Deriving Biological Meaning and Clinical Application for Pediatric Sepsis with Data-driven Analysis

by

Yidi Qin

BS, China Agricultural University, 2017

MS, University of Michigan, 2019

Submitted to the Graduate Faculty of the

School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2024

UNIVERSITY OF PITTSBURGH

SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Yidi Qin

It was defended on

April 5, 2024

and approved by

Daniel E. Weeks, PhD, Professor, Department of Human Genetics, Department of Biostatistics, School of Public Health, University of Pittsburgh

John R. Shaffer, PhD, Assistant Professor, Department of Human Genetics School of Public Health, Department of Oral Biology School of Dental Medicine, University of Pittsburgh

Joseph A. Carcillo, MD, Professor, Department of Critical Care Medicine, School of Medicine, University of Pittsburgh

Dissertation Director: Hyun-Jung Park, PhD, Assistant Professor, Department of Human Genetics, Department of Biostatistics, School of Public Health, University of Pittsburgh

Deriving Biological Meaning and Clinical Application for Pediatric Sepsis with Data-driven Analysis

Yidi Qin, PhD

University of Pittsburgh, 2024

Pediatric sepsis is a life-threatening syndrome characterized by abnormal immune response to infection, resulting in organ failure and mortality. However, the success of regular therapies of pediatric sepsis has been hindered by the unavoidable heterogeneity within the patient population. To enable advanced precision medicine treatment, it is of great importance to identify patients at high risk and unravel the potential biological mechanisms driving the heterogeneity. In line with this need, this study leveraged clinical, genetic, and epigenetic data to first identify pediatric sepsis patients at high risk of severe outcomes and then detect biological markers associated with the phenotype of interest.

Beyond the conventional empirical pediatric sepsis phenotypes, Aim 1 of this study applied a machine learning approach to bedside clinical features and derived four computable pediatric sepsis phenotypes, PedSep-A, B, C, and D, which exhibited distinct infection resources and sites, inflammations, metabolisms, organ failures, and mortalities. Among them, PedSep-D was distinguished by significantly more severe outcomes compared to other phenotypes. Following this discovery, gene-based analysis in Aim 2 identified several deleterious variants in one exome-wide significant (*LTBP4*, $p < 5E-8$) and two suggestive (*PLA2G4E* and *CCDC157*, $p < 5E-7$) genes associated with PedSep-D, demonstrating the contribution of rare variants in pediatric sepsis severity. Finally, epigenome-wide association analysis in Aim 3 identified one genome-wide significant (cg16704797, $p < 9E-8$) and 24 suggestive significant ($p < 1E-5$) differentially

methylated CpGs (DMCs), and one significant differentially methylated region (DMR) associated with PedSep-D. Functional analysis of the identified DMCs indicated their roles in regulating gene expression, immune cell activation, and lipid metabolisms.

This study has promoted our current knowledge of heterogeneity in pediatric sepsis and forwarded our understanding of disease pathology from perspectives of genetics and epigenetics. Furthermore, the accomplishment of this work contributed to addressing several gaps between current results from established studies and future applications in clinical programs to inform better development of precision medicine. The public health significance of findings gained from this study is particularly profound, offering the potential to revolutionize the way sepsis is diagnosed and treated in children, ultimately leading to more effective and efficient healthcare interventions.

Table of Contents

List of Tables

List of Figures

Preface

I would like to express my deepest gratitude to my supervisor and dissertation committee chair, Dr. Hyun-Jung Park, for his exceptional guidance and unwavering support through my Ph.D. training. His comprehensive oversight and insightful feedback have been crucial to my work. His expertise, understanding, and patience have significantly enriched my graduate experience. I could not make it without his support.

I am also sincerely thankful to the members of my dissertation committee, Dr. Daniel Weeks, Dr. John Shaffer, and Dr. Joseph Carcillo. Dr. Weeks is thoroughly knowledgeable in statistics and has helped me substantially in improving the quality of my dissertation. Dr. Shaffer has offered invaluable advice to help me refine my message delivery and concept illustration. Dr. Carcillo's profound expertise in pediatric sepsis has been essential in integrating clinical perspectives into my dissertation. Their collective guidance has been instrumental in both my personal and professional development, preparing me for future career endeavors.

I extend my thanks to my collaborators Dr. Soyeon Kim and Dr. Kate Kernan, who serve as exemplary role models for female researchers. I would also like to express my appreciation to my two academic advisors, Dr. Candace Kammerer and Dr. John Shaffer, for their guidance regarding my academic and career development. Special thanks to Dr. Ryan Minister and Dr. Zsolt Urban, for their great advice to help me improve the quality of my work. I want to thank all my lab mates and friends, Yulong, Zhenjiang, Jie, Rebecca, and Aditya. It has been a great pleasure to work with so many talented people during my Ph.D. program. Finally, my heartfelt appreciation goes to beloved family members for their understanding and unconditional support throughout of my Ph.D. journey. Their love and encouragement have been my anchor.

1.0 Overall Research Goal and Specific Aims

Sepsis is a life-threatening syndrome leading to a global health burden on pediatric populations. It is generally acknowledged that pediatric sepsis is characterized by a dysregulated immune response to infection (Schlapbach et al. 2024). Despite a growing understanding of the disease mechanism, the unavoidable heterogeneity of pediatric sepsis has hindered the success of regular therapies for critically ill patients (Stanski and Wong 2020). Hence, the field has made progress in developing approaches to stratify heterogeneous patients into homogeneous groups with shared features, aiming toward precision medicine. Furthermore, the development of next-generation sequencing and microarray techniques brought more possibilities to elucidate the role of genetics and epigenetics in sepsis pathology.

However, the absence of several critical components in previous studies impedes establishing a comprehensive, personalized medicine approach. First, although heterogeneity of adult sepsis patients has been recognized (Seymour et al. 2019), knowledge of pediatric sepsis subtypes remains limited. Given the immune response variation between children and adults, most current pediatric sepsis diagnoses and treatments should only be extrapolated from adult studies with caution (Weiss et al. 2020). Thus, studies of pediatric sepsis stratification are essential for exploring the heterogeneity in pediatric sepsis and guiding subtype-specific research on children. Second, the impact of sepsis host genetics (including adults and children) has been well studied in genome-wide association studies (GWASs) designed to discover common variants (Rautanen et al. 2015, Hernandez-Beeftink et al. 2022). Nevertheless, relatively little is known about the role of rare variants in the field. Additionally, most published exome-wide association studies focused on rare variants responsible for sepsis susceptibility instead of its internal heterogeneity (Backman et

al. 2021), ignoring the potential multifariousness of the biological basis for distinct subclasses. Third, prior studies suggested the involvement of epigenetic factors in pediatric sepsis and indicated the possibility of utilizing them as a diagnostic biomarker. However, these studies are either conducted in adult cohorts whose results may not be transferrable to pediatric sepsis or suffer from power issues with a limited sample size (Binnie et al. 2020, Lorente-Pozo et al. 2021). In line with this, it is of great importance to conduct methylation analysis in larger-scale pediatric cohorts to enable the detection of epigenetic signatures related to pediatric sepsis subgroups.

Therefore, the overall goal of this project is to gain a better understanding of pediatric sepsis pathology and enable progress toward precision therapeutics by addressing the absent critical components. Specifically, this dissertation consists of three components:

Aim 1: Identify pediatric sepsis subtypes in a data-driven manner via applying unsupervised consensus k-means clustering on clinical features

Aim 2: Uncover the role of rare variants in pediatric sepsis through an exome-wide gene-based association analysis

Aim 3: Detect diagnostic methylation markers of pediatric sepsis groups by an epigenome-wide association study (EWAS)

The phenotype derived in Aim 1 facilitated the identification of pediatric sepsis patients at high risk to improve clinical trial design. The gene-based rare variant analysis of Aim 2 broadened the spectrum of the genetic basis of pediatric sepsis severity. Epigenome analysis in Aim 3 strengthened the understanding of host-environment interplay in pediatric sepsis and provided more possibilities for biomarker target discovery. Altogether, the knowledge derived from this study can provide a roadmap for unraveling heterogeneity lying inside pediatric sepsis and pave the way for developing advanced precision medicine.

2.0 Introduction

2.1 Overview of pediatric sepsis

Pediatric sepsis is a syndrome in children characterized by a dysregulated host immune response to pathogen infection, with the potential to cause life-threatening organ dysfunction (Goldstein et al. 2005, Schlapbach et al. 2024). In normal scenarios, inflammation serves as a crucial response of the immune system to infection. However, when inflammation becomes uncontrolled, unregulated, self-sustaining, and intravascular, it can trigger a sequence of events that ultimately results in end-organ dysfunction in tissues far from the original insult. Without prompt treatment, pediatric sepsis can lead to severe sepsis, septic shock, multiple organ dysfunction (MOD), or death within the first 72 hours of admission to the pediatric intensive care unit (PICU) (Figure 2.1) (Miranda and Nadel 2023).

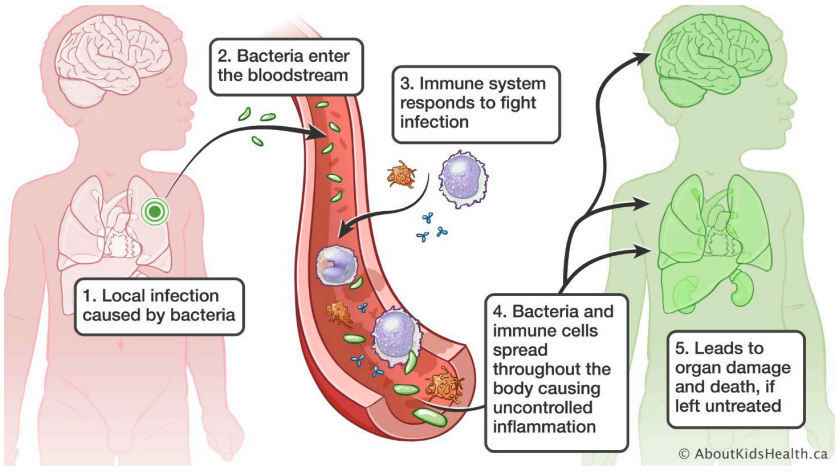

Figure 2.1 Illustration of pediatric sepsis progression

Bacterial infection represents one of the prevalent sources of pathogens, as depicted for illustrative purposes in the figure © 2004-2023 AboutKidsHealth.

2.2 Pathophysiology

In the context of the normal inflammatory response to infection such as bacteria, innate immune cells, especially macrophages, recognize and bind microbial components with their surface receptors (Takeuchi and Akira 2010, Watanabe et al. 2019). The interaction between immune cells and pathogens initiates the release of proinflammatory cytokines from macrophages and the recruitment of other inflammatory cells, such as polymorphonuclear leukocytes (PMNs), to the site (Wade and Mandell 1983, Arango Duque and Descoteaux 2014). This orchestrated process is regulated by a mixture of proinflammatory and anti-inflammatory mediators secreted from macrophages. With this balanced regulation, the host immune system undergoes a normal inflammatory process consisting of chemotaxis, adherence, ingestion, phagocytosis, and killing of invading pathogens (Rosales and Uribe-Querol 2017).

However, when the proinflammatory cytokines exceed the confines of the local environment and spread throughout the body, the normal inflammatory response escalates into pediatric sepsis. This process is usually described as a malignant intravascular inflammation (Pinsky and Matuschak 1989). The term “malignant” highlights the uncontrolled, unregulated, and self-sustaining nature of pediatric sepsis, while “intravascular” emphasizes the role of the bloodstream in distributing proinflammatory cytokines. Most importantly, “inflammation” signifies the essence of pediatric sepsis as an exaggeration of the normal inflammatory response. Although the exact mechanism triggering the transition from a normal response towards pediatric sepsis remains elusive, it is widely attributed to factors such as the toxic products of pathogens (Kellum and Ronco 2023), over-release of proinflammatory cytokines (Angurana et al. 2021), complement activation (Li et al. 2021), and host genetic predispositions (Wong 2012, Lu et al. 2019).

Pediatric sepsis may have multiple effects on cells and organs of the human body. Widespread cellular injury may take place as a result of the dysregulated inflammatory response. Although it is uncertain how pediatric sepsis causes cellular injury, multiple mechanisms have been proposed, including tissue ischemia characterized by an insufficient oxygen supply due to derangement in the metabolic autoregulation (Bateman et al. 2003); cytopathic injury caused by impaired mitochondrial electron transport (Weiss et al. 2019); and an alternate rate of apoptosis of activated immune cells resulting from elevated levels of proinflammatory cytokines (Cheng et al. 2020). In severe pediatric sepsis, the cellular damage often evolves into organ dysfunction, a condition where organs are unable to maintain their essential functions (Lelubre and Vincent 2018). The organs that are most commonly impacted include the cardiovascular, respiratory (pulmonary), neurologic, hematologic, renal, and hepatic organ systems (Figure 2.2).

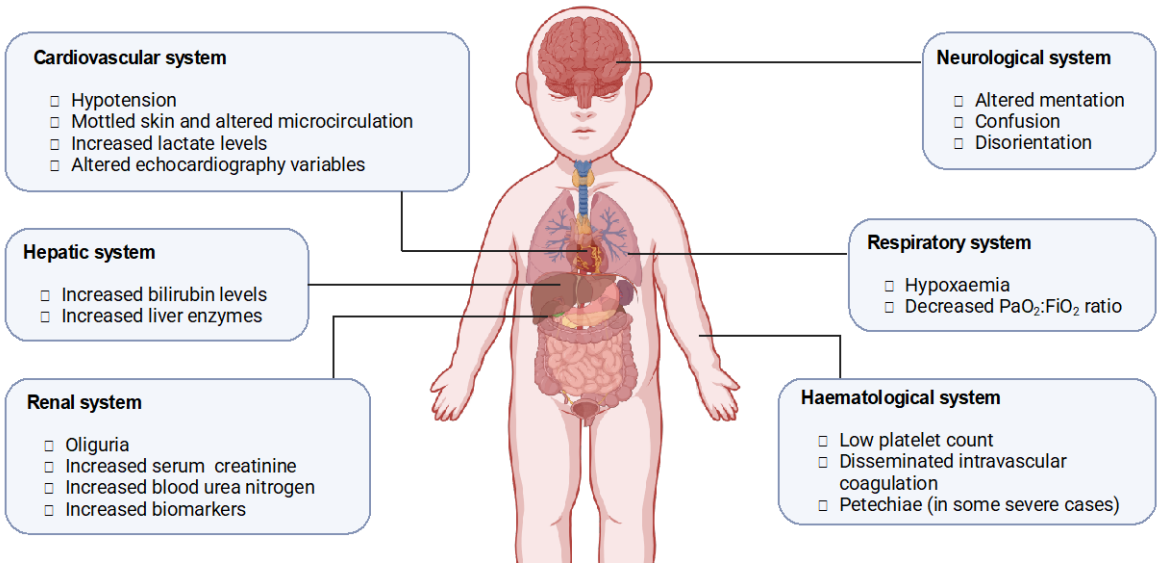

Figure 2.2 Six organ systems in which dysfunction commonly occurs

Adapted from Christophe Lelubre et. al. (Lelubre and Vincent 2018) using BioRender.com (2024)

2.3 Epidemiology

The overall burden of pediatric sepsis is high globally with important regional differences. According to a meta-analysis of 15 studies, the estimated incidence of pediatric sepsis is approximately 1.2 million cases per year, including cases from both PICU and hospital settings (Fleischmann-Struzek et al. 2018). Varying across by countries, mortality rates ranged from 1% to 5% in the pediatric sepsis population, and from 9% to 20% in the severe pediatric sepsis population. In the United States, approximately 75,000 children are hospitalized for severe sepsis each year with an annual incidence of about 1 case per 1000 population (Hartman et al. 2013).

2.4 Prevention, diagnosis, and management

In 2020, the Surviving Sepsis Campaign (SSC) released the most updated international guidelines for managing sepsis and septic shock in children, which proposed several recommendations for clinical care (Weiss et al. 2020). Essentially, these recommendations indicate that early identification and timely intervention are key principles to patient survival.

Several critical stages are crucial in the clinical care of pediatric sepsis. To prevent pediatric sepsis development, advanced practices reducing infection incidences are particularly significant in high-risk populations. There are three primary types of infection resources: bacterial, viral, and fungal, each characterized by its distinct mechanism of action (Dolin et al. 2019). In addition to infection, the major risk factors include age, chronic and serious illness, impaired immunity, and breach of natural barriers (Gaines et al. 2012, Mercurio et al. 2023). In the diagnosis and prognosis of pediatric sepsis, a variety of clinical features and criteria are widely applied, including systemic

inflammatory response syndrome (SIRS) criteria (Goldstein et al. 2005), Sequential Organ Failure Assessment (SOFA) criteria (Matics and Sanchez-Pinto 2017), quick Sequential Organ Failure Score (qSOFA) (Eun et al. 2021), and Phoenix Criteria Score (Schlapbach et al. 2024). Additionally, laboratory tests are required to complement the clinical examination, where traditional laboratory hematological, biochemical, and microbiological tests play essential roles. Pediatric sepsis management generally consists of three components: infection control, hemodynamic stabilization, and modulation of the septic response, during which empirical antibiotic therapies and organ-supportive treatment are at the core of all the interventions (Weiss et al. 2020). Unfortunately, notwithstanding endeavors of multiple preclinical studies over the last three decades, no new effective drug emerged that has improved sepsis patient outcomes, implying the challenges in the development of the treatment for the disease.

2.5 Heterogeneity and precision medicine

Despite the accumulation of knowledge of pediatric sepsis mechanisms, the underlying heterogeneity of the disease has arisen as the primary barrier to achieving optimal treatment outcomes (Leligdowicz and Matthay 2019). While clinical diversity in pediatric sepsis is widely recognized, there is no comprehensive approach to measuring it. This gap can be partly explained by the multifaceted source of heterogeneity at the individual patient level, encompassing factors such as infection etiologies, individual comorbidities, environmental exposures, genetic backgrounds, and the timeline of diagnosis and treatments (Marshall 2014, Hotchkiss et al. 2016). As these factors impact both the disease evolution and intervention response, subgroup-based studies that focus on a small subset of patients of interest offer valuable insight for both scientific

inquiry and clinical application. This targeted approach, known as precision medicine, has proven effective in managing other complex diseases such as cancer, by tailoring prevention, diagnosis, and treatment strategies to individual patient characteristics (Dienstmann et al. 2017). Thus, implementing precision medicine in the context of pediatric sepsis is advocated as a solution to the heterogeneity of the disease.

2.6 Rationale for the specific aims

2.6.1 Aim 1: Derivation of pediatric sepsis subtypes

To tackle the therapeutic challenge brought by heterogeneity, researchers have proposed various approaches to identify adult and pediatric sepsis subclasses within the broad syndrome. Generally, these approaches strive to classify patients into homogeneous groups by measuring the similarity of patients from different perspectives. Sepsis subtyping strategies can be further categorized based on distinct research goals (exploratory, prognostic, predictive), computational methods (unsupervised, supervised), data types (clinical, biological), or targeted populations (adult, pediatric) (DeMerle et al. 2021).

Usually, there are three study objectives for sepsis patients subtyping: exploration, prognosis, and prediction (DeMerle et al. 2021). Depending on the specific objective, researchers have been using diverse computational methods on an extensive range of data sources. The first category of subtyping analysis is designed to reveal the hidden pathological or molecular pattern in a high-dimensional space of patient features, during which a set of subclasses with shared characteristics can be naturally produced. Since this objective requires uncovering latent

mechanisms from the data automatically and does not involve outcome-related information, methods such as unsupervised clustering fit the requirement well and thus have been broadly applied. Commonly used unsupervised clustering methods include the k-means clustering (Lloyd 1982), hierarchical clustering (Jr 1963), mixture models (McLachlan and Basford 1988), and so on. Taking advantage of these unsupervised clustering algorithms, researchers have derived multiple novel sepsis subtypes from clinical and transcriptomics data (Table 2.1).

Table 2.1 Sepsis subtypes identified through unsupervised learning methods

Subtypes	Method	Data	Population	Refs
A, B, C	hierarchical clustering	gene expression	pediatric	(Wong et al. 2009)
'Shock with elevated creatinine', 'Minimal multi-organ dysfunction syndrome', 'Shock with hypoxemia and altered mental status', 'Hepatic disease'	self-organizing maps, k-means clustering	clinical	adult	(Knox et al. 2015)
SRS1, SRS2	hierarchical clustering	gene expression	adult	(Davenport et al. 2016)
Mars1-4	hierarchical clustering	gene expression	adult	(Scicluna et al. 2017)
Inflammopathic, Adaptive, Coagulopathic	k-means and PAM clustering	gene expression	adult and pediatric	(Sweeney et al. 2018)
Profile1-4	latent profile analysis (mixture models)	clinical	adult	(Zhang et al. 2018)
'hyperthermic, slow resolvers', 'hyperthermic, fast resolvers', 'normothermic', hypothermic	group-based trajectory modeling (mixture models)	clinical	adult	(Bhavani et al. 2019)
α, β, γ, δ	k-means clustering	clinical	adult	(Seymour et al. 2019)
'Severe, persistent encephalopathy', 'Moderate, resolving hypoxemia', 'Severe, persistent hypoxemia and shock', 'Moderate, persistent thrombocytopenia and shock'	subgraph augmented nonnegative matrix factorization	clinical	pediatric patients with multiple organ dysfunction syndrome (MODS)	(Sanchez-Pinto et al. 2020)

In contrast to the first objective, subtyping conducted for prognostic or predictive purposes focuses on subgroups of patients who are more likely to develop a severe outcome or benefit from a given therapy, respectively. The outcome-driven prerequisite of these two objectives can be well met by supervised learning methods that generalize a model from training data to predict outcomes with the highest accuracy. Notably, while supervised methods align with the prognostic and predictive objectives well and contribute to health care decisions directly, they cannot replace the role of unsupervised methods in exploratory analysis. This is because supervised methods highly rely on prior knowledge of outcomes or targeted treatment, so they lack the power to discover new subtypes from the patients.

Although most stratification research concentrates on the adult population, remarkable progress has been made in pediatric sepsis subtyping. In most pediatric targeted studies, transcriptomic data were analyzed to explain the dynamic outcome of patients. For instance, Wong et al. leveraged gene expression profiles to describe transcriptionally distinct subclasses from PICU patients (Wong et al. 2009). Following pediatric subtyping, the team further classified adult sepsis patients using the same strategy and observed similar biological features in some endotypes, albeit not replicable for other endotypes, suggesting a potential interaction between age and subtype assignment (Wong et al. 2017). Another important work was accomplished by Carcillo and colleagues, where researchers assessed three inflammation phenotypes derived from adult sepsis studies and found associations with severe outcomes such as macrophage activation syndrome and mortality (Carcillo et al. 2019). Altogether, these achievements proved the intrinsic heterogeneity in pediatric sepsis and built a solid foundation for future studies.

Compared to adult sepsis, the classification of pediatric sepsis patients is in a stage of rapid development, while several domains remain elusive. First, regardless of the achievement made in

transcriptomic subtyping, gene expression quantification techniques must be performed rapidly at a reasonable cost to meet the need for real-time diagnosis. This requirement restricts proof-of-concept molecular subtypes from being extended to clinical feasibility. Therefore, it is of great importance to incorporate routinely available clinical data into pediatric patient stratification. Second, when studying high-dimensional pediatric clinical data, conventional knowledge-based and supervised subtyping strategies may fail to make full use of data and lose the opportunity to uncover a subtype with unrealized biological meaning. Hence, how to choose the proper subtyping method is a topic waiting for further study. Third, as various researchers have identified multiple subtypes, it raises a question shared by adult sepsis studies: is there a way to assess the validity of newly discovered subtypes and examine their relevance to established subtypes? Integrating external resources and different types of data into follow-up analysis may serve as a key to unraveling this question. Last but not least, a logical next step after subtyping will be transforming knowledge gained from pediatric subtypes into practical tools for risk stratification and group-based management. To our knowledge, there has not been such a tool for pediatric sepsis diagnosis and treatment, implying a great potential to be explored in the field. To address the questions proposed above to help inform trial design for targeted pediatric sepsis groups, a large-scale stratification pipeline using first-day clinical features of pediatric sepsis patients is needed.

2.6.2 Aim 2: Exome-wide gene-based rare variants analysis on pediatric sepsis patients

Although sepsis is initiated by infection and is considered a complex disease under the effect of multiple factors, the heterogeneity in the host response to sepsis is thought to be partially explained by host genetic factors (Villar et al. 2004). The evidence first stems from the landmark study published by Sørensen et al. in 1988, where the authors followed 960 families including

children placed with adoptive parents unrelated to them, and reported an over five-fold increased risk of death observed in infected adoptees with one biological parent who had died from infection (Sorensen et al. 1988).

Stimulated by Sorensen's work, genetic association studies targeting candidate genes have made some progress in connecting well-known functional genes with sepsis (Sutherland and Walley 2009). For example, the A allele of the G-to-A polymorphism at the -308 in the promoter region of the tumor necrosis factor-α (TNF-α), one of the most important proinflammatory cytokines regulating the immune response to infection, was found to be associated with mortality in patients with septic shock (Mira et al. 1999). Similarly, several SNPs localized in or adjacent to genes such as inflammatory cytokines, (e.g., IL-1 and IL-6), pathogen-recognition receptors (e.g., TLR4 and LBP), and immunity genes (e.g., FcγRIIA) have been reported to be associated with sepsis susceptibility and outcomes (Lorenz et al. 2002, Schluter et al. 2002, Zhang et al. 2014, Beppler et al. 2016, Lu et al. 2018). Notably, Lu et al. conducted a comprehensive literature review followed by a meta-analysis to investigate 405 variants within 176 distinct genes. The study showed that 29 variants of 23 genes were significantly associated with the risk of sepsis (Lu et al. 2019).

Beyond the candidate gene analysis, it is important to unravel novel genetic signals to broaden the current knowledge of sepsis. Taking advantage of the high-throughput sequencing technique, researchers have conducted several genome-wide association studies (GWAS) on adult and pediatric populations to investigate genetic influences on sepsis susceptibility and outcome as summarized in Table 2.2 (Rautanen et al. 2015, Srinivasan et al. 2017, Butler-Laporte et al. 2020, D'Urso et al. 2020, Rosier et al. 2021, Hernandez-Beeftink et al. 2022). Among the identified SNPs from GWAS, several SNPs are located in or near genes that present potential functions in the

immune response to infection. For example, the FER gene identified by Rautanen et al. in 2015 encodes a non-transmembrane receptor protein tyrosine kinase that regulates cell adhesion and mediates signaling (Hao et al. 1991). It can impact leucocyte recruitment and intestinal barrier dysfunction in response to bacterial lipopolysaccharide (McCafferty et al. 2002, Qi et al. 2005). Moreover, a mice study found that FER can inhibit neutrophil chemotaxis, thereby hindering pathogens clearance and causing tissue damage (Khajah et al. 2013). As another example, the SAMD9 gene identified by Hernandez-Beeftink et al. in 2022 encodes a protein that belongs to the SAM domain-containing protein family (Aviv et al. 2003), which plays multiple roles in cellular processes and regulates inflammatory response during tissue injury and apoptosis (Mekhedov et al. 2017).

Table 2.2 SNPs identified associated with sepsis and septic shock

variant-risk allele	chr: position	mapped gene	associated trait	ref
rs4957796-C	5:109066439	FER	Sepsis mortality from pneumonia	(Rautanen et al. 2015)
rs79423885-?	6:103362128	R3HDM2P2, GRIK2		
rs13380717-?	16:86870529	LINC02188, LINC02181	Neonatal sepsis[1]	(Srinivasan et al. 2017)
rs2412930-?	4:58722063	LINC02619, LINC02494		
rs946883-?	6:163778989	RN7SL366P, QKI		
rs41461846-?	2:218448631	VIL1		
NA	2:219344165	NA		
rs6717433-?	2:218510988	USP37		
rs114078858-C	1:201742457	NAV1, IPO9-AS1	Sepsis hospital admission	(Butler-Laporte et al. 2020)[2]
rs113187813-C	3:44767224	KIF15		
rs9257270-A	6:28853318	NOP56P1, LINC01623		
rs577432066-G	5:66388602	LINC02229, LINC02065	Sepsis 28-day mortality	
rs79422343-T	4:175635426	GPM6A		
rs142021422-C	7:136773299	PSMMC1P3, CHRM2		
rs181021474-T	7:130000947	Y_RNA, ZC3HC1		
rs140871186-C	9:133030315	EEF1A1P5, GTF3C5		
rs149187226-G	14:92621755	RIN3		
rs147296048-A	7:1953164	MAD1L1		
rs35597084-C	6:31555893	NFKBIL1	Sepsis 90-day mortality	
rs9489328-G	6:97656699	MIR2113, MMS22L	Septic shock	(D'Urso et al. 2020)
rs11167801-C	5:143090769	ARHGAP26	Septic shock resolution	
rs7698838-T	4:57222516	RPS26P24, IGFBP7-AS1		
rs17128291-A	14:92416482	SLC24A4		
rs368584-G	13:110492726	COL4A2	Septic shock 28-day mortality	

1 The only study conducted in non-adult sepsis population in the table.

2 The GWAS results were extracted from a Mendelian Randomization study framework.

Table 2.2 Continued

variant-risk allele	chr: position	mapped gene	associated trait	ref
rs16857698-G	3:145685067	NA	Septic shock mortality before day 7	(Rosier et al. 2021)
rs5029231-T	3:145701146			
rs6763296-C	3:145709314			
rs16857836-T	3:145752473			
rs4544-C	8:143994806	CYP11B2		
rs11991278-T	8:144001245			
rs6981918-A	8:144007939			
rs956727-G	9:868446933	SLC28A3		
rs79744468-A	12:112927208	PTPNN11		
rs10849640-A	12:119712137	NA		
rs10849641-T	12:119721354			
rs10849642-T	12:119725314			
rs12491812-T	3:50556581	CACNA2D2	Septic shock mortality between day 7 and day 28	
rs2239753-C	3:50645158	CISH		
rs2239752-T	3:50645413			
rs2239751-C	3:50647888			
rs743753-T	3:50651395	MAPKAPK3		
rs616689-A	3:50668532			
rs9879397-A	3:50685642			
rs2170840-C	3:50686517			
rs12492982-T	3:50698155			
rs2035484-G	3:50721892	DOCK3		
rs17051403-A	3:50751643			
rs17072628-A	3:65229760	NA		
rs7804669-G	8:89929277	NA		
rs7953683-T	12:79993704	PAWR		
rs1502522-G	17:51544776	NA		
rs1393467-C	17:51560869			
rs146257041-G	11:26962266	SLC5A12, FIBIN	Sepsis 28-day mortality	(Hernandez-Beeftink et al. 2022)
rs138347802-G	13:60793063	LINC00378		
rs34896991-T	7:93101432	SAMD9		
rs113925942-C	13:92899923	LINC00363, GPC5		
rs183364907-T	14:57642742	SLC35F4		
rs76805442-A	5:85104034	RBBP4P6, PPIAP79		
rs114658749-T	3:196048273	TFRC		
rs114581095-T	3:36005805	RPL36AP17, RFC3P1		

Although genetic factors contributing to sepsis-related phenotypes have been identified from GWAS research, identified risk variants in most of the published GWAS fall outside coding regions, limiting a straightforward translation to the disease mechanism (Flores 2015). This is because GWAS was initially designed based on the "common disease, common variants" hypothesis, in which more focus is put on the higher end of the allele frequency spectrum (Manolio et al. 2009). Theoretically, there is an inverse relationship between effect size and population frequency of the variant, which results from natural selection. As the high frequency of common variants in the population implies less selection pressure from disease, common variants are more likely to confer small increments in risk and can only explain a limited proportion of heritability.

To address the so-called "missing heritability" problem and enable the detection of functional genetic factors, several studies have taken advantage of the exome-wide sequencing (WES) technique to concentrate on rare variants with a minor allele frequency (MAF) lower than 1% (Backman et al. 2021). Protein-affecting mutations can heavily predispose individuals to disease. Consequentially, disease-related variants are selected against during human evolution, leading to a low frequency in the population. Constructed on this principle, studies targeting rare variants in coding regions have a high potential to unravel the functional genetic components of sepsis. For instance, the aforementioned GWAS of 28-day sepsis mortality followed up their best findings in an independent WES study (Taudien et al. 2016), where researchers leveraged deleterious rare variants to predict sepsis courses in two ethnic groups. Another attempt is an analysis conducted on previously healthy children with bacterial sepsis (Borghesi et al. 2020). Applying WES on 176 pediatric patients, the study detected 41 rare variants of uncertain significance in a group of targeted primary immunodeficiency genes.

Despite the benefits regarding function interpretation and direct biological insight, performing analysis with rare variants is much more challenging compared to common variants. As one of the dominant issues, the scarcity of rare variants limited the power of classical individual-variant-based association tests to discover single novel signals from an exome-wide exploratory analysis. A commonly adopted strategy to boost power is to collapse rare variants into biologically relevant groups, such as genes, regions, or pathways, and then test cumulatively the effects of multiple variants in a group (Povysil et al. 2019). Employing this aggregation strategy in the general framework of association tests, researchers have developed numerous methods. Most of them can be categorized into two major types: burden test and variance-component test (Lee et al. 2014). Burden tests include the CMC method (Li and Leal 2008), WSS (Madsen and Browning 2009), MZ test (Morris and Zeggini 2010), etc., in which genetic scores are generated through collapsing rare variants. These burden tests assume that a vast majority of variants in a tested group are causal, and their effects are in the same direction. Unlike the burden tests, variance-component tests, including SKAT (Wu et al. 2011), SSU test (Pan 2009), C-alpha (Neale et al. 2011), etc., examine the variance of genetic effects and suppose the existence of both trait-increasing and trait decreasing variants or a limited number of causal variants. Concerning the moderate percentage of causal variants with opposite directions, combined tests such as SKAT-O (Lee et al. 2012), Fisher method (Derkach et al. 2013), or MiST (Sun et al. 2013), can take advantage of both burden and variance- component tests to gain the largest power. The development of these approaches based on the principle of aggregation enables the success of causal gene identification for multiple diseases.

In the field of sepsis, the role of rare variants has not been fully understood. First, although GWASs enable linking common variants with sepsis, common variants are often located in non-

coding regions where the function is currently poorly understood, thereby studies targeting coding regions are merited. Second, most works concentrated on targeted regions or gene sets according to prior knowledge (Borghesi et al. 2020), limiting the capacity to discover novel functional genes related to sepsis. Third, in addition to studying the genetic factors related to sepsis susceptibility and mortality, identification of genes related to the patients at high risk of developing severe outcomes may provide a deeper insight into disease pathology and yield new therapeutic strategies, yet no works have been done for this purpose. Given the three problems proposed above, the Aim 2 project, which is based on whole-exome sequencing and aggregation test approaches, focused on the role of rare variants and revealed their effects on pediatric sepsis subtypes. Insights from this project will enhance our knowledge of the genetic basis of sepsis and inform more tailored applications in precision medicine.

2.6.3 Aim 3: Epigenome-wide association analysis on pediatric septic patients

In addition to host genetics, epigenetics is another crucial source of biological factors contributing to disease heterogeneity. Epigenetics encompasses a series of regulation activities that show plasticity in response to exogenous environmental exposures and govern gene expression without changing the DNA sequence (Goldberg et al. 2007). Therefore, despite the identical genetic information shared across organisms, specific regulated genes can be expressed in various manners under the mediation of epigenetics. Under the broad terminology of "epigenetics", there are multiple regulatory mechanisms, including but not limited to DNA methylation, post-translational histone modification, non-coding RNAs, and others (Portela and Esteller 2010). Among them, DNA methylation is one of the most widely analyzed modifications, partially because of the fast development of its quantification methods (Kurdyukov and Bullock 2016).

DNA methylation describes a chemical process in which a methyl or a hydroxymethyl group is added to the fifth carbon of a cytosine under the mediation of DNA methyltransferase (DNMT) enzyme to form a 5-methylcytosine (Bird 2002). In human cells, this process occurs predominantly at cytosine-guanine dinucleotides (CpGs) characterized by a cytosine nucleotide immediately followed by a guanine nucleotide along the 5' to 3' direction (Jones 2012). CpGs are generally depleted in the human genome, except for enrichment in small genomic regions such as CpG islands (CGI) which are rarely methylated (Gardiner-Garden and Frommer 1987). Since over half of genes contain CGI in their promoters, transcription of genes is frequently maintained in a permissive state. However, hypermethylation in gene promoters can lead to local chromatin structural alternation and thereby prevent the binding of transcriptional factors, resulting in gene repression. In contrast, hypomethylation in gene promoters is associated with gene activation (Jones and Takai 2001).

Like genetics studies, array- and sequencing-based methylation profiling technologies enable large-scale epigenome-wide association studies (EWASs) to detect differentially methylated sites or regions associated with diseases (Rakyan et al. 2011). In site-based studies, the general task is to perform genome-wide DNA methylation scanning and compare mean methylation levels of each CpG site across groups (such as cases and controls) to identify differentially methylated positions (DMPs) (Campagna et al. 2021). Linear regression models are frequently applied by researchers in these studies, allowing for adjusting potential confounders such as sex, age, cell types, and other potential batch effects. However, site-based studies suffer from the limited power issue, especially in studies with small sample sizes. As a solution, region-based studies have been developed to search for differential methylations at the region level (DMR) as opposed to the site level (Campagna et al. 2021). In light of this design, DMR analysis

takes high correlation across adjacent CpG sites into account and also offers superior power over DMP analysis by reducing the multiple testing burden. Most DMR analysis tools can be categorized into two groups, unsupervised and supervised methods (Mallik et al. 2019). The unsupervised methods involve PCA (Jolliffe 2002), SPCA (Chen et al. 2008), KPCA (Gao et al. 2011) , SKAT (Wu et al. 2011), and others. They first utilize annotation information to group CpGs into biological meaningful sets such as CGI, CGI shores, and TSS200, then test the association between each set and the trait of interest (Zhang et al. 2016). Contrary to unsupervised methods that depend on annotation, supervised methods involve bump hunting (Jaffe et al. 2012), Comb-p (Pedersen et al. 2012), DMRcate (Peters et al. 2015), Probe Lasso (Butcher and Beck 2015), and Dmrff (Suderman et al. 2018). These methods take computed p-values from site-based analysis as input and scan for regions with consecutive small p-values based on customized criteria.

EWAS analysis has been widely applied in complex diseases such as cancer, diabetes, asthma, and so on, while studies in the field of pediatric sepsis remain limited. First, although EWAS studies conducted in adult sepsis by Binnie and colleagues recently found hundreds of DMR distinguishing septic patients and non-septic controls (Binnie et al. 2020), pediatric sepsis methylation studies are restricted by limited cohort size. The two published pediatric studies until now included 3 and 17 pediatric sepsis cases, which restricted the power of DMP and DMR detection (D. Benet Bosco Dhas 2015, Lorente-Pozo et al. 2021). The second gap to be filled involves revealing methylation signals related to pediatric sepsis heterogeneity. Despite the attempt of prior studies to identify DMR distinguishing pediatric subgroups such as early and late-onset pediatric sepsis (Lorente-Pozo et al. 2021), only a few significant regions differentially methylated between the two groups have been found due to insufficient power. To fill these

knowledge gaps, the Aim 3 project relying on epigenome-wide association analysis with various analysis strategies can facilitate understanding how epigenomic components affect pathology in pediatric sepsis. The understanding gained from this aim enables uncovering the impact of methylation on pediatric sepsis heterogeneity and provides more possibility to further ascertain potential therapeutic targets.

2.7 Public health relevance

Even though the WHO declared sepsis a global health priority in 2017, the disorder remains a major contributor of health loss in adult and pediatric populations. The failure of previous therapeutic attempts stems from insufficient knowledge of sepsis pathology. With the development of data-driven methods and next-generation sequencing techniques, comprehensively studying disease heterogeneity, genetic basis, and transcriptomic dynamics is of great significance in a real-world application. First, analysis using early-hour bedside clinical data can establish an immediate risk stratification system, allowing timely intervention for critically ill patients. Second, association studies with whole-exome sequencing data can broaden the spectrum of genetic components ever investigated and pave the way for integrating genetic information into clinical management. Third, the utilization of methylation microarray data provides us with opportunities to capture the diversity at the epigenetic level and determine the environmental factors contributing to disease heterogeneity. Therefore, we expect to formulate this study as a roadmap to elucidate sepsis pathology, improve disease intervention, guide clinical trial design, and offer hope of increasing favorable sepsis outcomes.

3.0 Cohort and Data

3.1 Patient enrollment

The demographic data, clinical records, and blood samples of all three aims were obtained from pediatric sepsis patients of a multicenter cohort PHENOMS (PHENOtyping pediatric sepsis-induced Multiple organ failure Study) (Carcillo et al. 2019). The enrollment procedure is illustrated in Figure 3.1.

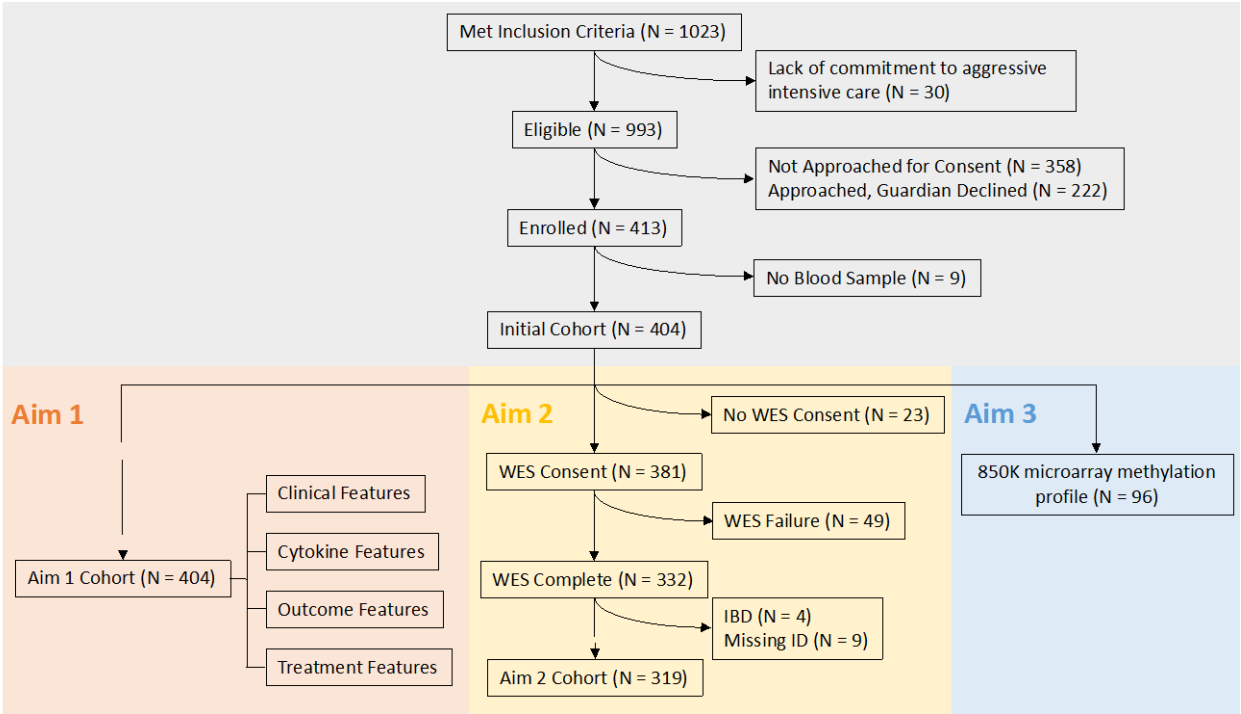

Figure 3.1 Patient enrollment diagram
IBD: identity by descent; Missing ID: fail to match WES sample ID and patient ID.

Specifically, the cohort enrolled pediatric patients from 2015 to 2017 with written informed consent from at least one of the guardians. Children were qualified for enrollment if they met all of the four criteria: (1) at the ages between 44 weeks to 18 years old; (2) were suspected of infection meeting two or more SIRS (systemic inflammatory response) criteria, see Table 3.1 (Goldstein et al. 2005); (3) presented one or more organ failures, see Table 3.2 (Villeneuve et al. 2016); and (4) had an indwelling arterial or central venous catheter. Patients without a commitment to aggressive PICU care or lack of blood samples were further excluded from the enrollment. With the above enrollment criteria, the initial cohort contained 404 individuals.

Table 3.1 SIRS criteria

	Criteria
1	Core temperature of 38.5 °C or < 36 °C
2	Tachycardia, defined as a mean heart rate > 2 SD above normal for age in the absence of external stimulus, chronic drugs, or painful stimuli; or otherwise unexplained persistent elevation over a 0.5- or 4-hr time period OR for children < 1 yr old: bradycardia, defined as a mean heart rate < 10^{th} percentile for age in the absence of external vagal stimulus, β-blocker drugs, or congenital heart disease; or otherwise unexplained persistent depression over a 0.5-hr time period
3	Mean respiratory rate > 2 SD above normal for age or mechanical ventilation for an acute process not related to underlying neuromuscular disease or the receipt of general anesthesia
4	Leukocyte count elevated or depressed for age (not secondary to chemotherapy-induced leukopenia) or > 10% immature neutrophils

Table 3.2 Criteria of organ failure

Organ failure	Criteria
Cardiovascular	need for cardiovascular agent infusion support
Hepatic	total bilirubin > 1.0 mg/dL and alanine aminotransferase (ALT) > 100 units/L
Hematologic	thrombocytopenia < 100,000/mm3 and prothrombin time INR > 1.5 × normal
Respiratory	need for mechanical ventilation support with the ratio of the arterial partial pressure of oxygen and the fraction of inspired oxygen (PaO2/FiO2) < 300 without this support
Neurological	Glasgow Coma Scale (GCS) Score < 12 in the absence of sedatives
Renal	serum creatinine > 1.0 mg/dL and oliguria (urine output < 0.5 mL/kg/h)

3.2 Data collection

After enrollment, multifaceted data were collected from each patient according to the different goals of the three aims. In the Aim 1 project, six demographic features, 46 bedside clinical variables, 33 cytokines measured from blood samples, 41 anti-inflammatory/immunomodulatory and three organ support therapy records, and infection and in-hospital mortality information were obtained from all 404 individuals of the initial cohort. Among them, clinical data and blood samples measuring C-reactive protein, Ferritin, sFASL, ADAMTS 13 activity, and whole blood ex vivo TNF response to endotoxin were obtained on day one and twice weekly until 28 days of PICU. Organ failure and in-hospital mortality were obtained daily until 28 days of PICU.

In the Aim 2 project, 381 out of 404 parents provided WES consent, and 2 mL of whole blood was collected for DNA extraction using standard methods. WES (Whole-exome sequencing) was completed on 332 patients from 2018 to 2020. The University of Pittsburgh Genomics Research Core performed WES on the Ion Torrent platform. Libraries were constructed by the Ampliseq Exome RDY (Thermo Fisher Scientific) with 100 × target coverage. FASTQ files were aligned to Homo sapiens reference sequence GRCh37/hg19 to generate VCF files.

In the Aim 3 project, the 850K DNA methylation microarray profiles (Illumina Infinium Methylation EPIC array) were completed from PBMC of 96 patients of the original cohort, based on their disease severity (i.e. whether developed immunoparalysis associated MOF, thrombocytopenia associated MOF, sequential liver failure associated MOF, or macrophage activation syndrome). In Aims 2 and 3, samples were chosen before Aim 1 and no parental DNA samples were collected.

4.0 Identify Pediatric Sepsis Subtypes through Unsupervised Clustering on Day One Bedside Features

4.1 Forward

Most parts of the writings, figures, and tables of this chapter are based on a previously published manuscript (Qin et al. 2022):

Qin Y, Kernan KF, Fan Z, Park HJ, Kim S, Canna SW, Kellum JA, Berg RA, Wessel D, Pollack MM, Meert K, Hall M, Newth C, Lin JC, Doctor A, Shanley T, Cornell T, Harrison RE, Zuppa AF, Banks R, Reeder RW, Holubkov R, Notterman DA, Michael Dean J, Carcillo JA. Machine learning derivation of four computable 24-h pediatric sepsis phenotypes to facilitate enrollment in early personalized anti-inflammatory clinical trials. Crit Care. 2022 May 7;26(1):128. doi: 10.1186/s13054-022-03977-3. PMID: 35526000; PMCID: PMC9077858.

4.2 Introduction

Sepsis defined by infection and organ failure contributes to 1 of 5 deaths globally, with 3 million per year occurring in children (Rudd et al. 2020). While there is evidence that sepsis mortality increases if treatment is delayed (Weiss et al. 2015), several studies in high-income countries where rapid access to intensive care support has been provided, have demonstrated patterns of mortality even in previously healthy children with timely treatment (Workman et al. 2016, Ames et al. 2018, Evans et al. 2018). This indicates that dysregulated host immune activation

could be targetable in the pediatric intensive care unit (PICU). Among such conditions are immune depression leading to immunoparalysis-associated MOF (IPMOF), thrombotic microangiopathy leading to thrombocytopenia-associated MOF (TAMOF), and hyperinflammatory macrophage activation syndrome (MAS) driven either by uncontrolled lymphoproliferation manifest as sequential liver failure-associated MOF (SMOF) or by macrophage activation without lymphoproliferation manifest as combined hepatobiliary dysfunction and disseminated intravascular coagulation (Doughty et al. 2002, Nguyen et al. 2008, Hall et al. 2011, Shakoory et al. 2016, Wong et al. 2016, Carcillo et al. 2017, Kyriazopoulou et al. 2017, Muszynski et al. 2018). The PHENOtyping pediatric sepsis-induced Multiple organ failure Study (PHENOMS) previously reported that these conditions developed at a median of day 3 to 7 of sepsis, with TAMOF and MAS demonstrating 46% mortality, and IPMOF 16% mortality (Carcillo et al. 2019). Anti-inflammatory therapies used to reverse TAMOF and MAS include methylprednisolone, intravenous immunoglobulin (IVIG), and plasma exchange (Emmenegger et al. 2001, Demirkol et al. 2012, Sevketoglu et al. 2014, Fortenberry et al. 2019). Therefore, the current clinical trial challenge is to identify these at-risk children for early enrollment when personalized therapies have the greatest likelihood to succeed.

The NIGMS sepsis research working group recommendations call for the use of new clinical research approaches in extant clinical data sets to characterize septic patients and improve the efficiency of early trials of new sepsis treatments. In this study, we test the hypothesis that machine learning methods previously used in adults could be applied to available bedside clinical variables including C-reactive protein and ferritin in the extant PHENOMS dataset to derive 24-h computable sepsis phenotypes that identify children at risk for the development of TAMOF and

MAS for enrollment in early personalized anti-thrombotic and anti-inflammatory clinical trials (Horvat et al. 2019, Seymour et al. 2019, Taylor et al. 2020).

4.3 Methods

The overall study workflow is illustrated in Fig 4.1, where each of the three colored sections corresponds to one or more major analysis steps described in the following subsections. The first section in the plot covers the methods from subsection 4.3.2 to 4.3.4. The second section covers the methods in subsections 4.3.5 and 4.3.6. The third section covers the methods in subsection 4.3.7. Other subsections (4.3.1, 4.3.8, and 4.3.9) described below are not presented in the plot.

