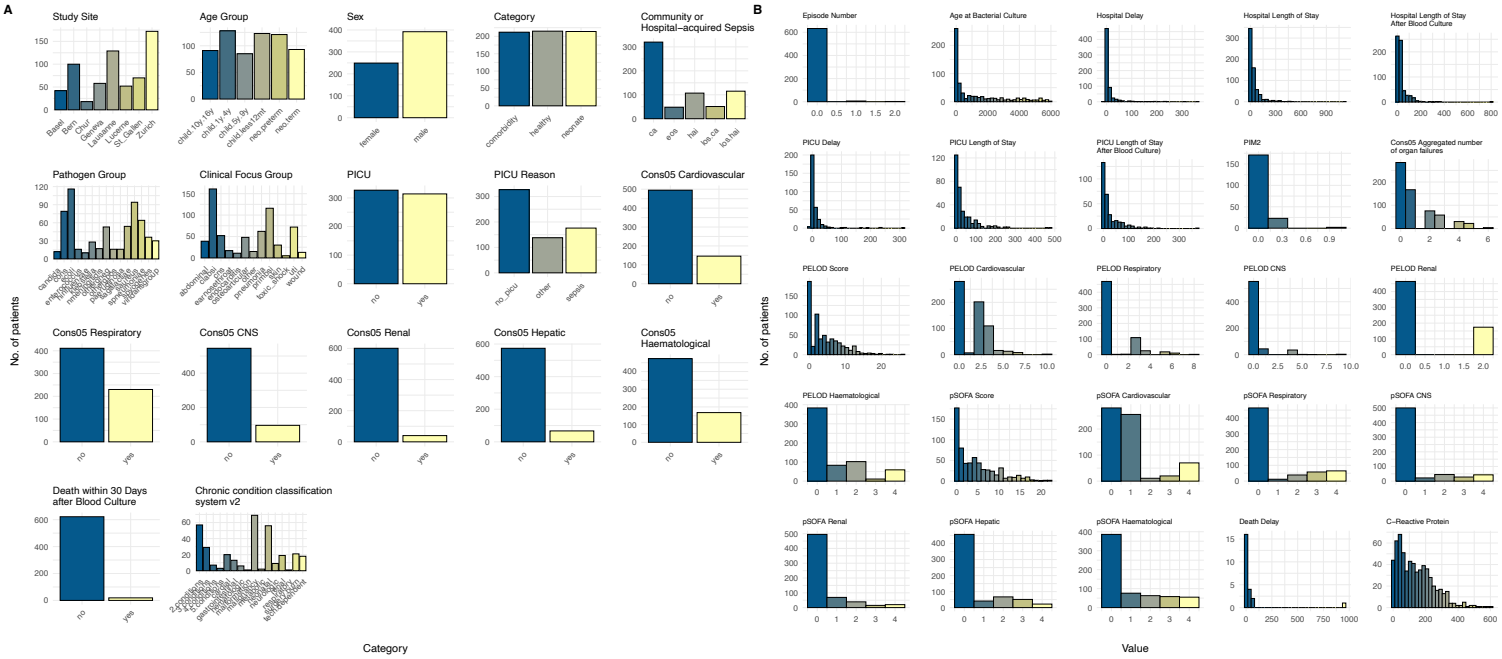
A
Study Site
Age Group
Sex
Category
Community or Hospital−acquired Sepsis
Pathogen Group
Clinical Focus Group
PICU
PICU Reason
Cons05 Cardiovascular
Cons05 Respiratory
Cons05 CNS
Cons05 Renal
Cons05 Hepatic
Cons05 Haematological
Death within 30 Days after Blood Culture
Chronic condition classification system v2
No. of patients
Category
B
Episode Number
Age at Bacterial Culture
Hospital Delay
Hospital Length of Stay
Hospital Length of Stay After Blood Culture
PICU Delay
PICU Length of Stay
PICU Length of Stay After Blood Culture)
PIM2
Cons05 Aggregated number of organ failures
PELOD Score
PELOD Cardiovascular
PELOD Respiratory
PELOD CNS
PELOD Renal
PELOD Haematological
pSOFA Score
pSOFA Cardiovascular
pSOFA Respiratory
pSOFA CNS
pSOFA Renal
pSOFA Hepatic
pSOFA Haematological
Death Delay
C−Reactive Protein
No. of patients
Value

