Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Gene Expression Analysis for Modelling Immune Trajectory and Endotype Transition During Pediatric Septic Shock

Shani Shameem
Imperial College London
Claire Dunican
Imperial College London https://orcid.org/0000-0001-5835-263X
Elisabeth Robinson
London School of Hygiene & Tropical Medicine https://orcid.org/0000-0002-0528-0561
Yael Feinstein
Reichman University
Joshua Hageman
Imperial College London
Shunmay Yeung
London School of Hygiene & Tropical Medicine
Jethro Herberg
Imperial College London
Ashleigh Green
Imperial College London
Victoria Wright
Imperial College London
Samuel Nichols
Imperial College London
Nazima Pathan
University of Cambridge
Naomi Edmonds
Barts Health NHS Trust
Katie Burnham
Wellcome Sanger Institute
Soukaina Timouma
University of Oxford
Julian C Knight
University of Oxford
Michael Levin

Imperial College London
Aubrey Cunnington
Imperial College London
Myrsini Kaforou
Imperial College London
Simon Nadel
Imperial College Healthcare NHS Trust
Mark Peters
University College London
Padmanabhan Ramnarayan
Imperial College London
Michael Carter
michael.carter@well.ox.ac.uk

University of Oxford https://orcid.org/0000-0003-3403-4746

Research Article

Keywords: Sepsis, immunology, pediatrics, critical care, gene expression

Posted Date: November 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11541283/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare potential competing interests as follows: JCK reports a grant to his institution from the Danaher Beacon Programme for work on RNA biomarker point-of-care test development in sepsis for endotype assignment in adult patients. Danaher had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. No other authors declare competing interests.

Abstract

Differential gene expression can describe groups of patients (endotypes) with differential outcome in sepsis based on immune (dys)function. However, limited mechanistic understanding of immune trajectory in sepsis hampers potential modulation of immune dysfunction. We aimed to characterise the immune trajectory of 19 children with septic shock from first interaction with pediatric critical care teams (time point, TP, 1) to 48 hours afterwards (TP2). We used differential gene expression, including weighted gene co-expression network analysis (WGCNA), digital cytometry and RNA velocity from bulk RNA-Seq data to compared modelled future gene expression with measured change in gene expression. We discovered key gene modules (WGCNAs, named by “hub” gene) associated with myeloid cell expansion between at TP1 (*CMTM6*) and poor outcome in previously described BASIC Endotype 1 patients (*RAB2B*), and modules associated with adaptive immune function at TP2 (*BCL2*) among others. RNA velocity analysis showed that immune trajectory was predictable from TP1 samples, including the ability to identify patients that subsequently transitioned from BASIC Endotype 1 (poor outcome) to BASIC Endotype 2 (better outcome). These data provide proof of concept for the identification of immune trajectory (including potentially treatable traits) from a single early sample in pediatric septic shock.

INTRODUCTION

Sepsis is a dysregulated host response to infection that leads to life-threatening organ dysfunction[1,2]. Hallmarks of sepsis include the release of circulating mediators of inflammation, an activated vascular endothelium (contributing to increased vascular permeability and capillary fluid leak), impaired antigen presentation, neutrophil activation and cell death, and exhausted and apoptotic T cells (contributing to immune suppression)[3]. Outcomes in sepsis are therefore determined by the relationship between host response and pathogen and timely access to appropriate treatment and organ supportive care. Although sepsis may occur in all age groups, the incidence is greatest in young children and the elderly. Childhood sepsis represents a unique clinical and immunological challenge, distinct from adults, due to the relative immaturity of innate and adaptive immunity[3].

Gene expression profiling – transcriptomics – has emerged as a powerful tool to understand the molecular mechanisms underlying sepsis[4]. In adult sepsis, gene expression profiling has led to the development and validation of “endotypes” (groups of patients characterised by shared immune responses, typically derived from gene expression data)[5–9] and mechanistic insights that may inform treatment decisions[10,11]. In childhood sepsis and other inflammatory conditions, endotypes predictive of outcome have also been described[12–15], as have gene expression-based severity markers[16] and aetiology-specific gene signatures to enhance diagnostic testing[17–19]. Many studies have examined gene expression at single time points, typically 24–48 hours following presentation to hospital, to measure the whole blood immune response. The use of a single time point limits both the ability to interpret likely highly dynamic gene expression and understanding of the immune trajectory in sepsis[20].