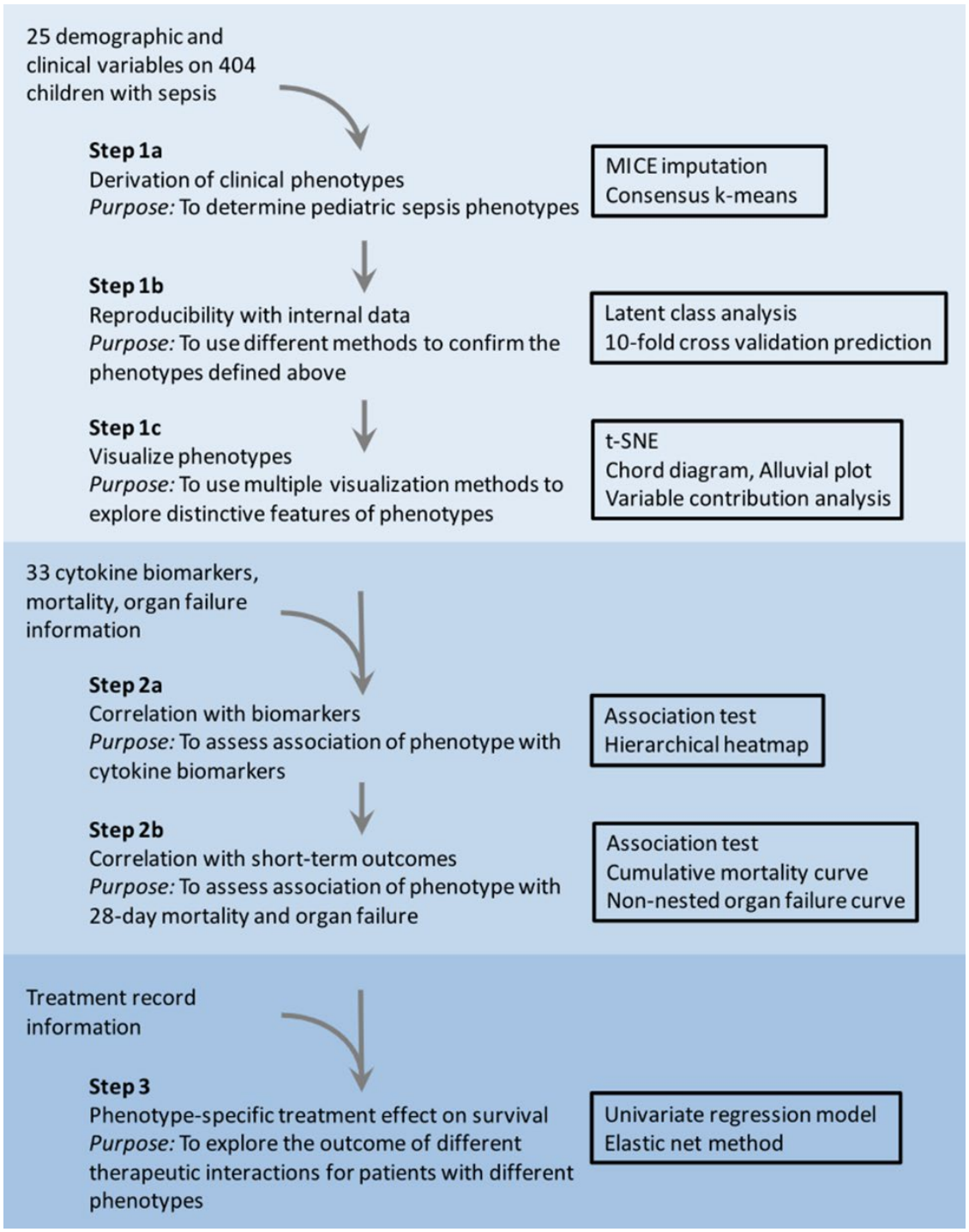

Figure 4.1 Overview of the method
Qin et al. (Qin et al. 2022)

4.3.1 Variable selection and missing data imputation

To prepare features for clustering, six demographic and 46 clinical day-one variables were selected as candidate variables from the initial study. Within this set of features, we further selected variables based on correlation and missingness. First, one of the highly correlated variable pairs (Pearson correlation > 0.6) was removed based on the missingness (i.e., remove the one with a higher missingness) (Akoglu 2018). Among the remaining variables, variables with missingness < 20% were selected and imputed to derive completed datasets. The missing pattern was investigated by performing Little's MCAR test, which showed no significant pattern of missing completely at random (MCAR) in variables with high missingness (> 2 missing values across all samples) (Little 1988). Then, we performed missing data imputation by multiple imputations with chained equations (MICE) (Van Buuren S 2011), assuming missing data is conditional on observed data and follows the pattern of "missing at random". MICE is based on the Fully Conditional Specification (FCS) approach, where each incomplete variable is imputed by a separate model. Using FCS, MICE can impute mixes of continuous, binary, unordered categorical, and ordered categorical data.

4.3.2 Consensus k-means clustering

After determining the final set for clustering, we first took log transformation on highly skewed variables and scaled all variables to prepare data for clustering. For categorical variables, we used one-hot representation to encode them to be used in the K-means clustering. Then, we used Consensus k-means clustering to identify the optimal number of phenotypes (clusters) and

derive the phenotype membership of patients based on the different number of clusters (k) (Wilkerson and Hayes 2010).

The consensus clustering framework increases the robustness of small-sized data clustering by providing a solution to represent the most common assignment across multiple runs of a clustering algorithm. It takes advantage of subsampling techniques so that perturbations of the original data can be simulated. In each subsampling run, a clustering algorithm, i.e. k-means, was applied to the perturbed data sets for a given k. Therefore, the employment of the consensus frame can provide qualitative and quantitative measurements for internal validation purposes, which contributes to the stability of the discovered clusters.

The consensus k-means method provides multiple visualization approaches to assist in determining the optimal number of clusters and evaluating clustering performance. For instance, the consensus matrix heatmap is a plot having patients as both rows and columns. The consensus value is the frequency the two patients are assigned to the same phenotype among 1000 iterations. It ranges from 0 (white, interpreted as two patients are never clustered together) to 1 (dark blue interpreted as two patients are always clustered together). A clear separation of white and dark blue blocks in the heatmap is an indicator of good partitioning. The Consensus CDF plot shows the cumulative distribution functions of the consensus matrix for each k, estimated by a histogram of 100 bins. It is used to determine at what number of clusters the CDF reaches an approximate maximum; thus, consensus and cluster confidence is at a maximum at this k. It is usually used together with the Delta area plot to determine the optimal k. Usually, an “elbow” in the Delta area plot is an indicator of the optimal k. The Cluster-consensus plot shows the cluster-consensus value of clusters at each k. This is the mean of all pairwise consensus values between a

cluster's members. High values indicate a cluster has high stability and low values indicate a cluster has low stability. We used 0.5 as a cut-off for diagnostic purposes.

4.3.3 Validation of clustering

We used multiple approaches to validate the consensus k-means clustering. First, we used KmeansInference method to test if the means are the same across clusters, given different numbers of k (Chen and Witten 2023). The kmeansInference method can take the fact that the clusters are generated based on the same data used for testing into account and thereby estimate the adjusted p-values to avoid bias from the k-means clustering. Next, we used two other methods, latent class analysis (LCA) and X-means, to perform clustering and compare with the results from the consensus k-means clustering (Pelleg 2000, Hagenaars J.A. 2002). LCA is a model-based unsupervised clustering method, which determines the best model (i.e., the optimal number of clusters) by the Bayesian information criterion (BIC) (Hagenaars J.A. 2002). X-means is a variation of k-means clustering, while it uses BIC as a criterion to determine the best number of clusters (Pelleg 2000). The usage of BIC in these two methods enables providing an alternative way of elbow plot.

4.3.4 Dissimilarity visualization of phenotypes

To assess dissimilarity among the derived consensus k means phenotypes we used 1) a t-distributed stochastic neighbor embedding (t-SNE) plot labeled by outcomes of interest; 2) chord diagrams in terms of a priori clinical characteristics and organ dysfunction patterns; and 3) variable contributions to pairwise phenotype discrimination.

4.3.5 Heterogeneity of biomarkers across phenotypes

Following the consensus k means phenotype determination, we correlated the derived phenotypes with 33 biomarkers, including 31 cytokines and two functional assays (ex vivo TNF response to endotoxin as a marker of immune depression and ADAMTS 13 activity as a marker of microvascular thrombosis in the presence of thrombocytopenia). All cytokines were measured on a BioPlex 200 System (Bio-Rad) as previously described. To determine the correlation of the derived 24-hour phenotypes with biomarkers of the host response, we compared mean and standard deviation, median and interquartile range (IQR) for continuous data, and the ratio of the cases in binary data. A cytokine heatmap was used to present the log ratio of the median biomarker values for various markers of the host response. Red represents a greater median biomarker value for that phenotype compared with the median for the entire study cohort, whereas blue represents a lower median biomarker value compared with the median for the entire study cohort. Hierarchical clustering was used to identify similarities in cytokine patterns across the phenotypes.

4.3.6 Heterogeneity of outcomes across phenotypes

Phenotype relationships to primary and secondary outcomes were investigated. The primary outcome was in-hospital mortality. The secondary outcomes involved the development of new or progressive multiple organ failure, length of stay in the PICU, development of immunoparalysis, thrombocytopenia-associated MOF, sequential liver failure-associated MOF, macrophage activation syndrome, use of mechanical ventilation, and extracorporeal therapies. To investigate the relationship of the derived phenotypes to outcomes (mortality, MOF groups), we estimated the association between derived phenotype and outcomes with a multivariate model

adjusting for demographic variables (age, sex, race, ethnicity) and PRISM score. We also generated phenotype-specific curve plots to assess differences in mortality and the number of organ failures over time.

4.3.7 Exploratory analysis of treatment heterogeneity with phenotypes

Within each of the derived phenotypes, we first evaluated which of the 41 anti-inflammatory/immunomodulatory and three organ support therapies given in the parent study by bedside clinicians were associated with survival in univariate logistic regression among the subset of patients who were given anti-inflammatory therapies by the bedside clinicians. For the significant individual treatments, we further applied Elastic Net regression to investigate their interactive effects on survival within each of the derived phenotypes. Finally, we performed traditional multivariable logistic regression to validate findings from the Elastic Net regression model while adjusting for age, sex, ethnicity, race, and total PRISM score.

4.3.8 Computable prediction of individual membership in phenotypes

We also developed a computable tool (https://pedsepsis.pitt.edu) that allows one to categorize the phenotype of a new individual patient into one of the four derived phenotypes at the bedside. Specifically, we first used the same pipeline to standardize and normalize the 25 input variables of the new patient as described above. Then we calculated the Euclidean distance from this patient to the centroid of each phenotype derived from the unsupervised consensus k-means clustering. Comparing the distances to four phenotype centroids, we assign the patient to the phenotype with the shortest distance.

4.3.9 Other information

For summary analysis, we presented continuous data as mean (SD) or median (IQR) and categorical data as count number (%). For comparison, we used the Kruskal-Wallis tests for continuous data and the chi-square test for categorical data. Fisher exact tests were applied for cells containing less than 5 samples. The threshold for statistical significance was less than 0.05 for two-sided tests after adjustment for multiple testing. Holm–Bonferroni correction was applied to correct for multiple testing. Analyses were performed with R version 3.6.2.

4.4 Results

4.4.1 Data preparation

Out of the 52 bedside variables collected within 24 hours in the PICU, 25 variables had less than 20% missingness and less than 60% correlation with any other variable (Appendix Table 1). These included demographic variables (age, gender, ethnicity, previous health status, post-op status), PRISM-related vital signs and laboratory values (systolic blood pressure, heart rate, Glasgow Coma Scale Score, hemoglobin, creatinine, platelet count, intubation status), markers of inflammation (highest temperature, lowest temperature, number of SIRS criteria, lymphocyte count, C-reactive protein level, ferritin level), and organ failures (Organ failure Index, Central Nervous System = Glasgow Coma Scale < 12 not explained by use of sedation; Cardiovascular = Requirement for vasoactive agents for Systolic Blood Pressure < 5th percentile for age; Respiratory = PaO2/FiO2 ratio < 300 requiring mechanical ventilation; Renal = oliguria

and serum creatinine > 1 mg/dL; Hepatic = ALT > 100 and Bilirubin > 1 mg / dL; Hematologic = Platelet Count < 100 K and INR > 1.5).

4.4.2 Derivation of clinical sepsis phenotypes

The derived consensus *k*-means clustering models found a 4-cluster model was the optimal fit, with phenotypes we named PedSep-A, B, C, and D (Fig 4.2). The consensus matrices heat map (Fig 4.2A) shows a relatively good partition of patients when k=4, where a clear separation between blue and white chunks is observed. The relative change under the CDF (cumulative distribution function) curve in Fig 4.2C implied little statistical gain by increasing to a 5- or 6-class model, with the penalty of overfitting.

We applied the KmeansInference method to infer the differences between clusters when k ranges from 2 to 6 by obtaining valid p-values adjusting for bias from k-means. (Table 4.1). As a result, there are significant differences between clusters when k equals 4, indicating that the 4-cluster model enables to distinguish the patients, although a significant difference was also observed when k equals 2. By applying LCA to the same dataset, lower Bayesian information criteria (BIC), higher Entropy, as well as adequate group size (> 10% of the cohort) confirmed that four underlying phenotypes are optimal (Table 4.2). With the side-by-side rank of variable contributions based on both methods (Fig 4.3), we observed similar variable contribution patterns between LCA and consensus k-means clustering, implying the identical phenotype characteristics derived from the two methods. Moreover, we observed consistency of phenotype membership between the two cluster methods. While a subset of patients is reassigned in the confirmatory LCA method, the majority of assignments remain robust to the method of clustering. X-means clustering found that the optimal number of clusters is two instead of four. To investigate the reason for the

inconsistency between consensus k-means clustering and X-means clustering, we used the *MASS* r package to simulate four-dimensional data for four groups, each adhering to a multivariate normal distribution (W. N. Venables 2002). Specifically, the four simulated groups were modeled with sample sizes corresponding to the PedSep-A, B, C, and D phenotypes. The covariance matrix of the four-dimensional data was pre-defined by a positive-definite symmetric matrix, Sigma. The means for the four variables were set as follows: (0, 0, 0, 0) for group 1, (0, 0, i, i) for group 2, (i, i, 0, 0) for group 3, and (i, i, i, i) for group 4. Then we changed the value of i from 0 to 8 and compared the optimal number of clusters inferred separately by k-means and X-means using the *RWeka* r package (Hornik K 2009). As a result, we found K-means-based algorithm is more sensitive than X-means when the difference across clusters is small (Table 4.3). The two methods agree with each other when difference across clusters is large enough (i.e. larger than 4 in our case). The small differences across four phenotypes explain the shape of the CDF in Figure 4.2 where we did not observe an ideal step-wise curve as expected as indicated by simulations of Monti et al. (Monti 2003). Another method we used to visualize the structure of the data is OPTICS plot (Fig 4.4) which shows the reachability distance of each individual. The reachability distances of individuals from the same phenotype can imply the density of the phenotype. Although we observed several dense samples within PedSep-C phenotype, the majority of PedSep phenotype patients were segmented into two valleys rather than four in the OPTICS plot.

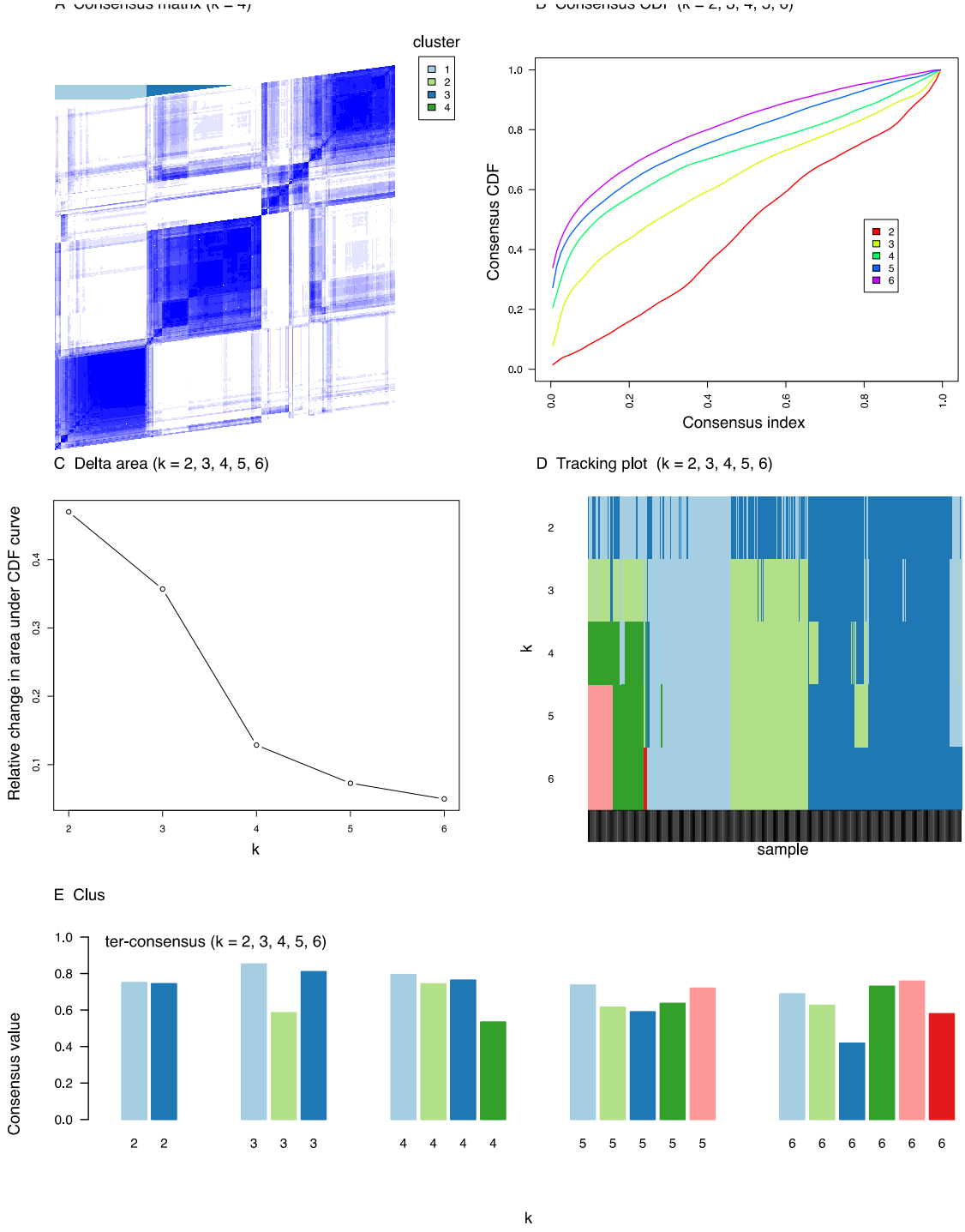

Figure 4.2 Consensus k-means clustering results

(A) The consensus matrices heat map (B) Consensus CDF plot (C) Delta area plot presenting relative change in area under the CDF curve (D) Tracking plot showing the cluster assignment of patients (columns) for each k (rows) by color (E) Cluster-consensus plot. Dark blue: PedSep-A, light green: PedSep-B, light blue: PedSep-C, dark green: PedSep-D. (Qin et al. 2022)

Table 4.1 Selective p-value from kmeans inference when k = 2-6

	K=2	K=3	K=4	K=5	K=6
1 vs 2	3.371E-02	9.122E-01	1.605E-04	9.300E-01	2.086E-03
1 vs 3		9.492E-03	8.689E-09	5.498E-03	1.889E-05
2 vs 3		9.580E-01	1.209E-09	8.567E-01	2.238E-10
1 vs 4			1.549E-02	7.285E-01	4.037E-04
2 vs 4			1.544E-04	9.016E-01	1.152E-03
3 vs 4			6.872E-08	3.493E-01	6.985E-11
1 vs 5				2.946E-01	6.363E-01
2 vs 5				9.483E-01	2.032E-01
3 vs 5				4.893E-01	2.662E-01
4 vs 5				4.579E-01	6.819E-01
1 vs 6					1.068E-02
2 vs 6					3.119E-04
3 vs 6					1.061E-05
4 vs 6					1.585E-02
5 vs 6					6.790E-01

Table 4.2 Statistical ouput from latent class analysis

	Statistic[a]			Class size[b] N, (%)					
Class number	AIC	BIC	Entropy[a]	1	2	3	4	5	6
2	44875	45199	0.917	308(76)	96(24)	-	-	-	-
3	44329	44817	0.872	212(52)	129(32)	63(16)	-	-	-
4	43774	44426	0.904	144 (36)	142(35)	73(18)	45(11)	-	-
5	38542	39359	0.999	146(36)	114(28)	87(22)	34(8)	23(6)	-
6	39540	40520	0.999	98(24)	87(22)	83(21)	74(18)	42(10)	20(5)

Abbreviations: AIC, Akaike information criterion; BIC, Bayesian information criteria; IQR, interquartile range. a: AIC and BIC are information criteria for comparing models, where a lower value suggests a better fit; Entropy is a measure between 0 and 1 measures the success of classification, where a value closer to 1 implies a better fit. b: class size shows the number of samples assigned to each cluster, relatively large size of each cluster is preferred. (Qin et al. 2022)

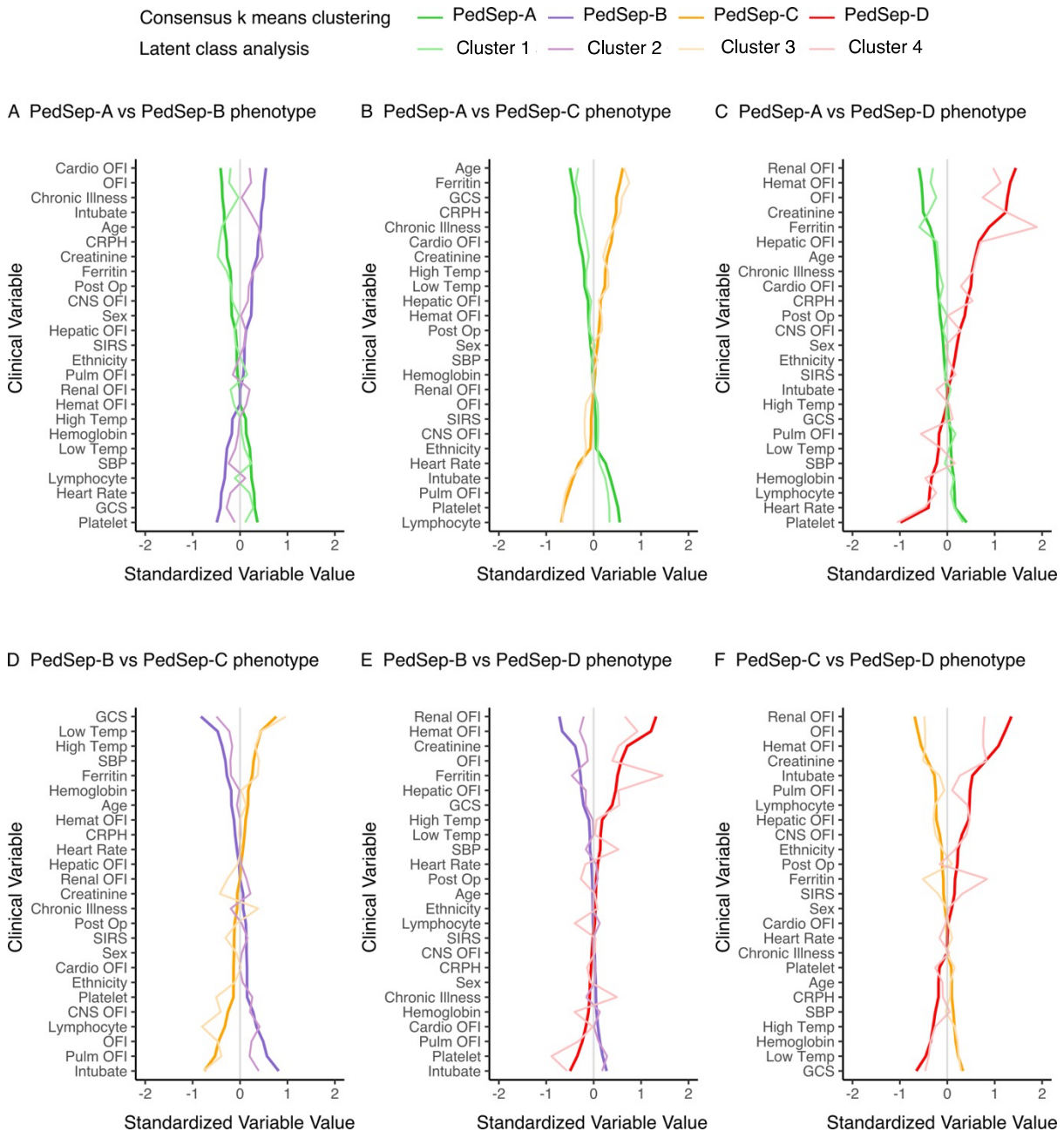

Figure 4.3 Comparison of variables contributing to phenotypes derived from consensus k-means and LCA
In all panels, the variables are standardized such that all means are scaled to 0 and SDs to 1. A value of 1 for the standardized variable value (x-axis) signifies that the mean value for the phenotype was 1 SD higher than the mean value for both phenotypes shown in the graph as a whole. CNS - central nervous system; CRP - C-reactive protein; GCS - Glasgow Coma Scale; Hemat - Hematologic; Intubate- Intubation with endotracheal tube; OFI- organ failure index; Post Op - post-surgery; Pulm- pulmonary; Temp- temperature; SBP- systolic blood pressure; Chronic illness – not previously healthy; Ethnicity – higher number with more non-Hispanic; Sex – higher with more males in group. (Qin et al. 2022)

Table 4.3 Identified number of clusters using two methods based on simulated data

Method	Difference of means (see methods)								
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
k-means	2	3	3	4	4	4	4	4	4
X-means	2	2	2	2	4	4	4	4	4

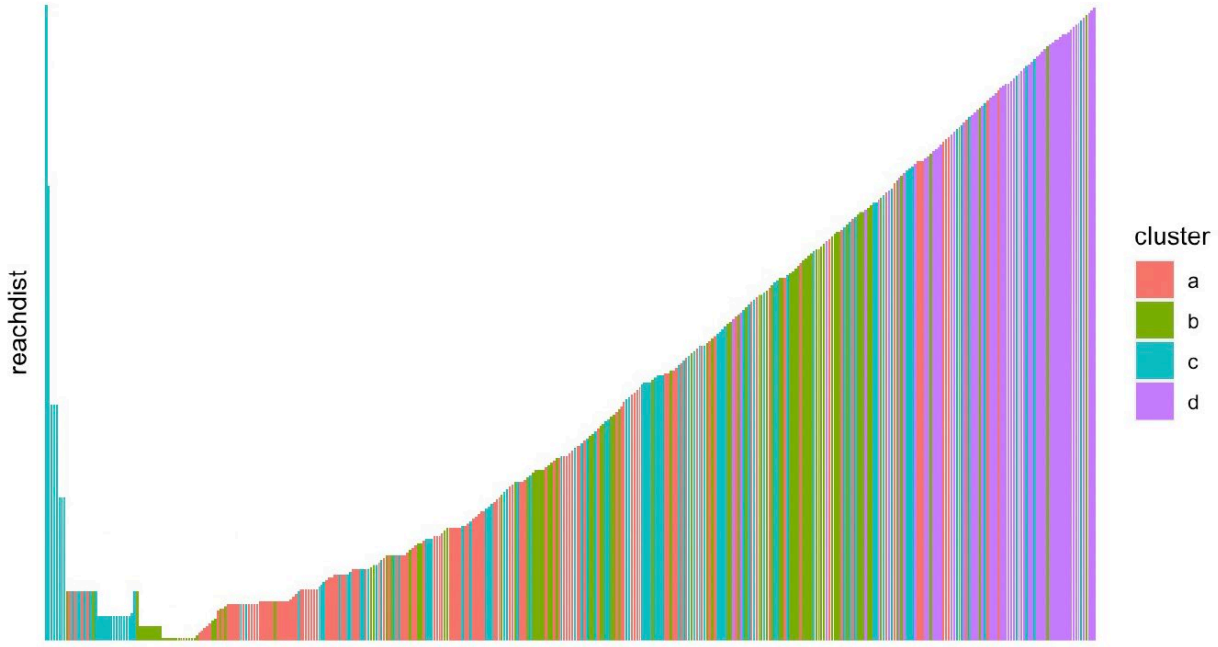

Figure 4.4 OPTICS plot colored by phenotype

The x-axis represents samples linearly ordered based on their spatial distance. The y-axis represents the reachability distance of the sample.

The size and characteristics of the 4-class model are given in Table 4.3 and Figure 4.4. Four derived phenotypes ranged in size from 14% to 34% of the cohort and differed in clinical characteristics and organ dysfunction patterns. With the exception of the SIRS criteria number, all of the other 24 variables differed among the phenotypes. Compared to all other phenotypes, PedSep-A patients were younger and previously healthy, with the lowest CRP and ferritin levels, the highest lymphocyte and platelet counts, highest heart rate, and lowest creatinine; PedSep-B patients were most likely to be intubated and had the lowest Glasgow Coma Scale Score; PedSep-C patients had the highest temperature and Glasgow Coma Scale Score, least pulmonary failure, and lowest lymphocyte count; and PedSep-D patients had the highest creatinine and number of organ failures, including renal, hepatic, and hematologic organ failure, with the lowest platelet count. On average, PedSep-B and D patients had multiple organ failure, whereas PedSep-A and C patients did not. Ferritin levels were highest in PedSep-C and PedSep-D distinguishing them from PedSep-A and B.

Table 4.4 Demographic and day one clinical characteristics of the four phenotypes

Characteristic[1]	Total	PedSep-A	PedSep-B	PedSep-C	PedSep-D
No. of patients, N (%)	404 (100)	136 (34)	102 (25)	110 (27)	56 (14)
Demographic					
Age years* mean (SD)	7 (6)	3 (4)	8 (6)	10 (5)	8 (6)
Male* N (%)	224 (55.4)	63 (46.3)	68 (66.7)	59 (53.6)	34 (60.7)
Female* N (%)	180 (44.6)	73 (53.7)	34 (33.3)	51 (46.4)	22 (39.3)
Hispanic* N (%)	67 (16.6)	28 (20.6)	12 (11.8)	23 (20.9)	4 (7.1)
Non-Hispanic* N (%)	323 (80.0)	100 (73.5)	86 (84.3)	86 (78.2)	51 (91.1)
Previous healthy* N (%)	180 (44.6)	96 (70.6)	28 (27.5)	37 (33.6)	19 (33.9)
Surgery* N (%)	49 (12.1)	6 (4.4)	19 (18.6)	12 (10.9)	12 (21.4)
Organ Dysfunction					
SIRS criteria, mean (SD)[2]	2.9 (0.8)	2.9 (0.8)	3.0 (0.8)	2.8 (0.8)	3 (0.8)
OFI* mean (SD)[3]	1.8 (0.9)	1.4 (0.5)	2.1 (0.6)	1.4 (0.6)	3.1 (1.0)
Inflammation					
CRP mg/dL* mean (SD)	11.7 (10.4)	7.3 (7.3)	13.2 (11.5)	15.2 (10.4)	13.1 (11.2)
Low Temperature ^{o}C* mean	36.6 (1.2)	36.7 (0.9)	36.0 (1.6)	37.1 (0.9)	36.3 (1.0)
High Temperature ^{o}C* mean	37.8 (1.3)	37.8 (1.1)	37.4 (1.3)	38.3 (1.2)	37.8 (1.4)
ALC /mm^3 * median (IQR)	1.2 (0.6-2.1)	1.9(1.3-3.2)	1.1(0.6-1.9)	0.6 (0.2-1.0)	1.1(0.6-2.1)
Ferritin ng/mL* median (IQR)	218 (98.0-625.3)	125(69.8-207.8)	223(116.5-544.2)	405(176.2-1485.7)	610 (221.1-2482.0)
Pulmonary					
Pulmonary OFI* N (%)	270 (66.8)	108 (79.4)	87 (85.3)	37 (33.6)	38 (67.9)
Intubation* N (%)	211 (52.2)	72 (52.9)	94 (92.2)	15 (13.6)	30 (53.6)
Cardiovascular or Hemodynamic					
Heart rate bpm* mean (SD)	155.4 (31.3)	168.1 (30.8)	146.5 (27.9)	150.4 (27.6)	150.6 (35.8)
Systolic blood pressure* mean (SD) mmHg	81.9 (19.3)	85.0 (15.7)	74.8 (22.0)	86.3 (17.2)	78.9 (21.9)
Cardiovascular OFI* N (%)	284 (70.3)	63 (46.3)	92 (90.2)	85 (77.3)	44 (78.6)
Renal					
Creatinine mg/dL* median (IQR)	0.5 (0.3-0.8)	0.3 (0.2-0.4)	0.6 (0.4-1.0)	0.6 (0.4-0.7)	1.4 (0.6-2.6)
Renal OFI* N (%)	30 (7.4)	0 (0.0)	0 (0.0)	0 (0.0)	30 (53.6)

Hepatic					
Hepatic OFI* N (%)	40 (9.9)	3 (2.2)	9 (8.8)	11 (10.0)	17 (30.4)
Hematologic					
Hemoglobin g/dL * mean (SD)	9.8 (2.0)	10.1 (1.8)	9.4 (2.1)	10.2 (2.1)	9.1 (1.8)
Platelets K/mm^3 * mean (SD)	171.1 (123.2)	260.1 (122.0)	154.3 (95.1)	118.8 (83.5)	88.2 (108.0)
Hematologic OFI* N (%)	39 (9.7)	0 (0.0)	0 (0.0)	8 (7.3)	31 (85.7)
Neurologic					
Glasgow Coma Scale score* mean (SD)[4,5]	8.7 (5.3)	8.5 (5.2)	4.7 (3.4)	13.2 (3.1)	7.9(5.5)
CNS OFI N (%)	54 (13.4)	12 (8.8)	24 (23.5)	6 (5.5)	12 (21.4)

(Qin et al. 2022)

Abbreviations: IQR, interquartile range; SIRS, systemic inflammatory response syndrome; OFI, organ failure index; ALC, absolute lymphocyte count; CNS, central nervous system

SI conversion factors: To convert alanine transaminase and aspartate aminotransferase to μkat/L, multiply by 0.0167; bilirubin to μmol/L, multiply by 17.104; C-reactive protein to nmol/L, multiply by 9.524; creatinine to μmol/L, multiply by 88.4.

1 The variables in this Table were log-transformed for modeling.

2 Indicates SIRS criteria ranging from 0 to 4 including abnormal heart rate, respiratory rate, temperature, and white blood cell count.

3 OFI is an integer score reflecting the number of organ failures. Scores are either 0 or 1 for cardiovascular, hepatic, hematologic, respiratory, neurological, and renal, and summed for total range of 0 to 6. Cardiovascular, need for cardiovascular agent infusion support; Pulmonary, need for mechanical ventilation support with the ratio of the arterial partial pressure of oxygen and the fraction of inspired oxygen (PaO2/FiO2) < 300 without this support; Hepatic, total bilirubin > 1.0 mg/dL and alanine aminotransferase (ALT) > 100 units/L; Renal, serum creatinine > 1.0 mg/dL and oliguria (urine output < 0.5 mL/kg/hr); Hematologic, thrombocytopenia < 100,000/mm3 and prothrombin time INR > 1.5 × normal; Central Nervous System, Glasgow Coma Scale (GCS) score < 12 in the absence of sedatives.

4 Corresponds to the minimum or maximum value within 6 hours of hospital presentation.

5 GCS ranges from 3 to 15.

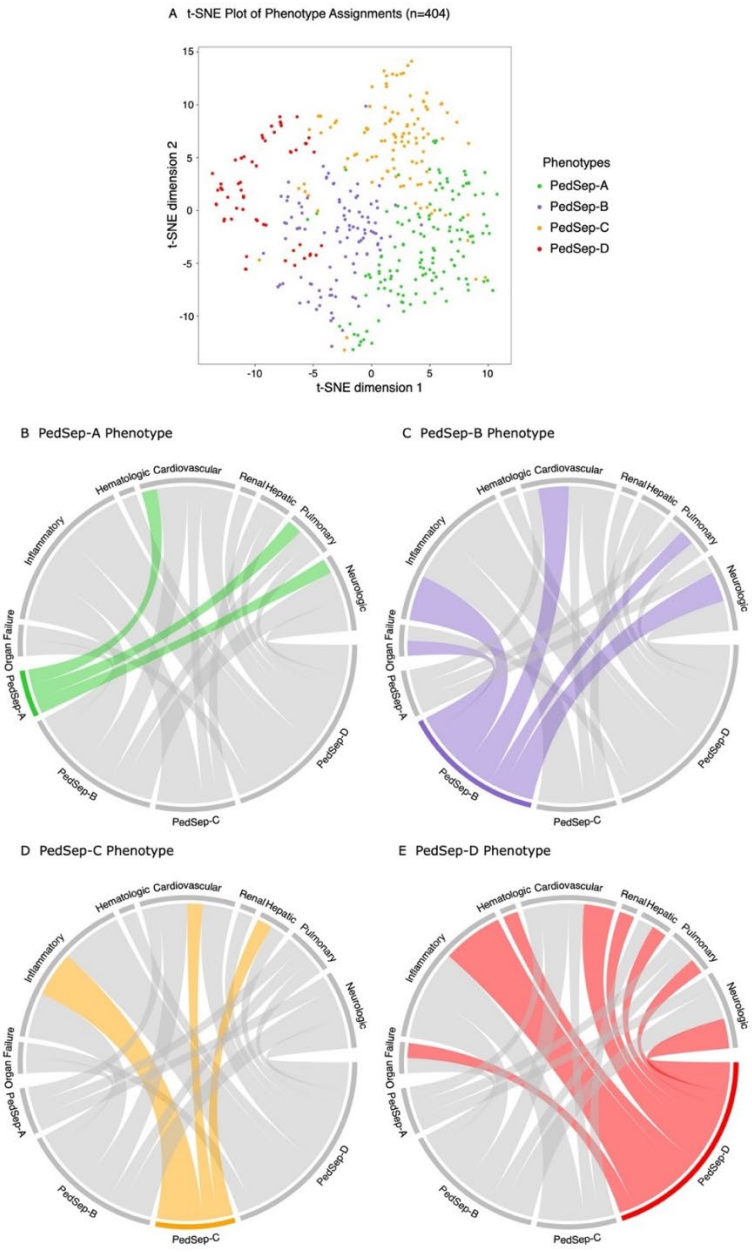

Figure 4.5 Sample distribution and chord plot

In panel **A**, visualization of phenotypes using t-distributed stochastic neighbor embedding (t-SNE) technique with phenotypes shown in color from the consensus *k*-means clustering analysis visualizes distinction among four phenotypes. In panels **B**–**E**, each phenotype is highlighted separately and the ribbons connect to the different patterns of clinical variables and organ system dysfunctions on the top of the circle (inflammation = low temperature, high temperature, max CRP, max ferritin; organ failure = total OFI; pulmonary = pulmonary OFI, intubation; cardiovascular = high heart rate, low systolic blood pressure, cardiovascular OFI; renal = high creatinine, renal OFI; hepatic = hepatic OFI; hematologic = low hemoglobin, low platelets, hematologic OFI; neurologic = Low Glasgow Coma Score Scale, central nervous system OFI). The chords connect from an individual phenotype to a category if the group mean involvement of the variables differs from the overall mean for the entire cohort specifically lower for low temperature, systolic blood pressure, hemoglobin, platelets, and Glasgow Coma Scale Score, but higher for all other variables. (Qin et al. 2022)

4.4.3 Correlation of phenotype with biomarker profiles

The inflammatory biomarker profiles differed across the four derived computable phenotypes. Inflammation evidenced by cytokine signature increased, and immune response (whole blood ex vivo TNF response to endotoxin) and coagulation function (ADAMTS13 activity) decreased going across PedSep-A, B, C, and D (Table 4.5, Table 4.6, Fig 4.5). PedSep-A showed the least inflammation with the lowest M-CSF, IL-8, IL-6, sCD163, MCP1/CCL2, ferritin, C-reactive protein, IL-10, IL-22, IL-18, IL-18BP, and MIP 1α levels overall; lower CXCL9 than PedSep-C and D; lower IL-17a than PedSep-B and C; lower IP10/CXCL10 than PedSep-C; and lower IL2Ra than PedSep-D. PedSep-A had the best immune and coagulation function with normal whole blood ex vivo TNF response to endotoxin (> 200 pg/mL) and ADAMTS 13 activity. In contrast, PedSep-D had the most profound inflammatory response with highest M-CSF, IL-8, SCF, sCD163, IL-16, IL-10, TNF, and MIP1α levels, and thrombotic microangiopathic response with lowest ADAMTS13 activity decreased to < 57% of control with thrombocytopenia. Consistent with this increased inflammation response, the macrophage inhibitor TRAIL was reduced in PedSep-D compared to PedSep-C. PedSep-D also had higher CXCL9 than PedSep-B but not PedSep-C.

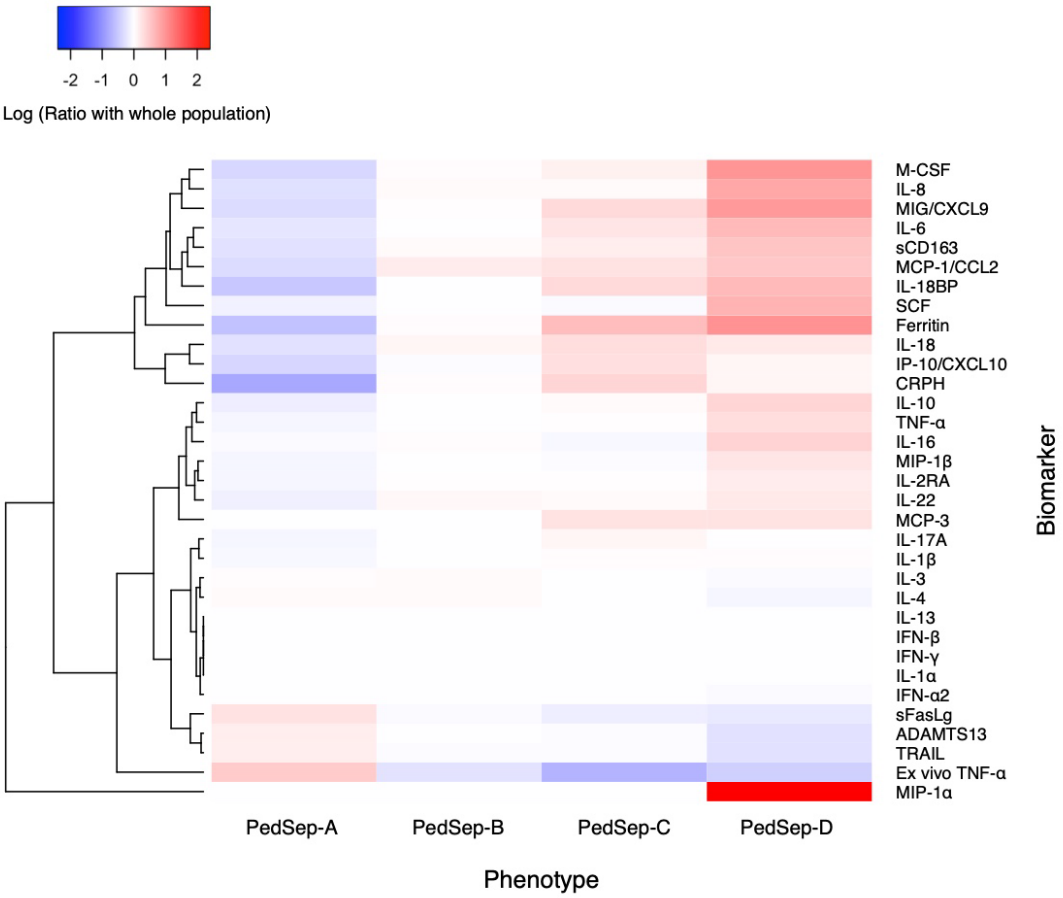

Figure 4.6 Patterns of inflammatory biomarker across phenotypes

The cytokine heatmap shows the log ratio of the median biomarker values for various markers of the host response and their hierarchical cluster relationships. Red represents a greater median biomarker value for that phenotype compared with the median for the entire study cohort, whereas blue represents a lower median biomarker value compared with the median for the entire study cohort. For example, M-CSF is lower in PedSep-A than the entire study cohort and is higher in PedSep-D than the entire study cohort. (Qin et al. 2022)

Table 4.5 Biomarker measured at day one by phenotype

Biomarker[a]	Total	Phenotype			
		PedSep-A (n = 136)	PedSep-B (n = 102)	PedSep-C (n = 110)	PedSep-D (n = 56)
ADAMTS13, %	71.0 (56.0, 88.0)	82.5 (65.0, 95.0)	71.0 (57.0, 89.3)	69.0 (52.5, 84.5)	54.0 (38.0, 66.5)
SFasLg, pg/ml	44.9 (29.0, 73.2)	58.4 (37.2, 84.6)	43.2 (31.1, 78.1)	38.0 (25.0, 65.9)	36.7 (20.9, 49.3)
Ex vivo TNF-α, pg/ml	427.8 (97.0, 1023.3)	689.7 (347.1, 1049.2)	331.1 (99.2, 806.8)	212.3 (35.7, 668.0)	278.0 (53.0, 1049.2)
TNF-α, pg/ml	74.9 (56.2, 105.7)	69.0 (51.9, 85.1)	74.9 (55.4, 101.8)	76.2 (55.4, 108.6)	102.2 (81.4, 131.7)
sCD163, pg/ml	294096 (195700, 496348)	223123 (163123, 323365)	309800 (185699, 508775)	345238 (248348, 572766)	668162 (280407, 897784)
IFN-β, pg/ml	6.4 (6.4, 8.2)	6.4 (6.4, 7.2)	6.4 (6.4, 9.9)	6.4 (6.4, 10.8)	6.4 (6.4, 6.4)
IL-22, pg/ml	26.0 (20.1, 34.2)	22.4 (17.8, 29.5)	28.0 (21.3, 36.9)	27.1 (20.1, 34.2)	31.9 (24.8, 49.2)
IL-18, pg/ml	424.5 (255.2, 732.9)	326.4 (217.2, 480.9)	461.1 (279.4, 792.1)	576.4 (300.2, 1100.5)	518.4 (344.6, 857.1)
IL-18BP, pg/ml	16083.6 (9107.0, 29173.7)	9649.8 (6141.4, 16025.1)	16084 (9513, 26918)	22654 (13694, 34513)	30713 (18114, 40878)
MIG/CXCL9, pg/ml	801.2 (428.4, 2013.6)	576.2 (378.0, 1008.5)	809.8 (432.4, 1963.7)	1125.3 (496.6, 2470.9)	2047.0 (649.8, 4354.5)
IL-1β, pg/ml	2.8 (2.4, 3.3)	2.6 (2.1, 3.2)	2.8 (2.3, 3.2)	2.9 (2.4, 3.3)	2.9 (2.5, 3.3)
IL-4, pg/ml	4.7 (3.5, 6.5)	4.9 (3.5, 6.3)	4.9 (3.9, 6.8)	4.7 (3.5, 6.7)	4.3 (3.5, 6.4)
IL-6, pg/ml	8.8 (6.5, 19.0)	6.9 (5.8, 10.0)	8.7 (6.5, 25.2)	11.1 (7.0, 27.2)	16.8 (8.4, 43.8)
IL-8, pg/ml	55.0 (31.4, 108.6)	41.2 (27.7, 66.7)	57.1 (34.7, 113.7)	57.5 (35.1, 127.9)	123.5 (71.2, 468.6)
IL-10, pg/ml	22.5 (17.5, 33.4)	19.3 (15.4, 24.6)	22.5 (18.6, 37.1)	23.7 (18.1, 37.5)	32.8 (24.6, 73.3)
IL-13, pg/ml	3.1 (3.1, 3.9)	3.1 (3.1, 4.2)	3.1 (3.1, 3.4)	3.1 (3.1, 4.3)	3.1 (3.1, 3.4)
IL-17A, pg/ml	19.1 (16.5, 23.4)	17.4 (15.6, 21.7)	19.1 (16.5, 26.0)	20.9 (16.7, 26.8)	19.1 (16.5, 23.4)
IFN-γ, pg/ml	2.8 (2.8, 2.8)	2.8 (2.8, 3.0)	2.8 (2.8, 3.0)	2.8 (2.8, 3.0)	2.8 (2.8, 3.2)
IP-10/CXCL10, pg/ml	727.8 (343.9, 1963.7)	494.0 (287.3, 1725.7)	705.8 (259.7, 1585.7)	967.7 (409.0, 2334.1)	789.5 (466.8, 2109.7)
MCP-1/CCL2, pg/ml	142.5 (70.9, 367.6)	103.6 (49.7, 197.2)	169.8 (83.9, 383.3)	184.4 (88.7, 474.6)	240.7 (114.5, 1623.0)
MIP-1α, pg/ml	0.6 (0.6, 7.7)	0.6 (0.6, 0.6)	0.6 (0.6, 8.1)	0.6 (0.6, 9.0)	6.6 (2.8, 16.1)
MIP-1β, pg/ml	45.9 (31.4, 70.4)	42.4 (28.1, 56.3)	45.7 (31.9, 73.9)	44.6 (32.9, 77.8)	58.4 (47.9, 95.0)
MCP-3, pg/ml	92.4 (92.4, 166.0)	92.4 (92.4, 147.8)	92.4 (92.4, 166.0)	119.5 (92.4, 166.0)	119.5 (92.4, 180.6)
IFN-α2, pg/ml	125.7 (105.8, 140.2)	124.3 (105.8, 140.2)	125.7 (108.8, 144.4)	125.7 (105.8, 140.2)	120.0 (105.8, 137.9)
IL-1α, pg/ml	9.4 (9.4, 13.2)	9.4 (9.4, 9.9)	9.4 (9.4, 16.4)	9.4 (9.4, 15.6)	9.4 (9.4, 13.2)

IL-2RA, pg/ml	378.8 (243.0, 623.2)	345.2 (237.6, 511.6)	385.3 (206.6, 683.1)	380.6 (243.1, 731.5)	449.7 (307.9, 747.5)
IL-3, pg/ml	612.2 (529.0, 724.4)	624.4 (496.1, 734.6)	636.6 (529.0, 724.4)	612.2 (529.0, 724.4)	586.4 (496.1, 693.0)
IL-16, pg/ml	569.8 (410.2, 763.0)	544.4 (391.2, 677.2)	590.4 (435.2, 770.2)	529.4 (382.7, 704.3)	858.0 (592.6, 1246.4)
M-CSF, pg/ml	30.0 (17.0, 55.3)	20.7 (14.2, 33.6)	30.6 (20.3, 54.6)	34.4 (20.3, 58.9)	79.9 (46.2, 122.7)
SCF, pg/ml	160.2 (118.8, 244.4)	141.6 (115.0, 199.9)	158.2 (113.8, 226.8)	151.5 (114.2, 238.8)	326.1 (227.5, 504.1)
TRAIL, pg/ml	36.6 (27.9, 54.2)	42.9 (32.9, 65.5)	35.4 (25.4, 54.5)	35.4 (29.1, 45.4)	27.9 (24.1, 40.4)
CRP, mg/dL	9.8 (3.3, 17.1)	4.3 (1.2, 12.4)	10.1 (4.8, 19.3)	14.3 (7.5, 21.7)	10.7 (3.4, 20.7)
Ferritin, ng/mL	218.0 (98.0, 625.3)	125.4 (69.8, 207.8)	223.1 (116.5, 544.2)	405.5 (176.2, 1485.7)	610.0 (221.1, 2482.0)

[a] All biomarkers are measured one time concomitantly in the first day. Values in the table are summarized as median (IQR). (Qin et al. 2022)

Table 4.6 Statistical test p-values of differences of biomarkers among phenotypes

Biomarker[a]	General	Pairwise					
		PedSep-A vs PedSep-B	PedSep-A vs PedSep-C	PedSep-A vs PedSep-D	PedSep-B vs PedSep-C	PedSep-B vs PedSep-D	PedSep-C vs PedSep-D
ADAMTS13	<0.001	0.033	0.001	<0.001	1.000	<0.001	<0.001
sFasLg	<0.001	0.140	<0.001	<0.001	0.419	0.111	1.000
Ex vivo TNF-α	<0.001	0.002	<0.001	0.061	0.506	1.000	1.000
TNF-α	<0.001	0.321	0.152	<0.001	1.000	<0.001	0.002
sCD163	<0.001	0.008	<0.001	<0.001	0.680	0.0015	0.021
IFN-β	0.068	0.400	0.320	1.000	1.000	0.400	0.430
IL-22	<0.001	0.005	0.004	<0.001	1.000	0.270	0.139
IL-18	<0.001	0.005	<0.001	<0.001	0.695	1.000	1.000
IL-18BP	<0.001	<0.001	<0.001	<0.001	0.016	<0.001	0.227
MIG/CXCL9	<0.001	0.053	<0.001	<0.001	1.000	0.025	0.207
IL-1β	0.071	1.000	0.130	0.250	1.000	1.000	1.000
IL-4	0.590	1.000	1.000	1.000	1.000	1.000	1.000
IL-6	<0.001	0.020	<0.001	<0.001	0.759	0.023	0.258
IL-8	<0.001	0.008	0.002	<0.001	1.000	<0.001	<0.001
IL-10	<0.001	0.003	<0.001	<0.001	1.000	0.003	0.012
IL-13	0.824	1.000	1.000	1.000	1.000	1.000	1.000
IL-17A	<0.001	0.010	<0.001	0.862	1.000	1.000	0.577
IFN-γ	0.998	1.000	1.000	1.000	1.000	1.000	1.000
IP-10/CXCL10	0.013	1.000	0.024	0.133	0.267	0.817	1.000
MCP-1/CCL2	<0.001	0.003	<0.001	<0.001	1.000	0.079	0.371
MIP-1α	<0.001	<0.001	<0.001	<0.001	1.000	0.006	0.007
MIP-1β	<0.001	0.168	0.093	<0.001	1.000	0.037	0.084
MCP-3	0.309	1.000	0.930	0.660	1.000	1.000	1.000
IFN-α2	0.803	1.000	1.000	1.000	1.000	1.000	1.000
IL-1α	0.500	1.000	1.000	1.000	1.000	1.000	1.000
IL-2RA	0.021	1.000	0.462	0.007	1.000	0.565	1.000
IL-3	0.596	1.000	1.000	1.000	1.000	1.000	1.000
IL-16	<0.001	0.318	1.000	<0.001	0.488	<0.001	<0.001
M-CSF	<0.001	<0.001	<0.001	<0.001	1.000	<0.001	<0.001
SCF	<0.001	0.760	0.850	<0.001	1.000	<0.001	<0.001
TRAIL	<0.001	0.032	0.003	<0.001	1.000	0.104	0.047
CRP	<0.001	<0.001	<0.001	0.002	0.137	0.952	0.265
Ferritin	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.187

Qin et al. (Qin et al. 2022)

4.4.4 Relationship with infection, organ support, and hospital mortality

PedSep-A had more viral infection, PedSep-B had more pneumonia, and PedSep-C and D had more blood infections (Table 4.7, Table 4.8). Patients in PedSep-C had the least mechanical ventilation and the shortest length of stay. Patients in PedSep-D required more extracorporeal membrane oxygenation than in PedSep-A, and the most continuous renal replacement therapy (CRRT) overall. PedSep-A patients required the least CRRT. PICU free days were highest in PedSep-C and lowest in PedSep-D (Table 4.7, Table 4.8).