Figure S1

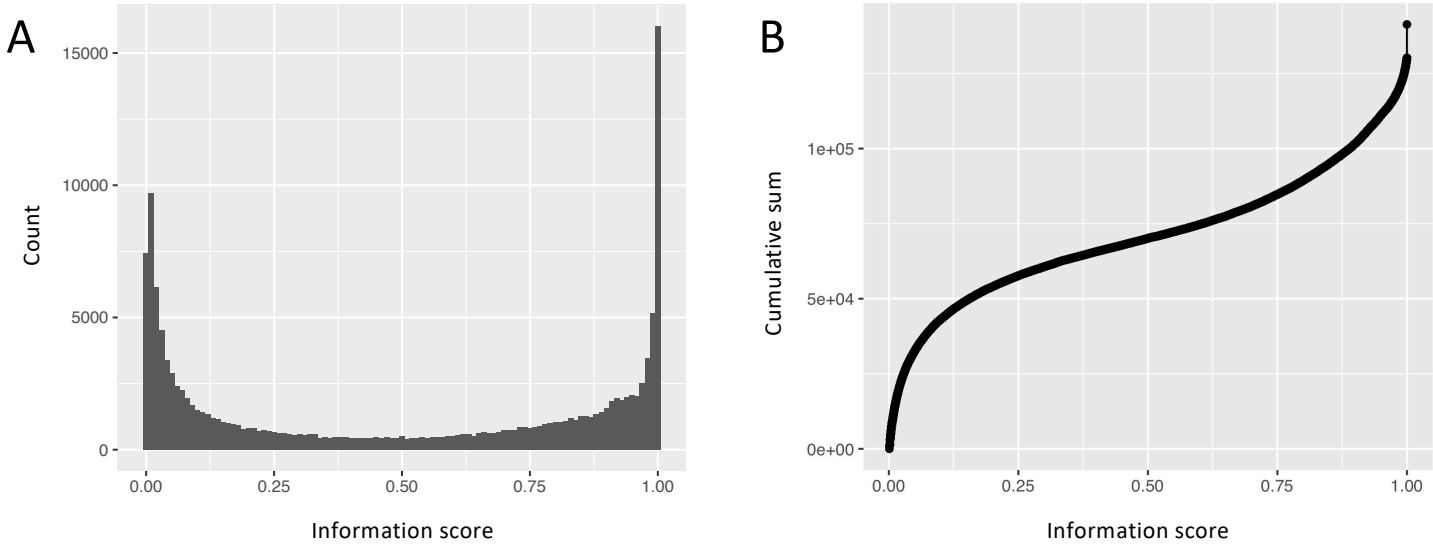
A
15000
10000
5000
0
Count
0.00
0.25
0.50
0.75
1.00
Information score
B
1e+05
5e+04
0e+00
Cumulative sum
0.00
0.25
0.50
0.75
1.00
Information score

Figure S2

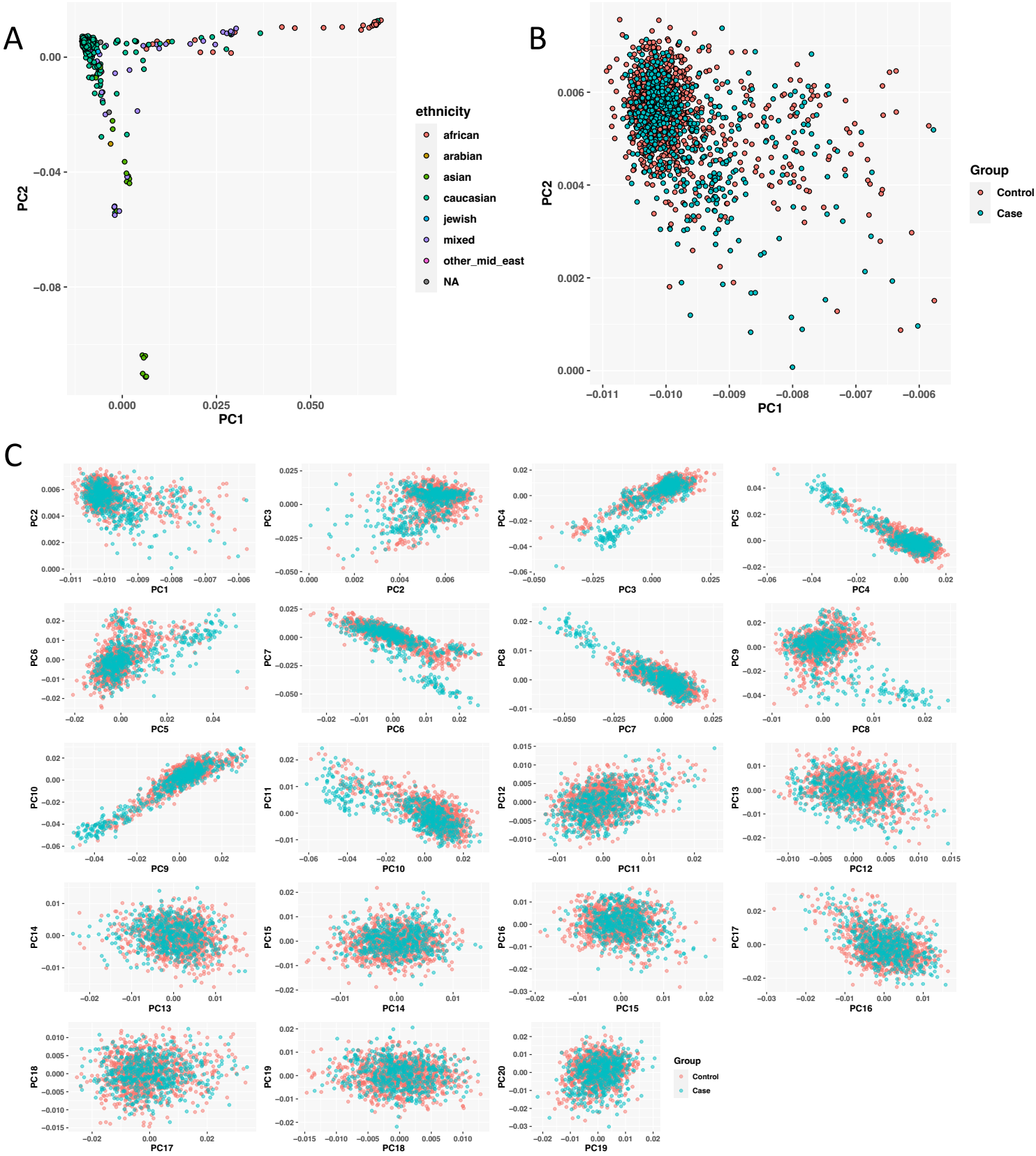
A
PC2
PC1
ethnicity
african
arabian
asian
caucasian
jewish
mixed
other_mid_east
NA
B
PC2
PC1
Group
Control
Case
C
PC1
PC2
PC3
PC4
PC5
PC6
PC7
PC8
PC9
PC10
PC11
PC12
PC13
PC14
PC15
PC16
PC17
PC18
PC19
PC20
Group
Control
Case