Hospital mortality was 2% in PedSep-A, 12% in PedSep-B, 10% in PedSep-C, and 34% in PedSep-D (PedSep-B vs. A Adj OR 4.11 95% CI (1.11–19.96) $p = 0.048$; PedSep-C vs. A Adj OR 4.35 95% CI (1.23–20.43) $p = 0.034$; PedSep-D vs. A Adj OR 17.25 95% CI (4.93–92.06) $p = 4.42E{-}05$; PedSep-D vs B Adj OR 4.20 95% CI (1.84–9.97) $p = 0.0008$; and PedSep-D vs. C Adj OR 3.97 95% CI (1.62–10.14) $p = 0.003$) (Table 4.6, Table 4.7).

The derived mortality curves show all deaths in PedSep-A occurred before seven days, whereas deaths in PedSep-B, C, and D continued to accrue after seven days (Fig 4.6). Mortality was associated with Central nervous system organ failure, decreased TNF and IL-2Ra levels, and increased MCP3 levels in PedSep-A; increased IL-6, IL-8, and MCP1/CCL2 levels in PedSep-B; high ferritin, lymphopenia, lower temperature, higher blood pressure, and increased IL-8 levels in PedSep-C; and hyperferritinemia, chronic illness, increased MIP-1α, IL-8, and IL-10 levels, and decreased IL-18 and sFASL levels in PedSep-D (Fig 4.7, Fig 4.8).

Table 4.7 Subsequent outcome characteristics of the four phenotypes

Characteristic[e]	Total	PedSep-A	PedSep-B	PedSep-C	PedSep-D
No. of patients, N (%)	404 (100)	136 (34)	102 (25)	110 (27)	56 (14)
Development of Subsequent MOF Empirical Phenotypes					
SMOF, N (%)	7 (1.7)	0 (0.0)	0 (0.0)	1 (0.9)	6 (10.7)[a,b,c]
TAMOF, N (%)	37 (9.2)	0 (0.0)	6 (5.9)[a]	3 (2.7)	28 (50.0)[a,b,c]
IPMOF, N (%)	85 (21.0)	12 (8.8)	29 (28.4)[a]	22 (20)	22 (39.3)[a]
MAS, N (%)	24 (5.5)	0 (0.0)	3 (2.9)	2 (1.8)	19 (33.9)[a,b,c]
NPMOF, N (%)	117 (29.0)	28 (20.6)	25 (24.5)	32 (29.1)	32 (57.1)[a,b,c]
Infections					
Bacterial infection, N (%)	141 (34.9)	43 (31.6)	33 (32.4)	45 (40.9)	20 (35.7)
Viral infection, N (%)	114 (28.2)	60 (44.1)[b,c,d]	21 (20.6)	24 (21.8)	9 (16.1)
Fungal infection, N (%)	4 (1.0)	0 (0.0)	1 (1.0)	0 (0.0)	3 (5.4)
Culture negative, N (%)	177 (43.8)	47 (34.6)	52 (51.0)	50 (45.5)	28 (50.0)
Sites of Infections [f]					
Blood, N (%)	51 (12.6)	10 (7.4)	6 (5.9)	22 (20.0)[a,b]	13 (23.2)[a,b]
Lung, N (%)	76 (18.8)	28 (20.6)	29 (28.4) [a,c,d]	12 (10.9)	7 (12.5)
Urine, N (%)	16 (4.0)	4 (2.9)	5 (4.9)	6 (5.5)	1 (1.8)
Organ Support					
MechVent, N (%)	366 (90.6)	134 (98.5)[c]	101 (99.0)[c]	79 (71.8)	52 (92.9)[c]
ECMO, N (%)	30 (7.4)	5 (3.7)	9 (8.8)	6 (5.5)	10 (17.9)[a]
CRRT, N (%)	52 (12.9)	1 (0.7)	7 (6.9)	7 (6.4)	37 (66.1)[a,b,c]
Anti-inflammatory Therapies of Interest					
Decadron, N (%)	94 (23.3)	50 (36.8)[c,d]	22 (21.6)	14 (12.7)	8 (14.3)
Methylprednisolone, N (%)	117 (29.0)	54 (39.7)[b]	23 (22.5)	24 (21.8)	16 (28.6)
IVIG, N (%)	51 (12.6)	6 (4.4)	10 (9.8)	19 (17.3)[a]	16 (28.6)[a]
IVIG + Methylprednisolone	23 (5.7)	3 (2.2)	4 (3.9)	9 (8.2)[a]	7 (12.5)[a]
Plasma exchange, N (%)	25 (6.2)	5 (3.7)	4 (3.9)	4 (3.6)	12 (21.4)[a,b,c]
Plasma exchange + ECMO	6 (1.5)	1 (0.7)	1 (1.0)	1 (0.9)	3 (5.4)
Outcome					
Length of Stay,	9.0 (5.0-17.)	9.0 (5.8-15)[c]	10.5 (5.3-17)[c]	6 (2.3-15)	12.5 (7-26.5)[c]

median (IQR), d					
Mortality, N (%)	45 (11.1)	3 (2.2)	12 (11.7)[a]	11 (10.0)[a]	19 (33.9)[a,b,c]
PICU free days, median (IQR), d	20.0 (8.0-25.0)	21.0 (14.8- 24.0)[d]	19.0 (9.8-24.0) [d]	24.0 (13.3-27)[a,b,d]	4.5 (0.0-21.0)

Qin et al. (Qin et al. 2022)

Abbreviations: SMOF, sequential liver failure associated multiple organ failure; TAMOF, thrombocytopenia associated multiple organ failure; IPMOF, immunoparalysis associated multiple organ failure; MAS, macrophage activation syndrome; NPMOF, new or progressive multiple organ failure; IQR, interquartile range; MechVent, Mechanical Ventilation; ECMO, Extracorporeal Membrane Oxygenation; CRRT, Continuous Renal Replacement Therapies; IVIG, intravenous gamma globulin

[a]The outcome characteristic of this computable phenotype is significantly higher than PedSep-A (p-value < 0.05)

[b]The outcome characteristic of this computable phenotype is significantly higher than PedSep-B (p-value < 0.05)

[c]The outcome characteristic of this computable phenotype is significantly higher than PedSep-C (p-value < 0.05)

[d]The outcome characteristic of this computable phenotype is significantly higher than PedSep-D (p-value < 0.05)

[e] Comparisons across all 4 computable phenotypes were performed using the Kruskal-Wallis test, the $\chi 2$ test, or the Fisher's exact test.

[f] Obtained at the first 3 days

Table 4.8 Statistical test results of differences in subsequent outcome characteristics among the phenotypes

Characteristic[a]	Statistical test p-value						
	General	Pairwise					
		PedSep-A vs PedSep-B	PedSep-A vs PedSep-C	PedSep-A vs PedSep-D	PedSep-B vs PedSep-C	PedSep-B vs PedSep-D	PedSep-C vs PedSep-D
MOF Empirical Phenotypes							
SMOF	<0.001	1.000	1.000	0.003	1.000	0.008	0.027
TAMOF	<0.001	0.017	0.173	<0.001	0.321	<0.001	<0.001
IPMOF	<0.001	0.00083	0.064	<0.001	0.446	0.446	0.064
MAS	<0.001	0.240	0.390	<0.001	0.680	<0.001	<0.001
NPMOF	<0.001	1.000	0.489	<0.001	1.000	<0.001	0.003
Infections							
Bacterial infection	0.440	1.000	1.000	1.000	1.000	1.000	1.000
Viral infection	<0.001	0.002	0.002	0.002	1.000	1.000	1.000
Fungal infection	0.003	1.000	1.000	0.140	1.000	0.510	0.190
Culture negative	0.049	0.096	0.432	0.336	1.000	1.000	1.000
Sites							
Blood	<0.001	1.000	0.023	0.023	0.023	0.019	1.000
Lung	0.006	0.634	0.245	0.634	0.014	0.185	0.963
Urine	0.644	1.000	1.000	1.000	1.000	1.000	1.000
Organ Support							
MechVent	<0.001	1.000	<0.001	0.302	<0.001	0.302	0.013
ECMO	0.006	0.629	0.983	0.015	0.983	0.629	0.112
CRRT	<0.001	0.067	0.067	<0.001	1.000	<0.001	<0.001
Anti-inflammatory therapies							
Dexamethasone	<0.001	0.057	<0.001	0.002	0.561	0.561	0.926
Methylprednisolone	0.009	0.034	0.086	0.203	1.000	1.000	1.000
IVIG	<0.001	0.300	0.002	<0.001	0.231	0.054	0.496
Methylprednisolone + IVIG	0.019	0.460	0.029	0.012	0.154	0.105	0.787
Plasma exchange	<0.001	1.000	1.000	0.002	1.000	0.006	0.002
ECMO + Plasma exchange	0.207	1.000	1.000	0.094	1.000	0.298	0.299
Outcome							
Length of Stay	<0.001	0.365	0.010	0.052	0.003	0.259	<0.001

Mortality	<0.001	0.015	0.023	<0.001	0.826	0.006	0.002
PICU-free Days	<0.001	0.329	0.038	<0.001	0.012	0.003	<0.001

Qin et al. (Qin et al. 2022)

Abbreviations: SMOF, sequential liver failure associated multiple organ failure; TAMOF, thrombocytopenia associated multiple organ failure; IPMOF, immunoparalysis associated multiple organ failure; MAS, macrophage activation syndrome; NPMOF, new or progressive multiple organ failure; IQR, interquartile range; MechVent, Mechanical Ventilation; ECMO, Extracorporeal Membrane Oxygenation; CRRT, Continuous Renal Replacement Therapies; IVIG, intravenous gamma globulin

[a] Comparisons across all 4 phenotypes were performed using the Kruskal-Wallis test for continuous variables, the $\chi 2$ test for categorical variables, or the Fisher's exact test for cells with less than 5 patients

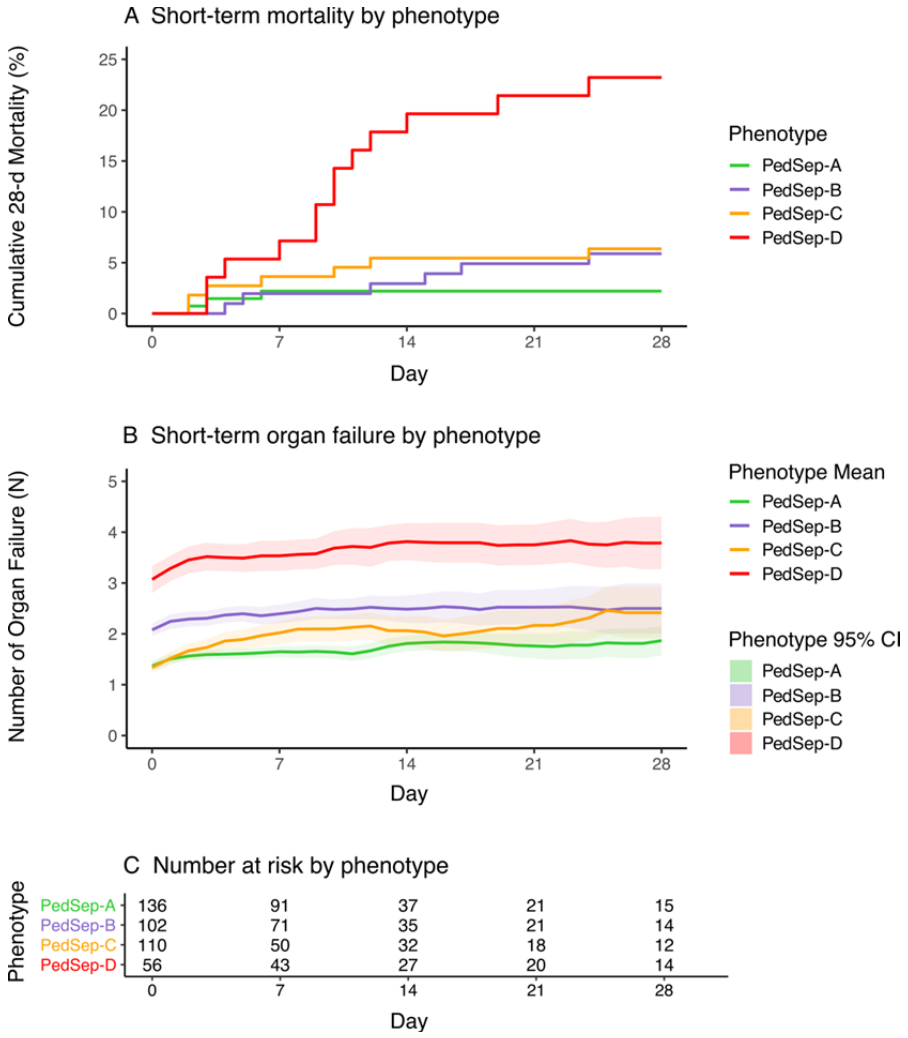

Figure 4.7 Mortality and organ failure curves over 28 days among the four phenotypes

Number of organ failures and mortality according to PedSep-A, B, C, and D phenotype over 28 days. Both short-term mortality (panel **A**) and organ failure (panel **B**) show significant differences by phenotype ($p < 0.001$). The mean numbers of organ failures and 95% confidence intervals (CI) are calculated each day by non-nested observation, where we do not carry forward the OFI at the time the patient leaves the PICU alive or dead. As a reference for patients at risk for Panel **B**, Panel **C** shows the number of children remaining in the PCU at day 0, 7, 14, 21, and 28. (Qin et al. 2022)

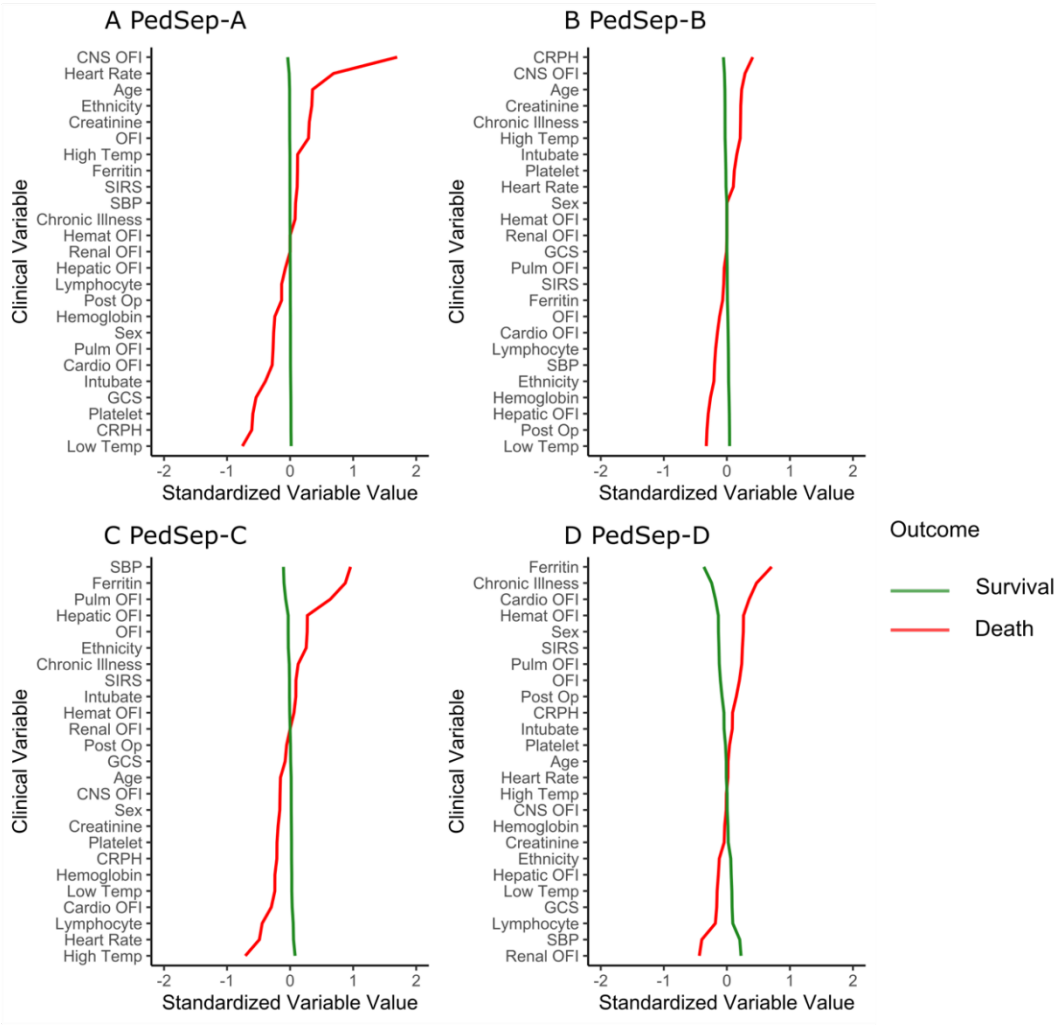

Figure 4.8 Comparison of relationships of 25 variables to mortality in the four phenotypes
In all panels, the variables are standardized such that all means are scaled to 0 and SDs to 1. A value of 1 for the standardized variable value (*x*-axis) signifies that the mean value for the phenotype was 1 SD higher, or lower for − 1, than the mean value for the phenotypes shown in the graph as a whole. *CNS* central nervous system, *CRP* C-reactive protein, *GCS* Glasgow Coma Scale, *Hemat* hematologic, *Intubate* intubation with endotracheal tube, *OFI* organ failure index, *Post-Op* post-surgery, *Pulm* pulmonary, *Temp* temperature, *SBP* systolic blood pressure, *Chronic illness* those who are not recorded as previous healthy, *Ethnicity* value is higher with more non-Hispanics in group, *Sex* value is higher with more males in group. (Qin et al. 2022)

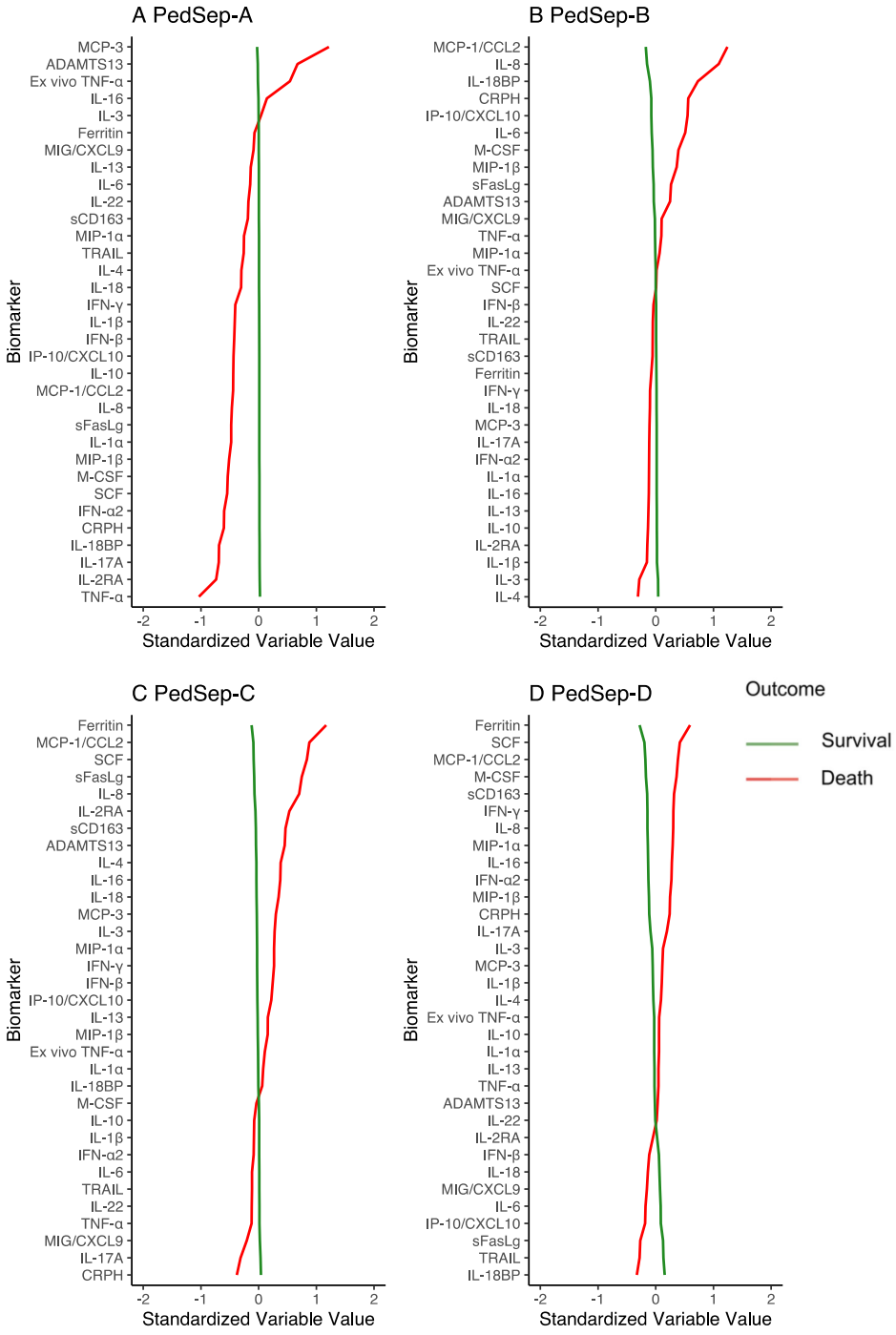

Figure 4.9 Comparison of biomarkers that related to mortality in each phenotype
In all panels, the variables are standardized such that all means are scaled to 0 and SDs to 1. A value of 1 for the standardized variable value (x-axis) signifies that the mean value for the phenotype was 1 SD higher than the mean value for both phenotypes shown in the graph as a whole. (Qin et al. 2022)

4.4.5 Relationship with empirical phenotypes

On average, children in PedSep-A and PedSep-C developed less than two organ failures; children in PedSep-B developed more than two organ failures; and children in PedSep-D developed more than three organ failures over 28 days (Fig 4.6). Children in PedSep-D had the highest proclivity to develop immunoparalysis (Adj OR 2.40 95% CI (1.25–4.53); $p = 7.20E-03$), new and progressive organ failure (Adj OR 4.03 95% CI (2.19–7.55); $p = 9.48E-06$), thrombocytopenia-associated MOF (Adj OR 47.51 95% CI (18.83–136.83); $p = 1.25E-14$), sequential liver failure-associated MOF (Adj OR 61.56 95% CI (8.93–1,282.58); $p = 3.80E-04$), and macrophage activation syndrome (Adj OR 38.63 95% CI (13.26–137.75); $p = 4.61E-10$). Immunoparalysis- and thrombocytopenia-associated MOF also occurred more commonly in children in PedSep-B and D compared to those in PedSep-A (Table 4.6, Table 4.7).

4.4.6 Heterogeneous treatment effect

All 3 organ support therapies and 11 of 41 anti-inflammatory therapies were associated with outcomes in univariable analysis (Table 4.9, Fig 4.9) among the children who received anti-inflammatory therapies and were included in the derived exploratory elastic net regression analysis (Fig 4.9). This analysis was not performed in PedSep-A because mortality was very low at 2%. The constructed elastic net regression heatmaps visualize heterogeneous mortality association patterns across PedSep-B, C, and D (Fig 4.10). Unadjusted mortality odds ratios < 0.1 with use of anti-inflammatory agents were not observed with any single therapy; however, unadjusted interactions were observed with use of methylprednisolone and IVIG.

In adjusted logistic regression modeling, neither methylprednisolone nor IVIG treatment alone, nor the combination, was associated with reduced odds of mortality. The interaction term identified in elastic net regression analysis between methylprednisolone and IVIG therapies in PedSep-D patients remained statistically significant in logistic regression analysis (Methylprednisolone * IVIG interaction = 0.03; 95% CI (0.00058–0.66), $p = 0.04$), indicating that the association of IVIG with mortality was modified by exposure to methylprednisolone in PedSep-D patients. There was also a significant interaction between PedSep-D membership and use of combined methylprednisolone plus IVIG therapy in logistic regression analysis (PedSep-D * Methylprednisolone + IVIG combination interaction = 0.04 95% CI (0.001–0.56), $p = 0.026$) interpreted as meaning that the mortality association with exposure to combined methylprednisolone plus IVIG use is modified by PedSep-D membership.

Table 4.9 Univariable association of 44 therapies with mortality in the subset of patients given anti-inflammatory therapies

Therapy	General	PedSep-A	PedSep-B	PedSep-C	PedSep-D
ANAKINRA, p-value (No.)	0.017 (5)	- (0)	1 (1)	0.055 (3)	0.367 (1)
BECLOMETHASONE, p-value (No.)	1 (2)	- (0)	- (0)	1 (2)	- (0)
BORTEZOMIB, p-value (No.)	1 (1)	- (0)	- (0)	- (0)	1 (1)
CAMPATH, p-value (No.)	1 (2)	- (0)	1 (1)	1 (1)	- (0)
CARBOPLATIN, p-value (No.)	1 (1)	1 (1)	- (0)	- (0)	- (0)
CELLCEPT, p-value (No.)	1 (3)	- (0)	- (0)	1 (1)	0.526 (2)
CISPLATIN, p-value (No.)	1 (1)	- (0)	- (0)	1 (1)	- (0)
CYCLOPHOSPHAMIDE, p-value (No.)	1 (1)	- (0)	- (0)	1 (1)	- (0)
CYCLOSPORINE, p-value (No.)	0.017 (2)	- (0)	- (0)	0.147 (1)	0.367 (1)
CYTARABINE, p-value (No.)	0.342 (3)	- (0)	- (0)	0.274 (2)	1 (1)
CYTOGAM, p-value (No.)	0.130 (1)	- (0)	- (0)	- (0)	0.367 (1)
DAUNORUBICIN, p-value (No.)	1 (1)	- (0)	- (0)	- (0)	1 (1)
DEXAMETHASON, p-value (No.)	0.580 (94)	1 (50)	1 (22)	0.004 (14)	1 (8)
DOXORUBICIN, p-value (No.)	1 (1)	- (0)	- (0)	1 (1)	- (0)
ENBREL, p-value (No.)	0.130 (1)	- (0)	- (0)	0.147 (1)	- (0)
EPOETIN ALFA, p-value (No.)	1 (1)	- (0)	- (0)	- (0)	1 (1)
ETANERCEPT, p-value (No.)	1 (1)	1 (1)	- (0)	- (0)	- (0)
ETOPOSIDE, p-value (No.)	0.128 (5)	1 (1)	- (0)	0.020 (2)	0.526 (2)
FLUDROCORTISONE, p-value (No.)	1 (1)	1 (1)	- (0)	- (0)	- (0)
FLUOROURACIL, p-value (No.)	1 (1)	- (0)	- (0)	1 (1)	- (0)
HYDROCORTISONE, p-value (No.)	0.061 (172)	1 (36)	1 (48)	1 (50)	0.724 (38)
HYDROXYCHLOROQUINE, p-value (No.)	1 (2)	1 (1)	- (0)	1 (1)	- (0)
HYDROXYUREA, p-value (No.)	0.017 (2)	- (0)	- (0)	- (0)	0.130 (2)
IMMUNOGLOBULIN G, p-value (No.)	0.001 (51)	1 (6)	0.113 (10)	0.025 (19)	0.537 (16)
INFLIXIMAB, p-value (No.)	0.130 (1)	- (0)	- (0)	0.147 (1)	- (0)
METHYLPREDNISOLONE, p-value (No.)	0.862 (117)	1 (54)	0.714 (23)	0.004 (24)	0.754 (16)
MYCOPHENOLATE, p-value (No.)	0.001 (8)	- (0)	1 (1)	0.009 (4)	0.546 (3)
NEUPOGEN, p-value (No.)	0.001 (23)	- (0)	0.241 (2)	0.012 (12)	0.708 (9)
PREDNISOLONE, p-value (No.)	0.781 (32)	1 (14)	1 (5)	0.612 (9)	1 (4)
PROGRAF, p-value (No.)	0.243 (2)	- (0)	- (0)	0.274 (2)	- (0)
PULMICORT, p-value (No.)	1 (3)	1 (1)	1 (1)	1 (1)	- (0)
RASBURICASE, p-value (No.)	1 (1)	- (0)	- (0)	- (0)	1 (1)
RITUXIMAB, p-value (No.)	1 (2)	1 (1)	1 (1)	- (0)	- (0)
SARGRAMOSTIM, p-value (No.)	0.130 (1)	- (0)	- (0)	- (0)	0.367 (1)
SIROLIMUS, p-value (No.)	1 (2)	- (0)	- (0)	1 (1)	1 (1)
SYMBICORT, p-value (No.)	1 (1)	1 (1)	- (0)	- (0)	- (0)
TACROLIMUS, p-value (No.)	0.042 (16)	1 (1)	1 (3)	0.021 (5)	1 (7)
THYMOGLOBULIN, p-value (No.)	0.130 (1)	- (0)	- (0)	0.147 (1)	- (0)
TOCILIZUMAB, p-value (No.)	1 (1)	- (0)	- (0)	1 (1)	- (0)
VINCRISTINE, p-value (No.)	1 (1)	- (0)	- (0)	1 (1)	- (0)
Plasma exchange, p-value (No.)	0.047 (25)	1 (5)	0.067 (4)	1 (4)	1 (12)
MechVent, p-value (No.)	0.10 (366)	1 (134)	1 (101)	0.032 (79)	1 (52)
ECMO, p-value (No.)	0.001 (30)	0.11 (5)	0.001 (9)	0.110 (6)	0.073 (10)
CRRT, p-value (No.)	0.001 (52)	1 (1)	0.003 (7)	0.14 (7)	0.23 (37)

2-6 columns of table present p values from statistical tests and number of patients treated by each therapy in each phenotype. A p value less than 0.05 indicates a significant association between individual therapy and mortality. "-" indicates no patient from a specific phenotype treated by this therapy. (Qin et al. 2022)

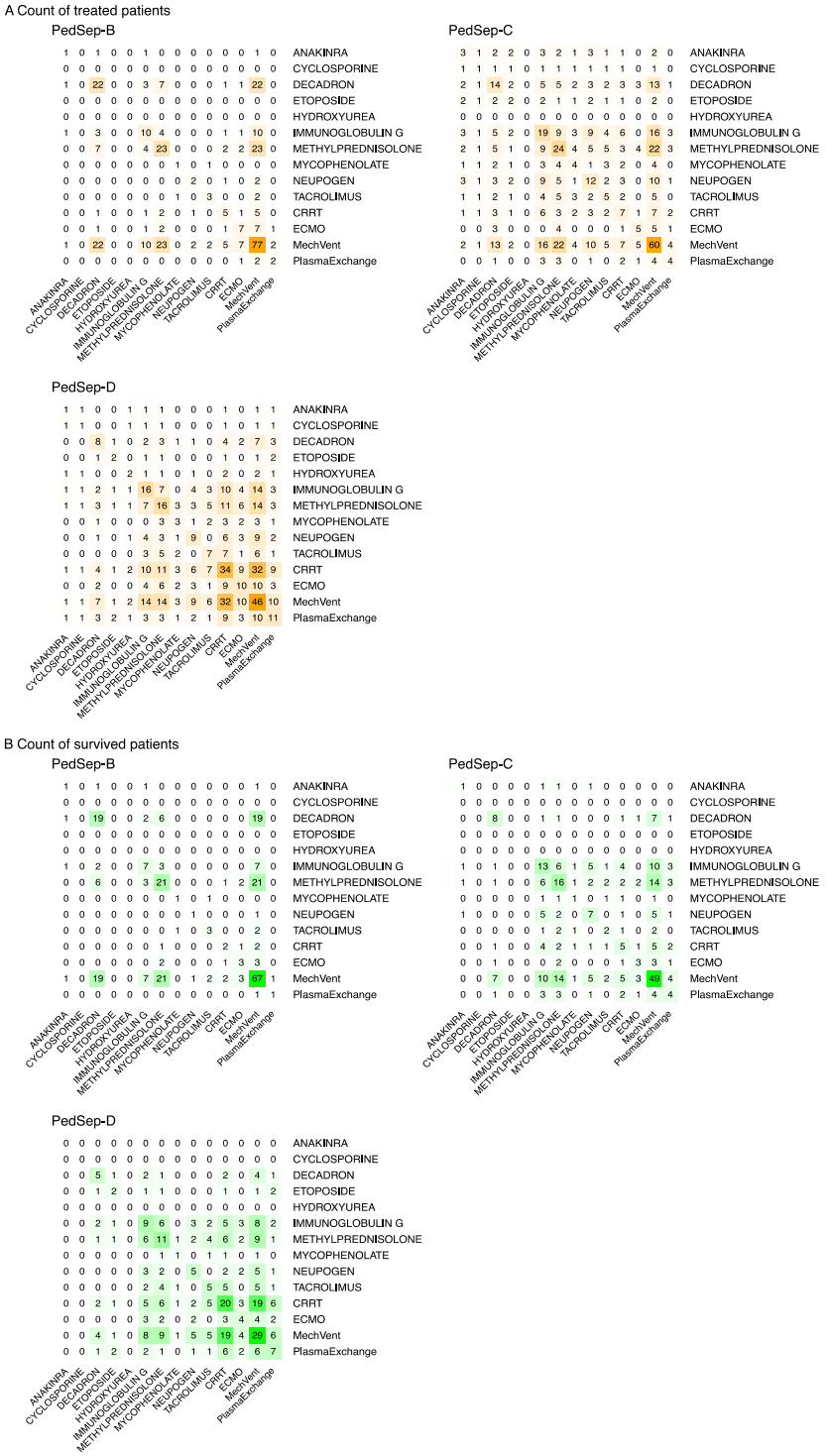

Figure 4.10 Counts of patients receiving 14 therapies alone and in combination

This heatmap shows the count of treated patients (A) and survivors among treated patients (B) in PedSep-B, C, and D. Therapies not associated with outcomes in univariable analysis (eTable 12) were not included. Drugs and Organ Support treatments are sorted in alphabetical order. (Qin et al. 2022)

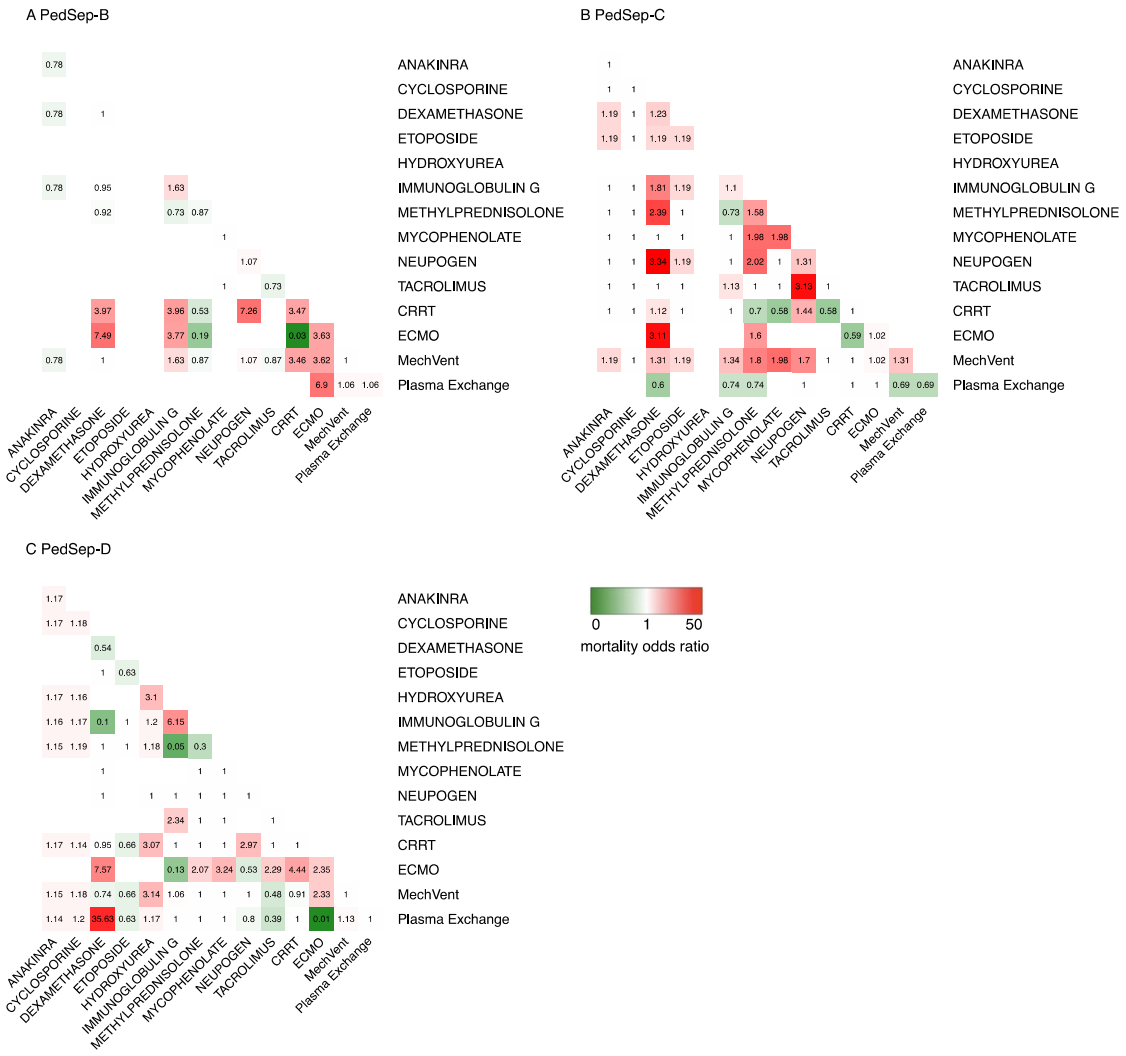

Figure 4.11 Heterogeneous treatment interactions and mortality risks among the phenotypes

Heatmap of Elastic Net Regression analysis shows the association between 14 individual therapies (diagonal values) and their 91 combination interactions (total cells = 105) with mortality in PedSep-B, C, and D among children who received anti-inflammatory therapies. Blank cells have no patients. Values in each cell represent odds ratios of mortality, where 1 represents no association with mortality. Color in each cell represents direction of effect, where red represents mortality direction, green represents survival direction. Cells located at the diagonal are odds ratio of association from the 14 individual therapies. (Qin et al. 2022)

4.5 Discussion

Machine learning analysis of patients from the extant PHENOMS study derived four computable 24-h phenotypes meeting three of five 'path forward' criteria providing impetus for their further evaluation in new pediatric sepsis studies (DeMerle et al. 2021). The derived computable phenotypes demonstrated clinical relevance with differences in types of presenting diagnoses, infections, organ failures, need for organ support therapies, outcomes, and proclivity to the development of TAMOF and MAS (Carcillo et al. 2019). Derived consensus k-means clustering and t-SNE analyses demonstrated that the computable phenotypes are nonsynonymous. The differences in cytokine profiles provide biological plausibility for these derived computable phenotypes having different inflammation responses, highlighted in PedSep-D by decreased ADAMTS13 with TAMOF and increased MIP 1α with MAS. Exploratory modeling of interactions between therapies among patients receiving anti-inflammatory treatments, derived computable phenotypes, and mortality demonstrated no reduction in mortality odds with methylprednisolone, IVIG or the combination; however, it identified a signal for methylprednisolone affecting the relationship of IVIG therapy to outcome in PedSep-D patients. We speculate that this interaction is reminiscent of the report that addition of methylprednisolone to IVIG improves cardiac function in children with COVID19-related multisystem inflammatory syndrome (MIS-C) compared to IVIG alone (Son et al. 2021). The very wide confidence intervals provide impetus to further evaluate this interaction signal in larger sample sizes using new study cohorts. We are presently assessing treatment responsiveness and reproducibility of the four derived phenotypes in our NICHD network's 1000-patient Personalized Immunomodulation in Pediatric Sepsis and Multiple Organ Dysfunction trial testing interleukin 1 antagonist protein for

hyper-inflammatory sepsis; and, also in the observational 500 patient Second Argentinian Pediatric Sepsis Epidemiology Study (PI Roberto Jabornisky).

PedSep-A is characterized by younger previously healthy children with respiratory failure and the least increased inflammation. This resembles the adult α phenotype in the SENECA trial (Seymour et al. 2019), and also the MARS 3 and sepsis response signature 2 endotypes, which found predominant expression of adaptive immune and B-cell developmental pathways (Davenport et al. 2016, Scicluna et al. 2017, Sweeney et al. 2018). Mortality in PedSep-A was low at 2% and did not increase after 7 days, making anti-inflammatory clinical trials directed to survival less feasible.

PedSep-B is characterized by multiple organ failure requiring intubation for more severe respiratory failure, shock, and central nervous system dysfunction with increased C-reactive protein levels and 12% mortality. This is reminiscent of children reported in the Life After Pediatric Sepsis Evaluation study (Zimmerman et al. 2020); the shock with hypoxia phenotype in adult sepsis-induced MOF (Knox et al. 2015); and the severe hypoxia, altered mental status, and shock phenotype in pediatric MOF (Ye and Sanchez-Pinto 2020).

PedSep-C is distinguished by cardiovascular failure and relative absence of need for intubation (14%) with the least pulmonary failure (34%) and need for mechanical ventilation (71%), in the presence of elevated C-reactive protein, high ferritin, and lymphopenia, with 10% mortality. This is reminiscent of the Toxic Shock (TSS)—Kawasaki syndrome phenotype currently being considered as PMIS/MIS-C syndrome (Ebato et al. 2017, Ma et al. 2018, Carter et al. 2020, Cook et al. 2020, Son et al. 2021). Similar to TSS and Kawasaki's, our PedSep-C patients showed elevated IL-17a and IP10/CXCL10 levels (Szabo et al. 2017, Chang et al. 2020).

PedSep-D patients had cardiovascular, respiratory, liver, renal, hematologic, and neurologic dysfunction with 34% mortality; clinical features shared by the adult δ phenotype characterized in the SENECA study using electronic health record criteria for Sepsis-3 (Seymour et al. 2019); the shock with thrombocytopenia pediatric MOF phenotype (Ye and Sanchez-Pinto 2020); and previously reported subclasses including the hyperinflammatory sub-phenotype reported in acute respiratory distress syndrome, a condition commonly related to sepsis (Calfee et al. 2014, Sinha et al. 2020, Yasin et al. 2020). It also resembles sepsis endotypes derived using transcriptomic analyses of circulating immune cells, specifically the inflammopathic cluster known as sepsis signature 1, or the Molecular Diagnosis and Risk Stratification of Sepsis (MARS) 2 cluster (Davenport et al. 2016, Scicluna et al. 2017, Sweeney et al. 2018). PedSep-D is specifically characterized by hyperferritinemic (ferritin > 500 ng/mL), thrombocytopenic (platelet count < 100 K) multiple organ failure with the highest likelihood of new or progressive multiple organ failure accruing mortality after 7 days and the lowest number of PICU free days. PedSep-D membership identifies children with the highest proclivity for decreased ADAMTS 13 activity with thrombocytopenia-associated MOF, and increased MIP 1α with macrophage activation syndrome.

An ongoing external validation analysis was conducted by Caldino Bohn et al to investigate pediatric sepsis phenotypes derived by the same method in an Argentina cohort, LMIC (N = 428). Variable sets and distributions were similar between PHENOMS and LMIC cohorts. Four phenotypes were validated as the optimal number of clusters, where the PedSep-D phenotype was the most similar phenotype between the two cohorts. The mortality rates of the four phenotypes were similar between the two cohorts , where the PedSep-D phenotype had the highest mortality (Argentina 33%; USA 34%).

There are limitations to consider in this post hoc machine learning analysis of the parent PHENOMS study and its inherent selection bias risks. Although the PHENOMS study represents the largest longitudinal multiple center pediatric sepsis-induced MOF cohort with concomitant CRP and ferritin levels available (Carcillo et al. 2019), it is small compared to adult standards because sepsis occurs 15 times more commonly in adults than in children. Definitions of pediatric sepsis and organ failures are also evolving and behind the changes in adult sepsis. Definitions of sepsis and organ failure were necessarily limited to those used in the extant study. Only 25 out of 52 available clinical and laboratory variables available in this parent study had < 20% missingness and were included in the machine learning derivation. Only 33 additional biomarkers were performed to assess biological plausibility for the computable phenotypes having different inflammatory responses. Lactate was not recorded and may be an important missing variable (Tonial et al. 2021). Interactions could only be assessed for those therapies given by bedside clinicians in a 'natural experiment' setting. Our models did not capture all confounders, comorbidities, therapies used, reasons for therapies, or site differences in clinical practice. Furthermore, combined methylprednisolone plus IVIG and ECMO plus plasma exchange therapies were rarely administered. The reproducibility of the derived computable phenotypes cannot be assessed in a single extant multiple-center resource-rich study. We are presently assessing reproducibility in two ongoing independent cohort studies.

5.0 Uncover the Role of Rare Variants in Pediatric Sepsis through an Exome-wide Gene-based Association Analysis

5.1 Forward

Most parts of the writings, figures, and tables of this chapter are under review as of this moment. The current version of the manuscript under review is "Yidi Qin, Kate F Kernan, Yulong Bai, John R Shaffer, Zsolt Urban, Scott Canna, Robert A. Berg, David Wessel, Murray M. Pollack, Kathleen Meert, Mark Hall, Christopher J. Newth, John C. Lin, Tom Shanley, Rick E. Harrison, Joseph A Carcillo, and Hyun-Jung Park1. Deleterious variants in LTBP4 are associated with a severe pediatric sepsis phenotype established by agnostic method. "

5.2 Introduction

Pediatric sepsis is a life-threatening condition associated with organ failure in children predominantly due to a dysregulated host immune response to infection. It is a recognized global public health problem that affects 20.3 million children and causes 2.9 million deaths in those under five years old every year (World Health Organization 2020). Despite global efforts to improve clinical outcomes for pediatric sepsis, its phenotypic heterogeneity remains a significant barrier to therapeutic advancement (Cavaillon et al. 2020). Several recent analyses have attempted to explore pediatric sepsis phenotypes using either empirical or data-driven analyses (Wong et al. 2009, Sweeney et al. 2018, Carcillo et al. 2019, Sanchez-Pinto et al. 2020). However, most of

these analyses were limited by small sample size or bedside relevance, failing to incorporate statistically significant and clinically reasonable features. Recently, by applying machine learning approaches to 25 first-day bedside clinical variables of 404 pediatric sepsis patients with organ dysfunction enrolled as part of a multicenter cohort, PHENOtyping sepsis-induced Multiple organ failure Study (PHENOMS), between 2015 to 2017, we derived four agnostic phenotypes PedSep-A, B, C, and D (Qin et al. 2022). The four phenotypes display significantly distinct patterns in terms of sources of infections (e.g., bacteria, viremia, fungal), cytokines, organ failure, clinical outcomes (e.g., mortality, length of stay in Pediatric Intensive Care Unit (PICU), PICU free days), and therapeutic responses, suggesting that they are biologically meaningful, clinically relevant, and potentially targetable phenotypes.

Several studies suggested that host genetic factors contribute to the heterogeneity of pediatric sepsis. Evidence for the genetic basis of sepsis stems from the landmark study of Sørensen et al. in 1988 (Sorensen et al. 1988), where the authors reported a five-fold higher risk of death in infected adoptees with one biological parent who had died from infection. Building on this observation, subsequent research has attempted to reveal the genetic factors of sepsis susceptibility and outcomes. Many of these studies have focused on known genetic loci or gene sets of interest due to their impact on human disease or immunologic function (Asgari et al. 2016, Borghesi et al. 2020, Kernan et al. 2022). Despite known associations, these targeted inquiries limit the capacity to discover novel functional sepsis-related genes. Alternatively, researchers have conducted several genome-wide association studies (GWAS) on adult and pediatric populations to identify common variants underlying sepsis susceptibility and outcomes (Rautanen et al. 2015, Butler-Laporte et al. 2020, Hernandez-Beeftink et al. 2022). However, although common variants have been used to understand the genetic basis of clinical outcomes (e.g., survival from sepsis or

hospital admission), they are limited in elucidating the genetic architecture for agnostic pediatric sepsis phenotypes with poor outcomes. First, common variants usually have small effects on complex traits, making them difficult to detect, requiring a prohibitively large sample size and of limited clinical benefit in a large fraction of the population (Manolio et al. 2009). Indeed, the only GWAS on 351 extremely premature infants with sepsis could not identify common variants separating them from 406 healthy controls, demonstrating the limitation of common variant analysis in delineating sepsis susceptibility (Srinivasan et al. 2017). Second, common variants are often located in non-coding regions of the genome where the function is currently poorly understood (Tam et al. 2019). As a result, out of 27 common variants reported in GWAS Catalog as associated with sepsis to date, 26 variants are located in non-coding regions without clearly annotated consequences. The only reported GWAS variant in a coding region is a missense variant within SAMD9 (rs34896991) with an allele frequency close to 0.01 (Hernandez-Beeftink et al. 2022). Without involving further functional validation, the large fraction of findings in non-coding regions challenges the interpretation of the sepsis GWAS results.

To address these limitations and delineate the genetic architecture of the agnostic pediatric sepsis phenotype with poor outcomes, we performed a gene-based exome-wide rare variant analysis using data from the PHENOMS study of severe pediatric sepsis. Using whole-exome sequencing data allowed us to investigate rare variants lying in coding regions with potentially larger effects. Specifically, to investigate the role of rare variants in distinguishing a previously identified PedSep-D phenotype from others (i.e. PedSep-A, B, and C), we performed a gene-based rare variant analysis on the pediatric sepsis phenotypes using whole exome sequencing (WES) data collected from a subset of the 404 children with sepsis and evidence of organ dysfunction (n=319, Figure 1). Additionally, as pediatric sepsis morbidity and mortality were incurred by only

a small subset of children, we focused our inquiry on the phenotype characterized by the worst outcomes by contradicting it with the rest of the cohort, to identify novel genotype associations and provide insights into potential therapies. Altogether, we present the first rare variant burden test in the phenotype with the worst outcomes in pediatric sepsis.

5.3 Method

5.3.1 DNA extraction and genotyping

Out of all pediatric patients enrolled in the cohort, a total of 381 parents of the children provided WES consent, and 2 mL of whole blood was collected for DNA extraction using standard methods. Whole-exome sequencing was successfully completed on 332 patients from 2018 to 2020 by the University of Pittsburgh Genomics Research Core performed on the Ion Torrent platform. Libraries were constructed by the Ampliseq Exome RDY (Thermo Fisher Scientific) with 100 × target coverage. FASTQ files were aligned to Homo sapiens reference sequence GRCh37/hg19 to generate VCF files. Variant calling was performed by GATK (Genome Analysis Toolkit) (Van der Auwera GA 2020).

5.3.2 Quality control

Two levels of quality control were conducted on 332 samples with completed whole-exome sequencing data, patient-level, and variant-level. At the patient level, we excluded nine individuals without phenotype information. Four pairs of individuals were identified as relatives

based on IBD (identity by descent). In each IBD pair, the individual with the higher missingness was removed from the analysis. In terms of variant-level quality control, we filtered sites with SOR (Strand Odds Ratio) > 3, MQ (root mean square Mapping Quality) < 40, QD (variant confidence normalized by depth) < 2.0, average GQ (Genotyping confidence) < 20, average DP (Depth) < 10, missingness > 0.05, HWE (Hardy-Weinberg equilibrium p) < 1e-06, and those located on sex chromosomes. No imputation of missing genotypes was performed due to concerns for potentially low imputation quality of rare variants in datasets with small sample sizes. Quality control was performed by software bcftools (v1.9) (Li 2011), VCFtools (v0.1.16) (Danecek et al. 2011), and PLINK (v1.9) (Purcell et al. 2007). Then variant function was annotated by ANNOVAR (Wang et al. 2010).

5.3.3 Gene-based analysis

Variants that passed quality control were included if they were in hg19 annotated exon regions and had a MAF (minor allele frequency) lower than 1%. Genes with less than three qualified variants were excluded from the analysis to ensure there is no inflation in test statistics. The final number of genes tested was 3,846. Therefore, the p-value threshold for declaring whole-exome level significance was 0.05/3846 = 1.3e-05. The suggestive p-value threshold is 0.5/3846 = 1.3e-04.

We aggregately examined the relationships between the rare variants and the binary indicator of phenotype membership by gene-based association test SKAT (Sequence Kernel Association Test) (Wu et al. 2011). SKAT is a widely-employed method to test the association between a group of variants and the trait, which increases the power to detect rare variant associations by pooling rare variants across a given region of interest, such as chromosome region

or gene. In running the SKAT test, a single null model was fitted containing only the covariates to be adjusted (i.e., age, sex, and the first four ancestry PCs constructed from common linkage disequilibrium (LD)-pruned SNPs). Then the effect of SNPs from each gene was tested between PedSep-D group and non-PedSep-D group by variance-component score tests in a mixed model, and their statistics were aggregated with weights through a kernel matrix to form a gene-level statistic. Compared to other gene-based tests such as the Burden Test and SKAT-O, one advantage of applying SKAT in our analysis is that it makes few assumptions about rare-variant effects and retains statistical power when variants within a gene have different directions and magnitude of effects (Lee et al. 2012). This property aligns with the study design that contrasts one phenotype with others and allows us to better account for potential heterogeneity in phenotypes.

Genes showing whole-exome level significance and suggestive significance were further investigated to query the gene function (GeneCards) (Stelzer et al. 2016), common variant evidence from previous GWAS analysis (GWAS Catalog) (Sollis et al. 2023), gene enrichment in GO biological process (FUMA, Enrichr) (Chen et al. 2013, Watanabe et al. 2017), and gene expression level in the GTEx database (Consortium 2013). Rare variants that in the top gene were annotated with four different types of score (CADD (Rentzsch et al. 2019), GERP (Davydov et al. 2010), SIFT (Ng and Henikoff 2001), and Polyphen2 (Adzhubei et al. 2010)) to indicate the effect of each variant.

5.3.4 Comparison of cytokine profiles between rare variants carriers and non-carriers

To further investigate the effect of variations on inflammation, levels of the 33 pre-collected biomarkers of the rare variant carriers were further visualized and compared with non-carrier (Qin et al. 2022). The cytokine heatmap was used to present the log ratio of the median

biomarker values of the host response. The red color represents a greater median value for the carrier group compared to the median value of the entire cohort, while the blue color represents a lower median value for the group compared to the median value oof the entire cohort. Hierarchical clustering was used to visualize the similarity of cytokine patterns between rare variant carriers. Additionally, we calculated p-values from a pairwise t-test comparing cytokine values of rare variant carriers and non-carriers.

5.3.5 Sensitivity Analysis

To further validate the top genes identified from the gene-based analysis, we investigated the influence of the ancestry information on the genes with a sensitivity analysis performed as follows. Briefly, we relabeled the phenotypes of the original PedSep-D samples as "non-PedSep-D". Then we selected the same number of samples as PedSep-D group from the original non-PedSep-D group, maintaining the same racial composition as the original PedSep-D group. Specifically, from the 279 samples in the original non-PedSep-D group, we randomly selected 28 samples with reported race as White, 7 samples with the reported race as Black, 3 samples with the reported race as Asian, and 2 samples with the reported race as Unknown. Then we relabeled the phenotypes of these randomly selected samples as "PedSep-D". With this approach, we ran 100,000 iterations of random selection step, coupled with SKAT gene-based analysis, to calculate how many times the test statistic value is greater than the test statistic value from observed data. Thus, we generated an empirical p-value for each gene.

5.3.6 Plasma protein quantification

Blood samples were drawn from 110 patients by venipuncture into EDTA tubes and plasma supernatant was collected. Plasma protein was measured by SomaScan proteomic platform based on a previously established protocol (SomaLogic, lnc) (Gold et al. 2010). Briefly, the Slow Off-Rate Modified Aptamer (SOMAmer)-based capture array enables reagents to bind target peptides and transform protein signals to nucleotide signals that can be quantified with relative fluorescence on a custom Agilent hybridization chip. Quality control involving hybridization controls and calibration samples was performed at the sample and SOMAmer levels to detect and mitigate technical variations. The plasma protein abundance from the SomeScan assay is reported as relative fluorescent units (RFU).

5.3.7 Serum protein profiling

The serum protein level of 33 biomarkers, including 31 cytokines and two functional assays, were measured at day one of sepsis. The procedures for the functional assays were conducted as outlined in prior studies (Qin et al. 2022). Specifically, these assays were used to measure the whole blood ex vivo TNF response to endotoxin, indicative of immune suppression, and the activity of ADAMTS 13, which signals microvascular clotting in cases of low platelet counts. The plasma designated for cytokine analysis was split across three separate tests. Measurements of IL-18, IL-18BP, and CXCL9 were carried out with a 25-fold dilution. IFN α, sCD163, and IL-22 levels were determined using the Bioplex inflammatory flex-set assay according to the guidelines provided by Bio-Rad. The analysis of the remaining cytokines was

performed using the Bioplex Group I/II flex-set assay from Bio-Rad. These cytokine levels were all assessed using the BioPlex 200 System by Bio-Rad.

5.4 Results

5.4.1 PedSep-D phenotype has the highest mortality with unique clinical presentation and immune system profile

Out of 404 pediatric sepsis patients enrolled in the PHENOM cohort, the blood samples of 319 patients underwent whole-exome sequencing and passed quality control (Figure 5.1, Appendix Table 2). These 319 patients were assigned to one of four established phenotypes (PedSep-A, B, C, D) as previously determined by the consensus k-means clustering of 25 first-day bedside features (Qin et al. 2022). Among them, the sample sizes of PedSep-A, B, C, and D are 116 (36%), 86 (27%), 77 (24 %), and 40 (13%), respectively (Table 5.1). The proportions of patients in each phenotype are close to our original study, in which PedSep-A, B, C, and D contained 34, 25, 27, and 14 percent of the 404 patients, respectively.

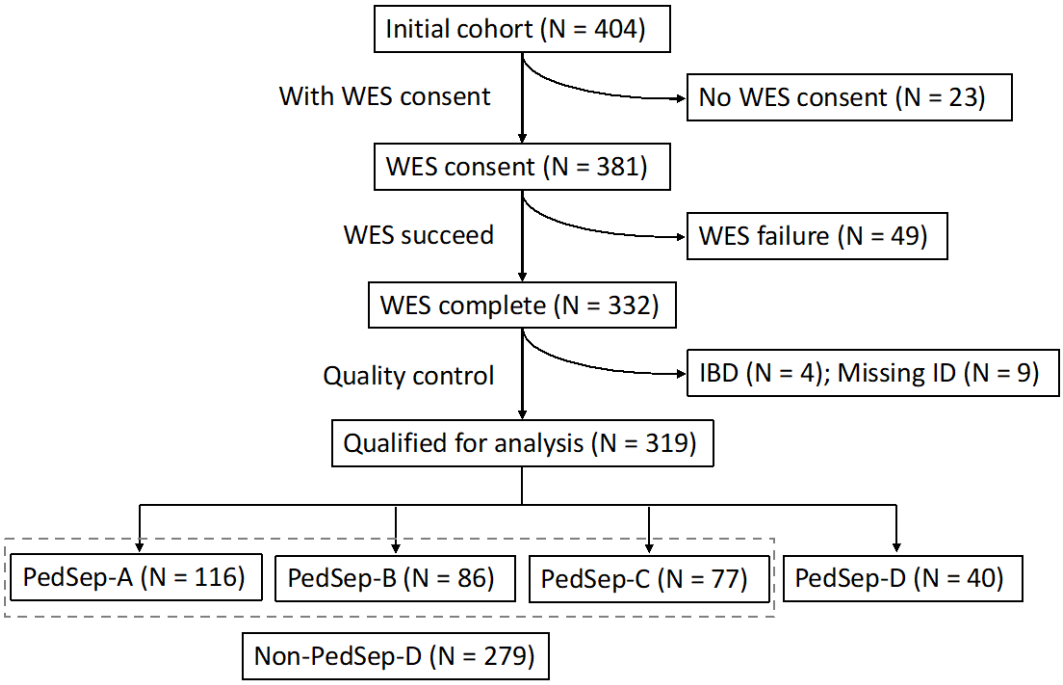

Figure 5.1 Workflow
IBD: identity by descent; Missing ID: fail to match WES sample ID and patient ID

Table 5.1 Demographic and day one clinical characteristics of PedSep-D and Non-PedSep-D patients

Characteristics	PedSep-D	Non-PedSep-D
No. of Patients, *N* (%)	40 (12.539)	279 (87.461)
Demographic		
Age, years mean (SD)	8 (6)	6 (6)
Male, N (%)	25 (62.5)	150 (53.8)
Hispanic, N (%)	3 (7.7)	47 (17.6)
Previous healthy, N (%)	17 (42.5)	136 (48.7)
Surgery, N (%)	8 (20.0)	30 (10.8)
Organ Dysfunction		
SIRS criteria[1], mean (SD)	3.0 (0.8)	2.9 (0.8)
OFI[2], mean (SD)	3.0 (1.1)	1.6 (0.6)
Inflammation		
CRP, mg/dL mean (SD)	12.8 (11.6)	11.5 (9.9)
Low temperature, °C mean (SD)	36.4 (1.0)	36.6 (1.3)
High temperature, °C mean (SD)	37.9 (1.4)	37.8 (1.2)
ALC, /mm^3 median (IQR)	1.3 (0.7-2.4)	1.3 (0.7-2.2)
Ferritin, ng/mL mean (IQR)	575.0 (195.6-1628.8)	180.0 (87.4-403.0)
Pulmonary		
Pulmonary OFI, N (%)	25 (62.5)	188 (67.4)
Intubation, N (%)	22 (55.0)	156 (55.9)
Cardiovascular or Hemodynamic		
Heart rate, bpm mean (SD)	145.5 (38.8)	156.3 (30.6)
Systolic blood pressure, mmHg mean (SD)	79.1 (22.2)	81.7 (19.3)
CV OFI, N (%)	30 (75.0)	189 (67.7)
Renal		
Creatinine, mg/dL median (IQR)	1.5 (1.0-3.0)	0.4 (0.3-0.7)
Renal OFI, N (%)	26 (65.0)	0 (0.0)
Hepatic		
Hepatic OFI, N (%)	12 (30.0)	19 (6.8)
Hematologic		
Hemoglobin, g/dL mean (SD)	9.4 (1.8)	10.0 (1.9)
Platelets, K/mm^3 mean (SD)	84.0 (73.7)	192.9 (112.9)
Hematologic OFI, N (%)	19 (47.5)	7 (2.5)
Other		
Glasgow Coma Scale score[3,4], mean (SD)	7.5 (5.6)	8.5 (5.3)
CNS OFI, N (%)	9 (22.5)	33 (11.8)

IQR interquartile range, SIRS systemic inflammatory response syndrome, OFI organ failure index, CRP high-sensitivity cardiac C-Reactive protein, ALC absolute lymphocyte count, CNS central nervous system

SI conversion factors: to convert alanine transaminase and aspartate aminotransferase to μkat/L, multiply by 0.0167; bilirubin to μmol/L, multiply by 17.104; C-reactive protein to nmol/L, multiply by 9.524; creatinine to μmol/L, multiply by 88.4

1 Indicates SIRS criteria ranging from 0 to 4, including abnormal heart rate, respiratory rate, temperature, and white blood cell count

2 OFI is an integer score reflecting the number of organ failures. Scores are either 0 or 1 for cardiovascular, hepatic, hematologic, respiratory, neurological, and renal, and summed for a total range of 0 to 6. Cardiovascular, need for cardiovascular agent infusion support; Pulmonary, need for mechanical ventilation support with the ratio of the arterial partial pressure of oxygen and the fraction of inspired oxygen (PaO2/FiO2) < 300 without this support; Hepatic, total bilirubin > 1.0 mg/dL and alanine aminotransferase (ALT) > 100 units/L; Renal, serum creatinine > 1.0 mg/dL and oliguria (urine output < 0.5 mL/kg/h); Hematologic, thrombocytopenia < 100,000/mm3 and prothrombin time INR > 1.5 × normal; Central Nervous System, Glasgow Coma Scale (GCS) Score < 12 in the absence of sedatives

3 Corresponds to the minimum or maximum value (as appropriate) within six h of hospital presentation

4 GCS ranges from 3 to 15

To distinguish each phenotype in its characteristics, biomarkers, and outcomes, we compared each phenotype and the other phenotypes in terms of clinical characteristics (Table 5.1, Appendix Table 3, 4, 5) and cytokine biomarkers (Table 5.2, Appendix Table 6, 7, 8). Comparing patients from PedSep-A vs. all the other patients, PedSep-A is characterized by younger age (3 versus 9 years old), lower organ failure in all organs (Appendix Table 3), and lower levels of inflammation biomarkers including sCD163, IL-22, IL-18, IL-18BP, MIG/CXCL9, IL-6, IL-8, IL-10, IL-17A, MCP-1/CCL2, MIP-1α, MIP-1β, IL-16, M-CSF, SCF, CRP, and Ferritin (Appendix Table 6). PedSep-B is characterized by relatively more organ dysfunction (Appendix Table 4), and PedSep-C is characterized by higher values of IL-18, IL-18BP, MIF/CXCL9, IL-6, IL-17A, IP-10/CXCL10, CRP, and Ferritin (Appendix Table 8), representing higher inflammation, PedSep-D is characterized by more organ failure (Table 5.1, higher OFI, more renal, hepatic, and hematologic organ failure), lower level of ADAMTS13, and a higher level of sCD163, IL-22, IL-18BP, MIG/CXCL9, IL-6, IL-8, IL-10, MIP-1α, IL-16, M-CSF, SCF, CRP, and Ferritin (Table 5.2),

showing a profound inflammatory response in the PedSep-D group. Subsequently, PedSep-D showed the longest PICU stay length, the fewest PICU-free days, and the highest mortality (Table 5.3). Altogether, the results highlight that PedSep-D patients had a high risk of developing severe outcomes.

Table 5.2 Biomarkers measured at day one by phenotype (N = 319)

Biomarker[a]	PedSep-D (N = 40)	Non-PedSep-D (N = 279)	p-value	fdr
ADAMTS13, %	55.0 (38.0, 66.2)	74.0 (58.0, 93.0)	1.219E-6	6.705E-6
SFasLg, pg/ml	42.4 (32.0, 62.3)	48.1 (31.2, 80.2)	0.175	0.251
Ex vivo TNF-α, pg/ml	623.3 (187.7, 1049.2)	476.6 (139.4, 1049.2)	0.355	0.457
TNF-α, pg/ml	728.7 (602.0, 1049.2)	1049.2 (726.5, 1161.9)	0.074	0.122
sCD163, pg/ml	464600 (270060, 741798)	267704 (174016, 417048)	4.439E-5	1.465E-4
IFN-β, pg/ml	6.4 (6.4, 6.4)	6.4 (6.4, 8.2)	0.356	0.457
IL-22, pg/ml	34.2 (25.4, 59.0)	24.8 (20.1, 33.0)	5.013E-4	1.379E-3
IL-18, pg/ml	518.4 (344.6, 744.9)	398.0 (236.1, 694.2)	0.074	0.122
IL-18BP, pg/ml	30713 (18613, 40878)	14388 (8034, 24711)	3.347E-7	2.209E-6
MIG/CXCL9, pg/ml	2462.4 (695.9, 4430.3)	753.6 (412.8, 1528.2)	7.866E-5	2.360E-4
IL-1β, pg/ml	3.0 (2.6, 3.3)	2.8 (2.3, 3.3)	0.155	0.233
IL-4, pg/ml	4.3 (3.5, 6.3)	4.7 (3.5, 6.5)	0.360	0.457
IL-6, pg/ml	17.1 (8.4, 43.8)	8.4 (6.2, 15.0)	6.400E-4	1.625E-3
IL-8, pg/ml	113.3 (60.7, 316.4)	46.5 (30.3, 77.8)	2.219E-7	1.831E-6
IL-10, pg/ml	29.9 (24.8, 71.0)	21.1 (16.3, 31.0)	1.051E-5	4.195E-5
IL-13, pg/ml	3.1 (3.1, 3.4)	3.1 (3.1, 4.3)	0.626	0.688
IL-17A, pg/ml	19.1 (16.5, 23.0)	18.3 (16.5, 23.4)	0.592	0.674
IFN-γ, pg/ml	2.8 (2.8, 2.8)	2.8 (2.8, 3.0)	0.089	0.140
IP-10/CXCL10, pg/ml	960.5 (526.4, 3160.0)	692.5 (315.9, 2007.7)	0.041	0.075
MCP-1/CCL2, pg/ml	190.9 (107.7, 488.8)	129.7 (56.6, 288.6)	0.016	0.031
MIP-1α, pg/ml	5.4 (1.7, 13.9)	0.6 (0.6, 5.7)	1.144E-5	4.195E-5
MIP-1β, pg/ml	56.8 (47.9, 89.3)	43.8 (30.6, 64.2)	2.093E-3	4.934E-3
MCP-3, pg/ml	92.4 (92.4, 147.8)	92.4 (92.4, 166.0)	0.780	0.784
IFN-α2, pg/ml	120.0 (105.8, 140.2)	125.7 (105.8, 144.4)	0.522	0.615
IL-1α, pg/ml	9.4 (9.4, 14.8)	9.4 (9.4, 16.4)	0.784	0.784
IL-2RA, pg/ml	456.0 (347.0, 660.9)	357.6 (223.9, 579.7)	0.014	0.028
IL-3, pg/ml	612.2 (496.1, 708.7)	612.2 (529.0, 724.4)	0.428	0.523
IL-16, pg/ml	1146.9 (756.5, 1346.6)	556.5 (416.4, 702.2)	2.674E-9	2.941E-8
M-CSF, pg/ml	79.9 (46.2, 117.2)	26.8 (15.4, 43.3)	8.266E-10	1.364E-8

SCF, pg/ml	344.6 (226.2, 560.4)	145.5 (111.1, 203.7)	1.370E-11	4.521E-10
TRAIL, pg/ml	30.4 (25.4, 42.9)	38.5 (30.4, 55.4)	8.634E-3	1.899E-2
CRP, mg/dL	10.1 (2.6, 20.7)	9.6 (3.7, 16.3)	0.767	0.784
Ferritin, ng/mL	575.0 (195.6, 1628.8)	180.0 (87.3, 403.0)	7.697E-6	3.628E-5

[a] All biomarkers are measured one time concomitantly in the first day. Values in the table are summarized as median (IQR)

Table 5.3 Outcome by phenotype (N = 319)

Outcome	PedSep-D	Non-PedSep-D	p-value
Length of stay, median (IQR), d	13 (5, 31)	8 (5, 15)	0.018
Mortality, N (%)	11 (24.5)	17 (6.1)	2.969E-5
PICU free day, median (IQR), d	11 (0, 21)	21 (13, 25)	4.917E-4

5.4.2 Gene-based test associates *LTBP4*, *PLA2G4E*, and *CCDC157* with PedSep-D

To detect genetic factors associated with the sepsis phenotypes, we performed a whole exome-wide rare variant analysis. To increase power in detecting associations, we aggregated the rare variant association signals by gene and estimated the significance in the following steps (see Methods). First, we performed quality control and selected a total of 3,864 genes (see Methods) that had more than three variants. Then, we ran SKAT on the WES data to compare PedSep-D phenotypes versus the other three phenotypes while adjusting for age, sex, and the first four PCs constructed based on common variants.

As a result, variation in *LTBP4* was significantly associated with PedSep-D phenotype at the exome-wide level (p-value = 1.069E-05), while variations in *PLA2G4E* and *CCDC157* were suggestively associated with PedSep-D phenotype (p-value = 3.288E-05, p = 6.192E-05, respectively) (Fig. 5.2, Table 5.4). Four, 4, and 8 rare variants are located in exon regions of genes *LTBP4*, *PLA2G4E*, and *CCDC157*, respectively (Table 5.5). All variants encode missense variants except one in the *LTBP4* gene. However, this silent variant, rs370696272, replaces a common leucine codon (CTG, 0.361) with a less common codon (TTG, 0.134) based on the CoCoPUT database (Alexaki et al. 2019), explaining its high CADD score (17.55). Most variants in three genes were predicted to be deleterious based on their CADD score (12 out of 16 with CADD > 10), among which SNP rs573310430 in *LTBP4* had the highest CADD score of 34, ranked over

the top 0.1% in terms of predicted deleteriousness among variants across the whole genome. This variant creates an unpaired cysteine in the 14th calcium-binding epidermal growth factor-like (cbEGF) domain of *LTBP4*, a domain stabilized by 3 pairs of cysteines forming intradomain disulfide bonds particularly sensitive to the removal or addition of cysteine residues (Downing et al. 1996).

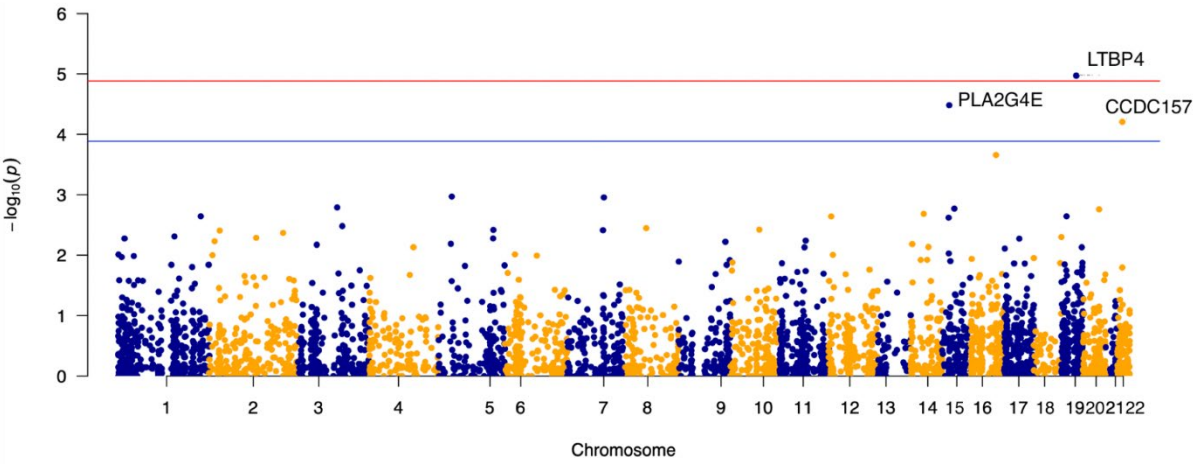

Figure 5.2 Manhattan plot for PedSep-D vs others (No. of genes = 3,846)
Red line: whole-exome wide significant –log10(P) value; Blue line: suggested significant –log10(P) value.

Table 5.4 SKAT gene-based association test result for PedSep-D vs non-PedSep-D

Gene	Chr	# carriers in PedSep-D (%)	# carriers in non-PedSep-D (%)	Odds Ratio	P-value
LTBP4	19	6 (15)	2 (0.7)	24.4	1.069E-5
PLA2G4E	15	4 (10)	2 (0.7)	15.4	3.288E-5
CCDC157	22	7 (17.5)	10 (3.6)	5.7	6.192E-5

Table 5.5 Information and functional prediction for variants contributing to the gene-level significance

Gene	Variant	SNP Information[a]	Amino acid change[b]	CADD score[c]	GERP score[d]	SIFT score[e]	Polyphen2 score[f]
LTBP4	rs370696272	19:41105311:C:T	Leu27Leu	17.55	1.63	-	-
	rs573310430	19:41122842:C:T	Arg984Cys	34	4.63	0.044 (D)	1.0 (D)
	-	19:41132970:C:T	Pro1388Leu	25.8	4.58	0.68 (T)	0.998 (D)
	rs200607327	19:41133005:G:A	Gly1437Arg	27	4.58	0.38 (T)	1.0 (D)
PLA2G4E	-	15:42276733:T:G	Lys387Gln	23	4.48	0.275 (T)	0.26 (B)
	rs764494895	15:42278161:G:A	Ala693Val	11.04	0.591	0.25 (T)	0.004 (B)
	rs143966595	15:42293394:C:T	Val212Ile	23.1	5.34	0.099 (T)	0.05 (B)
	rs776016335	15:42298270:T:C	Asp148Gly	27.1	5.66	0.002 (D)	1.0 (D)
CCDC157	rs9606721	22:30762035:A:G	Thr16Ala	12.47	1.76	0.28 (T)	0.001 (B)
	rs540507025	22:30762080:C:T	Arg31Cys	23.1	2.89	0.002 (D)	1.0 (D)
	rs143249037	22:30766366:G:A	Glu158Lys	14.06	1.36	0.282 (T)	0.035 (B)
	-	22:30766438:C:A	Gln182Lys	8.248	2.89	0.931 (T)	0.009 (B)
	rs139609945	22:30766496:C:T	Thr201Met	8.755	2.19	0.107 (T)	0.155 (B)
	rs1235664314	22:30766672:G:T	Asp260Tyr	28.4	5.29	0.008 (D)	1.0 (D)
	rs148283823	22:30766868:G:A	Arg325Gln	6.266	0.566	0.712 (T)	0.093 (B)
	rs202178544	22:30772567:T:C	Ser698Pro	0.246	-2.83	0.339 (T)	0.0 (B)

[a]SNPs are listed as chromosome: position (hg19): reference allele: alternative allele.
[b]Amino acid substitutions caused by SNPs.
[c]CADD (Combined Annotation-Dependent Depletion) score measures the predicted variant effect rank, higher value implies a greater damaging effect throughout the human genome reference assembly. A score of 10 indicates that the SNP is predicted to be in the top 10% most deleterious substitutions in the human genome, a score 0f 20 indicates that the SNP is predicted to be in the top 1% most deleterious substitutions, a score of 30 indicates that the SNP is predicted to be in the top 0.1% most deleterious substitutions and so forth.
[d]GERP (Genomic Evolutionary Rate Profiling) score indicates position-specific estimates of evolutionary constraint. A positive score scale with the level of constraint, a greater score suggests a greater level of evolutionary constraint. A negative score indicates that a site is probably evolving neutrally.
[e]SIFT (Sorting Intolerant from tolerant) score ranges from 0 to 1. A value less than 0.05 is classified as damaging (D), whereas a higher score is classified as tolerated (T).
[f]Polyphen2 (Polymorphism Phenotyping v2) score ranges from 0 to 1. Value implies probably damaging" ("D") for scores in (0.957, 1); "possibly damaging" ("D") for scores in (0.453, 0.956); "benign" ("B") for scores in (0, 0.452)

We observed well-calibrated test statistics and little evidence of inflation (Fig. 5.3, lambda = 0.98) for the exome-wide association analysis. To explore the effect of infections and diagnosed chronic diseases on the top signals, we compared the distribution of bacteria, viral, and fungal infections, as well as 14 diseases between carriers and non-carriers of rare variants (Table 5.6, Table 5.7). After multiple testing correction, no significant differences were observed, implying that the top signals are less likely driven by infection or comorbidity. To explore the allele

frequency (AF) of variants contributing to significant and suggestive genes, we compared the AF of all variants across three populations (Black, non-finnish European, and Asian) in the gnomAD database (Table 5.8). No large difference was observed between the AFs across populations, except one SNP rs370696272 with an allele frequency higher than 0.01 in Blacks. To further investigate if there is an ancestry difference driving the top signals, we conducted a sensitivity analysis by randomly selecting the same number of samples as PedSep-D group while keeping the same ancestry distribution, and labeling selected samples as "PedSep-D". This approach enables swapping the phenotype labels with similar ancestry information to keep the ancestry makeup of the groups the same while generating a meaningful empirical p-value. With 100,000 iterations of permutation for each of the three genes, we observed 0 times that permutated statistics were larger than the previously estimated statistic. This further indicates the significance of the three genes is not likely driven by ancestry differences.

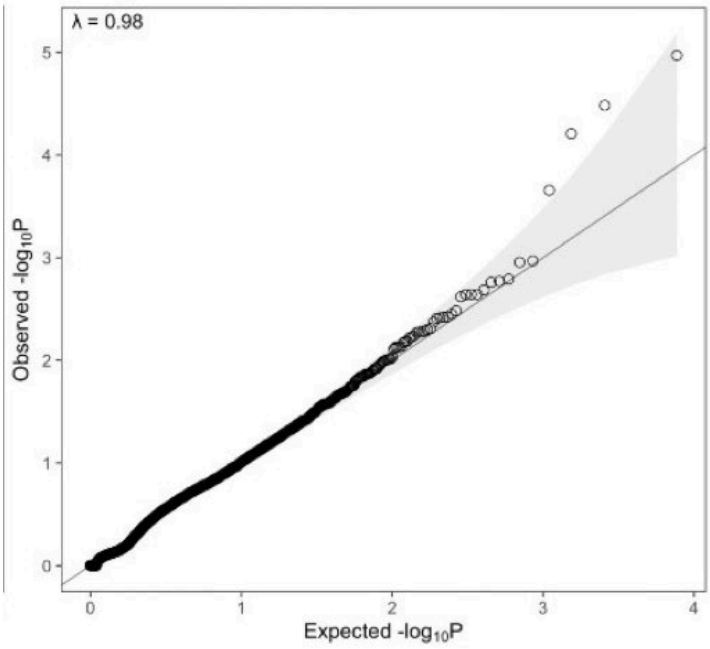

Figure 5.3 QQ plot

Table 5.6 Characteristics of rare variant carriers and non-carriers

characteristic	LTBP4		PLA2G4E		CCDC157	
	carriers	Non-carriers	carriers	Non-carriers	carriers	Non-carriers
No. of Patients	8	311	6	313	17	302
Age, median (IQR), y	8 (5, 12)	5 (1, 12)	1 (1, 2)	5 (1, 12)*	3 (2, 10)	5 (1, 12)
Sex, N (%)						
Female	3 (37.5)	141 (45.3)	4 (66.7)	140 (44.7)	7 (41.1)	137 (45.4)
Male	5 (62.5)	170 (54.7)	2 (33.3)	173 (55.3)	10 (58.9)	165 (54.6)
Race, N (%)						
White	3 (37.0)	210 (67.5)	5 (83.3)	208 (66.5)	12 (70.6)	201 (66.6)
Black	4 (50.0)	63 (20.3)	1 (16.7)	67 (21.4)	4 (23.5)	63 (20.9)
Asian	0 (0.0)	14 (4.5)	0 (0.0)	14 (4.5)	1 (5.9)	14 (4.6)
Other	1 (13.0)	24 (7.7)	0 (0.0)	24 (7.7)	0 (0.0)	24 (7.9)
Ethnicity, N (%)						
Non-Hispanic	6 (75.0)	250 (80.4)	6 (100.0)	250 (79.9)	13 (76.5)	243 (80.5)
Hispanic	2 (25.0)	48 (15.4)	0 (0.0)	50 (16.0)	2 (11.8)	48 (15.9)
Unknown	0 (0.0)	13 (4.2)	0 (0.0)	13 (4.2)	2 (11.8)	11 (3.6)
Previous healthy	5 (62.5)	148 (47.6)	3 (50.0)	150 (47.9)	8 (47.1)	145 (48.0)
Immunocompromised, N (%)	1 (12.5)	58 (18.6)	1 (16.7)	58 (18.5)	3 (17.6)	56 (18.5)
PRISM Score, median (IQR)	8.5 (7.25, 15.75)	8 (3, 15)	18.5 (15.75, 19.00)*	8 (3, 14)	10 (3, 15)	8 (3, 15)
OFI, median (IQR)	2.5 (2, 3)*	2 (1, 2)	2 (1, 3.75)	2 (1, 2)	2 (2, 3)*	2 (1, 2)

Infection, N (%)						
Bacterial infection	1 (12.5)	113 (36.3)	1 (16.7)	113 (36.1)	7 (41.1)	107 (35.4)
Viral infection	0 (0.0)	88 (28.3)	2 (33.3)	86 (27.5)	2 (11.8)	86 (28.5)
Fungal infection	0 (0.0)	2 (0.6)	0 (0.0)	2 (0.6)	0 (0.0)	2 (0.7)
No infection	7 (87.5)	108 (34.7)	3 (50.0)	110 (35.1)	8 (47.1)	107 (35.4)
Diagnosis, N (%)						
Leukemia	0 (0.0)	11 (3.5)	1 (16.7)	10 (3.2)	1 (5.9)	10 (3.3)
Hemo Anemia	0 (0.0)	2 (0.6)	1 (16.7)*	1 (0.3)	1 (5.9)	1 (0.3)
Rheuma Disease	0 (0.0)	7 (2.3)	0 (0.0)	7 (2.2)	1 (5.9)	6 (2.0)
IBD	1 (12.5)	2 (0.6)	0 (0.0)	3 (1.0)	0 (0.0)	3 (1.0)
RenalDisease	0 (0.0)	5 (1.6)	0 (0.0)	5 (1.6)	0 (0.0)	5 (1.7)
ChromAbnormal	2 (25.0)	45 (14.5)	1 (16.7)	46 (14.7)	4 (23.5)	43 (14.2)
MetabolicDisease	0 (0.0)	12 (3.9)	0 (0.0)	12 (3.8)	1 (5.9)	11 (3.6)
Diabetes	1 (12.5)	3 (1.0)	0 (0.0)	4 (1.3)	0 (0.0)	4 (1.3)
CardiovascularDisease	1 (12.5)	49 (15.8)	0 (0.0)	50 (16.0)	4 (23.5)	46 (15.2)
Cardio_PostOP	1 (12.5)	15 (4.8)	0 (0.0)	16 (5.1)	2 (11.8)	14 (4.6)
Trauma	0 (0.0)	3 (1.0)	0 (0.0)	3 (1.0)	0 (0.0)	3 (1.0)
ShortGut	0 (0.0)	7 (2.3)	0 (0.0)	7 (2.2)	1 (5.9)	6 (2.0)
LiverDisease	0 (0.0)	9 (2.9)	1 (16.7)	8 (2.6)	1 (5.9)	8 (2.6)
Bronchiolitis	0 (0.0)	9 (2.9)	0 (0.0)	9 (2.9)	0 (0.0)	9 (3.0)
Mortality, N (%)	2 (25.0)	26 (8.4)	1 (16.7)	27 (8.6)	1 (5.9)	27 (8.9)

*: The value in this group is significantly higher than the compared group ($p < 0.05$). Comparisons were performed using the Kruskal-Wallis test for continuous variables, the $\chi 2$ test for categorical variables, or the Fisher's exact test for cells with less than 5 patients.

Table 5.7 P-values comparing rare variant carriers with non-carriers

	LTBP4	**PLA2G4E**	**CCDC157**
Age, median (IQR), y	0.453	0.042	0.673
Sex, N (%)	0.936	0.512	0.931
Race, N (%)	0.177	0.491	0.809
Ethnicity, N (%)	0.664	0.471	0.245
Previous healthy	0.635	1.000	1.000
Immunocompromised, N (%)	1.000	1.000	1.000
PRISM Score, median (IQR)	0.380	0.021	0.960
OFI, median (IQR)	0.016	0.384	0.009
Infection, N (%)	0.083	0.321	0.400
Mortality, N (%)	0.150	0.426	1.000

The tests were performed using the Kruskal–Wallis test, the $\chi 2$ test, or the Fisher's exact test.

Table 5.8 Single variant association and allele frequency for variants contributing to the gene-level significance

Gene	Variant	SNP Information[a]	Allele Frequency in PedSep-D[b]			Allele Frequency in non-PedSep-D[c]			Allele Frequency in population[d]		
			Black (n = 7)	White (n =28)	Asian (n = 3)	Black (n = 60)	White (n = 185)	Asian (n = 11)	Black	White	Asian
LTBP4	rs370696272	19:41105311:C:T	0.429 (3/7)	0.071 (2/28)	0	0	0	0	0.01187	0.00005504	0.000
	rs573310430	19:41122842:C:T	0.143 (1/7)	0	0	0	0	0	0.00004134	0.000007788	0.000
	-	19:41132970:C:T	0	0	0	0	0.005 (1/185)	0	-	-	-
	rs200607327	19:41133005:G:A	0	0	0	0	0.005 (1/185)	0	0.00004143	0.0006489	0.000
PLA2G4E	-	15:42276733:T:G	0	0	0	0	0.005 (1/185)	0	-	-	-
	rs764494895	15:42278161:G:A	0	0	0	0	0.005 (1/185)	0	0.000	0.00006737	0.000
	rs143966595	15:42293394:C:T	0	0.107 (3/28)	0	0	0	0	0.0001654	0.0008021	0.000
	rs776016335	15:42298270:T:C	0	0.036 (1/28)	0	0	0	0	0.000	0.000	0.00005561
CCDC157	rs9606721	22:30762035:A:G	0	0.036 (1/28)	0	0	0.011 (2/185)	0	0.003390	0.02384	0.000
	rs540507025	22:30762080:C:T	0	0	0	0	0.005 (1/185)	0	0.000	0.00002669	0.0004897
	rs143249037	22:30766366:G:A	0	0	0	0	0.005 (1/185)	0	0.00004008	0.0007048	0.001535
	-	22:30766438:C:A	0	0.143 (4/28)	0	0.017 (1/60)	0.011 (2/185)	0	-	-	-
	rs139609945	22:30766496:C:T	0	0	0	0.017 (1/60)	0	0	0.000	0.0001055	0.000
	rs1235664314	22:30766672:G:T	0	0	0	0.017 (1/60)	0	0	0.0001147	0.000	0.000
	rs148283823	22:30766868:G:A	0	0.036 (1/28)	0	0	0	0	0.006199	0.00008130	0.000
	rs202178544	22:30772567:T:C	0	0	0	0.033 (2/60)	0	0	0.0004134	0.000007778	0.0001507

[a] SNPs are listed as chromosome: position (hg19): reference allele: alternative allele.
[b] The total number of patients in PedSep-D is 40, including 2 patients with reported race as unknown.
[c] The total number of patients in PedSep-D is 279, including 28 patients with reported race as unknown.
[d] Allele frequency of three populations according to gnomAD database (v4.0.0).

5.4.3 Rare variants of *LTBP4* are associated with plasma protein levels of *TGF-β*

Aiming to investigate the function of top signals found from the gene-based analysis, we measured day-one plasma protein level (reported as relative fluorescence) of *LTBP4* and *TGF-β*, a crucial cytokine regulated by the *LTBP4* gene, from 110 patients of the same cohort. The 110 patients include six *LTBP4* rare variants carriers and nine *CCDC157* rare variants carriers. By comparing protein levels of *LTBP4* and *TGF-β* between carriers and non-carriers, we discovered a significant association between *LTBP4* rare variant carriers and an increased plasma protein level of *TGF-β* (Figure 5.4A). No significant association was found between *LTBP4* rare variant carriers and the plasma protein level of *LTBP4* (Figure 5.4B). No significant associations were found between plasma protein and *CCDC157* rare variant carriers (Figure 5.4C and Figure 5.4D).

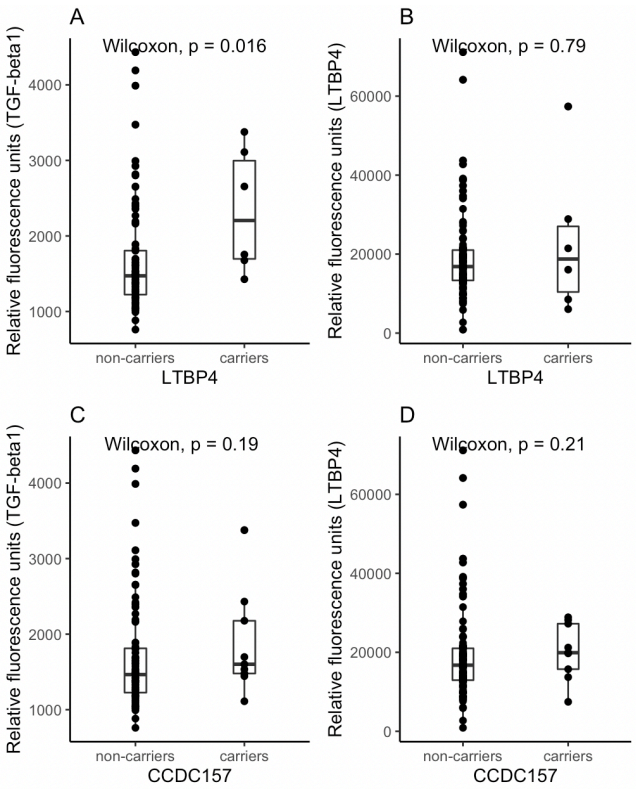

Figure 5.4 Comparison of plasma protein relative fluorescence of TGF-β1 and LTBP4 between rare variant carriers and non-carriers

5.4.4 *LTBP4*, *PLA2G4E*, and *CCDC157* underlie distinct serum cytokine patterns in patients

To explore the genes' association with inflammation status, we grouped patients based on whether they carried rare variants in one of the three genes of interest, generated a heatmap showing the normalized levels of 33 cytokines (Figure 5.5), and statistically tested the group-wise differences (Table 5.9, Table 5.10). Comparison between the rare variant carriers and non-carriers indicated some similarity shared by carrier groups. For example, a higher level of IL-6 is significantly related to both *LTBP4* and *CCDC157* rare variant carriers (fdr = 0.032 and 0.043, respectively), and higher level of M-CSF is significantly related to both *PLA2G4E* and *CCDC157* rare variant carriers (fdr = 0.013 and 0.011 separately). Simultaneously, several cytokines

important in regulating inflammation showed distinct patterns when comparing carriers to non-carriers with rare variants in *LTBP4*, *PLA2G4E,* and *CCDC157.* For instance, IL-4 is significantly higher in *LTBP4* rare variant carriers (fdr = 0.025) but is significantly lower in *CCDC157* rare variant carriers (fdr = 0.035). Compared to non-carriers, *PLA2G4E* rare variant carriers presented significantly higher levels of IL-16, and SCF, but significantly lower levels of CRP (fdr = 0.034, 0.005, and 0.022, separately), while *LTBP4* and *CCDC157* rare variant carrier groups showed no significant difference with non-carriers for these biomarkers. The ferritin level is uniquely higher in *LTBP4* rare variant carriers (fdr = 0.023). These results imply that the three genes might be involved in different pathological mechanisms driving the phenotype.

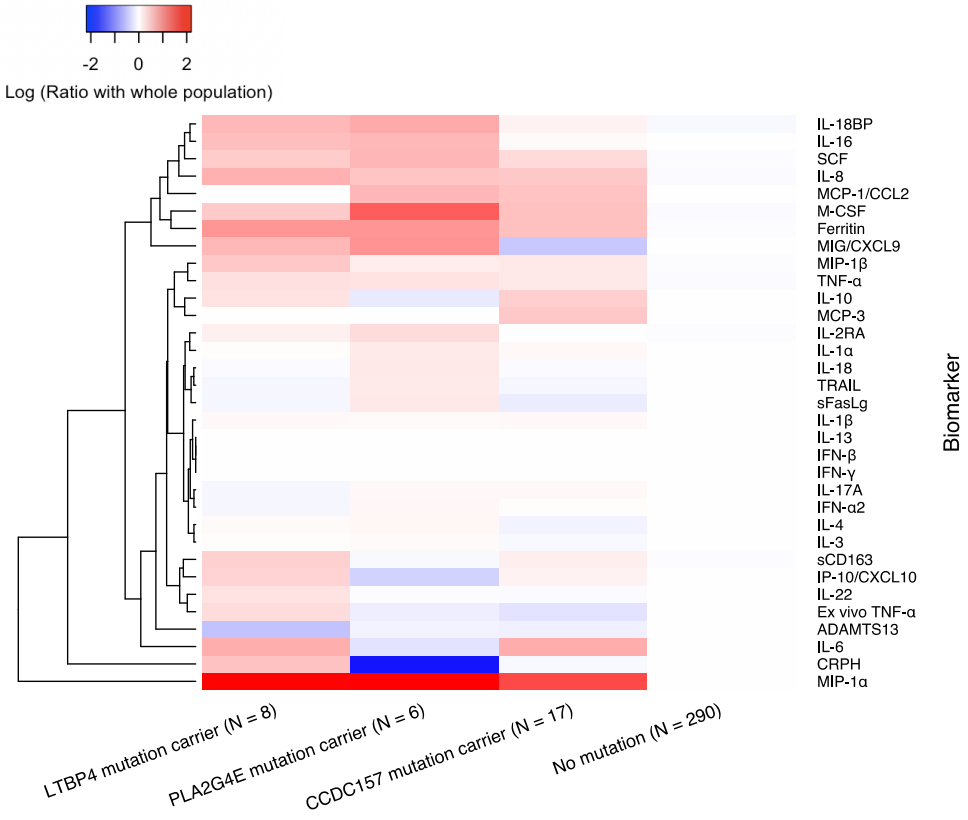

Figure 5.5 Biomarker heatmap of three genes' carriers and non-carriers
The log ratio of the median values of 33 inflammatory biomarkers by rare variants carriers and non-carriers. Red represents a greater median biomarker value for that group compared with the median for the entire study cohort, whereas blue represents a lower median biomarker value compared with the median for the entire study cohort.

Table 5.9 Median cytokine levels of rare variant carriers and non-carriers

Biomarker	LTBP4		PLA2G4E		CCDC157	
	carriers	Non-carriers	carriers	No-carriers	carriers	Non-carriers
ADAMTS13, %	42.50	71.50*	64.50	71.00	61.00	72.00
SFasLg, pg/ml	44.21	47.92	58.03	47.30	42.47	48.12
Ex vivo TNF-α, pg/ml	658.97	484.30	420.97	484.30	312.50	490.30
TNF-α, pg/ml	101.20	74.90	97.75	74.90	93.55	74.90
sCD163, pg/ml	425278.00	280412.00	265204.00	283880.00	327821.00	280412.00
IFN-β, pg/ml	6.40	6.40	6.40	6.40	6.40	6.40
IL-22, pg/ml	31.85	24.80	24.20	25.40	23.60	25.40
IL-18, pg/ml	397.90	411.20	504.40	408.90	398.00	412.80
IL-18BP, pg/ml	29541.00	15477.00	33202.00	15751.00	17799.00	15751.00
MIG/CXCL9, pg/ml	1482.00	779.00	2035.80	791.70	501.60	807.60
IL-1β, pg/ml	2.95	2.80	2.90	2.80	2.95	2.80
IL-4, pg/ml	4.90*	4.70	5.05	4.70	4.30	4.70*
IL-6, pg/ml	16.60	8.40	6.65	8.60	17.10*	8.40
IL-8, pg/ml	99.80	49.80	83.15	50.60	81.55*	49.40
IL-10, pg/ml	27.40	21.70	18.25	21.70	33.10	21.70
IL-13, pg/ml	3.10	3.10	3.10	3.10	3.10	3.10
IL-17A, pg/ml	16.95	18.70	19.55	18.30	19.55	18.30
IFN-γ, pg/ml	2.80	2.80	2.80	2.80	2.80	2.80
IP-10/CXCL10, pg/ml	1042.50	716.70	492.60	753.30	819.20	726.60
MCP-1/CCL2, pg/ml	131.90	133.30	248.30	133.35	220.80	131.55
MIP-1α, pg/ml	5.05	0.60	5.30	0.60	2.85	0.60
MIP-1β, pg/ml	73.25	45.10	53.85	45.45	55.30	45.10
MCP-3, pg/ml	92.40	92.40	92.40	92.40	147.80	92.40
IFN-α2, pg/ml	115.80	125.70	135.60	125.70	128.30	125.70
IL-1α, pg/ml	9.65	9.40	11.30	9.40	9.90	9.40
IL-2RA, pg/ml	428.00	371.80	503.60	367.90	380.30	373.60
IL-3, pg/ml	624.40	612.20	636.00	612.20	572.60	612.20
IL-16, pg/ml	1015.20	575.90	1055.30*	572.90	607.60	575.90
M-CSF, pg/ml	46.15	28.70	117.25*	28.70	50.35*	28.10
SCF, pg/ml	237.40	151.50	287.40*	151.50	210.70	151.50
TRAIL, pg/ml	35.40	37.90	45.40	37.90	35.40	37.90
CRP, mg/dL	16.58	9.73	1.34	10.11*	9.24	10.02
Ferritin, ng/mL	463.90*	187.00	463.50	187.70	321.60	187.00

* Significant higher value of cytokine in the group. Kruskal-Wallis tests were performed.

Table 5.10 Comparison of FDR of tests comparing cytokines between rare variant carriers and non-carriers

Biomarker	LTBP4	PLA2G4E	CCDC157
ADAMTS13, %	0.015	0.95	0.06
SFasLg, pg/ml	0.624	0.284	0.574
Ex vivo TNF-α, pg/ml	0.664	0.869	0.738
TNF-α, pg/ml	0.119	0.937	0.343
sCD163, pg/ml	0.812	0.703	0.852
IFN-β, pg/ml	0.131	0.534	0.22
IL-22, pg/ml	0.46	0.63	0.86
IL-18, pg/ml	0.903	0.25	0.767
IL-18BP, pg/ml	0.671	0.493	0.983
MIG/CXCL9, pg/ml	0.413	0.687	0.088
IL-1β, pg/ml	0.667	0.86	0.476
IL-4, pg/ml	0.025	0.186	0.035
IL-6, pg/ml	0.032	0.172	0.043
IL-8, pg/ml	0.286	0.396	0.044
IL-10, pg/ml	0.486	0.085	0.567
IL-13, pg/ml	0.315	0.801	0.952
IL-17A, pg/ml	0.097	0.152	0.089
IFN-γ, pg/ml	0.142	0.309	0.415
IP-10/CXCL10, pg/ml	0.715	0.319	0.059
MCP-1/CCL2, pg/ml	0.193	0.267	0.312
MIP-1α, pg/ml	0.087	0.361	0.184
MIP-1β, pg/ml	0.059	0.088	0.513
MCP-3, pg/ml	0.709	0.695	0.163
IFN-α2, pg/ml	0.418	0.73	0.94
IL-1α, pg/ml	0.846	0.741	0.287
IL-2RA, pg/ml	0.677	0.065	0.839
IL-3, pg/ml	0.825	0.819	0.24
IL-16, pg/ml	0.163	0.034	0.694
M-CSF, pg/ml	0.136	0.013	0.011
SCF, pg/ml	0.516	0.005	0.162
TRAIL, pg/ml	0.501	0.437	0.796
CRP, mg/dL	0.189	0.022	0.658
Ferritin, ng/mL	0.023	0.07	0.652

Kruskal-Wallis tests were performed.

In addition, GTEx tissue-specific expression analysis showed that the three genes are expressed in diverse tissues (Figure 5.6A). *LTBP4* is highly expressed in multiple tissues, including the lung, kidney, stomach, skin, and others. *PLA2G4E* is specifically expressed in the skin. *CCDC157* is specifically expressed in the testis. In terms of cell-type specific expression of three genes, all of them displayed expression in immune cells of the cardiovascular and pulmonary systems, both of which are highly affected by severe sepsis or septic shock (Figure 5.6B).

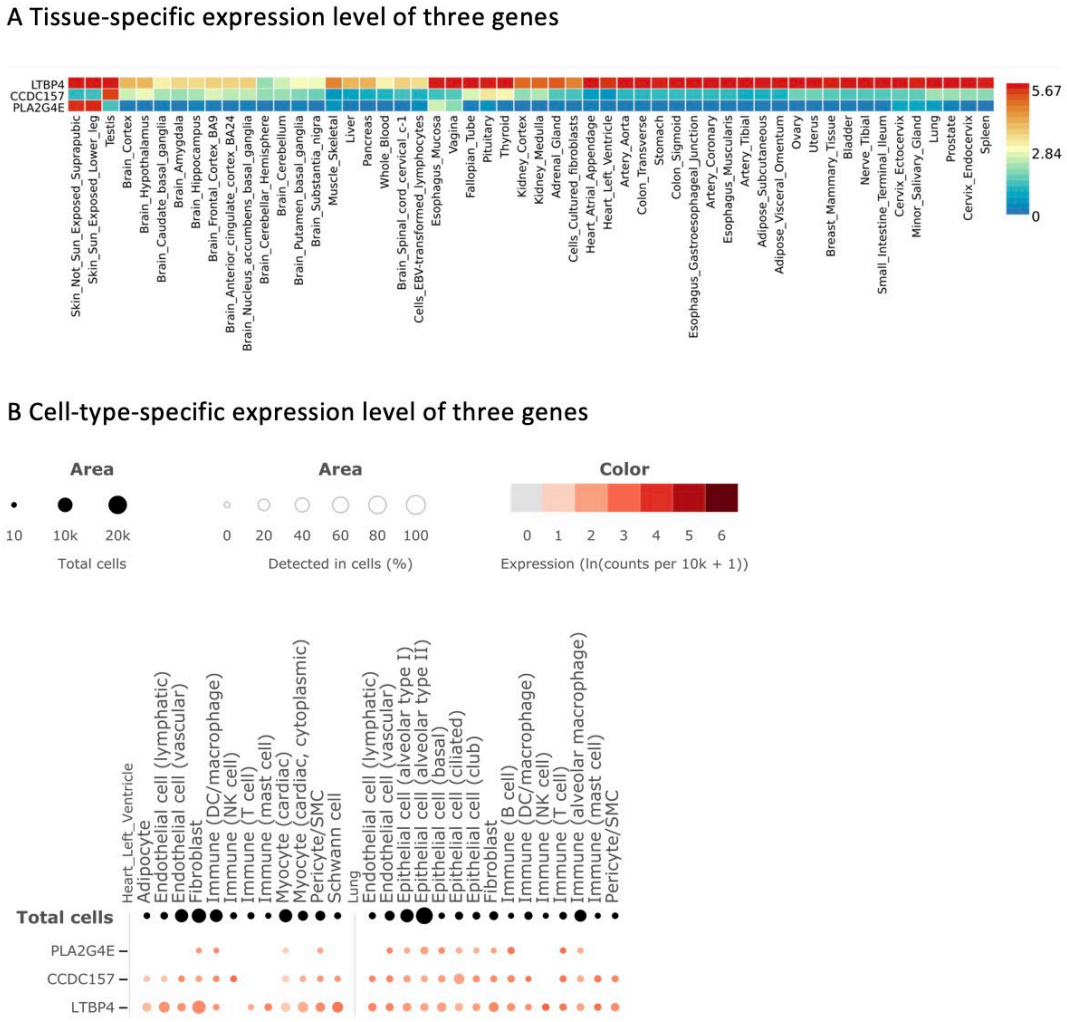

Figure 5.6 Expression of three genes in different tissues and cell types in public database (GTEx)
A. Heat map of tissue-specific log2 transformed average expression level for three genes based on GTEx v8 RNA data. Red color indicates a higher expression within a tissue compared to other tissues, whereas blue color indicates a lower expression within a tissue compared to other tissues. B. Dot plot of cell-type-specific expression level for

three genes in two sepsis-related tissues (GTEx Single Cell data). The dot color reports the mean expression value with each cell type. The dot size reports the fraction of cells in which a gene is detected. Black dot size reports the total detected level of all three genes.

5.5 Discussion

In this study, rare variants in *LTBP4* were significantly associated with the previously reported high-mortality PedSep-D phenotype, with additional suggestions of associations with rare variants in *PLA2G4E* and *CCDC157*. To our knowledge, this is the first time a rare variant burden test has been applied to pediatric sepsis with deep phenotyping.