Figure S3

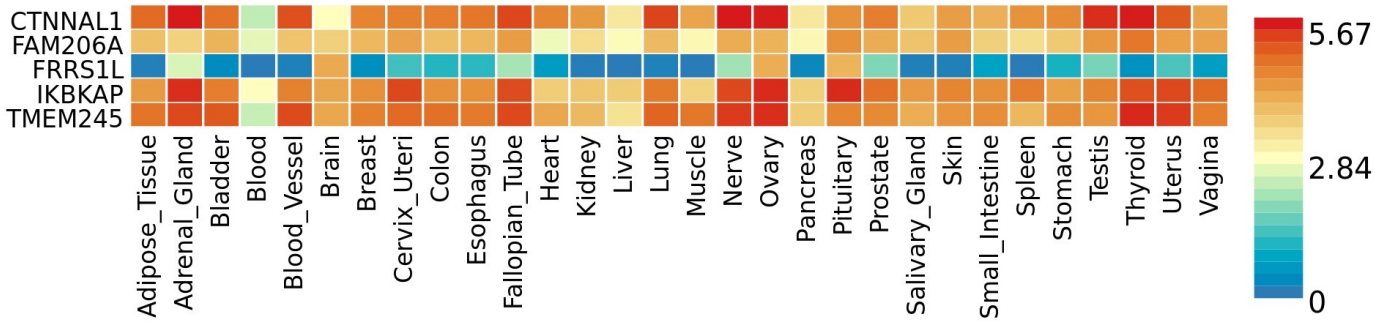
CTNNAL1
FAM206A
FRRS1L
IKBKAP
TMEM245
Adipose_Tissue
Adrenal_Gland
Bladder
Blood
Blood_Vessel
Brain
Breast
Cervix_Uteri
Colon
Esophagus
Fallopian_Tube
Heart
Kidney
Liver
Lung
Muscle
Nerve
Ovary
Pancreas
Pituitary
Prostate
Salivary_Gland
Skin
Small_Intestine
Spleen
Stomach
Testis
Thyroid
Uterus
Vagina
5.67
2.84
0

Figure S4

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract ✅
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported ✅
Objectives	3	State specific objectives, including any prespecified hypotheses ✅
Methods		
Study design	4	Present key elements of study design early in the paper ✅
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection ✅
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up ✅
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable ✅
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group ✅
Bias	9	Describe any efforts to address potential sources of bias ✅
Study size	10	Explain how the study size was arrived at ✅
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why ✅
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding ✅
		(*b*) Describe any methods used to examine subgroups and interactions ✅
		(*c*) Explain how missing data were addressed ✅
		(*d*) If applicable, explain how loss to follow-up was addressed ✅
		(*e*) Describe any sensitivity analyses ✅
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed ✅
		(b) Give reasons for non-participation at each stage ✅
		(c) Consider use of a flow diagram ✅
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders ✅
		(b) Indicate number of participants with missing data for each variable of interest ✅
		(c) Summarise follow-up time (eg, average and total amount)
Outcome data	15*	Report numbers of outcome events or summary measures over time ✅

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included ✅
		(*b*) Report category boundaries when continuous variables were categorized ✅
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses ✅
Discussion		
Key results	18	Summarise key results with reference to study objectives ✅
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias ✅
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence ✅
Generalisability	21	Discuss the generalisability (external validity) of the study results ✅
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based ✅

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.