The top signal found in our study, *LTBP4*, a member of the latent transforming growth factor β binding protein family, shares structural homology with fibrillin and is moderately expressed in plasma cells and immune cells (Su and Urban 2021). Mutations in *LTBP4* have been associated with autosomal recessive cutis laxa type 1C (Urban et al. 2009, Zhang et al. 2020, Mazaheri et al. 2022), Duchenne Muscular Dystrophy (DMD) (Kosac et al. 2022), fibrosis-related disorders (Lu et al. 2017), cancer (Li et al. 2020), pulmonary disorders, and cardiovascular disorders (Rocchiccioli et al. 2017). PedSep-D patients had the most severe kidney involvement, whereas *LTBP4* was found to protect against tubular interstitial fibrosis by strengthening angiogenesis, downregulating inflammatory gene expression, and facilitating the maintenance of mitochondrial structure in tubular epithelial cells (Su et al. 2021). Common variants in *LTBP4* have previously been reported in GWASs to be associated with several traits, including lung function (FEV1/FVC) (Shrine et al. 2023), peak expiratory flow (Shrine et al. 2019), hematocrit (Vuckovic et al. 2020), hemoglobin (Vuckovic et al. 2020), eosinophil counts (Chen et al. 2020), carotid intima-media thickness (Yeung et al. 2022), and diastolic blood pressure (Plotnikov et al. 2022).

Notably, *LTBP4* plays a regulatory role in the signaling of transforming growth factor β (*TGF-β*). Mouse studies found that gain of function in *LTBP4* increases activation of *TGF-β*, whereas loss of function in *LTBP4* decreases activation of *TGF-β* (Rifkin et al. 2022). In addition, *LTBP4* was shown to bind and stabilize *TGF-β* receptor 2 (TGFBR2), thus increasing *TGF-β* signaling (Su et al. 2015). *TGF-β* is a pleiotropic cytokine with crucial immunoregulatory properties. It plays an essential role in the pathogenesis of chronic fibrosis in diverse tissues, such as the lung, kidney, liver, and skin (Su and Urban 2021). Moreover, *TGF-β* is a multi-faceted cytokine with potent regulatory and inflammatory activity in response to infections. Although its function is still under discussion, several studies showed that *TGF-β* overexpression contributes to a wide array of metabolic disorders and dysfunction. Furthermore, overexpressed *TGF-β* triggers epithelial-mesenchymal transition and the overaccumulation of extracellular matrix (ECM), leading to immune dysfunction, fibrosis, and the development of cancers (Peng et al. 2022). In the field of sepsis, a couple of studies found elevated *TGF-β* levels in septic animal models and human patients (Ayala et al. 1993, Ahmad et al. 1997, Garcia-Lazaro et al. 2005, Huang et al. 2010, de Pablo et al. 2012, Xu et al. 2013, Nullens et al. 2018). Notably, Huang et al. examined serum levels of *TGF-β* in 106 burned, septic, and non-septic patients and revealed a significant increase of *TGF-β* levels in septic patients compared to non-septic patients, as well as in non-survivors compared to survivors, indicating the potential role of *TGF-β* in sepsis susceptibility and outcome (Huang et al. 2010). In addition to these findings, our findings unravel the relationships between sepsis-severity-related deleterious rare variants of LTBP4 and *TGF-β* protein level, serving as a functional validation of our gene-based analysis.

As one of the two genes with suggestive significance, the *PLA2G4E* gene encodes a member of the cytosolic phospholipase A2 group IV family involved in membrane tubule-

mediated transport regulation. It plays an important role in trafficking through the clathrin-independent endocytic pathway (Capestrano et al. 2014). *PLA2G4E* was also up-regulated in Alzheimer's disease APP-PS1 transgenic mice lacking CD8 T cells compared to the control group (Perez-Gonzalez et al. 2020). Common variants in *PLA2G4E* have been reported in previous GWASs to be associated with several sepsis clinical prognostic factors such as neutrophil count (Chen et al. 2020, Kachuri et al. 2021, Sakaue et al. 2021), white blood cell count (Astle et al. 2016), and mean platelet volume (Chen et al. 2020). As another gene displaying suggestive significance, *CCDC157* encodes a protein coiled-coil domain containing 157. Common variants in *CCDC157* have been reported in previous GWASs to be associated with sepsis risk factors such as hematocrit (Vuckovic et al. 2020), pulse pressure (Evangelou et al. 2018), and calcium levels (Sakaue et al. 2021). No clear function of immune dysregulation has been reported for *CCDC157* to date.

There are several limitations in this study. First, the tested sample size is small with 40 samples in the PedSep-D phenotype, limiting the power for detecting associations. Therefore, larger independent cohorts are required for result validation and meta-analysis purposes. Second, the heterogeneity of the non-PedSep-D group could limit the interpretability of the results. Third, only a small proportion of genes (n = 3846) were tested because there weren't enough rare variants present in the others. Fourth, due to the usage of SKAT, this will have low power when all the rare variants have effects in the same direction. Fifth, although we observed no clear relevance between previous diagnoses of 14 diseases and rare variant carriers of top genes, the possible confounding effect of comorbidity cannot be fully ruled out. Sixth, although we account for global ancestry by adjusting for top PCs in the association test, the signals from rare variants may be caused by ancestry differences. Seventh, although rare variant carriers exhibit elevated cytokine levels, such

as M-CSF, this may not conclusively indicate the impact of rare variants on cytokines, especially since M-CSF levels are significantly higher in the PedSep-D group compared to other groups. Additionally, although we utilized plasma protein to investigate potential functional consequences of identified rare variants, other functional analysis utilizing GTEx is based on public adult data which may fail to represent the expression of genes in pediatric sepsis patients. Thus, transcriptomic and other various types of data has to be collected from children with sepsis to corroborate our findings through QTL analyses and other functional genomics approaches (Rodenburg 2018). In summary, our pilot study identified rare variants that, if found to have functional effects in future studies, might play a role in pointing towards some possible mechanisms which may underlie pediatric septic patients at high risk.

6.0 Unravel Methylation Markers of Pediatric Sepsis Phenotype at High Risk through an Epigenome-wide Association Study (EWAS)

6.1 Introduction

Pediatric sepsis is a multifaceted disease leading to 1.2 million worldwide cases in children per year, with a mortality rate ranging from 1% to 5% (Massaud-Ribeiro et al. 2022). Resulting in life-threatening organ failure, the syndrome is characterized by dysregulation of immune response to infection. While significant progress has been made in regular disease management of the disease (Weiss et al. 2020), the heterogeneity in environmental exposures and host immune responses have hindered efforts to develop targeted therapies for pediatric sepsis patients (Atreya and Wong 2019). Acknowledging this heterogeneity, our recent published study reported four computational pediatric sepsis phenotypes, PedSep-A, B, C, and D, derived from a multicenter cohort through applying machine learning approaches to 25 first-day bedside clinical features (Qin et al. 2022). Comparing the four identified phenotypes, we observed reduced heterogeneity within phenotypes and differences between phenotypes in terms of their infection resources and sites, cytokine patterns, 28-day organ failures and mortalities, and therapeutic responses. These findings indicated that the diversity in environmental factors and the host response significantly influenced the severity of the disease. For example, children with the PedSep-D phenotype exhibits a high risk of abnormal inflammatory activity, leading to a subsequent higher organ failure and mortality rates. To improve our comprehension of the disease pathology and formulate precise treatment strategies for the high-risk patients, one of the logical next steps is to investigate the interplay between environmental and host biological factors which contribute to the disease severity.

In this context, there is a growing interest in elucidating the role of epigenetics in pediatric sepsis heterogeneity. This interest arises from the fact that epigenetic processes, such as DNA methylation, show plasticity in response to exogenous environment and govern gene expression without changing the DNA. Since pediatric sepsis heterogeneity is greatly affected by exogenous environment and gene expression changes (Mohammed et al. 2019, Yang et al. 2023), DNA methylation level analysis is expected to serve as a pivotal source of biological factors deserving thorough investigation. Moreover, array- and sequencing-based methylation profiling technologies enable large-scale epigenome-wide association studies (EWASs) to identify differentially methylated cytosine-phosphate-guanine (CpG) sites (DMCs) or regions (DMRs), thereby pinpointing the potential targets for epigenetic therapies (Dhas et al. 2015, Lorente-Pozo et al. 2021).

Despite the efforts of previous EWAS analyses for pediatric sepsis, there are several limitations that remain unattended. First, existing pediatric sepsis methylation studies are restricted by limited cohort size compared to adult sepsis. For instance, the EWAS study conducted by Binnie and colleagues recently found hundreds of DMR distinguishing adult septic patients and non-septic controls (Binnie et al. 2020), while the two published pediatric studies until now separately included 3 and 17 pediatric sepsis patients (Dhas et al. 2015, Lorente-Pozo et al. 2021). Thereby, the small size of the pediatric sepsis cohort constrained the ability to detect significant associations. The second gap to be filled involves revealing methylation signals related to pediatric sepsis patients at high risk. The majority of studies were designed for comparing pediatric sepsis patients with healthy controls, overlooking the potential pathological diversity within the patient population. Although one prior study of Lorente-Pozo et al identified DMRs distinguishing early

and late-onset neonatal sepsis, they focused on comparing mother- and self-acquiring infections of infants instead of pediatric septic populations with low and high-risk (Lorente-Pozo et al. 2021).

To tackle these deficiencies, we aim to investigate the role of the methylome in pediatric sepsis severity by performing an epigenome-wide association study on blood DNA methylation data from a multicenter cohort of pediatric sepsis (PHENOMS) (Carcillo et al. 2019). As the largest pediatric sepsis EWAS study to date, we hope to bring more power to detect novel signals in addition to the previous studies. By targeting an agnostic pediatric sepsis phenotype with high risk, PedSep-D, and comparing it with the other non-PedSep-D phenotypes, we expect to provide insights into pathology of disease severity. Consequently, this study could aid in identifying new diagnostic and prognostic biomarkers differentiating pediatric sepsis patients at risk and advancing precision medicine.

6.2 Methods

6.2.1 Agonistic pediatric sepsis phenotype generation

The pediatric sepsis phenotypes were derived using an unsupervised machine-learning approach as described in our previously published study (Qin et al. 2022). In general, we applied consensus k-means clustering to 404 children from the PHENOMS cohort and derived four computable phenotypes, PedSep-A, B, C, and D using 25 day-one bedside clinical features. The PedSep phenotype of each patient was extracted and used as the phenotype of interest in this study.

6.2.2 Methylation data preprocessing

The 850K DNA methylation microarray profiles (Illumina Infinium Methylation EPIC array) were collected from PBMC of 96 patients from the 404 children in the PHENOMS cohort (Fig. 6.1). The *minfi* r package was used to perform normalization and quality control (Aryee et al. 2014). Specifically, we first performed a series of quality control steps on the methylation data. Patient samples were excluded if the average detection p-value of all probes across the sample is greater than 0.01. Probes with the following criteria were excluded: (1) detection p-value of greater than 0.01 in one or more samples; (2) bead count lower than 3; (3) probes containing single nucleotide polymorphism-introduced artifacts and in cross-reactive regions (Pidsley et al. 2016); and (4) probes mapped to sex chromosomes (Appendix Table 9). Then, we normalized the data by Quantile normalization function to remove artificial batch effects. Finally, to fit the linear regression models with normally distributed dependent variable (i.e., methylation level of CpGs), we logit-transformed β values to M values, where M value measures the intensity of methylation of a CpG site.

6.2.3 Single-site-based differentially methylated CpGs analysis

To identify differentially methylated CpGs (DMCs) differentiating previously derived PedSep-D phenotype from non-PedSep-D phenotype, we fitted linear regression models using the *cate* r package. The *cate* package eliminates undesired variation while accounting for known variables in modeling and carries out a high dimensional factor analysis and confounder-adjusted multiple testing (Wang et al. 2017). Models compared methylation levels of PedSep-D and non-PedSep-D phenotype of pediatric sepsis patients, adjusting for the child's age, sex, EPIC array

chip, position of samples on the chip, and five cell type proportions (not including neutrophils to avoid collinearity) estimated by the *FlowSorted.Blood.EPIC* r package. The *FlowSorted.Blood.EPIC* r package utilizes blood cell reference data to deconvolute EPIC bulk methylation data and provide proportion estimation for six cell types, including T lymphocytes (CD4+ and CD8+), B cells (CD19+), monocytes (CD14+), NK cell (CD56+), and Neutrophils (Salas et al. 2018).

To minimize false positive rate while preserving sufficient power, we employed a Bayesian method implemented in the *BACON* r package to correct test-statistic bias and inflation in the association study. By taking advantage of prior knowledge of the distribution and the composition of test statistics, the *BACON* method has demonstrated superior performance in empirical null distribution estimation compared to other existing methods (Bird 2002). It also yielded the highest power when being used in combination with *cate*. QQ plot and inflation factor derived by the *BACON* method were used to visualize and quantify inflation.

For the identification of significant findings, we used a p-value of 9×10^{-8} as the whole-methylome-wide significance threshold and a p-value of 1×10^{-5} as the suggestive significance threshold following the guidance to control the false positive rate for 850K DNA methylation array data (Mansell et al. 2019). The annotations of the reported CpGs were obtained from the UCSC genome browser (hg19).

6.2.4 Region-based differentially methylated region analysis

Intending to discover differentially methylated regions (DMRs) where multiple methylation probes are consistently associated with the PedSep-D phenotype, we employed the *dmrff* r package to combine EWAS summary statistics from nearby sites while taking the

correlation between sites into consideration (Suderman et al. 2018). EWAS results derived from DMCs analysis were directly input into dmrff. P-values for each region were Bonferroni-adjusted for multiple tests. Results were visualized with a volcano plot. The top probes and regions were queried on the EWAS Catalog for relevant genes and functions. The annotations of the reported CpGs were obtained from the UCSC genome browser (hg19).

6.2.5 Cell-type-specific differential methylation analysis

To investigate whether differentially methylated CpGs (DMCs) identified in bulk methylation data exhibit hypermethylation or hypomethylation within distinct cell types, we applied the *EpiDISH* r package for cell-type-specific differential methylation analysis (Zheng et al. 2018). This analysis was based on estimated cell-type proportions obtained through blood cell-type deconvolution. Specifically, we used the *CellDMC* function of the *EpiDISH* package to examine the association of the DMCs with the PedSep-D/non-PedSep-D phenotype while adjusting for age, gender, ethnicity, batch effect, cell-type-proportions, and proportion-phenotype interactions to pinpoint cell-type-specific DMCs. The model was solved by least squares with the *lm* function and provides estimated coefficients and statistical significance via p-value for each tested CpG. The p-values were further adjusted for multiple testing using the Benjamini-Hochberg (BH) False Discovery Rate (FDR) approach.

6.2.6 Functional downstream analysis

To explore the function of DMCs identified from the association analysis with public databases, we first queried the DMCs in EWAS catalog to explore previously reported associations

with immunology traits (Battram et al. 2022). Then, we investigated the enrichment of the DMCs nearby genes in GTEx across 30 tissue types (Consortium 2013). We performed biological pathway analysis based on Gene Ontology (GO) (Ashburner et al. 2000, Gene Ontology et al. 2023), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000), and TRANSFAC (TF) (Matys et al. 2006) by using the *gprofiler2* r package (Kolberg et al. 2020). Additionally, we examined the association of the top DMCs with transcription of nearby genes using publicly available cis expression quantitative trait methylation (cis-eQTMs) from 823 children's blood in the HELIX Project (Ruiz-Arenas et al. 2022).

6.2.7 Correlation between methylation and cytokine levels

We investigated relationships between top signals and inflammation by conducting correlation analysis between methylation and cytokine data. Serum protein data of 33 cytokines were collected from the same cohort, detailed information of the data was described in our previously published study (Qin et al. 2022). Then we calculated the Spearman's correlations between the methylation level of each CpG and the protein serum level of each cytokine. Visualization approaches and statistical test (correlation test) were utilized to qualify and quantify the correlation directions and magnitudes.

6.2.8 Correlation between methylation and metabolite levels

We explored the function of PedSep-D-related differentially methylated CpGs (DMCs) in metabolisms by examining the correlation between methylation levels of DMCs and levels of serum metabolites collected from the same samples with available data (N = 26). To reduce the

multiple testing burden, we targeted two DMCs contributing to the metabolic pathway enrichment and major metabolites from the two pathways.

6.2.9 Mendelian randomization analysis

To investigate the potential causal relationships between severity-related DMCs and other autoimmune traits of interest, we performed a Mendelian randomization analysis using the *TwoSampleMR* r package (Hemani et al. 2018). To prepare exposure data, methylation quantitative trait loci (meQTLs) summary data were obtained from the Genetics of DNA Methylation Consortium (GoDMC) (Min et al. 2021). Independent significant instrument variables (SNPs with minor allele frequency > 0.01) for the exposure (CpGs) were selected after LD clumping. CpGs without associated significant SNPs ($p < 5\times10^{-8}$) were excluded from the MR analysis. To prepare outcome data, genome-wide analysis study (GWAS) summary data of three traits, including Sepsis 28-day mortality, Sepsis, and Rheumatoid arthritis, were extracted from the UK Biobank (Bycroft et al. 2018). No subjects overlapped between the exposure and outcome data. Then, we harmonized the effect of SNPs for both exposure and outcome data and removed SNPs with strand-ambiguity. Lastly, we applied MR-Egger regression method to conduct Mendelian randomization, given that it outperformed other methods when pleiotropy occurred (Bowden et al. 2015). FDR control was used to adjust for multiple testing.

6.3 Results

6.3.1 PedSep-D phenotype associated with severe outcomes

Out of the 404 pediatric sepsis patients enrolled into the PHENOM cohort, our study focused on 96 children with available blood DNA methylation data (Fig. 6.1). According to our computable phenotype assignment as described in the previous study (Qin et al. 2022), there are 12 patients in PedSep-A (12.5%), 34 patients in PedSep-B (35.4%), 19 patients in PedSep-C (19.8%), and 31 patients in PedSep-D (32.3%). The distribution of four phenotypes differs from the original cohort, with a lower proportion of PedSep-A patients and a higher proportion of PedSep-D patients (PedSep-A: 34% in original cohort; PedSep-D: 14% in original cohort). Given the subsampling from the original cohort, we first examined whether the subset of patients with available methylation data was representative of the original cohort. For PedSep-D patients, there is no significant difference between the subset of patients with and without available methylation data (Appendix Table 10). For non-PedSep-D patients, no significant difference was observed in most of the clinical characteristics except for OFI, ferritin, creatinine, hepatic OFI, and CNS OFI (Appendix Table 11).

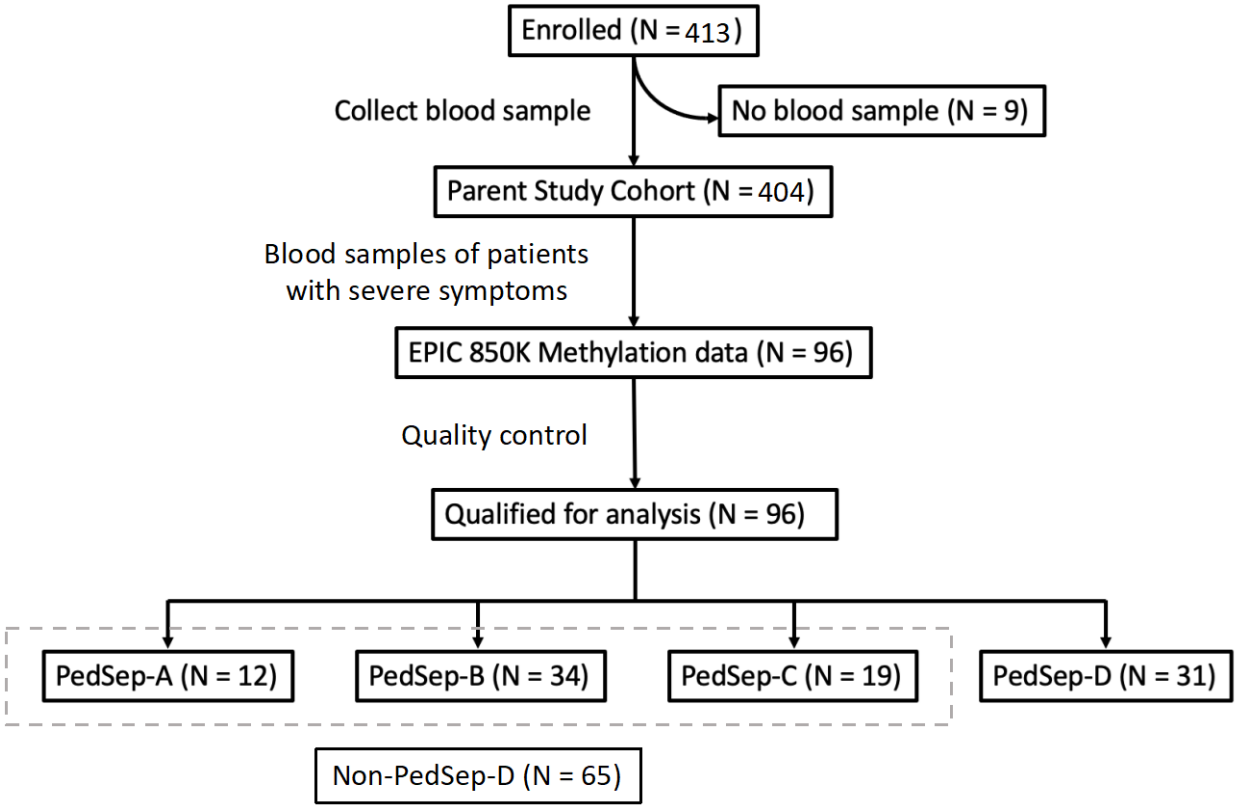

Figure 6.1 Workflow chart of the study

We then compared the differences between PedSep-D and non-PedSep-D groups among patients with available methylation data. In terms of demographic day-one clinical characteristics, the PedSep-D and non-PedSep-D groups differed in characteristics including organ failure index (OFI), day-one ferritin, creatinine, renal organ failure, platelets, and hematologic organ failure (Table 6.1). Among the subset of PedSep-D group with available methylation data, the median age was 4 (interquartile range (IQR, 1–14) years old, 67.7% were male, 9.5% were Hispanic. In terms of infection resources, no significant differences were observed between PedSep-D and non-PedSep-D groups (Table 6.2). In terms of outcomes, PedSep-D patients presented significantly higher mortality rate (41.9% vs 18.5%, p-value = 0.024) and lower PICU-free days (median 0 vs 17, p-value = 0.011) compared to non-PedSep-D patients (Table 6.3).

Table 6.1 Demographic and day 1 clinical characteristics of PedSep-D and Non-PedSep-D patients with methylation data

Characteristics	PedSep-D	Non-PedSep-D
No. of Patients, *N* (%)	31 (32.3)	65 (67.7)
Demographic		
Age, years median (IQR)	4 (1, 14)	6 (2, 12)
Male, N (%)	21 (67.7)	42 (64.6)
Hispanic, N (%)	2 (9.5)	10 (15.4)
Previous healthy, N (%)	10 (32.3)	25 (38.5)
Surgery, N (%)	5 (16.1)	5 (7.7)
Organ Dysfunction		
SIRS criteria[1], median (IQR)	3.0 (2.0, 4.0)	3.0 (2.0, 3.0)
OFI[2], median (IQR)	3.0 (3.0, 4.0)	2.0 (1.0, 2.0)
Inflammation		
CRP, mg/dL median (IQR)	10.8 (4.7, 23.3)	10.4 (4.8, 18.2)
Low temperature, °C median (IQR)	36.6 (36.0, 36.9)	36.6 (36.3, 37.2)
High temperature, °C median (IQR)	37.3 (36.8, 38.9)	37.7 (36.9, 38.5)
ALC, /mm^3 median (IQR)	1.1 (0.6, 2.2)	1.1 (0.6, 1.9)
Ferritin, ng/mL median (IQR)	639.0 (316.0, 2294.4)	275.0 (137.0, 942.2)
Pulmonary		
Pulmonary OFI, N (%)	22 (71.0)	43 (66.2)
Intubation, N (%)	17 (54.8)	39 (60.0)
Cardiovascular or Hemodynamic		
Heart rate, bpm median (IQR)	152.0 (124.5, 174.0)	153.0 (138.0, 174.0)
Systolic blood pressure, mmHg median (IQR)	75.0 (58.0, 93.5)	85.0 (72.0, 96.0)
CV OFI, N (%)	26 (83.9)	50 (76.9)
Renal		
Creatinine, mg/dL median (IQR)	1.5 (0.7, 2.9)	0.6 (0.3, 0.8)
Renal OFI, N (%)	18 (58.1)	0 (0.0)
Hepatic		
Hepatic OFI, N (%)	10 (32.3)	10 (15.4)
Hematologic		
Hemoglobin, g/dL median (IQR)	9.5 (8.1, 10.3)	9.5 (8.4, 10.3)
Platelets, K/mm^3 median (IQR)	49.0 (30.0, 89.5)	140.0 (89.0, 221.0)
Hematologic OFI, N (%)	18 (58.1)	3 (4.6)
Other		

Glasgow Coma Scale score[3,4], median (IQR)	7.0 (3.0, 15.0)	7.0 (3.0, 14.0)
CNS OFI, N (%)	8 (25.8)	17 (26.2)

IQR interquartile range, SIRS systemic inflammatory response syndrome, OFI organ failure index, ALC absolute lymphocyte count, CNS central nervous system

SI conversion factors: to convert alanine transaminase and aspartate aminotransferase to μkat/L, multiply by 0.0167; bilirubin to μmol/L, multiply by 17.104; C-reactive protein to nmol/L, multiply by 9.524; creatinine to μmol/L, multiply by 88.4

1 Indicates SIRS criteria ranging from 0 to 4 including abnormal heart rate, respiratory rate, temperature, and white blood cell count

2 OFI is an integer score reflecting the number of organ failures. Scores are either 0 or 1 for cardiovascular, hepatic, hematologic, respiratory, neurological, and renal, and summed for total range of 0 to 6. Cardiovascular, need for cardiovascular agent infusion support; Pulmonary, need for mechanical ventilation support with the ratio of the arterial partial pressure of oxygen and the fraction of inspired oxygen (PaO2/FiO2) < 300 without this support; Hepatic, total bilirubin > 1.0 mg/dL and alanine aminotransferase (ALT) > 100 units/L; Renal, serum creatinine > 1.0 mg/dL and oliguria (urine output < 0.5 mL/kg/h); Hematologic, thrombocytopenia < 100,000/mm3 and prothrombin time INR > 1.5 × normal; Central Nervous System, Glasgow Coma Scale (GCS) Score < 12 in the absence of sedatives

3 Corresponds to minimum or maximum value (as appropriate) within 6 h of hospital presentation

4 GCS ranges from 3 to 15

Table 6.2 Infections of PedSep-D and Non-PedSep-D patients

Infection	PedSep-D	Non-PedSep-D	p-value
Bacterial infection, N (%)	12 (38.7)	31 (47.7)	0.511
Viral infection, N (%)	4 (12.9)	7 (10.8)	0.743
Fungal infection, N (%)	2 (6.5)	1 (1.5)	0.243

Table 6.3 Outcomes of PedSep-D and Non-PedSep-D patients

Outcome	PedSep-D	Non-PedSep-D	p-value
Length of stay, median (IQR), d	14 (9, 32)	12 (6, 21)	0.361
Mortality, N (%)	13 (41.9)	12 (18.5)	**0.024**
PICU free day, median (IQR), d	0 (0, 18)	17 (0, 24)	**0.011**

6.3.2 Differential methylation analysis identified DMCs and DMRs associated with PedSep-D

The methylation intensity of 866,061 CpGs were measured in DNA methylation profiling using Illumina EPIC array technology. After quality control, a total of 699,145 CpGs remained and were included in the association analysis. Differential methylation analysis was conducted on single CpGs to test their associations with PedSep-D membership after adjusting for age, sex, chip, sample position, cell type proportion, and latent confounders (see Methods). From the QQ plot and the estimated inflation factor, we observed well-calibrated test statistics and little evidence of inflation (lambda = 1.02), suggesting that batch effect and potential confounders were well-controlled (Figure 6.2A). Examining the association between CpG methylation and PedSep-D membership across all sites, the EWAS identified one hypomethylated CpG (cg16704797, p-value = 1.67E-08) associated with PedSep-D compared to non-PedSep--D with genome-wide significance (p-value < 9×10^{-8}), and 24 CpGs with suggestive significance (p-value < 1×10^{-5}), with 13 being hypomethylated and 11 being hypermethylated (Figure 6.2B, Fig 6.2C, Table 6.4). Among them, most hypomethylated CpGs are located in the CpG open sea of gene body or 5 UTR regions, whereas most hypermethylated CpGs are located in the CpG islands or shores of gene promoter or transcription-start site (TSS) regions (Table 6.4). Then we performed differential methylation analysis on 32,846 CpG regions and discovered one significant differentially methylated region (chr19: 42342974 - 42343444, adjusted p-value = 0.02, Table 6.5).

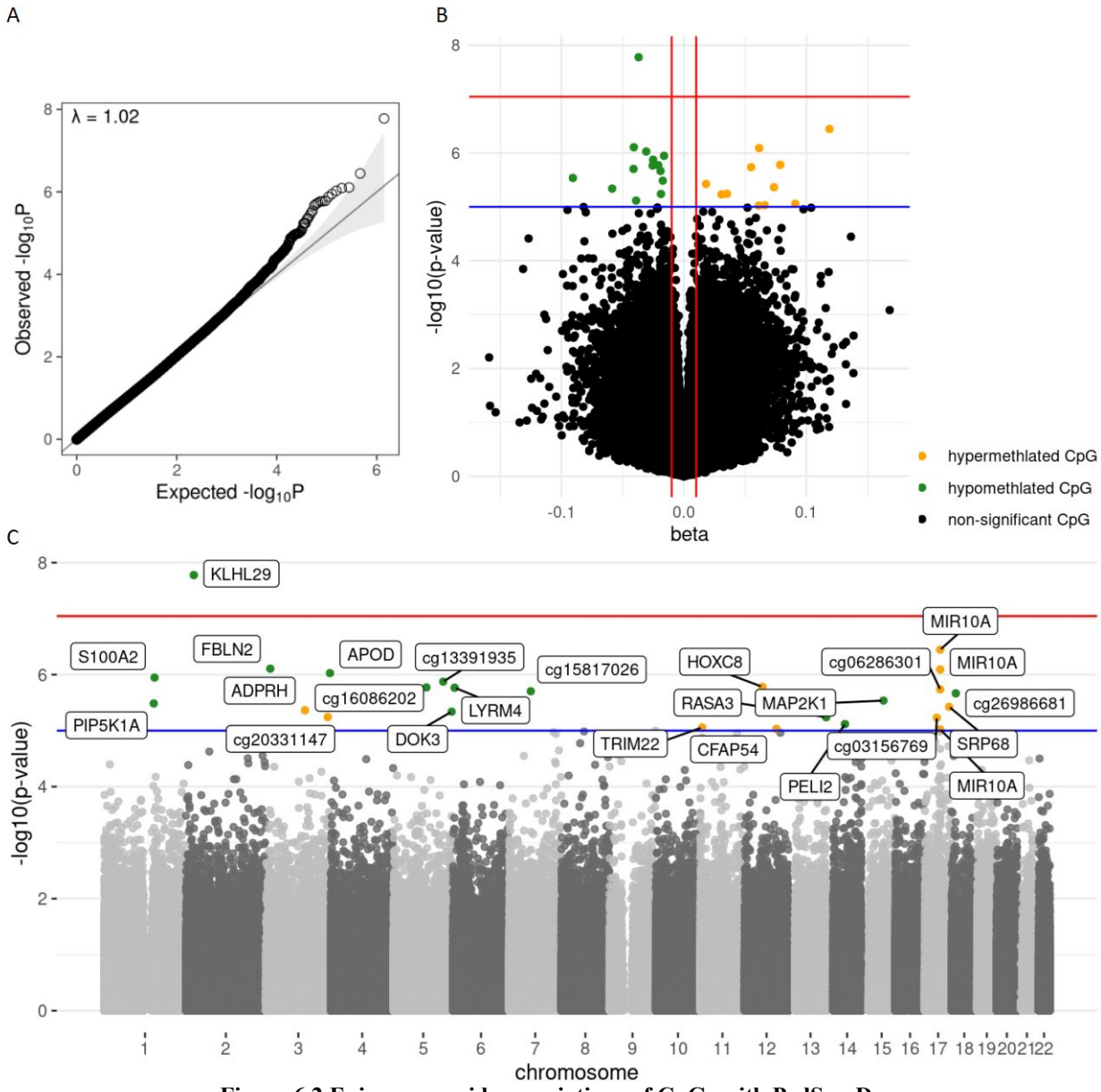

Figure 6.2 Epigenome-wide associations of CpGs with PedSep-D

A. Quantile-quantile plot shows a departure from the null hypothesis of no association. Confounding and batch effects were well-controlled with minimal inflation (λ = 1.02). The λ was calculated using BACON package. B. Volcano plot for the epigenome-wide association tests of PedSep-D shows 1 epigenome-wide DMC, and 24 suggestive DMC. C. Manhattan plot shows that the epigenome-wide DMCs were identified near KLHL29 gene on chromosome 2. DMCs with suggestive significance are labeled based on UCSC annotated genes or CpG ID if no gene annotation. In both the panel B and C, the epigenome-wide significance level (p-value < 9e-08) is denoted by the red line; the suggestive significant level (p-value < 1e-05) is denoted by the blue line. Hypermethylated CpGs are colored as orange dots, whereas hypomethylated CpGs are colored as green dots. DMC, differentially methylated CpG; EWAS, epigenome-wide association study.

Table 6.4 Detailed information of individual differentially methylated CpGs (p-value < 1E-05)

CpG Name	Beta	p-value	chr	position	Island[1]	Gene[2]	Region[3]
cg16704797	-0.04	1.67E-08	chr2	23838416	Open Sea	KLHL29	Body
cg01572694	0.12	3.58E-07	chr17	46657555	N Shore	MIR10A	TSS1500
cg03102887	-0.04	7.83E-07	chr3	13598359	Open Sea	FBLN2	5'UTR
cg13652985	0.06	8.11E-07	chr17	46658257	N Shore	MIR10A	TSS1500
cg23720929	-0.03	9.38E-07	chr3	195310887	Open Sea	APOD	5'UTR
cg27310485	-0.02	1.13E-06	chr1	153536563	Open Sea	S100A2	5'UTR
cg15601205	-0.03	1.33E-06	chr5	151196814	Open Sea	-	-
cg03037150	0.08	1.66E-06	chr12	54402717	Island	HOXC8	TSS200
cg10609524	-0.02	1.69E-06	chr5	100115478	Open Sea	-	-
cg23374256	-0.03	1.71E-06	chr6	5192344	Open Sea	LYRM4	Body
cg14285150	0.05	1.84E-06	chr17	46659019	Island	-	-
cg03233332	-0.04	1.98E-06	chr7	66118400	N Shore	-	-
cg09647390	-0.02	2.16E-06	chr18	13133407	N Shelf	-	-
cg22818074	-0.09	2.90E-06	chr15	66764860	Open Sea	MAP2K1	Body
cg24865494	-0.02	3.27E-06	chr1	151205142	Open Sea	PIP5K1A	Body
cg23672176	0.02	3.76E-06	chr17	74068680	Island	SRP68	TSS200
cg15564579	0.07	4.33E-06	chr3	119298083	N Shore	ADPRH	TSS200
cg23950714	-0.06	4.58E-06	chr5	176935364	S Shelf	DOK3	Body
cg17766219	0.04	5.72E-06	chr3	188696137	Open Sea	-	-
cg00216180	-0.02	5.77E-06	chr13	114778713	Island	RASA3	Body
cg07036914	0.03	5.88E-06	chr17	36577690	S Shore	-	-
cg19929409	-0.04	7.64E-06	chr14	56755226	Open Sea	PELI2	Body
cg26724018	0.09	8.76E-06	chr11	5716255	Open Sea	TRIM22	5'UTR
cg21150327	0.07	9.32E-06	chr12	96889770	Open Sea	CFAP54	Body
cg26916621	0.06	9.49E-06	chr17	46657346	N Shore	MIR10A	TSS200

1. Relation to CpG island. N Shore: north island shore; S Shore: south island shore.
2. Nearby gene annotated by UCSC database, "–" indicates no nearby genes annotated.
3. Regions related to the nearby gene annotated by UCSC database.

Table 6.5 Detailed information of differentially methylated regions (adjusted p-value < 0.1)

# CpG	Beta	Adj p-value	chr	position	Island[1]	Gene[2]	Region[3]
2	-0.04	0.025	chr19	42342974 - 42343444	Open Sea	LYPD4	5'UTR
2	-0.07	0.069	chr18	13133368 - 13133407	N Shelf	-	-

1. Relation to CpG island. N Shore: north island shore; S Shore: south island shore.
2. Nearby gene annotated by UCSC database, "–" indicates no nearby genes annotated.
3. Regions related to the nearby gene annotated by UCSC database.

6.3.3 Functional analysis revealed the connections of PedSep-D-related DMCs with immune cell types, gene expressions, and biological pathways

To conduct functional analyses using public databases, we analyze PedSep-D related DMCs to investigate their roles in different immune cells, cis gene expression regulation, tissue-specific gene sets, and pathways. To first explore relationships between PedSep-D related DMCs and cell types, we first estimated the proportion of six blood immune cell types deconvoluted from bulk methylation data using public data as reference. The six blood immune cell types include T lymphocytes (CD4+ and CD8+), B cells (CD19+), monocytes (CD14+), NK cell (CD56+), and Neutrophils. Among them, PedSep-D was associated with higher monocyte proportion and lower CD8+ T-cell proportion (Appendix Table 12, Figure 6.3A). Then we performed cell-type-specific differential methylation analysis on top DMCs and discovered a total of eight significant cell-type-specific DMCs in NK and Neutrophil cell types (four for each cell types), with six hypomethylated DMCs and two hypermethylated DMCs (Figure 6.3B). No cell-type-specific DMCs identified from cell types displaying different proportions between PedSep-D and other groups (i.e. monocytes and CD8 + T-cell).

To explore relationships between PedSep-D related DMCs and gene expression, we queried the summary statistics of cis expression quantitative trait methylation (cis-eQTM) from the public available Human Early Life Exposome (HELIX) project (Ruiz-Arenas et al. 2022). By mapping cis-eQTM and PedSep-D related DMCs, we identified a total of 661 candidate DMC-gene pairs. Among them, 16 DMC-gene pairs presented significant associations (FDR < 0.05, Appendix Table 13). Notably, three CpGs from a region within MIR10A gene on chr17 associated with expression of multiple genes from the homeobox (HOX) gene clusters encoding transcriptional factors.

Gene set enrichment analysis was conducted based on genes nearby the DMCs. A total of 16 unique genes were used in the analysis, including ten genes near the hypomethylated DMCs and six genes near the hypermethylated DMCs (Table 6.4). By separately testing the enrichment of two gene sets in differentially expressed genes (DEGs) across 30 tissues from GTEx (Consortium 2013), we found the genes near the hypomethylated DMCs were enriched in downregulated DEGs of kidney, liver, and ovary. By separately testing the enrichment of two gene sets in Gene Ontology (GO) (Ashburner et al. 2000, Gene Ontology et al. 2023), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000), and TRANSFAC (TF) (Matys et al. 2006) pathways, we found that the genes near the hypomethylated DMCs were enriched in seven pathways, including (positive) regulation of MAPK cascade, fc gamma r-mediated phagocytosis, choline metabolism in cancer, targets of transcription factor Osx, Yersinia infection, and phospholipase D signaling pathways (Figure 6.3C). The genes near the hypermethylated DMCs were enriched in targets of transcription factors p53 and ZSCAN11 (Figure 6.3D).

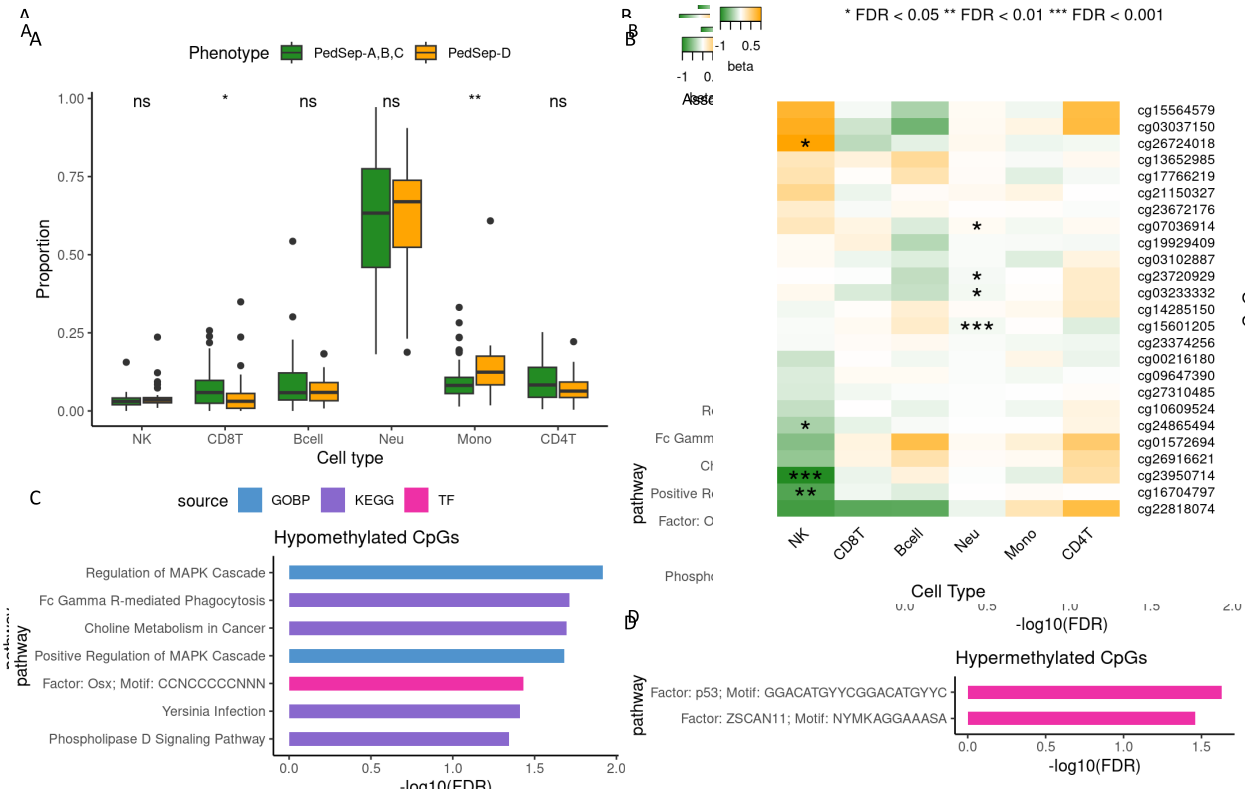

Figure 6.3 Functional analysis investigating the relationships between PedSep-D related DMCs and immune cell types (A and B) and pathways (C and D)

A. Estimated cell-type proportion of six immune cell types in PedSep-D and non-PedSep-D phenotype using 96 samples. Compared to non-PedSep-D group, PedSep-D group has higher proportion of monocyte and low proportion of CD8T. B. Estimated cell-type-specific differentially methylated CpGs (DMCs) using 96 samples. Six blood cell types were inferred from deconvolution analysis based on reference data. Then we test whether EWAS DMCs are specifically differentially methylated in these cell types. The "bulk" column represents the overall effect size estimated from EWAS of PedSep-D. Six cell type columns represent the cell-type-specific effect size of each CpG on PedSep-D. The color of the cell denotes the direction (green: hypomethylation; orange: hypermethylation) and the magnitude of the associations (the darker the stronger). Asterisks denotes the significance of association (*:FDR < 0.05; **: FDR < 0.01; ***: FDR < 0.001). C. Biological pathway analysis of genes near Hypomethylated CpGs using GO, KEGG, and TF databases (FDR < 0.05). D. Biological pathway of genes near Hypermethylated analysis using GO, KEGG, and TF databases (FDR < 0.05).

6.3.4 PedSep-D related DMCs presented high correlation with inflammatory biomarkers and metabolites

In addition to conducting functional analyses utilizing public databases, we further explored the role of PedSep-D related DMCs in inflammation and metabolism by employing cytokine and metabolite profiles collected from the same patients. Firstly, we validated the connections between identified DMCs and inflammation by examining the correlation between methylation levels of PedSep-D related DMCs and serum protein levels of 33 pro-inflammatory and anti-inflammatory cytokines collected at the first day of the patients' enrollment. Among them, we observed three significantly correlated CpG-cytokine pairs after multiple testing corrections (33 cytokines × 25 DMCs) (Figure 6.4A). All three significant pairs involve PedSep-D related hypomethylated DMCs. To keep consistency with the relationship between hypomethylation of DMCs and PedSep-D membership, we reported the correlations as the trend between lower methylation levels of DMCs and the corresponding change of serum protein level of cytokines. Specifically, the lower methylation level of cg00216180 is significantly correlated with lower serum protein level of anti-inflammatory cytokine ADAMTS13. Lower methylation level of cg23374256 is significantly correlated with lower serum protein level of anti-inflammatory cytokine TRAIL. Lower methylation level of cg27310485 is significantly correlated with higher serum protein level of pro-inflammatory cytokine IL-16. No PedSep-D related hypermethylated DMCs significantly correlated with cytokine levels.

Since the pathway enrichment analysis using public datasets implied the relationships between PedSep-D related DMCs and choline metabolism pathway and phospholipase D signaling pathway, we further investigated the connections by examining the correlation between methylation levels of PedSep-D related DMCs and levels of metabolites collected from the same

samples with available data (N = 26, Figure 6.4B). To reduce the multiple testing burden, we targeted two DMCs contributing to the metabolic pathway enrichment and major metabolites from the two pathways (2 metabolites × 2 DMCs). As a result, lower methylation level of cg24865494 is significantly correlated with higher level of metabolite choline. Lower methylation level of cg22818074 is significantly correlated with higher level of metabolite phosphatidylcholine (PC).

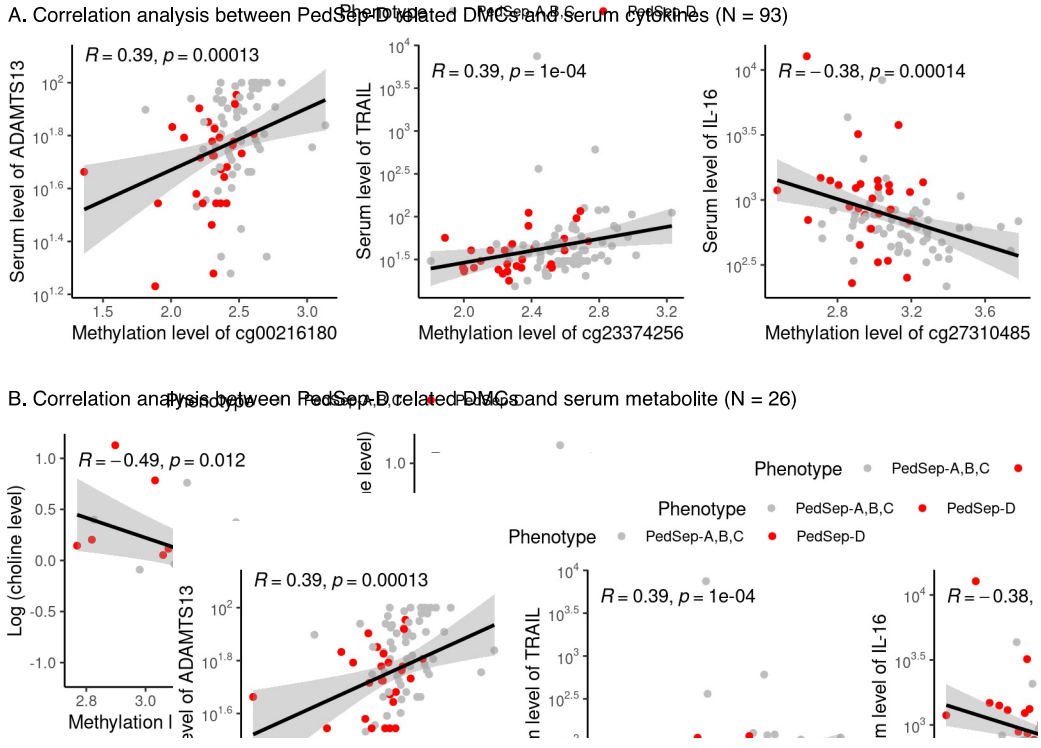

Figure 6.4 Correlation analysis

A. Spearman's correlation scatter plots of three significant CpGs-cytokine pairs. R: correlation coefficient; p: raw p-value of correlation test. Red dots indicate PedSep-D samples, whereas grey dots indicate non-PedSep-D samples. To keep consistency with the relationship between hypomethylation of DMCs and PedSep-D membership, we reported the correlations as the trend between lower methylation levels of DMCs and the corresponding change of serum protein level of cytokines. Specifically, lower methylation level of cg00216180 is significantly correlated with lower serum protein level of anti-inflammatory cytokine ADAMTS13. Lower methylation level of cg23374256 is significantly correlated with lower serum protein level of anti-inflammatory cytokine TRAIL. Lower methylation level of cg27310485 is significantly correlated with higher serum protein level of pro-inflammatory cytokine IL-16.

B. Spearman's correlation scatter plots of two significant CpGs-metabolite pairs. R: correlation coefficient; p: raw p-value of correlation test. Blue line indicates a negative correlation, whereas red line indicates a positive correlation. To keep consistency with the relationship between hypomethylation of DMCs and PedSep-D membership, we reported the correlations as the trend between lower methylation levels of DMCs and the corresponding change of metabolite. Specifically, lower methylation level of cg24865494 is significantly correlated with higher level of metabolite choline. Lower methylation level of cg22818074 is significantly correlated with higher level of metabolite phosphatidylcholine (PC).

6.3.5 PedSep-D related DMCs showed both association and causal relationships with sepsis development, mortality, and autoimmune diseases

To explore the associations between PedSep-D-related DMCs and autoimmune diseases, we searched the EWAS Catalog to identify all previously reported significant findings (Battram et al. 2022). Out of 25 DMCs with methylome-wide and suggestive significance, 19 DMCs had been reported in previous EWASs as being associated with at least one trait. Notably, PedSep-D-related DMCs were found to be associated with several autoimmune diseases, including Primary Sjogren's syndrome, Crohn's disease, Inflammatory Bowel disease (IBD), and Rheumatoid Arthritis (Appendix Table 14).

To further investigate causal relationships between PedSep-related-DMCs and sepsis as well as autoimmune diseases, we leveraged summary statistics from publicly available meQTL and GWAS biobanks to perform Mendelian randomization (MR) analysis using the MR-Egger regression method (see Method). Out of 25 PedSep-D-related DMCs, summary data were available for 10 DMCs. The MR analysis suggested a significant relationship between eight, eight, and three DMCs with sepsis development, sepsis 28-day mortality, and Rheumatoid arthritis (FDR < 0.05), respectively (Figure 6.5). The odds ratio in MR could be interpreted as the change in the odds of the phenotype for each unit increase of genetically controlled DNA methylation levels (in a beta-value scale). We observed some inconsistency between directions inferred using MR analysis based on public adult data and our EWAS based on pediatric sepsis patients, which might be explained by different population and phenotypes used in two analyses.

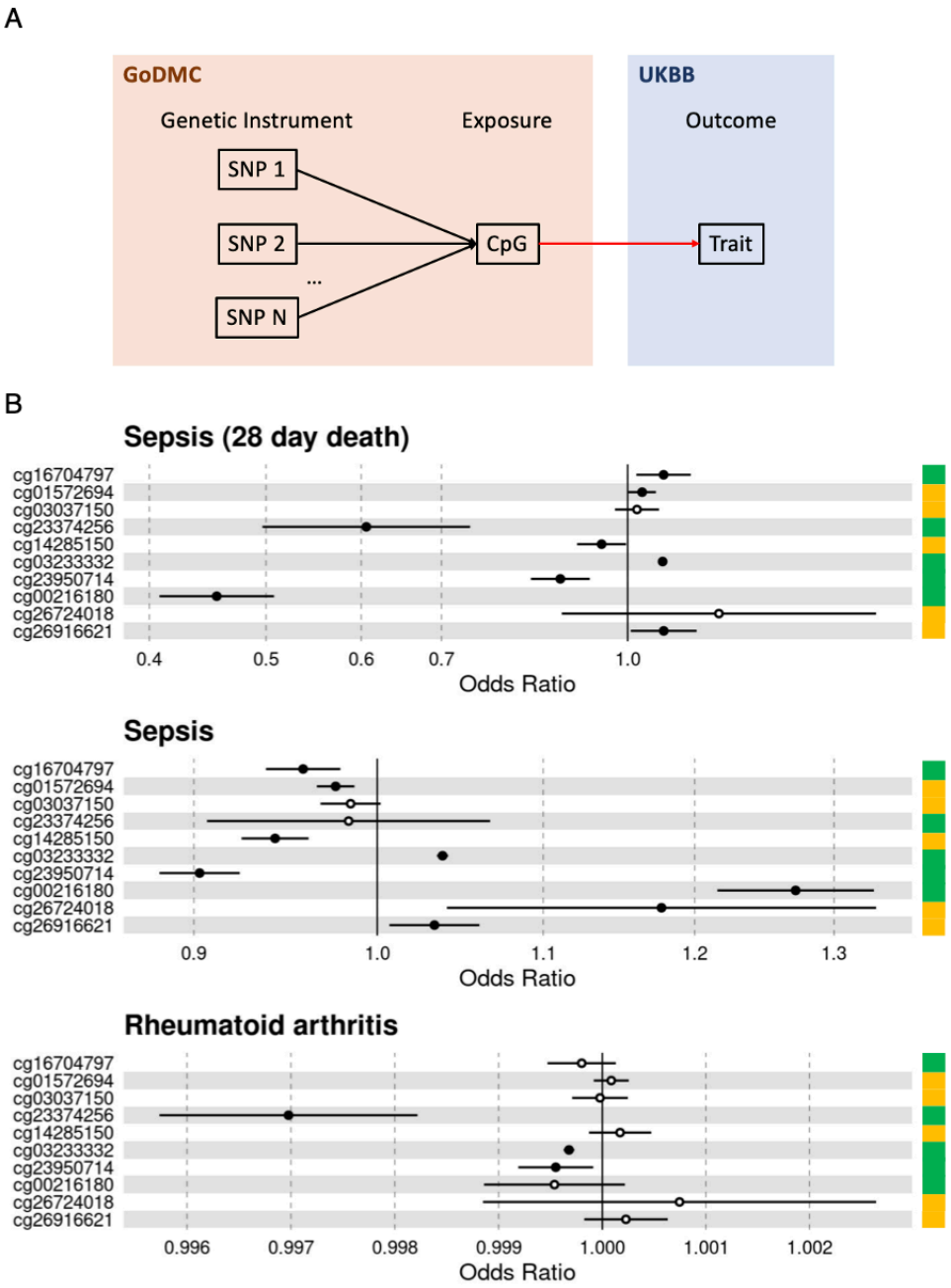

Figure 6.5 Mendelian randomization analysis of 10 out of 25 CpGs related to PedSep-D using GoDMC and UKB databases

A. Illustrative plot. Mendelian randomization analysis was performed using MR-Egger regression method to investigate the relationships between ten CpGs and four immune traits (A: sepsis; B: sepsis 28-day mortality; C: Rheumatoid arthritis). The meQTL data were retrieved from the GoDMC database. The GWAS data were retrieved from the UK Biobank. **B. Results of Mendelian randomization analysis.** The vertical color bar on the right side of CpG forest plot represents the direction of effect for PedSep-D. The colors (green or orange) will be helpful to compare the direction of effects for PedSep-D and the three immune disease traits. Orange denotes hypermethylation (i.e., CpG was positively associated with PedSep-D). Green denotes hypomethylation (i.e., CpG was negatively associated with PedSep-D).

6.4 Discussion

We conducted an Epigenome-Wide Association Study (EWAS) on a multicenter pediatric sepsis cohort comparing PedSep-D and non-PedSep-D phenotype of pediatric sepsis patients and identified one genome-wide significant and 24 suggestive significant DMCs associated with PedSep-D, an agnostic pediatric sepsis phenotype characterized by severe outcome. In subsequent functional analysis, PedSep-D related DMCs exhibited cell-type-specificity, were involved in gene expression regulation, and were enriched in multiple biological pathways. Additionally, we validated the relationship between DMCs and inflammatory response by leveraging serum protein levels of 33 cytokines and observing high correlations between PedSep-D related DMCs and inflammation biomarkers. Finally, we incorporated publicly available data and found that PedSep-D related DMCs depicted associations and potential causal relationships with adult sepsis and autoimmune diseases. Altogether, this study unveiled the role of DNA methylation in pediatric sepsis heterogeneity and contributed to the development of personalized therapies by targeting agnostic pediatric sepsis phenotype with severe outcome.

Our findings derived from differential methylation analysis were supported by several previous studies. For example, the most significant DMC (cg16704797) revealed in our study is hypomethylated in PedSep-D group. This CpG located in the body region of the gene KLHL29, which was previously uncovered as an up-regulated differentially expressed gene (DEG) in sepsis patient using whole blood RNA-seq data (Vastrad and Vastrad 2023). In terms of other suggestively significant DMCs, seven CpGs associated with multiple immune related diseases reported from previous EWAS analysis, including Primary Sjogren's syndrome, Crohn's disease, Inflammatory Bowel disease (IBD), and Rheumatoid Arthritis. For example, three PedSep-D related hypermethylated DMCs (cg01572694, cg13652985, and cg26916621) are located in the

TSS region of gene MIR10A. MIR10A gene encodes microRNA (miRNA) and is considered as a key post-transcriptional regulator that influences anti-inflammatory responses in several diseases, including Rheumatoid Arthritis (RA), Inflammatory Bowel Disease (IBD), Colitis, Acute Pancreatitis (AP), sepsis, atherosclerosis, and cancer (Das and Rao 2022). Additionally, the MIR10A gene is known to be co-expressed with and to target the nearby homeobox (HOX) gene clusters that encode transcriptional factors (TF). As one of the members of the HOX TF family, the HOX2 gene was found being involved in a regulatory network exclusively associating with pediatric sepsis instead of other unrelated inflammatory conditions (Oliveira et al. 2021). Furthermore, an in vivo study has discovered the role of DNA methylation in regulating MIR10A expression by knocking out DNA methyltransferases colon in cancer cell lines and observing an significant gene expression increasing of MIR10A gene and HOX gene cluster (Han et al. 2007). In differentially methylated region analysis, the top region (chr19:42342974-42343444) located within LYPD4 gene was previously reported being associated with eosinopenia (Kim et al. 2021, Gadd et al. 2022), which has been recognized as a reliable diagnostic and prognostic maker of sepsis in multiple studies (Abidi et al. 2008, Al Duhailib et al. 2021). In general, the EWAS results of our study align with earlier findings to validate the role of DNA methylation as a crucial biological factor associated with inflammatory diseases. Beyond this achievement, our findings also extend the role of disease-related signals to indicate their relationships with the severity of pediatric sepsis by targeting the patients with the highest risk.

The functional analysis of this study implies several potential mechanisms linking DNA methylation to pediatric sepsis severity. First, our findings suggest the potential role of methylation as a mediator of infection and disease severity at the innate immune stage of the sepsis manifestation. The cell-type-specific analysis of the study discovered severity-related DMCs

presenting differential methylation in neutrophils and natural killer (NK) cells, two immune cells that are recruited after the onset of infection and release large amounts of pro-inflammatory cytokines to induce pathogen clearance and further exacerbate inflammation (Chiche et al. 2011, Kovach and Standiford 2012). Notwithstanding the well-studied role of Neutrophils and natural killer (NK) cells in inflammatory diseases including sepsis, our analysis marks the first attempt to reveal the cell-type-specific relationship between methylation and the severity of pediatric sepsis. Second, the associations between PedSep-D related DMCs and their target genes were investigated from summary statistics of HELIX project, highlighting the potential role of identified CpGs in gene expression regulation. Third, the enrichment of CpG nearby genes in various tissues may reveal the relationship between the methylation alternation in blood and functional failure in multiple organ systems. Our tissue-specific DEG enrichment analysis found that genes near hypomethylated CpGs enriched in down-regulated DEGs of the kidney and liver, which is consistent with organ dysfunction patterns observed in PedSep-D phenotype. Lastly, our pathway enrichment analysis suggests the role of methylation regulated genes in biological process related to immune response to infections. For example, in the mitogen-activated protein kinase (MAPK) cascade regulation pathway, MAPK such as p38 and JNK regulate T cell activation and differentiation and mediate production of inflammatory cytokines including TNF-α, IL-1, IL-2, and IL-6 (Dong et al. 2002). As another well-established pathway, Fc gamma R-mediated phagocytosis pathway sculptures a process where activation of Fc gamma receptors in Macrophages leading to the release of products associated with inflammatory response, initiation of antibody-dependent cellular cytotoxicity, and phagocytosis, all of which play key roles in host immune defense against infections (Fitzer-Attas et al. 2000). Phospholipase D (PLD) signaling pathway was recently found being involves in immune responses where PLD isoforms expressed

in Neutrophils and Macrophages cells stimulate leukocyte migration, facilitate phagocytosis, and generate reactive oxygen species to enhance innate defense activity (Yoe-Sik Bae 2023). Overall, by leveraging the public data, our functional analysis provides insights of pathology mechanisms driven by methylation and demonstrates the potential biological targets in severe pediatric sepsis population.

In addition to functional analyses utilizing public databases, the relationships between inflammation and methylation levels of the identified PedSep-D related DMCs were further validated using measured serum protein level of 33 cytokines. Although many of the associations between tested CpG-cytokine pairs lost significance after correcting for multiple testing, this analysis highlighted three cytokines (ADAMTS13, TRAIL, and IL-16) that exhibited sensitivity to methylation alterations. ADAMTS13 (a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13) is a large protein involved in blood clotting and plays a vital role in preventing microvascular thrombosis and inflammation (Lu et al. 2020). TRAIL (TNF-related apoptosis-inducing ligand) induces cell apoptosis and suppress inflammation (Wiley et al. 1995). In contrast, IL-16 (Interleukin 16) is a pro-inflammatory cytokine functioning as a T-cell chemoattractant and inducing inflammation (Mathy et al. 2000). The correlations between severity-related DMCs and both anti-inflammation and pro-inflammation cytokines reaffirm the role of methylation in inflammation responses, emphasizing the need for future investigations.

Finally, our Mendelian randomization analysis employing large-scale public biobank data has indicated that 10 DMCs were associated with sepsis mortality, sepsis, and RA. This result potentially suggests that severe sepsis phenotype may be consequences resulting from the epigenetic regulatory impacts of the genetic variants. The results from our Mendelian randomization also align with the previous studies. For example, we found that cg00216180 on

RASA3 was associated with severe sepsis phenotype. Consistently, a recently published GWAS identified common variants in RASA3 significantly associated with EBV seropositivity, a common infection source of sepsis (Muckian et al. 2023). Subsequent Mendelian randomization analysis of this study indicated several environmental causal risk factors for EBV seropositivity. Despite the complexity of Mendelian randomization analysis, establishing causal relationships between DNA methylation and pediatric sepsis severity is crucial for the development of epigenetic therapies. Notably, there have been *in vivo* sepsis treatments implemented in mice and rats by modifying methylation factors (Falcao-Holanda et al. 2021). For instance, Cao et al, treated CLP mice with decitabine to degrade DNMTs, and attenuated NF-κB activation, thereby suppressing inflammatory cytokine levels and inhibiting sepsis progression (Cao et al. 2020). Combined with the existing potential therapy strategies, our findings are expected to propel research towards the methylation-based treatments for pediatric sepsis patients in order to preventing the development of severe outcomes.

Our study presents several potential limitations. First, the sample size of 96 is still modest, larger scale cohorts are needed for independent replication. Second, the non-PedSep-D group is heterogeneous and may limit the interpretability of the results. Third, the suggestive DMCs are not genome-wide significant and so many of them may simply represent false positives. Fourth, DNA methylation profiling was conducted with blood samples, restricting the generalizability of our findings to other tissue types. Fifth, the interpretation of DMC results in each cell type should be approached with caution due to the assumption made by the "CellDMC" function in the EpiDISH package. This function assumes that all other cell types are at 0% when estimating a specific cell type driving the methylation change, whereas our data involve mixed cell types. Subsequent studies utilizing single-cell-type based methylation date are warranted to validate our findings.

Sixth, although we have used the cis-eQTM data from the HELIX Project to explore the association of CpGs and gene expression, our study lacks paired transcriptome data to investigate the impact of DNA methylation on gene expression. Seventh, our study did not have mechanistic experiments to validate the function of identified CpGs. Yet, our study derives well-calibrated hypotheses that can guide future experiments. Eighth, the cross-sectional design limited us to investigate the precise causal link between the DNA methylation and pediatric sepsis severity. While our Mendelian randomization analysis showed potential causal relationship of severity-related DMCs with inflammatory diseases and outcomes, it is important to investigate the association in a longitudinal design. Finally, some functional analysis presented here are based on data in adults for a sepsis case/normal control scenario, which might not be directly relevant to the pediatric PedSep-D vs. non-PedSep-D results.

In summary, our EWAS identified blood DNA methylation signatures associated with the PedSep-D phenotype at the high risk of developing severe outcome. Downstream analysis revealed that identified CpGs play important regulatory roles in various tissues, immune cells, and pathways, and are highly relevant to the inflammatory response. Furthermore, the identified CpGs were associated with additional inflammatory and immune traits, such as Primary Sjogren's syndrome, Crohn's disease, Inflammatory Bowel disease (IBD), and Rheumatoid Arthritis. Our findings serve as a foundation for further exploration into the intricate interplay between environmental factors, epigenetics, host response, and disease pathobiology of pediatric sepsis. Together with other existing findings, our study provides insights into epigenetic components of pediatric sepsis and guidance of the clinical management of the high risk pediatric sepsis population.

7.0 Conclusion

7.1 Summary

This dissertation investigated heterogeneity within pediatric sepsis by integrating clinical, whole exome sequencing, methylation, protein, and metabolism data. In Chapter 4, we employed machine learning approaches to derive four pediatric sepsis phenotypes presenting distinct inflammatory patterns and clinical outcomes. Notably, the identification of PedSep-D phenotype characterized by the most severe inflammation and poor outcomes enables the efficient enrollment of early anti-inflammatory trials targeting patients at high risk. In Chapter 5, a comprehensive whole-exome-wide association analysis was conducted in a gene-based manner to investigate the role of rare variants in pediatric sepsis heterogeneity. This analysis revealed one exome-wide and two suggestive significant genes associated with the PedSep-D phenotype, highlighting several deleterious rare variants that may underlie these associations. Subsequent functional analysis of plasma protein suggested a potential role for the top-identified gene in modulating inflammatory cytokine activities. Chapter 6 delves into an epigenome-wide association analysis to explore the role of methylome in pediatric sepsis heterogeneity. By targeting the PedSep-D phenotype, the study found one genome-wide and 24 suggestive significant differentially methylated CpGs displaying cell-type-specificity and relating to immune cell regulation and lipid metabolism. The findings presented in the three chapters advance our understanding of pediatric sepsis heterogeneity, as well as the complex genetic and epigenetic landscape underlying the disease, offering new insights into sepsis pathology and potential avenues for personalized treatment strategies.

7.2 Significance

This dissertation contributed to bridging the knowledge gaps in pediatric sepsis in multiple aspects. First, although adult sepsis studies have unraveled several subtypes within the scope of its broad definition, the heterogeneity of sepsis was validated in the children population by applying machine learning methods to first-day bedside clinical variables, facilitating the development of precision medicine approaches tailored to pediatric sepsis patients at risk. Moreover, the whole-exome-wide association analysis showed for the first time the role of rare variants in sepsis severity, followed by functional analysis targeting the top signals. Additionally, the epigenome-wide association analysis served as the first pilot study to examine the relationships between methylation factors and pediatric sepsis severity. Altogether, the insights garnered from this dissertation enhance our understanding of pediatric sepsis heterogeneity and its potential biological underpinnings. These contributions are poised to guide the direction of future research and refine clinical practices in the realm of pediatric sepsis, marking a significant step forward in the field.

7.3 Future research

Beyond the scope of this dissertation, several potential research directions could be further explored. First, additional studies should be conducted to validate the four pediatric sepsis phenotypes reported in this study, which would contribute to confirming the universality and robustness of the four PedSep phenotypes. Second, longitudinal studies can be implemented to monitor the disease progression of patients in different phenotypes over time. These studies could indicate the possibility of observing four phenotypes as a result of different stages of disease

progression. Trajectory analysis could also provide insights into the transitions across phenotypes in the long term and help refine therapeutic approaches. Third, conducting genetic and epigenetic association studies to compare the four PedSep phenotypes in a pair-wise manner using a large scale cohort is desired to fully understand the heterogeneity of the disease. Fourth, in-depth functional studies with the usage of *in vitro* and *in vivo* models on the significant genes identified are merited. Finally, more comprehensive integrative analysis using multi-omics data (e.g. transcriptomic, proteomic, etc.) collected from the pediatric sepsis population could be beneficial to provide the new insights into the regulatory mechanisms and interactions that contribute to sepsis heterogeneity.

Appendix A Supplemental materials for Chapter 4

Appendix Table 1 list of 25 variables used to construct the clusters

	Category	Variable
1	demographic	Age
2		Gender
3		Ethnicity
4		Previous healthy
5		Post-surgery
6	PRISM-related vital signs and laboratory values	Systolic blood pressure
7		Heart rate
8		Glasgow Coma Scale Score
9		Hemoglobin
10		Creatinine
11		Platelet count
12		Intubation status
13	Markers of inflammation	Highest temperature
14		Lowest temperature
15		Number of SIRS criteria
16		Lymphocyte count
17		C-reactive protein level
18		Ferritin level
19	Organ failure	Organ failure index
20		Central Nervous System organ failure
21		Cardiovascular organ failure
22		Respiratory organ failure
23		Renal organ failure
24		Hepatic organ failure
25		Hematologic organ failure

Appendix B Supplemental materials for Chapter 5

Appendix Table 2 Demographic and clinical characteristics of patients with and without available WES data

Characteristics	With WES data	Without WES data	p-value[1]	fdr
No. of Patients, *N* (%)	319 (79.0)	85 (21.0)		
Demographic				
Age, years mean (SD)	7 (6)	8 (6)	0.154	0.428
Male, N (%)	175 (54.9)	49 (57.6)	0.736	0.969
Hispanic, N (%)	50 (15.7)	17 (20.0)	0.329	0.587
Previous healthy, N (%)	153 (48.0)	27 (31.8)	0.011	0.039
Surgery, N (%)	38 (11.9)	11 (12.9)	0.943	1.000
Organ Dysfunction				
SIRS criteria[2], mean (SD)	2.9 (0.8)	3.0 (0.8)	0.252	0.525
OFI[3], mean (SD)	1.7 (0.9)	1.9 (0.9)	0.200	0.465
Inflammation				
CRP, mg/dL mean (SD)	11.6 (10.1)	12.1 (11.5)	0.910	1.000
Low temperature, °C mean (SD)	36.6 (1.2)	36.6 (0.9)	0.884	1.000
High temperature, °C mean (SD)	37.8 (1.3)	37.9 (1.3)	0.477	0.745
ALC, /mm^3 median (IQR)	1.3 (0.7-2.2)	0.9 (0.2-1.5)	2.249E-4	0.002
Ferritin, ng/mL mean (IQR)	188.4 (94.1-493.0)	404.0 (183.6-2636.0)	9.507E-6	2.377E-4
Pulmonary				
Pulmonary OFI, N (%)	213 (66.8)	57 (67.1)	1.000	1.000
Intubation, N (%)	178 (55.8)	33 (38.8)	7.766E-3	0.039
Cardiovascular or Hemodynamic				
Heart rate, bpm mean (SD)	154.9 (31.9)	157.1 (29.1)	0.516	0.759
Systolic blood pressure, mmHg mean (SD)	81.4 (19.7)	83.9 (17.7)	0.321	0.587
CV OFI, N (%)	219 (68.7)	65 (76.5)	0.205	0.465
Renal				
Creatinine, mg/dL median (IQR)	0.5 (0.3-0.9)	0.5 (0.3-0.7)	0.731	0.969
Renal OFI, N (%)	26 (8.2)	4 (4.7)	0.357	0.595
Hepatic				
Hepatic OFI, N (%)	31 (9.7)	9 (10.6)	0.838	1.000
Hematologic				
Hemoglobin, g/dL mean (SD)	9.9 (1.9)	9.3 (2.2)	3.395E-3	0.021
Platelets, K/mm^3 mean (SD)	179.3 (114.5)	140.5 (148.5)	7.419E-5	9.274E-4
Hematologic OFI, N (%)	26 (8.2)	13 (15.3)	0.061	0.192

Other				
Glasgow Coma Scale score[4,5], mean (SD)	8.4 (5.3)	10.1 (5.0)	9.611E-3	0.039
CNS OFI, N (%)	42 (13.2)	12 (14.1)	0.960	1.000

IQR interquartile range, SIRS systemic inflammatory response syndrome, OFI organ failure index, ALC absolute lymphocyte count, CNS central nervous system

SI conversion factors: to convert alanine transaminase and aspartate aminotransferase to μkat/L, multiply by 0.0167; bilirubin to μmol/L, multiply by 17.104; C-reactive protein to nmol/L, multiply by 9.524; creatinine to μmol/L, multiply by 88.4

1 Comparisons across all 4 phenotypes were performed using the Kruskal–Wallis test, the $\chi 2$ test, or the Fisher's exact test

2 Indicates SIRS criteria ranging from 0 to 4 including abnormal heart rate, respiratory rate, temperature, and white blood cell count

3 OFI is an integer score reflecting the number of organ failures. Scores are either 0 or 1 for cardiovascular, hepatic, hematologic, respiratory, neurological, and renal, and summed for total range of 0 to 6. Cardiovascular, need for cardiovascular agent infusion support; Pulmonary, need for mechanical ventilation support with the ratio of the arterial partial pressure of oxygen and the fraction of inspired oxygen (PaO2/FiO2) < 300 without this support; Hepatic, total bilirubin > 1.0 mg/dL and alanine aminotransferase (ALT) > 100 units/L; Renal, serum creatinine > 1.0 mg/dL and oliguria (urine output < 0.5 mL/kg/h); Hematologic, thrombocytopenia < 100,000/mm3 and prothrombin time INR > 1.5 × normal; Central Nervous System, Glasgow Coma Scale (GCS) Score < 12 in the absence of sedatives

4 Corresponds to minimum or maximum value (as appropriate) within 6 h of hospital presentation

5 GCS ranges from 3 to 15

Appendix Table 3 Demographic and day 1 clinical characteristics of PedSep-A and Non-PedSep-A patients

Characteristics	PedSep-A	Non-PedSep-A
No. of Patients, *N* (%)	116 (36.4)	203 (63.6)
Demographic		
Age, years mean (SD)	3 (4)	9 (6)
Male, N (%)	52 (44.8)	123 (60.6)
Hispanic, N (%)	24 (22)	26 (13.2)
Previous healthy, N (%)	79 (68.1)	74 (36.5)
Surgery, N (%)	3 (2.6)	35 (17.2)
Organ Dysfunction		
SIRS criteria[1], mean (SD)	2.9 (0.8)	2.9 (0.8)
OFI[2], mean (SD)	1.3 (0.5)	2.0 (0.9)
Inflammation		
CRP, mg/dL mean (SD)	7.4 (7.3)	14.1 (10.7)
Low temperature, °C mean (SD)	36.7 (0.9)	36.5 (1.4)
High temperature, °C mean (SD)	37.8 (1.0)	37.8 (1.4)
ALC, /mm^3 median (IQR)	1.9 (1.3-3.4)	1.0 (0.5-1.7)
Ferritin, ng/mL mean (IQR)	204.2 (71.5-210.5)	260.7 (130.7-681.6)
Pulmonary		
Pulmonary OFI, N (%)	89 (76.7)	124 (61.1)
Intubation, N (%)	62 (53.4)	116 (57.1)
Cardiovascular or Hemodynamic		
Heart rate, bpm mean (SD)	168.6 (29.6)	147.1 (30.5)
Systolic blood pressure, mmHg mean (SD)	85.1 (15.9)	79.2 (21.3)
CV OFI, N (%)	54 (46.6)	165 (81.3)
Renal		
Creatinine, mg/dL median (IQR)	0.3 (0.2-0.4)	0.6 (0.4-1.2)
Renal OFI, N (%)	0 (0.0)	26 (12.8)
Hepatic		
Hepatic OFI, N (%)	3 (2.6)	28 (13.8)
Hematologic		
Hemoglobin, g/dL mean (SD)	10.1 (1.8)	9.8 (2.0)
Platelets, K/mm^3 mean (SD)	257.2 (110.3)	134.7 (90.7)
Hematologic OFI, N (%)	0 (0.0)	26 (12.8)
Other		
Glasgow Coma Scale score[3,4], mean (SD)	8.4 (5.2)	8.3 (5.4)
CNS OFI, N (%)	10 (8.6)	32 (15.8)

IQR interquartile range, SIRS systemic inflammatory response syndrome, OFI organ failure index, ALC absolute lymphocyte count, CNS central nervous system

SI conversion factors: to convert alanine transaminase and aspartate aminotransferase to μkat/L, multiply by 0.0167; bilirubin to μmol/L, multiply by 17.104; C-reactive protein to nmol/L, multiply by 9.524; creatinine to μmol/L, multiply by 88.4

1 Indicates SIRS criteria ranging from 0 to 4 including abnormal heart rate, respiratory rate, temperature, and white blood cell count

2 OFI is an integer score reflecting the number of organ failures. Scores are either 0 or 1 for cardiovascular, hepatic, hematologic, respiratory, neurological, and renal, and summed for total range of 0 to 6. Cardiovascular, need for cardiovascular agent infusion support; Pulmonary, need for mechanical ventilation support with the ratio of the arterial partial pressure of oxygen and the fraction of inspired oxygen (PaO2/FiO2) < 300 without this support; Hepatic, total bilirubin > 1.0 mg/dL and alanine aminotransferase (ALT) > 100 units/L; Renal, serum creatinine > 1.0 mg/dL and oliguria (urine output < 0.5 mL/kg/h); Hematologic, thrombocytopenia < 100,000/mm3 and prothrombin time INR > 1.5 × normal; Central Nervous System, Glasgow Coma Scale (GCS) Score < 12 in the absence of sedatives

3 Corresponds to minimum or maximum value (as appropriate) within 6 h of hospital presentation

4 GCS ranges from 3 to 15

Appendix Table 4 Demographic and day 1 clinical characteristics of PedSep-B and Non-PedSep-B patients

Characteristics	PedSep-B	Non-PedSep-B
No. of Patients, *N* (%)	86 (26.959)	233 (73.041)
Demographic		
Age, years mean (SD)	8 (6)	6 (6)
Male, N (%)	56 (65.1)	119 (51.1)
Hispanic, N (%)	10 (12.2)	40 (17.9)
Previous healthy, N (%)	26 (30.2)	127 (54.5)
Surgery, N (%)	18 (20.9)	20 (8.6)
Organ Dysfunction		
SIRS criteria[1], mean (SD)	3.0 (0.8)	2.9 (0.8)
OFI[2], mean (SD)	2.1 (0.6)	1.6 (0.9)
Inflammation		
CRP, mg/dL mean (SD)	13.5 (11.1)	11.0 (9.6)
Low temperature, °C mean (SD)	35.9 (1.7)	36.8 (0.9)
High temperature, °C mean (SD)	37.3 (1.3)	38.0 (1.2)
ALC, /mm^3 median (IQR)	1.1 (0.7-2.0)	1.3 (0.7-2.4)
Ferritin, ng/mL mean (IQR)	198.2 (111.6-535.2)	183.0 (89.0-481.7)
Pulmonary		
Pulmonary OFI, N (%)	73 (84.9)	140 (60.1)
Intubation, N (%)	80 (93.0)	98 (42.1)
Cardiovascular or Hemodynamic		
Heart rate, bpm mean (SD)	144.9 (29.4)	158.6 (32.0)
Systolic blood pressure, mmHg mean (SD)	73.5 (22.0)	84.3 (17.9)
CV OFI, N (%)	78 (90.7)	141 (60.5)
Renal		
Creatinine, mg/dL median (IQR)	0.6 (0.3-0.8)	0.4 (0.3-0.8)
Renal OFI, N (%)	0 (0.0)	26 (11.2)
Hepatic		
Hepatic OFI, N (%)	9 (10.5)	22 (9.4)
Hematologic		
Hemoglobin, g/dL mean (SD)	9.6 (2.1)	10.1 (1.9)
Platelets, K/mm^3 mean (SD)	155.7 (96.2)	187.9 (119.5)
Hematologic OFI, N (%)	0 (0.0)	26 (11.2)
Other		
Glasgow Coma Scale score[3,4], mean (SD)	4.5 (3.3)	9.8 (5.2)
CNS OFI, N (%)	19 (22.1)	23 (9.9)

IQR interquartile range, SIRS systemic inflammatory response syndrome, OFI organ failure index, ALC absolute lymphocyte count, CNS central nervous system

SI conversion factors: to convert alanine transaminase and aspartate aminotransferase to μkat/L, multiply by 0.0167; bilirubin to μmol/L, multiply by 17.104; C-reactive protein to nmol/L, multiply by 9.524; creatinine to μmol/L, multiply by 88.4

1 Indicates SIRS criteria ranging from 0 to 4 including abnormal heart rate, respiratory rate, temperature, and white blood cell count

2 OFI is an integer score reflecting the number of organ failures. Scores are either 0 or 1 for cardiovascular, hepatic, hematologic, respiratory, neurological, and renal, and summed for total range of 0 to 6. Cardiovascular, need for cardiovascular agent infusion support; Pulmonary, need for mechanical ventilation support with the ratio of the arterial partial pressure of oxygen and the fraction of inspired oxygen (PaO2/FiO2) < 300 without this support; Hepatic, total bilirubin > 1.0 mg/dL and alanine aminotransferase (ALT) > 100 units/L; Renal, serum creatinine > 1.0 mg/dL and oliguria (urine output < 0.5 mL/kg/h); Hematologic, thrombocytopenia < 100,000/mm3 and prothrombin time INR > 1.5 × normal; Central Nervous System, Glasgow Coma Scale (GCS) Score < 12 in the absence of sedatives

3 Corresponds to minimum or maximum value (as appropriate) within 6 h of hospital presentation

4 GCS ranges from 3 to 15

Appendix Table 5 Demographic and day 1 clinical characteristics of PedSep-C and Non-PedSep-C patients

Characteristics	PedSep-C	Non-PedSep-C
No. of Patients, *N* (%)	77 (24.138)	242 (75.862)
Demographic		
Age, years mean (SD)	10 (5)	6 (6)
Male, N (%)	42 (54.5)	133 (55.0)
Hispanic, N (%)	13 (17.1)	37 (16.1)
Previous healthy, N (%)	31 (40.3)	122.0 (50.4)
Surgery, N (%)	9 (11.7)	29 (12.0)
Organ Dysfunction		
SIRS criteria[1], mean (SD)	2.8 (0.8)	2.9 (0.8)
OFI[2], mean (SD)	1.3 (0.5)	1.9 (0.9)
Inflammation		
CRP, mg/dL mean (SD)	15.4 (9.6)	10.5 (10.0)
Low temperature, °C mean (SD)	37.2 (0.9)	36.4 (1.3)
High temperature, °C mean (SD)	38.4 (1.3)	37.6 (1.2)
ALC, /mm^3 median (IQR)	0.6 (0.3-1.1)	1.5 (1.0-2.7)
Ferritin, ng/mL mean (IQR)	260.7 (165.0-3616.4)	173.5 (87.1-443.8)
Pulmonary		
Pulmonary OFI, N (%)	26 (33.8)	187 (77.3)
Intubation, N (%)	14 (18.2)	164.0 (67.8)
Cardiovascular or Hemodynamic		
Heart rate, bpm mean (SD)	150.4 (26.7)	156.4 (33.3)
Systolic blood pressure, mmHg mean (SD)	85.7 (18.0)	80 (20.0)
CV OFI, N (%)	57 (74.0)	162 (66.9)
Renal		
Creatinine, mg/dL median (IQR)	0.5 (0.4-0.7)	0.4 (0.3-0.9)
Renal OFI, N (%)	0 (0.0)	26 (10.7)
Hepatic		
Hepatic OFI, N (%)	7 (9.1)	24 (9.9)
Hematologic		
Hemoglobin, g/dL mean (SD)	10.4 (1.9)	9.8 (1.9)
Platelets, K/mm^3 mean (SD)	137.6 (82.9)	192.5 (120.0)
Hematologic OFI, N (%)	7 (9.1)	19 (7.9)
Other		
Glasgow Coma Scale score[3,4], mean (SD)	13.1 (3.2)	6.9 (5.0)
CNS OFI, N (%)	4 (5.2)	38 (15.7)

IQR interquartile range, SIRS systemic inflammatory response syndrome, OFI organ failure index, ALC absolute lymphocyte count, CNS central nervous system

SI conversion factors: to convert alanine transaminase and aspartate aminotransferase to μkat/L, multiply by 0.0167; bilirubin to μmol/L, multiply by 17.104; C-reactive protein to nmol/L, multiply by 9.524; creatinine to μmol/L, multiply by 88.4

1 Indicates SIRS criteria ranging from 0 to 4 including abnormal heart rate, respiratory rate, temperature, and white blood cell count

2 OFI is an integer score reflecting the number of organ failures. Scores are either 0 or 1 for cardiovascular, hepatic, hematologic, respiratory, neurological, and renal, and summed for total range of 0 to 6. Cardiovascular, need for cardiovascular agent infusion support; Pulmonary, need for mechanical ventilation support with the ratio of the arterial partial pressure of oxygen and the fraction of inspired oxygen (PaO2/FiO2) < 300 without this support; Hepatic, total bilirubin > 1.0 mg/dL and alanine aminotransferase (ALT) > 100 units/L; Renal, serum creatinine > 1.0 mg/dL and oliguria (urine output < 0.5 mL/kg/h); Hematologic, thrombocytopenia < 100,000/mm3 and prothrombin time INR > 1.5 × normal; Central Nervous System, Glasgow Coma Scale (GCS) Score < 12 in the absence of sedatives

3 Corresponds to minimum or maximum value (as appropriate) within 6 h of hospital presentation

4 GCS ranges from 3 to 15

Appendix Table 6 Biomarkers measured at day 1 by phenotype PedSep-A (N = 319)

Biomarker[a]	PedSep-A (N = 116)	Non-PedSep-A (N = 203)	p-value
ADAMTS13, %	83.0 (65.0, 95.0)	66.0 (48.0, 83.5)	<0.001
SFasLg, pg/ml	59.8 (39.1, 87.0)	40.8 (29.4, 66.3)	<0.001
Ex vivo TNF-α, pg/ml	691.0 (334.0, 1049.2)	385.9 (97.1, 944.3)	<0.001
TNF-α, pg/ml	1049.2 (728.0, 1049.2)	1049.2 (604.1, 1049.2)	0.721
sCD163, pg/ml	212829 (155490, 300033)	340338 (215555, 580440)	<0.001
IFN-β, pg/ml	6.4 (6.4, 6.4)	6.4 (6.4, 8.2)	0.118
IL-22, pg/ml	22.4 (17.8, 29.5)	28.3 (22.1, 38.4)	<0.001
IL-18, pg/ml	329.6 (222.6, 533.6)	485.9 (294.6, 882.7)	<0.001
IL-18BP, pg/ml	9445 (6185, 16011)	20764 (12054, 33449)	<0.001
MIG/CXCL9, pg/ml	619.3 (406.0, 1061.2)	1026.0 (490.6, 2883.8)	<0.001
IL-1β, pg/ml	2.6 (2.1, 3.1)	2.8 (2.4, 3.3)	0.071
IL-4, pg/ml	4.7 (3.5, 6.3)	4.7 (3.5, 6.5)	0.799
IL-6, pg/ml	6.9 (5.8, 9.7)	10.9 (6.5, 30.2)	<0.001
IL-8, pg/ml	38.0 (26.4, 65.5)	59.5 (36.3, 130.5)	<0.001
IL-10, pg/ml	19.3 (15.4, 24.6)	25.0 (18.1, 40.3)	<0.001
IL-13, pg/ml	3.1 (3.1, 4.3)	3.1 (3.1, 3.5)	0.416
IL-17A, pg/ml	17.4 (15.1, 21.7)	20.0 (16.5, 25.1)	<0.001
IFN-γ, pg/ml	2.8 (2.8, 3.0)	2.8 (2.8, 2.8)	0.477
IP-10/CXCL10, pg/ml	492.2 (268.5, 1691.1)	850.5 (381.0, 2361.3)	0.008
MCP-1/CCL2, pg/ml	103.5 (48.3, 190.8)	178.6 (84.3, 400.5)	<0.001
MIP-1α, pg/ml	0.6 (0.6, 0.6)	2.0 (0.6, 10.7)	<0.001
MIP-1β, pg/ml	42.8 (28.1, 57.2)	50.7 (34.5, 81.1)	<0.001
MCP-3, pg/ml	92.4 (92.4, 147.8)	92.4 (92.4, 166.0)	0.269
IFN-α2, pg/ml	125.7 (105.8, 142.3)	125.7 (105.8, 142.8)	0.542
IL-1α, pg/ml	9.4 (9.4, 11.6)	9.4 (9.4, 16.4)	0.183
IL-2RA, pg/ml	343.5 (235.8, 504.6)	401.8 (243.4, 696.7)	0.014
IL-3, pg/ml	612.2 (496.1, 734.6)	612.2 (529.0, 724.4)	0.976
IL-16, pg/ml	544.1 (398.1, 660.5)	635.5 (453.6, 836.7)	<0.001
M-CSF, pg/ml	19.7 (13.8, 34.0)	36.2 (21.6, 78.6)	<0.001
SCF, pg/ml	138.2 (111.4, 192.7)	167.5 (115.4, 269.8)	<0.001
TRAIL, pg/ml	42.9 (32.9, 64.2)	35.4 (27.9, 48.5)	<0.001
CRP, mg/dL	4.6 (1.3, 12.4)	11.4 (6.2, 20.8)	<0.001
Ferritin, ng/mL	121.0 (71.0, 204.5)	260.9 (130.6, 682.4)	<0.001

[a] All biomarkers are measured one time concomitantly in the first day. Values in table are summarized as median (IQR).

Appendix Table 7 Biomarkers measured at day 1 by phenotype PedSep-B (N = 319)

Biomarker[a]	PedSep-B (N = 86)	Non-PedSep-B (N = 233)	p-value
ADAMTS13, %	70.5 (56.2, 89.2)	72.0 (53.8, 88.2)	0.710
SFasLg, pg/ml	38.4 (29.9, 67.9)	49.0 (31.9, 79.5)	0.160
Ex vivo TNF-α, pg/ml	336.3 (99.2, 806.8)	564.7 (184.3, 1049.2)	0.022
TNF-α, pg/ml	1049.2 (1049.2, 1500.0)	957.3 (602.2, 1049.2)	0.023
sCD163, pg/ml	309800 (179839, 508713)	276625 (177921, 427838)	0.363
IFN-β, pg/ml	6.4 (6.4, 10.5)	6.4 (6.4, 6.4)	0.014
IL-22, pg/ml	28.0 (21.3, 36.0)	24.8 (20.1, 34.2)	0.166
IL-18, pg/ml	444.0 (258.8, 820.7)	400.3 (250.4, 665.8)	0.386
IL-18BP, pg/ml	16142 (8938, 27901)	16000 (8699, 27428)	0.920
MIG/CXCL9, pg/ml	772.7 (403.0, 2033.0)	810.5 (458.8, 1993.6)	0.658
IL-1β, pg/ml	2.8 (2.1, 3.3)	2.8 (2.4, 3.3)	0.984
IL-4, pg/ml	4.7 (3.9, 6.8)	4.7 (3.5, 6.5)	0.293
IL-6, pg/ml	8.9 (6.5, 22.3)	8.4 (6.2, 17.1)	0.583
IL-8, pg/ml	54.7 (34.7, 113.7)	49.4 (30.1, 88.7)	0.242
IL-10, pg/ml	22.4 (18.0, 37.7)	21.7 (16.3, 32.8)	0.322
IL-13, pg/ml	3.1 (3.1, 3.5)	3.1 (3.1, 4.1)	0.787
IL-17A, pg/ml	19.1 (16.5, 25.1)	18.3 (15.6, 23.4)	0.060
IFN-γ, pg/ml	2.8 (2.8, 3.0)	2.8 (2.8, 2.8)	0.127
IP-10/CXCL10, pg/ml	705.8 (257.2, 1723.9)	769.1 (354.8, 2284.0)	0.409
MCP-1/CCL2, pg/ml	169.9 (83.5, 377.0)	130.8 (57.4, 300.8)	0.159
MIP-1α, pg/ml	0.6 (0.6, 10.2)	0.6 (0.6, 5.8)	0.210
MIP-1β, pg/ml	43.4 (32.3, 70.1)	46.7 (31.5, 66.7)	0.978
MCP-3, pg/ml	92.4 (92.4, 166.0)	92.4 (92.4, 147.8)	0.821
IFN-α2, pg/ml	125.7 (112.6, 148.3)	125.7 (105.8, 140.2)	0.181
IL-1α, pg/ml	9.4 (9.4, 16.4)	9.4 (9.4, 13.2)	0.211
IL-2RA, pg/ml	401.8 (199.2, 689.9)	364.7 (246.1, 561.6)	0.520
IL-3, pg/ml	624.4 (529.0, 724.4)	612.2 (512.5, 724.4)	0.412
IL-16, pg/ml	605.0 (444.1, 759.8)	575.9 (418.2, 778.0)	0.752
M-CSF, pg/ml	30.6 (19.8, 54.6)	28.1 (15.6, 50.8)	0.250
SCF, pg/ml	152.8 (109.2, 229.5)	154.2 (116.1, 232.2)	0.680
TRAIL, pg/ml	35.4 (25.4, 51.3)	39.1 (30.3, 53.5)	0.204
CRP, mg/dL	10.2 (6.0, 19.6)	9.4 (2.9, 16.2)	0.069
Ferritin, ng/mL	199.5 (111.0, 545.0)	182.3 (88.6, 481.8)	0.354

[a] All biomarkers are measured one time concomitantly in the first day. Values in table are summarized as median (IQR).

Appendix Table 8 Biomarkers measured at day 1 by phenotype PedSep-C (N = 319)

Biomarker[a]	PedSep-C (N = 77)	Non-PedSep-C (N = 242)	p-value
ADAMTS13, %	69.0 (47.0, 84.0)	73.0 (57.0, 93.0)	0.061
SFasLg, pg/ml	42.7 (28.2, 66.9)	49.3 (33.0, 79.5)	0.077
Ex vivo TNF-α, pg/ml	384.8 (78.2, 485.4)	526.7 (230.7, 1049.2)	0.021
TNF-α, pg/ml	938.9 (430.0, 1049.2)	1049.2 (728.0, 1049.2)	0.268
sCD163, pg/ml	334582 (206378, 489112)	262850 (174207, 422271)	0.028
IFN-β, pg/ml	6.4 (6.4, 8.2)	6.4 (6.4, 8.2)	0.946
IL-22, pg/ml	26.0 (20.1, 33.0)	24.8 (20.1, 34.2)	0.440
IL-18, pg/ml	510.7 (269.2, 918.1)	391.5 (240.6, 638.6)	0.012
IL-18BP, pg/ml	22755 (14543, 32016)	13654 (7423, 26090)	<0.001
MIG/CXCL9, pg/ml	1145.3 (535.7, 2394.1)	753.6 (428.4, 1781.8)	0.037
IL-1β, pg/ml	2.8 (2.4, 3.3)	2.8 (2.2, 3.2)	0.339
IL-4, pg/ml	4.7 (3.5, 6.5)	4.7 (3.5, 6.5)	0.506
IL-6, pg/ml	10.3 (6.9, 28.4)	8.1 (6.0, 16.4)	0.009
IL-8, pg/ml	49.4 (30.5, 92.2)	51.0 (31.4, 91.1)	0.919
IL-10, pg/ml	22.5 (17.5, 37.0)	21.7 (16.3, 30.6)	0.232
IL-13, pg/ml	3.1 (3.1, 4.3)	3.1 (3.1, 3.9)	0.795
IL-17A, pg/ml	20.9 (17.4, 26.8)	18.3 (15.6, 22.6)	0.014
IFN-γ, pg/ml	2.8 (2.8, 2.8)	2.8 (2.8, 3.0)	0.288
IP-10/CXCL10, pg/ml	985.1 (412.6, 2704.1)	668.5 (303.5, 1939.3)	0.025
MCP-1/CCL2, pg/ml	169.1 (83.9, 455.5)	125.8 (57.2, 290.6)	0.060
MIP-1α, pg/ml	0.6 (0.6, 9.0)	0.6 (0.6, 6.1)	0.164
MIP-1β, pg/ml	48.4 (33.0, 79.6)	45.5 (31.4, 64.5)	0.167
MCP-3, pg/ml	119.5 (92.4, 166.0)	92.4 (92.4, 147.8)	0.430
IFN-α2, pg/ml	125.7 (105.8, 140.2)	125.7 (105.8, 144.4)	0.838
IL-1α, pg/ml	9.4 (9.4, 16.4)	9.4 (9.4, 16.4)	0.678
IL-2RA, pg/ml	367.7 (239.4, 696.7)	380.6 (243.0, 575.4)	0.844
IL-3, pg/ml	612.2 (529.0, 724.4)	612.2 (529.0, 724.4)	0.840
IL-16, pg/ml	563.6 (413.5, 743.3)	590.2 (432.5, 795.6)	0.391
M-CSF, pg/ml	31.9 (17.3, 54.7)	28.1 (15.6, 51.1)	0.269
SCF, pg/ml	148.9 (112.1, 216.6)	154.2 (115.4, 233.8)	0.397
TRAIL, pg/ml	35.4 (30.4, 45.4)	39.1 (27.9, 56.4)	0.336
CRP, mg/dL	14.4 (9.2, 21.2)	8.1 (2.6, 15.5)	<0.001
Ferritin, ng/mL	260.7 (165.0, 664.6)	172.5 (87.1, 447.0)	<0.001

[a] All biomarkers are measured one time concomitantly in the first day. Values in table are summarized as median (IQR).

Appendix C Supplemental materials for Chapter 6

Appendix Table 9 Number of CpGs filtered in each quality control (QC) step

	QC steps	# probes filtered
1	Detection p-value of greater than 0.01 in one or more samples	11,142
2	Bead count lower than 3	86,391
3	Containing single nucleotide polymorphism-introduced artifacts and in cross-reactive regions	67,031
4	Mapped to sex chromosomes	19,627

Appendix Table 10 Demographic and day 1 clinical characteristics of PedSep-D patients with and without available methylation data

Characteristics	With methylation data	Without methylation data	p-value[1]
No. of Patients, *N*	31	25	
Demographic			
Age, years median (IQR)	4 (1, 14)	11 (4, 13)	0.138
Male, N (%)	21 (67.7)	13 (52.0)	0.278
Hispanic, N (%)	2 (9.5)	2 (8.0)	1.000
Previous healthy, N (%)	10 (32.3)	9 (36.0)	0.784
Surgery, N (%)	5 (16.1)	7 (28.0)	0.338
Organ Dysfunction			
SIRS criteria[2], median (IQR)	3.0 (2.0, 4.0)	3.0 (2.0, 4.0)	0.854
OFI[3], median (IQR)	3.0 (3.0, 4.0)	2.0 (3.0, 3.0)	0.077
Inflammation			
CRP, mg/dL median (IQR)	10.8 (4.7, 23.3)	10.5 (3.4, 16.5)	0.589
Low temperature, °C median (IQR)	36.6 (36.0, 36.9)	36.4 (36.2, 36.9)	0.773
High temperature, °C median (IQR)	37.3 (36.8, 38.9)	37.3 (36.9, 38.0)	0.895
ALC, /mm^3 median (IQR)	1.1 (0.6, 2.2)	1.3 (0.8, 2.1)	0.882
Ferritin, ng/mL median (IQR)	639.0 (316.0, 2294.4)	537.0 (202.0, 2906.0)	0.744
Pulmonary			
Pulmonary OFI, N (%)	22 (71.0)	16 (64.0)	0.774
Intubation, N (%)	17 (54.8)	13 (52.0)	1.000
Cardiovascular or Hemodynamic			
Heart rate, bpm median (IQR)	152.0 (124.5, 174.0)	160.0 (144.0, 169.0)	0.615
Systolic blood pressure, mmHg median (IQR)	75.0 (58.0, 93.5)	83.0 (65.0, 95.0)	0.335
CV OFI, N (%)	26 (83.9)	18 (72.0)	0.338
Renal			
Creatinine, mg/dL median (IQR)	1.5 (0.7, 2.9)	1.3 (0.6, 1.9)	0.644
Renal OFI, N (%)	18 (58.1)	12 (48.0)	0.591
Hepatic			
Hepatic OFI, N (%)	10 (32.3)	7 (28.0)	0.778
Hematologic			
Hemoglobin, g/dL median (IQR)	9.5 (8.1, 10.3)	8.7 (7.2, 10.7)	0.817
Platelets, K/mm^3 median (IQR)	49.0 (30.0, 89.5)	83.0 (17.0, 151.0)	0.526
Hematologic OFI, N (%)	18 (58.1)	13 (52.0)	0.788

Other			
Glasgow Coma Scale score[4,5], median (IQR)	7.0 (3.0, 15.0)	8.0 (3.0, 14.0)	0.913
CNS OFI, N (%)	8 (25.8)	4 (16.0)	0.516

IQR interquartile range, SIRS systemic inflammatory response syndrome, OFI organ failure index, ALC absolute lymphocyte count, CNS central nervous system

SI conversion factors: to convert alanine transaminase and aspartate aminotransferase to μkat/L, multiply by 0.0167; bilirubin to μmol/L, multiply by 17.104; C-reactive protein to nmol/L, multiply by 9.524; creatinine to μmol/L, multiply by 88.4

1 Comparisons across all 4 phenotypes were performed using the Kruskal–Wallis test, the $\chi 2$ test, or the Fisher's exact test

2 Indicates SIRS criteria ranging from 0 to 4 including abnormal heart rate, respiratory rate, temperature, and white blood cell count

3 OFI is an integer score reflecting the number of organ failures. Scores are either 0 or 1 for cardiovascular, hepatic, hematologic, respiratory, neurological, and renal, and summed for total range of 0 to 6. Cardiovascular, need for cardiovascular agent infusion support; Pulmonary, need for mechanical ventilation support with the ratio of the arterial partial pressure of oxygen and the fraction of inspired oxygen (PaO2/FiO2) < 300 without this support; Hepatic, total bilirubin > 1.0 mg/dL and alanine aminotransferase (ALT) > 100 units/L; Renal, serum creatinine > 1.0 mg/dL and oliguria (urine output < 0.5 mL/kg/h); Hematologic, thrombocytopenia < 100,000/mm3 and prothrombin time INR > 1.5 × normal; Central Nervous System, Glasgow Coma Scale (GCS) Score < 12 in the absence of sedatives

4 Corresponds to minimum or maximum value (as appropriate) within 6 h of hospital presentation

5 GCS ranges from 3 to 15

Appendix Table 11 Demographic and day 1 clinical characteristics of non-PedSep-D patients with and without available methylation data

Characteristics	With methylation data	Without methylation data	p-value[1]
No. of Patients, *N*	65	283	
Demographic			
Age, years median (IQR)	6 (2, 12)	6 (1, 12)	0.394
Male, N (%)	42 (64.6)	148 (52.3)	0.075
Hispanic, N (%)	10 (15.4)	53 (18.7)	0.783
Previous healthy, N (%)	25 (38.5)	136 (48.1)	0.171
Surgery, N (%)	5 (7.7)	32 (11.3)	0.506
Organ Dysfunction			
SIRS criteria[2], median (IQR)	3.0 (2.0, 3.0)	3.0 (2.0, 4.0)	0.582
OFI[3], median (IQR)	2.0 (1.0, 2.0)	1.0 (1.0, 2.0)	**< 0.001**
Inflammation			
CRP, mg/dL median (IQR)	10.4 (4.8, 18.2)	9.4 (2.9, 16.5)	0.210
Low temperature, °C median (IQR)	36.6 (36.3, 37.2)	36.7 (36.2, 37.3)	0.582
High temperature, °C median (IQR)	37.7 (36.9, 38.5)	37.7 (36.9, 38.7)	0.675
ALC, /mm^3 median (IQR)	1.1 (0.6, 1.9)	1.2 (0.6, 2.1)	0.876
Ferritin, ng/mL median (IQR)	275.0 (137.0, 942.2)	183.6 (87.0, 424.4)	**0.002**
Pulmonary			
Pulmonary OFI, N (%)	43 (66.2)	189 (66.8)	1.000
Intubation, N (%)	39 (60.0)	142 (50.2)	0.170
Cardiovascular or Hemodynamic			
Heart rate, bpm median (IQR)	153.0 (138.0, 174.0)	156.0 (136.5, 176.5)	0.602
Systolic blood pressure, mmHg median (IQR)	85.0 (72.0, 96.0)	83.0 (72.0, 93.0)	0.806
CV OFI, N (%)	50 (76.9)	190 (67.1)	0.139
Renal			
Creatinine, mg/dL median (IQR)	0.6 (0.3, 0.8)	0.4 (0.3, 0.7)	**0.011**
Renal OFI, N (%)	0 (0.0)	0 (0.0)	1.000
Hepatic			
Hepatic OFI, N (%)	10 (15.4)	13 (4.6)	**0.004**
Hematologic			
Hemoglobin, g/dL median (IQR)	9.5 (8.4, 10.3)	9.9 (8.6, 11.4)	0.082
Platelets, K/mm^3 median (IQR)	140.0 (89.0, 221.0)	177.0 (98.5, 254.5)	0.121
Hematologic OFI, N (%)	3 (4.6)	5 (1.8)	0.173
Other			

Glasgow Coma Scale score[4,5], median (IQR)	7.0 (3.0, 14.0)	10.0 (3.0, 15.0)	0.081
CNS OFI, N (%)	17 (26.2)	25 (8.8)	**< 0.001**

IQR interquartile range, SIRS systemic inflammatory response syndrome, OFI organ failure index, ALC absolute lymphocyte count, CNS central nervous system

SI conversion factors: to convert alanine transaminase and aspartate aminotransferase to μkat/L, multiply by 0.0167; bilirubin to μmol/L, multiply by 17.104; C-reactive protein to nmol/L, multiply by 9.524; creatinine to μmol/L, multiply by 88.4

1 Comparisons across all 4 phenotypes were performed using the Kruskal–Wallis test, the $\chi 2$ test, or the Fisher's exact test

2 Indicates SIRS criteria ranging from 0 to 4 including abnormal heart rate, respiratory rate, temperature, and white blood cell count

3 OFI is an integer score reflecting the number of organ failures. Scores are either 0 or 1 for cardiovascular, hepatic, hematologic, respiratory, neurological, and renal, and summed for total range of 0 to 6. Cardiovascular, need for cardiovascular agent infusion support; Pulmonary, need for mechanical ventilation support with the ratio of the arterial partial pressure of oxygen and the fraction of inspired oxygen (PaO2/FiO2) < 300 without this support; Hepatic, total bilirubin > 1.0 mg/dL and alanine aminotransferase (ALT) > 100 units/L; Renal, serum creatinine > 1.0 mg/dL and oliguria (urine output < 0.5 mL/kg/h); Hematologic, thrombocytopenia < 100,000/mm3 and prothrombin time INR > 1.5 × normal; Central Nervous System, Glasgow Coma Scale (GCS) Score < 12 in the absence of sedatives

4 Corresponds to minimum or maximum value (as appropriate) within 6 h of hospital presentation

5 GCS ranges from 3 to 15

Appendix Table 12 Estimated cell type proportion in PedSep-D and non-PedSep-D phenotypes

Cell type	PedSep-D	Non-PedSep-D	p-value*
NK cell, mean (sd)	0.046 (0.043)	0.033 (0.021)	0.175
Neutrophil, mean (sd)	0.623 (0.177)	0.617 (0.193)	0.925
Monocyte, mean (sd)	0.141 (0.102)	0.095 (0.062)	0.003
CD8 T-cell, mean (sd)	0.053 (0.075)	0.071 (0.062)	0.042
CD4 T-cell, mean (sd)	0.069 (0.046)	0.096 (0.063)	0.073
B-cell, mean (sd)	0.067 (0.045)	0.088 (0.087)	0.531

* t-test.

Appendix Table 13 Summary statistics of significant DMC-gene pairs from the HELIX project

CpG	CpG gene[1]	TC gene[2]	log2FC	p-value	FDR
cg01572694	MIR10A	HOXB2	0.14	4.56E-22	3.01E-19
cg01572694	MIR10A	HOXB3;HOXB4;MIR10A	0.03	4.64E-20	1.53E-17
cg14285150		HOXB2	0.21	4.48E-17	9.87E-15
cg26916621	MIR10A	HOXB2	0.19	8.55E-17	1.13E-14
cg26916621	MIR10A	HOXB3;HOXB4;MIR10A	0.05	8.55E-17	1.13E-14
cg01572694	MIR10A	HOXB-AS1	0.06	3.62E-12	3.98E-10
cg26916621	MIR10A	HOXB-AS1	0.08	1.32E-10	1.25E-08
cg14285150		HOXB3;HOXB4;MIR10A	0.04	2.01E-10	1.66E-08
cg14285150		SKAP1	0.13	3.61E-07	2.66E-05
cg01572694	MIR10A	SKAP1	0.07	5.28E-07	3.49E-05
cg14285150		HOXB-AS1	0.07	1.64E-06	9.88E-05
cg26916621	MIR10A	SKAP1	0.09	7.09E-05	3.90E-03
cg23950714	DOK3	FAM153A;FAM153C	-0.07	1.67E-04	7.89E-03
cg23950714	DOK3	FAM153C	-0.10	1.64E-04	7.89E-03
cg23950714	DOK3		0.05	7.92E-04	3.27E-02
cg23950714	DOK3	FAM153C	-0.05	7.87E-04	3.27E-02

1. CpG gene: nearest gene of the CpG; 2. TC gene: transcript cluster. All data in the table comes from summary statistics of the HELIX project. FDR were calculated based on p-value.

Appendix Table 14 DMCs associated traits in EWAS Catalog

CpG	Trait	Tissue	Age	beta	p-value
cg16704797 (hypo)	tissue	buccal cells and PBMC	children	0.158	2.70E-88
	age	whole blood	children	-0.009	1.80E-15
	age	whole blood	children	0.001	1.20E-12
	gestational age	cord blood	infants	-0.0003	5.02E-05
cg01572694 (hyper)	age	whole blood	children	-0.008	0
	age	whole blood	children	0.034	0
	tissue	buccal cells and PBMC	children	0.274	4.00E-116
	age 4 vs 0	whole blood	NA	NA	3.00E-86
	Primary Sjogren's syndrome	whole blood	NA	NA	3.81E-16
	sex	umbilical artery	infants	NA	2.80E-11
	smoking	whole blood	adults	-0.008	5.10E-09
	gestational age	cord blood	infants	0.008	1.80E-07
	smoking	CD4+ T cells, monocyte	adults	-0.008	7.90E-07
	Crohn's disease	whole blood	adults	0.035	9.20E-07
	incident COPD	whole blood	adults (18 - 65 years)	-0.019	5.10E-05
	Inflammatory Bowel disease	whole blood	adults	-0.026	6.00E-05
cg03102887 (hypo)	HNF4A protein level	whole blood	adults (18 - 65 years)	1.58	5.80E-07
	UBE2G2 protein level	whole blood	adults (18 - 65 years)	1.49	2.20E-06
	PRKCG protein level	whole blood	adults (18 - 65 years)	1.44	4.30E-06
	ARHGAP36 protein level	whole blood	adults (18 - 65 years)	1.38	1.30E-05
	TRAPPC3 protein level	whole blood	adults (18 - 65 years)	1.34	2.00E-05
	VAV3 protein level	whole blood	adults (18 - 65 years)	1.31	3.30E-05
	optimal NICU network neurobehavioral scale profile	buccal cells	infants	-0.039	8.30E-05
	VWA2 protein level	whole blood	adults (18 - 65 years)	1.23	9.70E-05
cg23720929 (hypo)	tissue	buccal cells and PBMC	children	0.007	3.50E-79
	APOD gene expression	whole blood	adults	NA	4.30E-15
	Primary Sjogren's syndrome	whole blood	NA	NA	1.92E-09

DMCs associated traits in EWAS Catalog (continued)

CpG	Trait	Tissue	Age	Beta	p-value
cg227310485 (hypo)	tissue	buccal cells and PBMC	children	0.633	3.90E-142
	age	whole blood	children	-0.0005	2.40E-07
	age	whole blood	children	0.0005	5.50E-05
cg15601205 (hypo)	eosinophilia	nasal polyp	adults	-0.01	2.70E-12
cg03037150 (hyper)	Rheumatoid arthritis	whole blood	adults	0.012	8.51E-12
cg23374256 (hypo)	tissue	buccal cells and PBMC	children	-0.021	6.70E-25
	human immunodeficiency virus	whole blood	adults	NA	1.64E-06
	age	whole blood	children	-0.003	5.40E-05
	LRIG1 protein level	whole blood	adults (18 -65 years)	0.86	9.70E-05
cg14285150 (hyper)	age	whole blood	children	-0.006	0
	age	whole blood	children	0.027	0
	tissue	buccal cells and PBMC	children	0.004	1.60E-152
	infant sex	umbilical artery	infants	NA	3.10E-12
	HOXB2 gene expression	whole blood	adults	NA	1.70E-10
	gestational age	cord blood	infants	0.0002	3.28E-06
	gestational age	cord blood	infants	0.0001	3.49E-05
cg03233332 (hypo)	tissue	buccal cells and PBMC	children	0.292	2.80E-71
	age	whole blood	children	0.001	4.20E-15
	alcohol consumption per day	CD14+ monocyte	adults	-0.0002	7.50E-05
cg22818074 (hypo)	incident COPD	whole blood	adults (18 -65 years)	0.024	1.10E-11
	smoking	whole blood	adults	NA	5.00E-07
cg23672176 (hyper)	age	whole blood	children	-3.40E-05	2.70E-06
	CTNNB1 protein level	whole blood	adults (18 -65 years)	-0.91	8.30E-05

DMCs associated traits in EWAS Catalog (continued)

CpG	Trait	Tissue	Age	Beta	p-value
cg23950714 (hypo)	age	whole blood	children	0.002	0
	tissue	buccal cells and PBMC	children	-0.215	2.10E-71
	clear cell renal carcinoma	clear cell renal carcinoma tumor cells, adjacent healthy cells	adults	NA	7.87E-28
	gestational age	cord blood	infants	-0.009	1.20E-14
	Alzheimer's disease braak stage	prefrontal cortex	adults	-0.069	2.00E-10
	gestational age	cord blood	infants	-0.001	4.94E-08
	GZMK protein level	whole blood	adults (18 -65 years)	0.81	4.20E-06
	braak stage	prefrontal cortex	geriatrics	-0.124	1.10E-05
cg17766219 (hyper)	KRT1 protein level	whole blood	adults (18 -65 years)	0.66	3.80E-05
cg00216180 (hypo)	tissue	buccal cells and PBMC	children	0.030	1.70E-27
	age	whole blood	children	-0.001	2.60E-11
	maternal body mass index	cord blood	infants	-0.0003	9.80E-11
	age	whole blood	children	0.002	4.00E-10
	maternal body mass index	cord blood	infants	-0.003	9.30E-09
cg07036914 (hyper)	tissue	buccal cells and PBMC	children	0.085	1.60E-40
	age	whole blood	children	0.002	4.30E-10
	CHST9 protein level	whole blood	adults (18 -65 years)	0.346	8.10E-05
cg19929409 (hypo)	age	whole blood	children	-0.002	0
	age	whole blood	children	0.002	3.90E-31
	Rheumatoid arthritis	whole blood	adults	-0.029	6.98E-10
	age	whole blood	children	0.007	6.30E-06

DMCs associated traits in EWAS Catalog (continued)

CpG	Trait	Tissue	Age	Beta	p-value
cg26724018 (hyper)	age	whole blood	children	-0.010	0
	age	whole blood	children	0.044	0
	tissue	buccal cells and PBMC	children	-0.452	1.50E-151
	gestational age	whole blood	infants	-0.007	4.20E-17
	gestational age	cord blood	infants	-0.001	8.33E-13
	gestational age	cord blood	infants	-0.001	1.18E-12
	gestational age	cord blood	infants	-0.007	1.20E-10
	Primary Sjogren's syndrome	whole blood	NA	NA	1.31E-08
	alcohol consumption per day	whole blood	adults	-7.60E-05	2.49E-07
	birthweight	cord blood	infants	-38.1	8.60E-06
	1-hour glucose	cord blood	infants	-0.306	1.30E-05
	1-hour glucose	cord blood	infants	-1.232	3.30E-05
cg21150327 (hyper)	Incident Type 2 Diabetes	whole blood	adults (18 -65 years)	-0.028	1.80E-05
	prevalent Rheumatoid arthritis (self-reported)	whole blood	adults (18 -65 years)	-0.048	6.40E-05

All data was queried from EWAS Catalog. Hypo: hypomethylated in PedSep-D; Hyper: hypermethylated in PedSep-D. NA indicates no available data.

Bibliography

Abidi, K., I. Khoudri, J. Belayachi, N. Madani, A. Zekraoui, A. A. Zeggwagh and R. Abouqal (2008). "Eosinopenia is a reliable marker of sepsis on admission to medical intensive care units." Crit Care **12**(2): R59.

Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov and S. R. Sunyaev (2010). "A method and server for predicting damaging missense mutations." Nat Methods **7**(4): 248-249.

Ahmad, S., M. A. Choudhry, R. Shankar and M. M. Sayeed (1997). "Transforming growth factor-beta negatively modulates T-cell responses in sepsis." FEBS Lett **402**(2-3): 213-218.

Akoglu, H. (2018). "User's guide to correlation coefficients." Turk J Emerg Med **18**(3): 91-93.

Al Duhailib, Z., M. Farooqi, J. Piticaru, W. Alhazzani and P. Nair (2021). "The role of eosinophils in sepsis and acute respiratory distress syndrome: a scoping review." Can J Anaesth **68**(5): 715-726.

Alexaki, A., J. Kames, D. D. Holcomb, J. Athey, L. V. Santana-Quintero, P. V. N. Lam, N. Hamasaki-Katagiri, E. Osipova, V. Simonyan, H. Bar, A. A. Komar and C. Kimchi-Sarfaty (2019). "Codon and Codon-Pair Usage Tables (CoCoPUTs): Facilitating Genetic Variation Analyses and Recombinant Gene Design." J Mol Biol **431**(13): 2434-2441.

Ames, S. G., C. M. Horvat, A. Zaritsky and J. A. Carcillo (2018). "The path to great pediatric septic shock outcomes." Crit Care **22**(1): 224.

Angurana, S. K., A. Bansal, J. Muralidharan, R. Aggarwal and S. Singhi (2021). "Cytokine Levels in Critically Ill Children With Severe Sepsis and Their Relation With the Severity of Illness and Mortality." J Intensive Care Med **36**(5): 576-583.

Arango Duque, G. and A. Descoteaux (2014). "Macrophage cytokines: involvement in immunity and infectious diseases." Front Immunol **5**: 491.

Aryee, M. J., A. E. Jaffe, H. Corrada-Bravo, C. Ladd-Acosta, A. P. Feinberg, K. D. Hansen and R. A. Irizarry (2014). "Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays." Bioinformatics **30**(10): 1363-1369.

Asgari, S., P. J. McLaren, J. Peake, M. Wong, R. Wong, I. Bartha, J. R. Francis, K. Abarca, K. A. Gelderman, P. Agyeman, C. Aebi, C. Berger, J. Fellay, L. J. Schlapbach and S. Swiss Pediatric Sepsis (2016). "Exome Sequencing Reveals Primary Immunodeficiencies in Children with Community-Acquired Pseudomonas aeruginosa Sepsis." Front Immunol **7**: 357.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nat Genet **25**(1): 25-29.

Astle, W. J., H. Elding, T. Jiang, D. Allen, D. Ruklisa, A. L. Mann, D. Mead, H. Bouman, F. Riveros-Mckay, M. A. Kostadima, J. J. Lambourne, S. Sivapalaratnam, K. Downes, K. Kundu, L. Bomba, K. Berentsen, J. R. Bradley, L. C. Daugherty, O. Delaneau, K. Freson, S. F. Garner, L. Grassi, J. Guerrero, M. Haimel, E. M. Janssen-Megens, A. Kaan, M. Kamat, B. Kim, A. Mandoli, J. Marchini, J. H. A. Martens, S. Meacham, K. Megy, J. O'Connell, R. Petersen, N. Sharifi, S. M. Sheard, J. R. Staley, S. Tuna, M. van der Ent, K. Walter, S. Y. Wang, E. Wheeler, S. P. Wilder, V. Iotchkova, C. Moore, J. Sambrook, H. G. Stunnenberg, E. Di Angelantonio, S. Kaptoge, T. W. Kuijpers, E. Carrillo-de-Santa-Pau, D. Juan, D. Rico, A. Valencia, L. Chen, B. Ge, L. Vasquez, T. Kwan, D. Garrido-Martin, S. Watt, Y. Yang, R. Guigo, S. Beck, D. S. Paul, T. Pastinen, D. Bujold, G. Bourque, M. Frontini, J. Danesh, D. J. Roberts, W. H. Ouwehand, A. S. Butterworth and N. Soranzo (2016). "The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease." Cell **167**(5): 1415-1429 e1419.

Atreya, M. R. and H. R. Wong (2019). "Precision medicine in pediatric sepsis." Curr Opin Pediatr **31**(3): 322-327.

Aviv, T., Z. Lin, S. Lau, L. M. Rendl, F. Sicheri and C. A. Smibert (2003). "The RNA-binding SAM domain of Smaug defines a new family of post-transcriptional regulators." Nat Struct Biol **10**(8): 614-621.

Ayala, A., J. B. Knotts, W. Ertel, M. M. Perrin, M. H. Morrison and I. H. Chaudry (1993). "Role of interleukin 6 and transforming growth factor-beta in the induction of depressed splenocyte responses following sepsis." Arch Surg **128**(1): 89-94; discussion 94-85.

Backman, J. D., A. H. Li, A. Marcketta, D. Sun, J. Mbatchou, M. D. Kessler, C. Benner, D. Liu, A. E. Locke, S. Balasubramanian, A. Yadav, N. Banerjee, C. E. Gillies, A. Damask, S. Liu, X. Bai, A. Hawes, E. Maxwell, L. Gurski, K. Watanabe, J. A. Kosmicki, V. Rajagopal, J. Mighty, C. Regeneron Genetics, DiscovEhr, M. Jones, L. Mitnaul, E. Stahl, G. Coppola, E. Jorgenson, L. Habegger, W. J. Salerno, A. R. Shuldiner, L. A. Lotta, J. D. Overton, M. N. Cantor, J. G. Reid, G. Yancopoulos, H. M. Kang, J. Marchini, A. Baras, G. R. Abecasis and M. A. R. Ferreira (2021). "Exome sequencing and analysis of 454,787 UK Biobank participants." Nature **599**(7886): 628-634.

Bateman, R. M., M. D. Sharpe and C. G. Ellis (2003). "Bench-to-bedside review: microvascular dysfunction in sepsis--hemodynamics, oxygen transport, and nitric oxide." Crit Care **7**(5): 359-373.

Battram, T., P. Yousefi, G. Crawford, C. Prince, M. Sheikhali Babaei, G. Sharp, C. Hatcher, M. J. Vega-Salas, S. Khodabakhsh, O. Whitehurst, R. Langdon, L. Mahoney, H. R. Elliott, G. Mancano, M. A. Lee, S. H. Watkins, A. C. Lay, G. Hemani, T. R. Gaunt, C. L. Relton, J.

R. Staley and M. Suderman (2022). "The EWAS Catalog: a database of epigenome-wide association studies." Wellcome Open Res **7**: 41.

Beppler, J., P. Koehler-Santos, G. Pasqualim, U. Matte, C. S. Alho, F. S. Dias, T. W. Kowalski, I. T. Velasco, R. C. Monteiro and F. Pinheiro da Silva (2016). "Fc Gamma Receptor IIA (CD32A) R131 Polymorphism as a Marker of Genetic Susceptibility to Sepsis." Inflammation **39**(2): 518-525.

Bhavani, S. V., K. A. Carey, E. R. Gilbert, M. Afshar, P. A. Verhoef and M. M. Churpek (2019). "Identifying Novel Sepsis Subphenotypes Using Temperature Trajectories." Am J Respir Crit Care Med **200**(3): 327-335.

Binnie, A., C. J. Walsh, P. Hu, D. J. Dwivedi, A. Fox-Robichaud, P. C. Liaw, J. L. Y. Tsang, J. Batt, G. Carrasqueiro, S. Gupta, J. C. Marshall, P. Castelo-Branco, C. C. Dos Santos and G. Epigenetic Profiling in Severe Sepsis Study of the Canadian Critical Care Translational Biology (2020). "Epigenetic Profiling in Severe Sepsis: A Pilot Study of DNA Methylation Profiles in Critical Illness." Crit Care Med **48**(2): 142-150.

Bird, A. (2002). "DNA methylation patterns and epigenetic memory." Genes Dev **16**(1): 6-21.

Borghesi, A., J. Truck, S. Asgari, V. Sancho-Shimizu, P. K. A. Agyeman, E. Bellos, E. Giannoni, M. Stocker, K. M. Posfay-Barbe, U. Heininger, S. Bernhard-Stirnemann, A. Niederer-Loher, C. R. Kahlert, G. Natalucci, C. Relly, T. Riedel, C. E. Kuehni, C. W. Thorball, N. Chaturvedi, F. Martinon-Torres, T. W. Kuijpers, L. Coin, V. Wright, J. Herberg, M. Levin, C. Aebi, C. Berger, J. Fellay and L. J. Schlapbach (2020). "Whole-exome Sequencing for the Identification of Rare Variants in Primary Immunodeficiency Genes in Children With Sepsis: A Prospective, Population-based Cohort Study." Clin Infect Dis **71**(10): e614-e623.

Bowden, J., G. Davey Smith and S. Burgess (2015). "Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression." Int J Epidemiol **44**(2): 512-525.

Butcher, L. M. and S. Beck (2015). "Probe Lasso: a novel method to rope in differentially methylated regions with 450K DNA methylation data." Methods **72**: 21-28.

Butler-Laporte, G., A. Harroud, V. Forgetta and J. B. Richards (2020). "Elevated body mass index is associated with an increased risk of infectious disease admissions and mortality: a mendelian randomization study." Clin Microbiol Infect.

Bycroft, C., C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N. Allen, P. Donnelly and J. Marchini (2018). "The UK Biobank resource with deep phenotyping and genomic data." Nature **562**(7726): 203-209.

Calfee, C. S., K. Delucchi, P. E. Parsons, B. T. Thompson, L. B. Ware, M. A. Matthay and N. A. Network (2014). "Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials." Lancet Respir Med **2**(8): 611-620.

Campagna, M. P., A. Xavier, J. Lechner-Scott, V. Maltby, R. J. Scott, H. Butzkueven, V. G. Jokubaitis and R. A. Lea (2021). "Epigenome-wide association studies: current knowledge, strategies and recommendations." Clin Epigenetics **13**(1): 214.

Cao, L., T. Zhu, X. Lang, S. Jia, Y. Yang, C. Zhu, Y. Wang, S. Feng, C. Wang, P. Zhang, J. Chen and H. Jiang (2020). "Inhibiting DNA Methylation Improves Survival in Severe Sepsis by Regulating NF-kappaB Pathway." Front Immunol **11**: 1360.

Capestrano, M., S. Mariggio, G. Perinetti, A. V. Egorova, S. Iacobacci, M. Santoro, A. Di Pentima, C. Iurisci, M. V. Egorov, G. Di Tullio, R. Buccione, A. Luini and R. S. Polishchuk (2014). "Cytosolic phospholipase A(2)epsilon drives recycling through the clathrin-independent endocytic route." J Cell Sci **127**(Pt 5): 977-993.

Carcillo, J. A., R. A. Berg, D. Wessel, M. Pollack, K. Meert, M. Hall, C. Newth, J. C. Lin, A. Doctor, T. Shanley, T. Cornell, R. E. Harrison, A. F. Zuppa, R. W. Reeder, R. Banks, J. A. Kellum, R. Holubkov, D. A. Notterman, J. M. Dean, H. Eunice Kennedy Shriver National Institute of Child and N. Human Development Collaborative Pediatric Critical Care Research (2019). "A Multicenter Network Assessment of Three Inflammation Phenotypes in Pediatric Sepsis-Induced Multiple Organ Failure." Pediatr Crit Care Med **20**(12): 1137-1146.

Carcillo, J. A., B. Podd, R. Aneja, S. L. Weiss, M. W. Hall, T. T. Cornell, T. P. Shanley, L. A. Doughty and T. C. Nguyen (2017). "Pathophysiology of Pediatric Multiple Organ Dysfunction Syndrome." Pediatr Crit Care Med **18**(3_suppl Suppl 1): S32-S45.

Carter, M. J., M. Fish, A. Jennings, K. J. Doores, P. Wellman, J. Seow, S. Acors, C. Graham, E. Timms, J. Kenny, S. Neil, M. H. Malim, S. M. Tibby and M. Shankar-Hari (2020). "Peripheral immunophenotypes in children with multisystem inflammatory syndrome associated with SARS-CoV-2 infection." Nat Med **26**(11): 1701-1707.

Cavaillon, J. M., M. Singer and T. Skirecki (2020). "Sepsis therapies: learning from 30 years of failure of translational research to propose new leads." EMBO Mol Med **12**(4): e10128.

Chang, S. F., S. F. Liu, C. N. Chen and H. C. Kuo (2020). "Serum IP-10 and IL-17 from Kawasaki disease patients induce calcification-related genes and proteins in human coronary artery smooth muscle cells in vitro." Cell Biosci **10**: 36.

Chen, E. Y., C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. V. Meirelles, N. R. Clark and A. Ma'ayan (2013). "Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool." BMC Bioinformatics **14**: 128.

Chen, M. H., L. M. Raffield, A. Mousas, S. Sakaue, J. E. Huffman, A. Moscati, B. Trivedi, T. Jiang, P. Akbari, D. Vuckovic, E. L. Bao, X. Zhong, R. Manansala, V. Laplante, M. Chen, K. S. Lo, H. Qian, C. A. Lareau, M. Beaudoin, K. A. Hunt, M. Akiyama, T. M. Bartz, Y. Ben-Shlomo, A. Beswick, J. Bork-Jensen, E. P. Bottinger, J. A. Brody, F. J. A. van Rooij, K. Chitrala, K. Cho, H. Choquet, A. Correa, J. Danesh, E. Di Angelantonio, N. Dimou, J. Ding, P. Elliott, T. Esko, M. K. Evans, J. S. Floyd, L. Broer, N. Grarup, M. H. Guo, A. Greinacher, J. Haessler, T. Hansen, J. M. M. Howson, Q. Q. Huang, W. Huang, E.

Jorgenson, T. Kacprowski, M. Kahonen, Y. Kamatani, M. Kanai, S. Karthikeyan, F. Koskeridis, L. A. Lange, T. Lehtimaki, M. M. Lerch, A. Linneberg, Y. Liu, L. P. Lyytikainen, A. Manichaikul, H. C. Martin, K. Matsuda, K. L. Mohlke, N. Mononen, Y. Murakami, G. N. Nadkarni, M. Nauck, K. Nikus, W. H. Ouwehand, N. Pankratz, O. Pedersen, M. Preuss, B. M. Psaty, O. T. Raitakari, D. J. Roberts, S. S. Rich, B. A. T. Rodriguez, J. D. Rosen, J. I. Rotter, P. Schubert, C. N. Spracklen, P. Surendran, H. Tang, J. C. Tardif, R. C. Trembath, M. Ghanbari, U. Volker, H. Volzke, N. A. Watkins, A. B. Zonderman, V. A. M. V. Program, P. W. F. Wilson, Y. Li, A. S. Butterworth, J. F. Gauchat, C. W. K. Chiang, B. Li, R. J. F. Loos, W. J. Astle, E. Evangelou, D. A. van Heel, V. G. Sankaran, Y. Okada, N. Soranzo, A. D. Johnson, A. P. Reiner, P. L. Auer and G. Lettre (2020). "Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations." Cell **182**(5): 1198-1213 e1114.

Chen, X., L. Wang, J. D. Smith and B. Zhang (2008). "Supervised principal component analysis for gene set enrichment of microarray data with continuous or survival outcomes." Bioinformatics **24**(21): 2474-2481.

Chen, Y. T. and D. M. Witten (2023). "Selective inference for k-means clustering." J Mach Learn Res **24**.

Cheng, Z., S. T. Abrams, J. Toh, S. S. Wang, Z. Wang, Q. Yu, W. Yu, C. H. Toh and G. Wang (2020). "The Critical Roles and Mechanisms of Immune Cell Death in Sepsis." Front Immunol **11**: 1918.

Chiche, L., J. M. Forel, G. Thomas, C. Farnarier, F. Vely, M. Blery, L. Papazian and E. Vivier (2011). "The role of natural killer cells in sepsis." J Biomed Biotechnol **2011**: 986491.

Consortium, G. T. (2013). "The Genotype-Tissue Expression (GTEx) project." Nat Genet **45**(6): 580-585.

Cook, A., S. Janse, J. R. Watson and G. Erdem (2020). "Manifestations of Toxic Shock Syndrome in Children, Columbus, Ohio, USA, 2010-2017(1)." Emerg Infect Dis **26**(6): 1077-1083.

D. Benet Bosco Dhas, A. H., B. VishnuBhat, S.Kalaivani, Subash Chandra Parija (2015). "Comparison of genomic DNA methylation pattern among septic and non-septic newborns — An epigenome wide association study." Genomics Data **3**: 36-40.

D'Urso, S., D. Rajbhandari, E. Peach, E. de Guzman, Q. Li, S. E. Medland, S. D. Gordon, N. G. Martin, C. I. W. Group, S. Ligthart, M. A. Brown, J. Powell, C. McArthur, A. Rhodes, J. Meyer, S. Finfer, J. Myburgh, A. Blumenthal, J. Cohen, B. Venkatesh, G. Cuellar-Partida and D. M. Evans (2020). "Septic Shock: A Genomewide Association Study and Polygenic Risk Score Analysis." Twin Res Hum Genet **23**(4): 204-213.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin and G. Genomes Project Analysis (2011). "The variant call format and VCFtools." Bioinformatics **27**(15): 2156-2158.

Das, K. and L. V. M. Rao (2022). "The Role of microRNAs in Inflammation." Int J Mol Sci **23**(24).

Davenport, E. E., K. L. Burnham, J. Radhakrishnan, P. Humburg, P. Hutton, T. C. Mills, A. Rautanen, A. C. Gordon, C. Garrard, A. V. Hill, C. J. Hinds and J. C. Knight (2016). "Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study." Lancet Respir Med **4**(4): 259-271.

Davydov, E. V., D. L. Goode, M. Sirota, G. M. Cooper, A. Sidow and S. Batzoglou (2010). "Identifying a high fraction of the human genome to be under selective constraint using GERP++." PLoS Comput Biol **6**(12): e1001025.

de Pablo, R., J. Monserrat, E. Reyes, D. Diaz, M. Rodriguez-Zapata, A. la Hera, A. Prieto and M. Alvarez-Mon (2012). "Sepsis-induced acute respiratory distress syndrome with fatal outcome is associated to increased serum transforming growth factor beta-1 levels." Eur J Intern Med **23**(4): 358-362.

DeMerle, K. M., D. C. Angus, J. K. Baillie, E. Brant, C. S. Calfee, J. Carcillo, C. H. Chang, R. Dickson, I. Evans, A. C. Gordon, J. Kennedy, J. C. Knight, C. J. Lindsell, V. Liu, J. C. Marshall, A. G. Randolph, B. P. Scicluna, M. Shankar-Hari, N. I. Shapiro, T. E. Sweeney, V. B. Talisa, B. Tang, B. T. Thompson, E. L. Tsalik, T. van der Poll, L. A. van Vught, H. R. Wong, S. Yende, H. Zhao and C. W. Seymour (2021). "Sepsis Subclasses: A Framework for Development and Interpretation." Crit Care Med **49**(5): 748-759.

Demirkol, D., D. Yildizdas, B. Bayrakci, B. Karapinar, T. Kendirli, T. F. Koroglu, O. Dursun, N. Erkek, H. Gedik, A. Citak, S. Kesici, M. Karabocuoglu, J. A. Carcillo and H. L. H. M. A. S. C. C. S. G. Turkish Secondary (2012). "Hyperferritinemia in the critically ill child with secondary hemophagocytic lymphohistiocytosis/sepsis/multiple organ dysfunction syndrome/macrophage activation syndrome: what is the treatment?" Crit Care **16**(2): R52.

Derkach, A., J. F. Lawless and L. Sun (2013). "Robust and powerful tests for rare variants using Fisher's method to combine evidence of association from two or more complementary tests." Genet Epidemiol **37**(1): 110-121.

Dhas, D. B., A. H. Ashmi, B. V. Bhat, S. Kalaivani and S. C. Parija (2015). "Comparison of genomic DNA methylation pattern among septic and non-septic newborns - An epigenome wide association study." Genom Data **3**: 36-40.

Dienstmann, R., L. Vermeulen, J. Guinney, S. Kopetz, S. Tejpar and J. Tabernero (2017). "Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer." Nat Rev Cancer **17**(4): 268.

Dolin, H. H., T. J. Papadimos, X. Chen and Z. K. Pan (2019). "Characterization of Pathogenic Sepsis Etiologies and Patient Profiles: A Novel Approach to Triage and Treatment." Microbiol Insights **12**: 1178636118825081.

Dong, C., R. J. Davis and R. A. Flavell (2002). "MAP kinases in the immune response." Annu Rev Immunol **20**: 55-72.

Doughty, L., R. S. Clark, S. S. Kaplan, H. Sasser and J. Carcillo (2002). "sFas and sFas ligand and pediatric sepsis-induced multiple organ failure syndrome." Pediatr Res **52**(6): 922-927.

Downing, A. K., V. Knott, J. M. Werner, C. M. Cardy, I. D. Campbell and P. A. Handford (1996). "Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders." Cell **85**(4): 597-605.

Ebato, T., S. Ogata, Y. Ogihara, M. Fujimoto, A. Kitagawa, M. Takanashi and M. Ishii (2017). "The Clinical Utility and Safety of a New Strategy for the Treatment of Refractory Kawasaki Disease." J Pediatr **191**: 140-144.

Emmenegger, U., U. Frey, A. Reimers, C. Fux, D. Semela, P. Cottagnoud, P. J. Spaeth and K. A. Neftel (2001). "Hyperferritinemia as indicator for intravenous immunoglobulin treatment in reactive macrophage activation syndromes." Am J Hematol **68**(1): 4-10.

Eun, S., H. Kim, H. Y. Kim, M. Lee, G. E. Bae, H. Kim, C. M. Koo, M. K. Kim and S. H. Yoon (2021). "Age-adjusted quick Sequential Organ Failure Assessment score for predicting mortality and disease severity in children with infection: a systematic review and meta-analysis." Sci Rep **11**(1): 21699.

Evangelou, E., H. R. Warren, D. Mosen-Ansorena, B. Mifsud, R. Pazoki, H. Gao, G. Ntritsos, N. Dimou, C. P. Cabrera, I. Karaman, F. L. Ng, M. Evangelou, K. Witkowska, E. Tzanis, J. N. Hellwege, A. Giri, D. R. Velez Edwards, Y. V. Sun, K. Cho, J. M. Gaziano, P. W. F. Wilson, P. S. Tsao, C. P. Kovesdy, T. Esko, R. Magi, L. Milani, P. Almgren, T. Boutin, S. Debette, J. Ding, F. Giulianini, E. G. Holliday, A. U. Jackson, R. Li-Gao, W. Y. Lin, J. Luan, M. Mangino, C. Oldmeadow, B. P. Prins, Y. Qian, M. Sargurupremraj, N. Shah, P. Surendran, S. Theriault, N. Verweij, S. M. Willems, J. H. Zhao, P. Amouyel, J. Connell, R. de Mutsert, A. S. F. Doney, M. Farrall, C. Menni, A. D. Morris, R. Noordam, G. Pare, N. R. Poulter, D. C. Shields, A. Stanton, S. Thom, G. Abecasis, N. Amin, D. E. Arking, K. L. Ayers, C. M. Barbieri, C. Batini, J. C. Bis, T. Blake, M. Bochud, M. Boehnke, E. Boerwinkle, D. I. Boomsma, E. P. Bottinger, P. S. Braund, M. Brumat, A. Campbell, H. Campbell, A. Chakravarti, J. C. Chambers, G. Chauhan, M. Ciullo, M. Cocca, F. Collins, H. J. Cordell, G. Davies, M. H. de Borst, E. J. de Geus, I. J. Deary, J. Deelen, M. F. Del Greco, C. Y. Demirkale, M. Dorr, G. B. Ehret, R. Elosua, S. Enroth, A. M. Erzurumluoglu, T. Ferreira, M. Franberg, O. H. Franco, I. Gandin, P. Gasparini, V. Giedraitis, C. Gieger, G. Girotto, A. Goel, A. J. Gow, V. Gudnason, X. Guo, U. Gyllensten, A. Hamsten, T. B. Harris, S. E. Harris, C. A. Hartman, A. S. Havulinna, A. A. Hicks, E. Hofer, A. Hofman, J. J. Hottenga, J. E. Huffman, S. J. Hwang, E. Ingelsson, A. James, R. Jansen, M. R. Jarvelin, R. Joehanes, A. Johansson, A. D. Johnson, P. K. Joshi, P. Jousilahti, J. W. Jukema, A. Jula, M. Kahonen, S. Kathiresan, B. D. Keavney, K. T. Khaw, P. Knekt, J. Knight, I. Kolcic, J. S. Kooner, S. Koskinen, K. Kristiansson, Z. Kutalik, M. Laan, M. Larson, L. J. Launer, B. Lehne, T. Lehtimaki, D. C. M. Liewald, L. Lin, L. Lind, C. M. Lindgren, Y. Liu, R. J. F. Loos, L. M. Lopez, Y. Lu, L. P. Lyytikainen, A. Mahajan, C. Mamasoula, J. Marrugat, J. Marten, Y. Milaneschi, A. Morgan, A. P. Morris, A. C. Morrison, P. J. Munson, M. A. Nalls, P. Nandakumar, C. P. Nelson, T. Niiranen, I. M. Nolte, T. Nutile, A. J. Oldehinkel, B. A. Oostra, P. F. O'Reilly, E. Org, S. Padmanabhan, W. Palmas, A. Palotie, A. Pattie, B. Penninx, M. Perola, A. Peters, O. Polasek, P. P. Pramstaller, Q. T. Nguyen, O. T. Raitakari, M. Ren, R. Rettig, K. Rice, P. M. Ridker, J. S. Ried, H. Riese, S. Ripatti, A. Robino, L. M. Rose, J. I. Rotter, I. Rudan, D. Ruggiero, Y. Saba, C. F. Sala, V. Salomaa, N. J. Samani, A. P. Sarin, R. Schmidt, H. Schmidt, N. Shrine, D. Siscovick, A. V. Smith,

H. Snieder, S. Sober, R. Sorice, J. M. Starr, D. J. Stott, D. P. Strachan, R. J. Strawbridge, J. Sundstrom, M. A. Swertz, K. D. Taylor, A. Teumer, M. D. Tobin, M. Tomaszewski, D. Toniolo, M. Traglia, S. Trompet, J. Tuomilehto, C. Tzourio, A. G. Uitterlinden, A. Vaez, P. J. van der Most, C. M. van Duijn, A. C. Vergnaud, G. C. Verwoert, V. Vitart, U. Volker, P. Vollenweider, D. Vuckovic, H. Watkins, S. H. Wild, G. Willemsen, J. F. Wilson, A. F. Wright, J. Yao, T. Zemunik, W. Zhang, J. R. Attia, A. S. Butterworth, D. I. Chasman, D. Conen, F. Cucca, J. Danesh, C. Hayward, J. M. M. Howson, M. Laakso, E. G. Lakatta, C. Langenberg, O. Melander, D. O. Mook-Kanamori, C. N. A. Palmer, L. Risch, R. A. Scott, R. J. Scott, P. Sever, T. D. Spector, P. van der Harst, N. J. Wareham, E. Zeggini, D. Levy, P. B. Munroe, C. Newton-Cheh, M. J. Brown, A. Metspalu, A. M. Hung, C. J. O'Donnell, T. L. Edwards, B. M. Psaty, I. Tzoulaki, M. R. Barnes, L. V. Wain, P. Elliott, M. J. Caulfield and P. Million Veteran (2018). "Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits." Nat Genet **50**(10): 1412-1425.

Evans, I. V. R., G. S. Phillips, E. R. Alpern, D. C. Angus, M. E. Friedrich, N. Kissoon, S. Lemeshow, M. M. Levy, M. M. Parker, K. M. Terry, R. S. Watson, S. L. Weiss, J. Zimmerman and C. W. Seymour (2018). "Association Between the New York Sepsis Care Mandate and In-Hospital Mortality for Pediatric Sepsis." JAMA **320**(4): 358-367.

Falcao-Holanda, R. B., M. K. C. Brunialti, M. G. Jasiulionis and R. Salomao (2021). "Epigenetic Regulation in Sepsis, Role in Pathophysiology and Therapeutic Perspective." Front Med (Lausanne) **8**: 685333.

Fitzer-Attas, C. J., M. Lowry, M. T. Crowley, A. J. Finn, F. Meng, A. L. DeFranco and C. A. Lowell (2000). "Fcgamma receptor-mediated phagocytosis in macrophages lacking the Src family tyrosine kinases Hck, Fgr, and Lyn." J Exp Med **191**(4): 669-682.

Fleischmann-Struzek, C., D. M. Goldfarb, P. Schlattmann, L. J. Schlapbach, K. Reinhart and N. Kissoon (2018). "The global burden of paediatric and neonatal sepsis: a systematic review." Lancet Respir Med **6**(3): 223-230.

Flores, C. (2015). "Host genetics shapes adult sepsis survival." Lancet Respir Med **3**(1): 7-8.

Fortenberry, J. D., T. Nguyen, J. R. Grunwell, R. K. Aneja, D. Wheeler, M. Hall, G. Fleming, R. Tarrago, S. Buttram, H. Dalton, Y. Han, K. A. Easley, A. Knezevic, T. Dai, M. Paden, J. A. Carcillo and G. Thrombocytopenia-Associated Multiple Organ Failure Network Study (2019). "Therapeutic Plasma Exchange in Children With Thrombocytopenia-Associated Multiple Organ Failure: The Thrombocytopenia-Associated Multiple Organ Failure Network Prospective Experience." Crit Care Med **47**(3): e173-e181.

Gadd, D. A., R. F. Hillary, D. L. McCartney, L. Shi, A. Stolicyn, N. A. Robertson, R. M. Walker, R. I. McGeachan, A. Campbell, S. Xueyi, M. C. Barbu, C. Green, S. W. Morris, M. A. Harris, E. V. Backhouse, J. M. Wardlaw, J. D. Steele, D. A. Oyarzun, G. Muniz-Terrera, C. Ritchie, A. Nevado-Holgado, T. Chandra, C. Hayward, K. L. Evans, D. J. Porteous, S. R. Cox, H. C. Whalley, A. M. McIntosh and R. E. Marioni (2022). "Integrated methylome

and phenome study of the circulating proteome reveals markers pertinent to brain health." Nat Commun **13**(1): 4670.

Gaines, N. N., B. Patel, E. A. Williams and A. T. Cruz (2012). "Etiologies of septic shock in a pediatric emergency department population." Pediatr Infect Dis J **31**(11): 1203-1205.

Gao, Q., Y. He, Z. Yuan, J. Zhao, B. Zhang and F. Xue (2011). "Gene- or region-based association study via kernel principal component analysis." BMC Genet **12**: 75.

Garcia-Lazaro, J. F., F. Thieringer, S. Luth, P. Czochra, E. Meyer, I. B. Renteria, P. R. Galle, A. W. Lohse, J. Herkel and S. Kanzler (2005). "Hepatic over-expression of TGF-beta1 promotes LPS-induced inflammatory cytokine secretion by liver cells and endotoxemic shock." Immunol Lett **101**(2): 217-222.

Gardiner-Garden, M. and M. Frommer (1987). "CpG islands in vertebrate genomes." J Mol Biol **196**(2): 261-282.

Gene Ontology, C., S. A. Aleksander, J. Balhoff, S. Carbon, J. M. Cherry, H. J. Drabkin, D. Ebert, M. Feuermann, P. Gaudet, N. L. Harris, D. P. Hill, R. Lee, H. Mi, S. Moxon, C. J. Mungall, A. Muruganugan, T. Mushayahama, P. W. Sternberg, P. D. Thomas, K. Van Auken, J. Ramsey, D. A. Siegele, R. L. Chisholm, P. Fey, M. C. Aspromonte, M. V. Nugnes, F. Quaglia, S. Tosatto, M. Giglio, S. Nadendla, G. Antonazzo, H. Attrill, G. Dos Santos, S. Marygold, V. Strelets, C. J. Tabone, J. Thurmond, P. Zhou, S. H. Ahmed, P. Asanitthong, D. Luna Buitrago, M. N. Erdol, M. C. Gage, M. Ali Kadhum, K. Y. C. Li, M. Long, A. Michalak, A. Pesala, A. Pritazahra, S. C. C. Saverimuttu, R. Su, K. E. Thurlow, R. C. Lovering, C. Logie, S. Oliferenko, J. Blake, K. Christie, L. Corbani, M. E. Dolan, H. J. Drabkin, D. P. Hill, L. Ni, D. Sitnikov, C. Smith, A. Cuzick, J. Seager, L. Cooper, J. Elser, P. Jaiswal, P. Gupta, P. Jaiswal, S. Naithani, M. Lera-Ramirez, K. Rutherford, V. Wood, J. L. De Pons, M. R. Dwinell, G. T. Hayman, M. L. Kaldunski, A. E. Kwitek, S. J. F. Laulederkind, M. A. Tutaj, M. Vedi, S. J. Wang, P. D'Eustachio, L. Aimo, K. Axelsen, A. Bridge, N. Hyka-Nouspikel, A. Morgat, S. A. Aleksander, J. M. Cherry, S. R. Engel, K. Karra, S. R. Miyasato, R. S. Nash, M. S. Skrzypek, S. Weng, E. D. Wong, E. Bakker, T. Z. Berardini, L. Reiser, A. Auchincloss, K. Axelsen, G. Argoud-Puy, M. C. Blatter, E. Boutet, L. Breuza, A. Bridge, C. Casals-Casas, E. Coudert, A. Estreicher, M. Livia Famiglietti, M. Feuermann, A. Gos, N. Gruaz-Gumowski, C. Hulo, N. Hyka-Nouspikel, F. Jungo, P. Le Mercier, D. Lieberherr, P. Masson, A. Morgat, I. Pedruzzi, L. Pourcel, S. Poux, C. Rivoire, S. Sundaram, A. Bateman, E. Bowler-Barnett, A. J. H. Bye, P. Denny, A. Ignatchenko, R. Ishtiaq, A. Lock, Y. Lussi, M. Magrane, M. J. Martin, S. Orchard, P. Raposo, E. Speretta, N. Tyagi, K. Warner, R. Zaru, A. D. Diehl, R. Lee, J. Chan, S. Diamantakis, D. Raciti, M. Zarowiecki, M. Fisher, C. James-Zorn, V. Ponferrada, A. Zorn, S. Ramachandran, L. Ruzicka and M. Westerfield (2023). "The Gene Ontology knowledgebase in 2023." Genetics **224**(1).

Gold, L., D. Ayers, J. Bertino, C. Bock, A. Bock, E. N. Brody, J. Carter, A. B. Dalby, B. E. Eaton, T. Fitzwater, D. Flather, A. Forbes, T. Foreman, C. Fowler, B. Gawande, M. Goss, M. Gunn, S. Gupta, D. Halladay, J. Heil, J. Heilig, B. Hicke, G. Husar, N. Janjic, T. Jarvis, S. Jennings, E. Katilius, T. R. Keeney, N. Kim, T. H. Koch, S. Kraemer, L. Kroiss, N. Le, D.

Levine, W. Lindsey, B. Lollo, W. Mayfield, M. Mehan, R. Mehler, S. K. Nelson, M. Nelson, D. Nieuwlandt, M. Nikrad, U. Ochsner, R. M. Ostroff, M. Otis, T. Parker, S. Pietrasiewicz, D. I. Resnicow, J. Rohloff, G. Sanders, S. Sattin, D. Schneider, B. Singer, M. Stanton, A. Sterkel, A. Stewart, S. Stratford, J. D. Vaught, M. Vrkljan, J. J. Walker, M. Watrobka, S. Waugh, A. Weiss, S. K. Wilcox, A. Wolfson, S. K. Wolk, C. Zhang and D. Zichi (2010). "Aptamer-based multiplexed proteomic technology for biomarker discovery." PLoS One **5**(12): e15004.

Goldberg, A. D., C. D. Allis and E. Bernstein (2007). "Epigenetics: a landscape takes shape." Cell **128**(4): 635-638.

Goldstein, B., B. Giroir, A. Randolph and S. International Consensus Conference on Pediatric (2005). "International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics." Pediatr Crit Care Med **6**(1): 2-8.

Hagenaars J.A., M. A. L. (2002). "Applied latent class analysis." Cambridge University Press.

Hall, M. W., N. L. Knatz, C. Vetterly, S. Tomarello, M. D. Wewers, H. D. Volk and J. A. Carcillo (2011). "Immunoparalysis and nosocomial infection in children with multiple organ dysfunction syndrome." Intensive Care Med **37**(3): 525-532.

Han, L., P. D. Witmer, E. Casey, D. Valle and S. Sukumar (2007). "DNA methylation regulates MicroRNA expression." Cancer Biol Ther **6**(8): 1284-1288.

Hao, Q. L., D. K. Ferris, G. White, N. Heisterkamp and J. Groffen (1991). "Nuclear and cytoplasmic location of the FER tyrosine kinase." Mol Cell Biol **11**(2): 1180-1183.

Hartman, M. E., W. T. Linde-Zwirble, D. C. Angus and R. S. Watson (2013). "Trends in the epidemiology of pediatric severe sepsis*." Pediatr Crit Care Med **14**(7): 686-693.

Hemani, G., J. Zheng, B. Elsworth, K. H. Wade, V. Haberland, D. Baird, C. Laurin, S. Burgess, J. Bowden, R. Langdon, V. Y. Tan, J. Yarmolinsky, H. A. Shihab, N. J. Timpson, D. M. Evans, C. Relton, R. M. Martin, G. Davey Smith, T. R. Gaunt and P. C. Haycock (2018). "The MR-Base platform supports systematic causal inference across the human phenome." Elife **7**.

Hernandez-Beeftink, T., B. Guillen-Guio, J. M. Lorenzo-Salazar, A. Corrales, E. Suarez-Pajes, R. Feng, L. A. Rubio-Rodriguez, M. L. Paynton, R. Cruz, M. I. Garcia-Laorden, M. Prieto-Gonzalez, A. Rodriguez-Perez, D. Carriedo, J. Blanco, A. Ambros, E. Gonzalez-Higueras, E. Espinosa, A. Muriel, E. Tamayo, M. M. Martin, L. Lorente, D. Dominguez, A. G. de Lorenzo, H. M. Giannini, J. P. Reilly, T. K. Jones, J. M. Anon, M. Soro, A. Carracedo, L. V. Wain, N. J. Meyer, J. Villar, C. Flores and N. Genetics of Sepsis (2022). "A genome-wide association study of survival in patients with sepsis." Crit Care **26**(1): 341.

Hornik K, B. C., Zeileis A (2009). "Open-Source Machine Learning: R Meets Weka." Computational Statistics **24**(2): 225-232.

Horvat, C. M., J. Bell, S. Kantawala, A. K. Au, R. S. B. Clark and J. A. Carcillo (2019). "C-Reactive Protein and Ferritin Are Associated With Organ Dysfunction and Mortality in Hospitalized Children." Clin Pediatr (Phila) **58**(7): 752-760.

Hotchkiss, R. S., L. L. Moldawer, S. M. Opal, K. Reinhart, I. R. Turnbull and J. L. Vincent (2016). "Sepsis and septic shock." Nat Rev Dis Primers **2**: 16045.

Huang, L. F., Y. M. Yao, N. Dong, Y. Yu, L. X. He and Z. Y. Sheng (2010). "Association between regulatory T cell activity and sepsis and outcome of severely burned patients: a prospective, observational study." Crit Care **14**(1): R3.

Jaffe, A. E., P. Murakami, H. Lee, J. T. Leek, M. D. Fallin, A. P. Feinberg and R. A. Irizarry (2012). "Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies." Int J Epidemiol **41**(1): 200-209.

Jolliffe, I. T. (2002). Principal component analysis. New York, Springer.

Jones, P. A. (2012). "Functions of DNA methylation: islands, start sites, gene bodies and beyond." Nat Rev Genet **13**(7): 484-492.

Jones, P. A. and D. Takai (2001). "The role of DNA methylation in mammalian epigenetics." Science **293**(5532): 1068-1070.

Jr, J. H. W. (1963). "Hierarchical Grouping to Optimize an Objective Function." Journal of the American Statistical Association **58**(301): 236-244.

Kachuri, L., S. Jeon, A. T. DeWan, C. Metayer, X. Ma, J. S. Witte, C. W. K. Chiang, J. L. Wiemels and A. J. de Smith (2021). "Genetic determinants of blood-cell traits influence susceptibility to childhood acute lymphoblastic leukemia." Am J Hum Genet **108**(10): 1823-1835.

Kanehisa, M. and S. Goto (2000). "KEGG: kyoto encyclopedia of genes and genomes." Nucleic Acids Res **28**(1): 27-30.

Kellum, J. A. and C. Ronco (2023). "The role of endotoxin in septic shock." Crit Care **27**(1): 400.

Kernan, K. F., L. Ghaloul-Gonzalez, J. Vockley, J. Lamb, D. Hollingshead, U. Chandran, R. Sethi, H. J. Park, R. A. Berg, D. Wessel, M. M. Pollack, K. L. Meert, M. W. Hall, C. J. L. Newth, J. C. Lin, A. Doctor, T. Shanley, T. Cornell, R. E. Harrison, A. F. Zuppa, R. Banks, R. W. Reeder, R. Holubkov, D. A. Notterman, J. M. Dean and J. A. Carcillo (2022). "Prevalence of Pathogenic and Potentially Pathogenic Inborn Error of Immunity Associated Variants in Children with Severe Sepsis." J Clin Immunol **42**(2): 350-364.

Khajah, M., G. Andonegui, R. Chan, A. W. Craig, P. A. Greer and D. M. McCafferty (2013). "Fer kinase limits neutrophil chemotaxis toward end target chemoattractants." J Immunol **190**(5): 2208-2216.

Kim, K. W., S. C. Park, H. J. Cho, H. Jang, J. Park, H. S. Shim, E. G. Kim, M. N. Kim, J. Y. Hong, Y. H. Kim, S. Lee, S. T. Weiss, C. H. Kim, S. Won and M. H. Sohn (2021). "Integrated genetic and epigenetic analyses uncover MSI2 association with allergic inflammation." J Allergy Clin Immunol **147**(4): 1453-1463.

Knox, D. B., M. J. Lanspa, K. G. Kuttler, S. C. Brewer and S. M. Brown (2015). "Phenotypic clusters within sepsis-associated multiple organ dysfunction syndrome." Intensive Care Med **41**(5): 814-822.

Kolberg, L., U. Raudvere, I. Kuzmin, J. Vilo and H. Peterson (2020). "gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler." F1000Res **9**.

Kosac, A., J. Pesovic, L. Radenkovic, M. Brkusanin, N. Radovanovic, M. Djurisic, D. Radivojevic, J. Mladenovic, S. Ostojic, G. Kovacevic, R. Kravljanac, D. Savic Pavicevic and V. Milic Rasic (2022). "LTBP4, SPP1, and CD40 Variants: Genetic Modifiers of Duchenne Muscular Dystrophy Analyzed in Serbian Patients." Genes (Basel) **13**(8).

Kovach, M. A. and T. J. Standiford (2012). "The function of neutrophils in sepsis." Curr Opin Infect Dis **25**(3): 321-327.

Kurdyukov, S. and M. Bullock (2016). "DNA Methylation Analysis: Choosing the Right Method." Biology (Basel) **5**(1).

Kyriazopoulou, E., K. Leventogiannis, A. Norrby-Teglund, G. Dimopoulos, A. Pantazi, S. E. Orfanos, N. Rovina, I. Tsangaris, T. Gkavogianni, E. Botsa, E. Chassiou, A. Kotanidou, C. Kontouli, P. Chaloulis, D. Velissaris, A. Savva, J. S. Cullberg, K. Akinosoglou, C. Gogos, A. Armaganidis, E. J. Giamarellos-Bourboulis and G. Hellenic Sepsis Study (2017). "Macrophage activation-like syndrome: an immunological entity associated with rapid progression to death in sepsis." BMC Med **15**(1): 172.

Lee, S., G. R. Abecasis, M. Boehnke and X. Lin (2014). "Rare-variant association analysis: study designs and statistical tests." Am J Hum Genet **95**(1): 5-23.

Lee, S., M. J. Emond, M. J. Bamshad, K. C. Barnes, M. J. Rieder, D. A. Nickerson, N. G. E. S. P.-E. L. P. Team, D. C. Christiani, M. M. Wurfel and X. Lin (2012). "Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies." Am J Hum Genet **91**(2): 224-237.

Lee, S., M. C. Wu and X. Lin (2012). "Optimal tests for rare variant effects in sequencing association studies." Biostatistics **13**(4): 762-775.

Leligdowicz, A. and M. A. Matthay (2019). "Heterogeneity in sepsis: new biological evidence with clinical applications." Crit Care **23**(1): 80.

Lelubre, C. and J. L. Vincent (2018). "Mechanisms and treatment of organ failure in sepsis." Nat Rev Nephrol **14**(7): 417-427.

Li, B. and S. M. Leal (2008). "Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data." Am J Hum Genet **83**(3): 311-321.

Li, H. (2011). "A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data." Bioinformatics **27**(21): 2987-2993.

Li, H., J. Chen, Y. Hu, X. Cai and P. Zhang (2021). "Elevated Serum C1q Levels in Children With Sepsis." Front Pediatr **9**: 619899.

Li, Y., B. Burgman, D. J. McGrail, M. Sun, D. Qi, S. A. Shukla, E. Wu, A. Capasso, S. Y. Lin, C. J. Wu, S. G. Eckhardt, G. B. Mills, B. Li, N. Sahni and S. S. Yi (2020). "Integrated Genomic Characterization of the Human Immunome in Cancer." Cancer Res **80**(21): 4854-4867.

Little, R. J. A. (1988). "A test of missing completely at random for multivariate data with missing values." Journal of the American Statistical Association **83**(404).

Lloyd, S. (1982). "Least squares quantization in PCM." IEEE Transactions on Information Theory **28**(2): 129-137.

Lorente-Pozo, S., P. Navarrete, M. J. Garzon, I. Lara-Canton, J. Beltran-Garcia, R. Osca-Verdegal, S. Mena-Molla, E. Garcia-Lopez, M. Vento, F. V. Pallardo and J. L. Garcia-Gimenez (2021). "DNA Methylation Analysis to Unravel Altered Genetic Pathways Underlying Early Onset and Late Onset Neonatal Sepsis. A Pilot Study." Front Immunol **12**: 622599.

Lorenz, E., J. P. Mira, K. L. Frees and D. A. Schwartz (2002). "Relevance of mutations in the TLR4 receptor in patients with gram-negative septic shock." Arch Intern Med **162**(9): 1028-1032.

Lu, H., D. Wen, X. Wang, L. Gan, J. Du, J. Sun, L. Zeng, J. Jiang and A. Zhang (2019). "Host genetic variants in sepsis risk: a field synopsis and meta-analysis." Crit Care **23**(1): 26.

Lu, H. X., J. H. Sun, D. L. Wen, J. Du, L. Zeng, A. Q. Zhang and J. X. Jiang (2018). "LBP rs2232618 polymorphism contributes to risk of sepsis after trauma." World J Emerg Surg **13**: 52.

Lu, J., Q. Liu, L. Wang, W. Tu, H. Chu, W. Ding, S. Jiang, Y. Ma, X. Shi, W. Pu, X. Zhou, L. Jin, J. Wang and W. Wu (2017). "Increased expression of latent TGF-beta-binding protein 4 affects the fibrotic process in scleroderma by TGF-beta/SMAD signaling." Lab Invest **97**(9): 1121.

Lu, K., L. Liu, X. Xu, F. Zhao, J. Deng, X. Tang, X. Wang, B. Q. Zhao, X. Zhang and Y. Zhao (2020). "ADAMTS13 ameliorates inflammatory responses in experimental autoimmune encephalomyelitis." J Neuroinflammation **17**(1): 67.

Ma, L., Y. Y. Zhang and H. G. Yu (2018). "Clinical Manifestations of Kawasaki Disease Shock Syndrome." Clin Pediatr (Phila) **57**(4): 428-435.

Madsen, B. E. and S. R. Browning (2009). "A groupwise association test for rare mutations using a weighted sum statistic." PLoS Genet **5**(2): e1000384.

Mallik, S., G. J. Odom, Z. Gao, L. Gomez, X. Chen and L. Wang (2019). "An evaluation of supervised methods for identifying differentially methylated regions in Illumina methylation arrays." Brief Bioinform **20**(6): 2224-2235.

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. Mackay, S. A. McCarroll and P. M. Visscher (2009). "Finding the missing heritability of complex diseases." Nature **461**(7265): 747-753.

Mansell, G., T. J. Gorrie-Stone, Y. Bao, M. Kumari, L. S. Schalkwyk, J. Mill and E. Hannon (2019). "Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array." BMC Genomics **20**(1): 366.

Marshall, J. C. (2014). "Why have clinical trials in sepsis failed?" Trends Mol Med **20**(4): 195-203.

Massaud-Ribeiro, L., P. Silami, F. Lima-Setta and A. Prata-Barbosa (2022). "Pediatric Sepsis Research: Where Are We and Where Are We Going?" Front Pediatr **10**: 829119.

Mathy, N. L., W. Scheuer, M. Lanzendorfer, K. Honold, D. Ambrosius, S. Norley and R. Kurth (2000). "Interleukin-16 stimulates the expression and production of pro-inflammatory cytokines by human monocytes." Immunology **100**(1): 63-69.

Matics, T. J. and L. N. Sanchez-Pinto (2017). "Adaptation and Validation of a Pediatric Sequential Organ Failure Assessment Score and Evaluation of the Sepsis-3 Definitions in Critically Ill Children." JAMA Pediatr **171**(10): e172352.

Matys, V., O. V. Kel-Margoulis, E. Fricke, I. Liebich, S. Land, A. Barre-Dirrie, I. Reuter, D. Chekmenev, M. Krull, K. Hornischer, N. Voss, P. Stegmaier, B. Lewicki-Potapov, H. Saxel, A. E. Kel and E. Wingender (2006). "TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes." Nucleic Acids Res **34**(Database issue): D108-110.

Mazaheri, M., H. R. Jahantigh, M. Yavari, S. R. Mirjalili and H. Vahidnezhad (2022). "Autosomal recessive cutis laxa type 1C with a homozygous LTBP4 splicing variant: a case report and update of literature." Mol Biol Rep **49**(5): 4135-4140.

McCafferty, D. M., A. W. Craig, Y. A. Senis and P. A. Greer (2002). "Absence of Fer protein-tyrosine kinase exacerbates leukocyte recruitment in response to endotoxin." J Immunol **168**(10): 4930-4935.

McLachlan, G. J. and K. E. Basford (1988). Mixture models: Inference and applications to clustering, M. Dekker New York.

Mekhedov, S. L., K. S. Makarova and E. V. Koonin (2017). "The complex domain architecture of SAMD9 family proteins, predicted STAND-like NTPases, suggests new links to inflammation and apoptosis." Biol Direct **12**(1): 13.

Mercurio, L., S. Pou, S. Duffy and C. Eickhoff (2023). "Risk Factors for Pediatric Sepsis in the Emergency Department: A Machine Learning Pilot Study." Pediatr Emerg Care **39**(2): e48-e56.

Min, J. L., G. Hemani, E. Hannon, K. F. Dekkers, J. Castillo-Fernandez, R. Luijk, E. Carnero-Montoro, D. J. Lawson, K. Burrows, M. Suderman, A. D. Bretherick, T. G. Richardson, J. Klughammer, V. Iotchkova, G. Sharp, A. Al Khleifat, A. Shatunov, A. Iacoangeli, W. L. McArdle, K. M. Ho, A. Kumar, C. Soderhall, C. Soriano-Tarraga, E. Giralt-Steinhauer, N. Kazmi, D. Mason, A. F. McRae, D. L. Corcoran, K. Sugden, S. Kasela, A. Cardona, F. R. Day, G. Cugliari, C. Viberti, S. Guarrera, M. Lerro, R. Gupta, S. Bollepalli, P. Mandaviya, Y. Zeng, T. K. Clarke, R. M. Walker, V. Schmoll, D. Czamara, C. Ruiz-Arenas, F. I. Rezwan, R. E. Marioni, T. Lin, Y. Awaloff, M. Germain, D. Aissi, R. Zwamborn, K. van Eijk, A. Dekker, J. van Dongen, J. J. Hottenga, G. Willemsen, C. J. Xu, G. Barturen, F. Catala-Moll, M. Kerick, C. Wang, P. Melton, H. R. Elliott, J. Shin, M. Bernard, I. Yet, M. Smart, T. Gorrie-Stone, B. Consortium, C. Shaw, A. Al Chalabi, S. M. Ring, G. Pershagen, E. Melen, J. Jimenez-Conde, J. Roquer, D. A. Lawlor, J. Wright, N. G. Martin, G. W. Montgomery, T. E. Moffitt, R. Poulton, T. Esko, L. Milani, A. Metspalu, J. R. B. Perry, K. K. Ong, N. J. Wareham, G. Matullo, C. Sacerdote, S. Panico, A. Caspi, L. Arseneault, F. Gagnon, M. Ollikainen, J. Kaprio, J. F. Felix, F. Rivadeneira, H. Tiemeier, I. M. H. van, A. G. Uitterlinden, V. W. V. Jaddoe, C. Haley, A. M. McIntosh, K. L. Evans, A. Murray, K. Raikkonen, J. Lahti, E. A. Nohr, T. I. A. Sorensen, T. Hansen, C. S. Morgen, E. B. Binder, S. Lucae, J. R. Gonzalez, M. Bustamante, J. Sunyer, J. W. Holloway, W. Karmaus, H. Zhang, I. J. Deary, N. R. Wray, J. M. Starr, M. Beekman, D. van Heemst, P. E. Slagboom, P. E. Morange, D. A. Tregouet, J. H. Veldink, G. E. Davies, E. J. C. de Geus, D. I. Boomsma, J. M. Vonk, B. Brunekreef, G. H. Koppelman, M. E. Alarcon-Riquelme, R. C. Huang, C. E. Pennell, J. van Meurs, M. A. Ikram, A. D. Hughes, T. Tillin, N. Chaturvedi, Z. Pausova, T. Paus, T. D. Spector, M. Kumari, L. C. Schalkwyk, P. M. Visscher, G. Davey Smith, C. Bock, T. R. Gaunt, J. T. Bell, B. T. Heijmans, J. Mill and C. L. Relton (2021). "Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation." Nat Genet **53**(9): 1311-1321.

Mira, J. P., A. Cariou, F. Grall, C. Delclaux, M. R. Losser, F. Heshmati, C. Cheval, M. Monchi, J. L. Teboul, F. Riche, G. Leleu, L. Arbibe, A. Mignon, M. Delpech and J. F. Dhainaut (1999). "Association of TNF2, a TNF-alpha promoter polymorphism, with septic shock susceptibility and mortality: a multicenter study." JAMA **282**(6): 561-568.

Miranda, M. and S. Nadel (2023). "Pediatric Sepsis: a Summary of Current Definitions and Management Recommendations." Curr Pediatr Rep **11**(2): 29-39.

Mohammed, A., Y. Cui, V. R. Mas and R. Kamaleswaran (2019). "Differential gene expression analysis reveals novel genes and pathways in pediatric septic shock patients." Sci Rep **9**(1): 11270.

Monti, S., Tamayo, P., Mesirov, J. et al. (2003). "Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data." Machine Learning(52): 91-118.

Morris, A. P. and E. Zeggini (2010). "An evaluation of statistical approaches to rare variant analysis in genetic association studies." Genet Epidemiol **34**(2): 188-193.

Muckian, M. D., J. F. Wilson, G. S. Taylor, H. R. Stagg and N. Pirastu (2023). "Mendelian randomisation identifies priority groups for prophylactic EBV vaccination." BMC Infect Dis **23**(1): 65.

Muszynski, J. A., R. Nofziger, M. Moore-Clingenpeel, K. Greathouse, L. Anglim, L. Steele, J. Hensley, L. Hanson-Huber, J. Nateri, O. Ramilo and M. W. Hall (2018). "Early Immune Function and Duration of Organ Dysfunction in Critically III Children with Sepsis." Am J Respir Crit Care Med **198**(3): 361-369.

Neale, B. M., M. A. Rivas, B. F. Voight, D. Altshuler, B. Devlin, M. Orho-Melander, S. Kathiresan, S. M. Purcell, K. Roeder and M. J. Daly (2011). "Testing for an unusual distribution of rare variants." PLoS Genet **7**(3): e1001322.

Ng, P. C. and S. Henikoff (2001). "Predicting deleterious amino acid substitutions." Genome Res **11**(5): 863-874.

Nguyen, T. C., Y. Y. Han, J. E. Kiss, M. W. Hall, A. C. Hassett, R. Jaffe, R. A. Orr, J. Janosky and J. A. Carcillo (2008). "Intensive plasma exchange increases a disintegrin and metalloprotease with thrombospondin motifs-13 activity and reverses organ dysfunction in children with thrombocytopenia-associated multiple organ failure." Crit Care Med **36**(10): 2878-2887.

Nullens, S., J. De Man, C. Bridts, D. Ebo, S. Francque and B. De Winter (2018). "Identifying Therapeutic Targets for Sepsis Research: A Characterization Study of the Inflammatory Players in the Cecal Ligation and Puncture Model." Mediators Inflamm **2018**: 5130463.

Oliveira, R. A. C., D. O. Imparato, V. G. S. Fernandes, J. V. F. Cavalcante, R. D. Albanus and R. J. S. Dalmolin (2021). "Reverse Engineering of the Pediatric Sepsis Regulatory Network and Identification of Master Regulators." Biomedicines **9**(10).

Pan, W. (2009). "Asymptotic tests of association with multiple SNPs in linkage disequilibrium." Genet Epidemiol **33**(6): 497-507.

Pedersen, B. S., D. A. Schwartz, I. V. Yang and K. J. Kechris (2012). "Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values." Bioinformatics **28**(22): 2986-2988.

Pelleg, D., Moore A.W. (2000). "X-means: Extending K-means with Efficient Estimation of the Number of Clusters." Proceedings of the 17th International Conference on Machine Learning **1**: 727-734.

Peng, D., M. Fu, M. Wang, Y. Wei and X. Wei (2022). "Targeting TGF-beta signal transduction for fibrosis and cancer therapy." Mol Cancer **21**(1): 104.

Perez-Gonzalez, M., M. Mendioroz, S. Badesso, D. Sucunza, M. Roldan, M. Espelosin, S. Ursua, R. Lujan, M. Cuadrado-Tejedor and A. Garcia-Osta (2020). "PLA2G4E, a candidate gene for resilience in Alzheimer s disease and a new target for dementia treatment." Prog Neurobiol **191**: 101818.

Peters, T. J., M. J. Buckley, A. L. Statham, R. Pidsley, K. Samaras, V. L. R, S. J. Clark and P. L. Molloy (2015). "De novo identification of differentially methylated regions in the human genome." Epigenetics Chromatin **8**: 6.

Pidsley, R., E. Zotenko, T. J. Peters, M. G. Lawrence, G. P. Risbridger, P. Molloy, S. Van Djik, B. Muhlhausler, C. Stirzaker and S. J. Clark (2016). "Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling." Genome Biol **17**(1): 208.

Pinsky, M. R. and G. M. Matuschak (1989). "Multiple systems organ failure: failure of host defense homeostasis." Crit Care Clin **5**(2): 199-220.

Plotnikov, D., Y. Huang, A. P. Khawaja, P. J. Foster, Z. Zhu, J. A. Guggenheim and M. He (2022). "High Blood Pressure and Intraocular Pressure: A Mendelian Randomization Study." Invest Ophthalmol Vis Sci **63**(6): 29.

Portela, A. and M. Esteller (2010). "Epigenetic modifications and human disease." Nat Biotechnol **28**(10): 1057-1068.

Povysil, G., S. Petrovski, J. Hostyk, V. Aggarwal, A. S. Allen and D. B. Goldstein (2019). "Rare-variant collapsing analyses for complex traits: guidelines and applications." Nat Rev Genet **20**(12): 747-759.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly and P. C. Sham (2007). "PLINK: a tool set for whole-genome association and population-based linkage analyses." Am J Hum Genet **81**(3): 559-575.

Qi, W., K. V. Ebbert, A. W. Craig, P. A. Greer and D. M. McCafferty (2005). "Absence of Fer protein tyrosine kinase exacerbates endotoxin induced intestinal epithelial barrier dysfunction in vivo." Gut **54**(8): 1091-1097.

Qin, Y., K. F. Kernan, Z. Fan, H. J. Park, S. Kim, S. W. Canna, J. A. Kellum, R. A. Berg, D. Wessel, M. M. Pollack, K. Meert, M. Hall, C. Newth, J. C. Lin, A. Doctor, T. Shanley, T. Cornell, R. E. Harrison, A. F. Zuppa, R. Banks, R. W. Reeder, R. Holubkov, D. A. Notterman, J. Michael Dean and J. A. Carcillo (2022). "Machine learning derivation of four computable 24-h pediatric sepsis phenotypes to facilitate enrollment in early personalized anti-inflammatory clinical trials." Crit Care **26**(1): 128.

Rakyan, V. K., T. A. Down, D. J. Balding and S. Beck (2011). "Epigenome-wide association studies for common human diseases." Nat Rev Genet **12**(8): 529-541.

Rautanen, A., T. C. Mills, A. C. Gordon, P. Hutton, M. Steffens, R. Nuamah, J. D. Chiche, T. Parks, S. J. Chapman, E. E. Davenport, K. S. Elliott, J. Bion, P. Lichtner, T. Meitinger, T. F. Wienker, M. J. Caulfield, C. Mein, F. Bloos, I. Bobek, P. Cotogni, V. Sramek, S. Sarapuu, M. Kobilay, V. M. Ranieri, J. Rello, G. Sirgo, Y. G. Weiss, S. Russwurm, E. M. Schneider, K. Reinhart, P. A. Holloway, J. C. Knight, C. S. Garrard, J. A. Russell, K. R. Walley, F. Stüber, A. V. Hill and C. J. Hinds (2015). "Genome-wide association study of survival from sepsis due to pneumonia: an observational cohort study." Lancet Respir Med **3**(1): 53-60.

Rautanen, A., T. C. Mills, A. C. Gordon, P. Hutton, M. Steffens, R. Nuamah, J. D. Chiche, T. Parks, S. J. Chapman, E. E. Davenport, K. S. Elliott, J. Bion, P. Lichtner, T. Meitinger, T. F. Wienker, M. J. Caulfield, C. Mein, F. Bloos, I. Bobek, P. Cotogni, V. Sramek, S. Sarapuu, M. Kobilay, V. M. Ranieri, J. Rello, G. Sirgo, Y. G. Weiss, S. Russwurm, E. M. Schneider, K. Reinhart, P. A. Holloway, J. C. Knight, C. S. Garrard, J. A. Russell, K. R. Walley, F. Stuber, A. V. Hill, C. J. Hinds and E. E. G. Investigators (2015). "Genome-wide association study of survival from sepsis due to pneumonia: an observational cohort study." Lancet Respir Med **3**(1): 53-60.

Rentzsch, P., D. Witten, G. M. Cooper, J. Shendure and M. Kircher (2019). "CADD: predicting the deleteriousness of variants throughout the human genome." Nucleic Acids Res **47**(D1): D886-D894.

Rifkin, D., N. Sachan, K. Singh, E. Sauber, G. Tellides and F. Ramirez (2022). "The role of LTBPs in TGF beta signaling." Dev Dyn **251**(1): 95-104.

Rocchiccioli, S., A. Cecchettini, P. Panesi, P. A. Farneti, M. Mariani, N. Ucciferri, L. Citti, M. G. Andreassi and I. Foffa (2017). "Hypothesis-free secretome analysis of thoracic aortic aneurysm reinforces the central role of TGF-beta cascade in patients with bicuspid aortic valve." J Cardiol **69**(3): 570-576.

Rodenburg, R. J. (2018). "The functional genomics laboratory: functional validation of genetic variants." J Inherit Metab Dis **41**(3): 297-307.

Rosales, C. and E. Uribe-Querol (2017). "Phagocytosis: A Fundamental Process in Immunity." Biomed Res Int **2017**: 9042851.

Rosier, F., A. Brisebarre, C. Dupuis, S. Baaklini, D. Puthier, C. Brun, L. C. Pradel, P. Rihet and D. Payen (2021). "Genetic Predisposition to the Mortality in Septic Shock Patients: From GWAS to the Identification of a Regulatory Variant Modulating the Activity of a CISH Enhancer." Int J Mol Sci **22**(11).

Rudd, K. E., S. C. Johnson, K. M. Agesa, K. A. Shackelford, D. Tsoi, D. R. Kievlan, D. V. Colombara, K. S. Ikuta, N. Kissoon, S. Finfer, C. Fleischmann-Struzek, F. R. Machado, K. K. Reinhart, K. Rowan, C. W. Seymour, R. S. Watson, T. E. West, F. Marinho, S. I. Hay, R. Lozano, A. D. Lopez, D. C. Angus, C. J. L. Murray and M. Naghavi (2020). "Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study." Lancet **395**(10219): 200-211.

Ruiz-Arenas, C., C. Hernandez-Ferrer, M. Vives-Usano, S. Mari, I. Quintela, D. Mason, S. Cadiou, M. Casas, S. Andrusaityte, K. B. Gutzkow, M. Vafeiadi, J. Wright, J. Lepeule, R. Grazuleviciene, L. Chatzi, A. Carracedo, X. Estivill, E. Marti, G. Escaramis, M. Vrijheid, J. R. Gonzalez and M. Bustamante (2022). "Identification of autosomal cis expression quantitative trait methylation (cis eQTMs) in children's blood." Elife **11**.

Sakaue, S., M. Kanai, Y. Tanigawa, J. Karjalainen, M. Kurki, S. Koshiba, A. Narita, T. Konuma, K. Yamamoto, M. Akiyama, K. Ishigaki, A. Suzuki, K. Suzuki, W. Obara, K. Yamaji, K. Takahashi, S. Asai, Y. Takahashi, T. Suzuki, N. Shinozaki, H. Yamaguchi, S. Minami, S. Murayama, K. Yoshimori, S. Nagayama, D. Obata, M. Higashiyama, A. Masumoto, Y. Koretsune, FinnGen, K. Ito, C. Terao, T. Yamauchi, I. Komuro, T. Kadowaki, G. Tamiya, M. Yamamoto, Y. Nakamura, M. Kubo, Y. Murakami, K. Yamamoto, Y. Kamatani, A. Palotie, M. A. Rivas, M. J. Daly, K. Matsuda and Y. Okada (2021). "A cross-population atlas of genetic associations for 220 human phenotypes." Nat Genet **53**(10): 1415-1424.

Salas, L. A., D. C. Koestler, R. A. Butler, H. M. Hansen, J. K. Wiencke, K. T. Kelsey and B. C. Christensen (2018). "An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray." Genome Biol **19**(1): 64.

Sanchez-Pinto, L. N., E. K. Stroup, T. Pendergrast, N. Pinto and Y. Luo (2020). "Derivation and Validation of Novel Phenotypes of Multiple Organ Dysfunction Syndrome in Critically Ill Children." JAMA Netw Open **3**(8): e209271.

Schlapbach, L. J., R. S. Watson, L. R. Sorce, A. C. Argent, K. Menon, M. W. Hall, S. Akech, D. J. Albers, E. R. Alpern, F. Balamuth, M. Bembea, P. Biban, E. D. Carrol, K. Chiotos, M. J. Chisti, P. E. DeWitt, I. Evans, C. Flauzino de Oliveira, C. M. Horvat, D. Inwald, P. Ishimine, J. C. Jaramillo-Bustamante, M. Levin, R. Lodha, B. Martin, S. Nadel, S. Nakagawa, M. J. Peters, A. G. Randolph, S. Ranjit, M. N. Rebull, S. Russell, H. F. Scott, D. C. de Souza, P. Tissieres, S. L. Weiss, M. O. Wiens, J. L. Wynn, N. Kissoon, J. J. Zimmerman, L. N. Sanchez-Pinto, T. D. Bennett and F. Society of Critical Care Medicine Pediatric Sepsis Definition Task (2024). "International Consensus Criteria for Pediatric Sepsis and Septic Shock." JAMA.

Schluter, B., C. Raufhake, M. Erren, H. Schotte, F. Kipp, S. Rust, H. Van Aken, G. Assmann and E. Berendes (2002). "Effect of the interleukin-6 promoter polymorphism (-174 G/C) on the incidence and outcome of sepsis." Crit Care Med **30**(1): 32-37.

Scicluna, B. P., L. A. van Vught, A. H. Zwinderman, M. A. Wiewel, E. E. Davenport, K. L. Burnham, P. Nurnberg, M. J. Schultz, J. Horn, O. L. Cremer, M. J. Bonten, C. J. Hinds, H. R. Wong, J. C. Knight, T. van der Poll and M. consortium (2017). "Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study." Lancet Respir Med **5**(10): 816-826.

Sevketoglu, E., D. Yildizdas, O. O. Horoz, H. S. Kihtir, T. Kendirli, S. Bayraktar and J. A. Carcillo (2014). "Use of therapeutic plasma exchange in children with thrombocytopenia-associated

multiple organ failure in the Turkish thrombocytopenia-associated multiple organ failure network." Pediatr Crit Care Med **15**(8): e354-359.

Seymour, C. W., J. N. Kennedy, S. Wang, C. H. Chang, C. F. Elliott, Z. Xu, S. Berry, G. Clermont, G. Cooper, H. Gomez, D. T. Huang, J. A. Kellum, Q. Mi, S. M. Opal, V. Talisa, T. van der Poll, S. Visweswaran, Y. Vodovotz, J. C. Weiss, D. M. Yealy, S. Yende and D. C. Angus (2019). "Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis." JAMA **321**(20): 2003-2017.

Shakoory, B., J. A. Carcillo, W. W. Chatham, R. L. Amdur, H. Zhao, C. A. Dinarello, R. Q. Cron and S. M. Opal (2016). "Interleukin-1 Receptor Blockade Is Associated With Reduced Mortality in Sepsis Patients With Features of Macrophage Activation Syndrome: Reanalysis of a Prior Phase III Trial." Crit Care Med **44**(2): 275-281.

Shrine, N., A. L. Guyatt, A. M. Erzurumluoglu, V. E. Jackson, B. D. Hobbs, C. A. Melbourne, C. Batini, K. A. Fawcett, K. Song, P. Sakornsakolpat, X. Li, R. Boxall, N. F. Reeve, M. Obeidat, J. H. Zhao, M. Wielscher, S. Weiss, K. A. Kentistou, J. P. Cook, B. B. Sun, J. Zhou, J. Hui, S. Karrasch, M. Imboden, S. E. Harris, J. Marten, S. Enroth, S. M. Kerr, I. Surakka, V. Vitart, T. Lehtimaki, R. J. Allen, P. S. Bakke, T. H. Beaty, E. R. Bleecker, Y. Bosse, C. A. Brandsma, Z. Chen, J. D. Crapo, J. Danesh, D. L. DeMeo, F. Dudbridge, R. Ewert, C. Gieger, A. Gulsvik, A. L. Hansell, K. Hao, J. D. Hoffman, J. E. Hokanson, G. Homuth, P. K. Joshi, P. Joubert, C. Langenberg, X. Li, L. Li, K. Lin, L. Lind, N. Locantore, J. Luan, A. Mahajan, J. C. Maranville, A. Murray, D. C. Nickle, R. Packer, M. M. Parker, M. L. Paynton, D. J. Porteous, D. Prokopenko, D. Qiao, R. Rawal, H. Runz, I. Sayers, D. D. Sin, B. H. Smith, M. Soler Artigas, D. Sparrow, R. Tal-Singer, P. Timmers, M. Van den Berge, J. C. Whittaker, P. G. Woodruff, L. M. Yerges-Armstrong, O. G. Troyanskaya, O. T. Raitakari, M. Kahonen, O. Polasek, U. Gyllensten, I. Rudan, I. J. Deary, N. M. Probst-Hensch, H. Schulz, A. L. James, J. F. Wilson, B. Stubbe, E. Zeggini, M. R. Jarvelin, N. Wareham, E. K. Silverman, C. Hayward, A. P. Morris, A. S. Butterworth, R. A. Scott, R. G. Walters, D. A. Meyers, M. H. Cho, D. P. Strachan, I. P. Hall, M. D. Tobin, L. V. Wain and G. Understanding Society Scientific (2019). "New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries." Nat Genet **51**(3): 481-493.

Shrine, N., A. G. Izquierdo, J. Chen, R. Packer, R. J. Hall, A. L. Guyatt, C. Batini, R. J. Thompson, C. Pavuluri, V. Malik, B. D. Hobbs, M. Moll, W. Kim, R. Tal-Singer, P. Bakke, K. A. Fawcett, C. John, K. Coley, N. N. Piga, A. Pozarickij, K. Lin, I. Y. Millwood, Z. Chen, L. Li, G. China Kadoorie Biobank Collaborative, S. R. A. Wijnant, L. Lahousse, G. Brusselle, A. G. Uitterlinden, A. Manichaikul, E. C. Oelsner, S. S. Rich, R. G. Barr, S. M. Kerr, V. Vitart, M. R. Brown, M. Wielscher, M. Imboden, A. Jeong, T. M. Bartz, S. A. Gharib, C. Flexeder, S. Karrasch, C. Gieger, A. Peters, B. Stubbe, X. Hu, V. E. Ortega, D. A. Meyers, E. R. Bleecker, S. B. Gabriel, N. Gupta, A. V. Smith, J. Luan, J. H. Zhao, A. F. Hansen, A. Langhammer, C. Willer, L. Bhatta, D. Porteous, B. H. Smith, A. Campbell, T. Sofer, J. Lee, M. L. Daviglus, B. Yu, E. Lim, H. Xu, G. T. O'Connor, G. Thareja, O. M. E. Albagha, C. Qatar Genome Program Research, K. Suhre, R. Granell, T. O. Faquih, P. S. Hiemstra, A. M. Slats, B. H. Mullin, J. Hui, A. James, J. Beilby, K. Patasova, P. Hysi, J. T. Koskela, A. B. Wyss, J. Jin, S. Sikdar, M. Lee, S. May-Wilson, N. Pirastu, K. A. Kentistou, P. K.

Joshi, P. Timmers, A. T. Williams, R. C. Free, X. Wang, J. L. Morrison, F. D. Gilliland, Z. Chen, C. A. Wang, R. E. Foong, S. E. Harris, A. Taylor, P. Redmond, J. P. Cook, A. Mahajan, L. Lind, T. Palviainen, T. Lehtimaki, O. T. Raitakari, J. Kaprio, T. Rantanen, K. H. Pietilainen, S. R. Cox, C. E. Pennell, G. L. Hall, W. J. Gauderman, C. Brightling, J. F. Wilson, T. Vasankari, T. Laitinen, V. Salomaa, D. O. Mook-Kanamori, N. J. Timpson, E. Zeggini, J. Dupuis, C. Hayward, B. Brumpton, C. Langenberg, S. Weiss, G. Homuth, C. O. Schmidt, N. Probst-Hensch, M. R. Jarvelin, A. C. Morrison, O. Polasek, I. Rudan, J. H. Lee, I. Sayers, E. L. Rawlins, F. Dudbridge, E. K. Silverman, D. P. Strachan, R. G. Walters, A. P. Morris, S. J. London, M. H. Cho, L. V. Wain, I. P. Hall and M. D. Tobin (2023). "Multi-ancestry genome-wide association analyses improve resolution of genes and pathways influencing lung function and chronic obstructive pulmonary disease risk." Nat Genet **55**(3): 410-422.

Sinha, P., M. M. Churpek and C. S. Calfee (2020). "Machine Learning Classifier Models Can Identify Acute Respiratory Distress Syndrome Phenotypes Using Readily Available Clinical Data." Am J Respir Crit Care Med **202**(7): 996-1004.

Sollis, E., A. Mosaku, A. Abid, A. Buniello, M. Cerezo, L. Gil, T. Groza, O. Gunes, P. Hall, J. Hayhurst, A. Ibrahim, Y. Ji, S. John, E. Lewis, J. A. L. MacArthur, A. McMahon, D. Osumi-Sutherland, K. Panoutsopoulou, Z. Pendlington, S. Ramachandran, R. Stefancsik, J. Stewart, P. Whetzel, R. Wilson, L. Hindorff, F. Cunningham, S. A. Lambert, M. Inouye, H. Parkinson and L. W. Harris (2023). "The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource." Nucleic Acids Res **51**(D1): D977-D985.

Son, M. B. F., N. Murray, K. Friedman, C. C. Young, M. M. Newhams, L. R. Feldstein, L. L. Loftis, K. M. Tarquinio, A. R. Singh, S. M. Heidemann, V. L. Soma, B. J. Riggs, J. C. Fitzgerald, M. Kong, S. Doymaz, J. S. Giuliano, Jr., M. A. Keenaghan, J. R. Hume, C. V. Hobbs, J. E. Schuster, K. N. Clouser, M. W. Hall, L. S. Smith, S. M. Horwitz, S. P. Schwartz, K. Irby, T. T. Bradford, A. B. Maddux, C. J. Babbitt, C. M. Rowan, G. E. McLaughlin, P. H. Yager, M. Maamari, E. H. Mack, C. L. Carroll, V. L. Montgomery, N. B. Halasa, N. Z. Cvijanovich, B. M. Coates, C. E. Rose, J. W. Newburger, M. M. Patel, A. G. Randolph and C.-I. Overcoming (2021). "Multisystem Inflammatory Syndrome in Children - Initial Therapy and Outcomes." N Engl J Med **385**(1): 23-34.

Sorensen, T. I., G. G. Nielsen, P. K. Andersen and T. W. Teasdale (1988). "Genetic and environmental influences on premature death in adult adoptees." N Engl J Med **318**(12): 727-732.

Srinivasan, L., G. Page, H. Kirpalani, J. C. Murray, A. Das, R. D. Higgins, W. A. Carlo, E. F. Bell, R. N. Goldberg, K. Schibler, B. G. Sood, D. K. Stevenson, B. J. Stoll, K. P. Van Meurs, K. J. Johnson, J. Levy, S. A. McDonald, K. M. Zaterka-Baxter, K. A. Kennedy, P. J. Sanchez, S. Duara, M. C. Walsh, S. Shankaran, J. L. Wynn, C. M. Cotten, H. Eunice Kennedy Shriver National Institute of Child and N. Human Development Neonatal Research (2017). "Genome-wide association study of sepsis in extremely premature infants." Arch Dis Child Fetal Neonatal Ed **102**(5): F439-F445.

Stanski, N. L. and H. R. Wong (2020). "Prognostic and predictive enrichment in sepsis." Nat Rev Nephrol **16**(1): 20-31.

Stelzer, G., N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, T. I. Stein, R. Nudel, I. Lieder, Y. Mazor, S. Kaplan, D. Dahary, D. Warshawsky, Y. Guan-Golan, A. Kohn, N. Rappaport, M. Safran and D. Lancet (2016). "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses." Curr Protoc Bioinformatics **54**: 1 30 31-31 30 33.

Su, C. T., J. W. Huang, C. K. Chiang, E. C. Lawrence, K. L. Levine, B. Dabovic, C. Jung, E. C. Davis, S. Madan-Khetarpal and Z. Urban (2015). "Latent transforming growth factor binding protein 4 regulates transforming growth factor beta receptor stability." Hum Mol Genet **24**(14): 4024-4036.

Su, C. T., T. M. Jao, Z. Urban, Y. J. Huang, D. H. W. See, Y. C. Tsai, W. C. Lin and J. W. Huang (2021). "LTBP4 affects renal fibrosis by influencing angiogenesis and altering mitochondrial structure." Cell Death Dis **12**(10): 943.

Su, C. T. and Z. Urban (2021). "LTBP4 in Health and Disease." Genes (Basel) **12**(6).

Suderman, M., J. R. Staley, R. French, R. Arathimos, A. Simpkin and K. Tilling (2018). "dmrff: identifying differentially methylated regions efficiently with power and control." bioRxiv.

Sun, J., Y. Zheng and L. Hsu (2013). "A unified mixed-effects model for rare-variant association in sequencing studies." Genet Epidemiol **37**(4): 334-344.

Sutherland, A. M. and K. R. Walley (2009). "Bench-to-bedside review: Association of genetic variation with sepsis." Crit Care **13**(2): 210.

Sweeney, T. E., T. D. Azad, M. Donato, W. A. Haynes, T. M. Perumal, R. Henao, J. F. Bermejo-Martin, R. Almansa, E. Tamayo, J. A. Howrylak, A. Choi, G. P. Parnell, B. Tang, M. Nichols, C. W. Woods, G. S. Ginsburg, S. F. Kingsmore, L. Omberg, L. M. Mangravite, H. R. Wong, E. L. Tsalik, R. J. Langley and P. Khatri (2018). "Unsupervised Analysis of Transcriptomics in Bacterial Sepsis Across Multiple Datasets Reveals Three Robust Clusters." Crit Care Med **46**(6): 915-925.

Szabo, P. A., A. Goswami, D. M. Mazzuca, K. Kim, D. B. O'Gorman, D. A. Hess, I. D. Welch, H. A. Young, B. Singh, J. K. McCormick and S. M. Haeryfar (2017). "Rapid and Rigorous IL-17A Production by a Distinct Subpopulation of Effector Memory T Lymphocytes Constitutes a Novel Mechanism of Toxic Shock Syndrome Immunopathology." J Immunol **198**(7): 2805-2818.

Takeuchi, O. and S. Akira (2010). "Pattern recognition receptors and inflammation." Cell **140**(6): 805-820.

Tam, V., N. Patel, M. Turcotte, Y. Bosse, G. Pare and D. Meyre (2019). "Benefits and limitations of genome-wide association studies." Nat Rev Genet **20**(8): 467-484.

Taudien, S., L. Lausser, E. J. Giamarellos-Bourboulis, C. Sponholz, F. Schoneweck, M. Felder, L. R. Schirra, F. Schmid, C. Gogos, S. Groth, B. S. Petersen, A. Franke, W. Lieb, K. Huse, P. F. Zipfel, O. Kurzai, B. Moepps, P. Gierschik, M. Bauer, A. Scherag, H. A. Kestler and M. Platzer (2016). "Genetic Factors of the Disease Course After Sepsis: Rare Deleterious Variants Are Predictive." EBioMedicine **12**: 227-238.

Taylor, M. D., V. Allada, M. L. Moritz, A. J. Nowalk, R. Sindhi, R. K. Aneja, K. Torok, M. J. Morowitz, M. Michaels and J. A. Carcillo (2020). "Use of C-Reactive Protein and Ferritin Biomarkers in Daily Pediatric Practice." Pediatr Rev **41**(4): 172-183.

Tonial, C. T., C. A. D. Costa, G. R. H. Andrades, F. Crestani, F. Bruno, J. P. Piva and P. C. R. Garcia (2021). "Performance of prognostic markers in pediatric sepsis." J Pediatr (Rio J) **97**(3): 287-294.

Urban, Z., V. Hucthagowder, N. Schurmann, V. Todorovic, L. Zilberberg, J. Choi, C. Sens, C. W. Brown, R. D. Clark, K. E. Holland, M. Marble, L. Y. Sakai, B. Dabovic, D. B. Rifkin and E. C. Davis (2009). "Mutations in LTBP4 cause a syndrome of impaired pulmonary, gastrointestinal, genitourinary, musculoskeletal, and dermal development." Am J Hum Genet **85**(5): 593-605.

Van Buuren S, G.-O. K. (2011). "mice: Multivariate Imputation by Chained Equations in R." Journal of Statistical Software **45**(3): 1-67.

Van der Auwera GA, O. C. B. (2020). "Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition)." O'Reilly Media.

Vastrad, B. and C. Vastrad (2023). "Bioinformatics and next generation data analysis reveals the potential role of inflammation in sepsis and its associated complications." bioRxiv: 2023.2008.2002.551653.

Villar, J., N. Maca-Meyer, L. Perez-Mendez and C. Flores (2004). "Bench-to-bedside review: understanding genetic predisposition to sepsis." Crit Care **8**(3): 180-189.

Villeneuve, A., J. S. Joyal, F. Proulx, T. Ducruet, N. Poitras and J. Lacroix (2016). "Multiple organ dysfunction syndrome in critically ill children: clinical value of two lists of diagnostic criteria." Ann Intensive Care **6**(1): 40.

Vuckovic, D., E. L. Bao, P. Akbari, C. A. Lareau, A. Mousas, T. Jiang, M. H. Chen, L. M. Raffield, M. Tardaguila, J. E. Huffman, S. C. Ritchie, K. Megy, H. Ponstingl, C. J. Penkett, P. K. Albers, E. M. Wigdor, S. Sakaue, A. Moscati, R. Manansala, K. S. Lo, H. Qian, M. Akiyama, T. M. Bartz, Y. Ben-Shlomo, A. Beswick, J. Bork-Jensen, E. P. Bottinger, J. A. Brody, F. J. A. van Rooij, K. N. Chitrala, P. W. F. Wilson, H. Choquet, J. Danesh, E. Di Angelantonio, N. Dimou, J. Ding, P. Elliott, T. Esko, M. K. Evans, S. B. Felix, J. S. Floyd, L. Broer, N. Grarup, M. H. Guo, Q. Guo, A. Greinacher, J. Haessler, T. Hansen, J. M. M. Howson, W. Huang, E. Jorgenson, T. Kacprowski, M. Kahonen, Y. Kamatani, M. Kanai, S. Karthikeyan, F. Koskeridis, L. A. Lange, T. Lehtimaki, A. Linneberg, Y. Liu, L. P. Lyytikainen, A. Manichaikul, K. Matsuda, K. L. Mohlke, N. Mononen, Y. Murakami, G. N. Nadkarni, K. Nikus, N. Pankratz, O. Pedersen, M. Preuss, B. M. Psaty, O. T. Raitakari,

S. S. Rich, B. A. T. Rodriguez, J. D. Rosen, J. I. Rotter, P. Schubert, C. N. Spracklen, P. Surendran, H. Tang, J. C. Tardif, M. Ghanbari, U. Volker, H. Volzke, N. A. Watkins, S. Weiss, V. A. M. V. Program, N. Cai, K. Kundu, S. B. Watt, K. Walter, A. B. Zonderman, K. Cho, Y. Li, R. J. F. Loos, J. C. Knight, M. Georges, O. Stegle, E. Evangelou, Y. Okada, D. J. Roberts, M. Inouye, A. D. Johnson, P. L. Auer, W. J. Astle, A. P. Reiner, A. S. Butterworth, W. H. Ouwehand, G. Lettre, V. G. Sankaran and N. Soranzo (2020). "The Polygenic and Monogenic Basis of Blood Traits and Diseases." Cell **182**(5): 1214-1231 e1211.

W. N. Venables, B. D. R. (2002). "Modern Applied Statistics with S, Fourth edition." Springer.

Wade, B. H. and G. L. Mandell (1983). "Polymorphonuclear leukocytes: dedicated professional phagocytes." Am J Med **74**(4): 686-693.

Wang, J., Q. Zhao, T. Hastie and A. B. Owen (2017). "Confounder Adjustment in Multiple Hypothesis Testing." Ann Stat **45**(5): 1863-1894.

Wang, K., M. Li and H. Hakonarson (2010). "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data." Nucleic Acids Res **38**(16): e164.

Watanabe, K., E. Taskesen, A. van Bochoven and D. Posthuma (2017). "Functional mapping and annotation of genetic associations with FUMA." Nat Commun **8**(1): 1826.

Watanabe, S., M. Alexander, A. V. Misharin and G. R. S. Budinger (2019). "The role of macrophages in the resolution of inflammation." J Clin Invest **129**(7): 2619-2628.

Weiss, S. L., J. C. Fitzgerald, J. Pappachan, D. Wheeler, J. C. Jaramillo-Bustamante, A. Salloo, S. C. Singhi, S. Erickson, J. A. Roy, J. L. Bush, V. M. Nadkarni, N. J. Thomas, O. Sepsis Prevalence, I. Therapies Study, I. Pediatric Acute Lung and N. Sepsis Investigators (2015). "Global epidemiology of pediatric severe sepsis: the sepsis prevalence, outcomes, and therapies study." Am J Respir Crit Care Med **191**(10): 1147-1157.

Weiss, S. L., M. J. Peters, W. Alhazzani, M. S. D. Agus, H. R. Flori, D. P. Inwald, S. Nadel, L. J. Schlapbach, R. C. Tasker, A. C. Argent, J. Brierley, J. Carcillo, E. D. Carrol, C. L. Carroll, I. M. Cheifetz, K. Choong, J. J. Cies, A. T. Cruz, D. De Luca, A. Deep, S. N. Faust, C. F. De Oliveira, M. W. Hall, P. Ishimine, E. Javouhey, K. F. M. Joosten, P. Joshi, O. Karam, M. C. J. Kneyber, J. Lemson, G. MacLaren, N. M. Mehta, M. H. Moller, C. J. L. Newth, T. C. Nguyen, A. Nishisaki, M. E. Nunnally, M. M. Parker, R. M. Paul, A. G. Randolph, S. Ranjit, L. H. Romer, H. F. Scott, L. N. Tume, J. T. Verger, E. A. Williams, J. Wolf, H. R. Wong, J. J. Zimmerman, N. Kissoon and P. Tissieres (2020). "Surviving Sepsis Campaign International Guidelines for the Management of Septic Shock and Sepsis-Associated Organ Dysfunction in Children." Pediatr Crit Care Med **21**(2): e52-e106.

Weiss, S. L., D. Zhang, J. Bush, K. Graham, J. Starr, F. Tuluc, S. Henrickson, T. Kilbaugh, C. S. Deutschman, D. Murdock, F. X. McGowan, Jr., L. Becker and D. C. Wallace (2019). "Persistent Mitochondrial Dysfunction Linked to Prolonged Organ Dysfunction in Pediatric Sepsis." Crit Care Med **47**(10): 1433-1441.

Wiley, S. R., K. Schooley, P. J. Smolak, W. S. Din, C. P. Huang, J. K. Nicholl, G. R. Sutherland, T. D. Smith, C. Rauch, C. A. Smith and et al. (1995). "Identification and characterization of a new member of the TNF family that induces apoptosis." Immunity **3**(6): 673-682.

Wilkerson, M. D. and D. N. Hayes (2010). "ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking." Bioinformatics **26**(12): 1572-1573.

Wong, H. R. (2012). "Genetics and genomics in pediatric septic shock." Crit Care Med **40**(5): 1618-1626.

Wong, H. R., N. Cvijanovich, R. Lin, G. L. Allen, N. J. Thomas, D. F. Willson, R. J. Freishtat, N. Anas, K. Meyer, P. A. Checchia, M. Monaco, K. Odom and T. P. Shanley (2009). "Identification of pediatric septic shock subclasses based on genome-wide expression profiling." BMC Med **7**: 34.

Wong, H. R., N. Z. Cvijanovich, N. Anas, G. L. Allen, N. J. Thomas, M. T. Bigham, S. L. Weiss, J. Fitzgerald, P. A. Checchia, K. Meyer, M. Quasney, M. Hall, R. Gedeit, R. J. Freishtat, J. Nowak, S. S. Raj, S. Gertz, K. Howard, K. Harmon, P. Lahni, E. Frank, K. W. Hart, T. C. Nguyen and C. J. Lindsell (2016). "Pediatric Sepsis Biomarker Risk Model-II: Redefining the Pediatric Sepsis Biomarker Risk Model With Septic Shock Phenotype." Crit Care Med **44**(11): 2010-2017.

Wong, H. R., T. E. Sweeney, K. W. Hart, P. Khatri and C. J. Lindsell (2017). "Pediatric Sepsis Endotypes Among Adults With Sepsis." Crit Care Med **45**(12): e1289-e1291.

Workman, J. K., S. G. Ames, R. W. Reeder, E. K. Korgenski, S. M. Masotti, S. L. Bratton and G. Y. Larsen (2016). "Treatment of Pediatric Septic Shock With the Surviving Sepsis Campaign Guidelines and PICU Patient Outcomes." Pediatr Crit Care Med **17**(10): e451-e458.

World Health Organization (2020). "Global report on the epidemiology and burden of sepsis: current evidence, identifying gaps and future directions." World Health Organization.

Wu, M. C., S. Lee, T. Cai, Y. Li, M. Boehnke and X. Lin (2011). "Rare-variant association testing for sequencing data with the sequence kernel association test." Am J Hum Genet **89**(1): 82-93.

Xu, F., S. H. Lin, Y. Z. Yang, R. Guo, J. Cao and Q. Liu (2013). "The effect of curcumin on sepsis-induced acute lung injury in a rat model through the inhibition of the TGF-beta1/SMAD3 pathway." Int Immunopharmacol **16**(1): 1-6.

Yang, J. O., M. S. Zinter, M. Pellegrini, M. Y. Wong, K. Gala, D. Markovic, B. Nadel, K. Peng, N. Do, S. Mangul, V. M. Nadkarni, A. Karlsberg, D. Deshpande, M. J. Butte, L. Asaro, M. Agus, A. Sapru and C.-P. Study Site Investigators for (2023). "Whole blood transcriptomics identifies subclasses of pediatric septic shock." Crit Care **27**(1): 486.

Yasin, S., N. Fall, R. A. Brown, M. Henderlight, S. W. Canna, C. Girard-Guyonvarc'h, C. Gabay, A. A. Grom and G. S. Schulert (2020). "IL-18 as a biomarker linking systemic juvenile

idiopathic arthritis and macrophage activation syndrome." Rheumatology (Oxford) **59**(2): 361-366.

Ye, J. and L. N. Sanchez-Pinto (2020). "Three Data-Driven Phenotypes of Multiple Organ Dysfunction Syndrome Preserved from Early Childhood to Middle Adulthood." AMIA Annu Symp Proc **2020**: 1345-1353.

Yeung, M. W., S. Wang, Y. J. van de Vegte, O. Borisov, J. van Setten, H. Snieder, N. Verweij, M. A. Said and P. van der Harst (2022). "Twenty-Five Novel Loci for Carotid Intima-Media Thickness: A Genome-Wide Association Study in >45 000 Individuals and Meta-Analysis of >100 000 Individuals." Arterioscler Thromb Vasc Biol **42**(4): 484-501.

Yoe-Sik Bae, G. H. B., Seon Hyang Park, Ji Hyeon Kang, Brian A. Zabel, Sung Ho Ryu (2023). "Chapter 3 - The role and regulation of phospholipase D in infectious and inflammatory diseases." Phospholipases in Physiology and Pathology: 43-77.

Zhang, A. Q., W. Pan, J. W. Gao, C. L. Yue, L. Zeng, W. Gu and J. X. Jiang (2014). "Associations between interleukin-1 gene polymorphisms and sepsis risk: a meta-analysis." BMC Med Genet **15**: 8.

Zhang, Q., Z. Qin, S. Yi, H. Wei, X. Z. Zhou and J. Su (2020). "Two novel compound heterozygous variants of LTBP4 in a Chinese infant with cutis laxa type IC and a review of the related literature." BMC Med Genomics **13**(1): 183.

Zhang, Q., Y. Zhao, R. Zhang, Y. Wei, H. Yi, F. Shao and F. Chen (2016). "A Comparative Study of Five Association Tests Based on CpG Set for Epigenome-Wide Association Studies." PLoS One **11**(6): e0156895.

Zhang, Z., G. Zhang, H. Goyal, L. Mo and Y. Hong (2018). "Identification of subclasses of sepsis that showed different clinical outcomes and responses to amount of fluid resuscitation: a latent profile analysis." Crit Care **22**(1): 347.

Zheng, S. C., C. E. Breeze, S. Beck and A. E. Teschendorff (2018). "Identification of differentially methylated cell types in epigenome-wide association studies." Nat Methods **15**(12): 1059-1066.

Zimmerman, J. J., R. Banks, R. A. Berg, A. Zuppa, C. J. Newth, D. Wessel, M. M. Pollack, K. L. Meert, M. W. Hall, M. Quasney, A. Sapru, J. A. Carcillo, P. S. McQuillen, P. M. Mourani, H. Wong, R. S. Chima, R. Holubkov, W. Coleman, S. Sorenson, J. W. Varni, J. McGalliard, W. Haaland, K. Whitlock, J. M. Dean, R. W. Reeder and I. Life After Pediatric Sepsis Evaluation (2020). "Critical Illness Factors Associated With Long-Term Mortality and Health-Related Quality of Life Morbidity Following Community-Acquired Pediatric Septic Shock." Crit Care Med **48**(3): 319-328.