A delay in sampling of 24–48 hours also may miss early gene expression that may represent peak immune dysfunction in sepsis and creates a survivor bias because most children who do not survive sepsis die within the first 48 hours following hospital presentation[3,21,22].

We therefore aimed to characterise temporal changes in gene expression, modules of co-regulated/expressed genes, and immune cell composition at two time points during the early stages of childhood sepsis. We selected a subset of patients with septic shock requiring invasive ventilation and vasoactive infusions from the Biomarkers of Acute Serious Illness in Children (BASIC) study[23] from whom paired samples taken on retrieval of children to pediatric intensive care units (PICUs) and ~ 48 hours following PICU admission were available. We explored whether there were differences in gene expression and immune trajectory by groups of patients, known as endotypes, that have been previously described at baseline admission to PICU in patients from the cohort[14]. We also explored whether *future* gene expression during the illness could be modelled from the spliced and unspliced transcript patterns of expressed genes at the earliest practical time point following presentation with sepsis using RNA velocity analysis[24]. This approach could hypothetically identify, and enable precision medicine for, a dysfunctional immune trajectory from a single early time point sample, leading to improved personalized treatment for critically ill patients.

RESULTS

Recruitment and study cohort

During the BASIC study period, 1017 eligible children were transported to participating PICUs via the retrieval team with any cause of critical illness, of whom 674 children were recruited. Children were sampled for gene expression analysis on first contact with retrieval team or admitting PICU (via a deferred consent approach[23,25]; Timepoint 1, TP1) and in a subset of children at 48 hours (following consent; Timepoint 2, TP2). From this cohort, we selected 19 children with a contemporaneous diagnosis of infection/sepsis who required both invasive mechanical ventilation and vasoactive infusions and from whom paired TP1 and TP2 samples for RNA-Seq analyses were available (Fig. 1a). Five (26%) children were 0–11 months of age, further 8 (42%) children were 12 months–4 years of age, and 6 (32%) were ≥ 5 years of age. The most prevalent causes of sepsis in this study cohort were septic shock without focus (7, 37%, children) and pneumonia (6, 32%, children). The median Phoenix Sepsis Score[2,26] was 3 at presentation (IQR 2–5), the median duration of mechanical ventilation was 133 (IQR 91–224) hours and the median duration of vasoactive infusions was 45 (IQR 16–102) hours. Of these 19 children, no participants died in-hospital. The duration of pre-hospital symptoms, duration of organ support and duration of PICU stay for each participant are shown in Fig. 1b. The microbiological diagnoses and etiological classification of patients (using a standardised phenotyping system)[27] are in Supplementary Table 1.

Differential expression analysis

Principal component analysis (PCA) constructed using all expressed genes showed no clear directionality between TP1 and TP2 on principal components (PC) 1 and PC2 (variance 16% and 13% respectively), and on other principal component combinations, suggesting that at the whole-transcriptome level, inter-individual differences dominate over TP-driven variation (Fig. 2a). Differential gene expression analysis identified 1880 differentially expressed genes (DEGs), of which 1050 had lower expression values, and 830 had higher expression values at TP2 relative to TP1 (Benjamini-Hochberg-adjusted $p < 0.05$ and $\log_2$ fold-change ≥ 0.5; Fig. 2b). PCA based on the DEGs captured a higher proportion of the variance on PC1 (38%) and showed greater separation by TP (Fig. 2c) compared to the PCA using all expressed genes (Fig. 2a).

To estimate the proportions of immune cell types in participants at TP1 and TP2, we used digital cytometry (CIBERSORTx, https://cibersortx.stanford.edu[28]). There was a significant decrease in the proportion of neutrophils ($p = 0.001$), and a significant increase in the proportions of naïve CD4 T cell ($p = 0.002$), CD8 T cell ($p = 0.005$) and naïve B cell ($p = 0.002$) proportions between TP1 and TP2 (Fig. 2d); other major immune cell subset proportions did not significantly alter over time. We used multivariable regression analyses, to explore the effect of TP on immune cell subsets adjusted for age group, etiological classification and BASIC endotype (see below). In these analyses, neutrophil proportions were higher in children > 12 years of age ($p = 0.039$) at TP1, and increased more in bacterial infection in comparison with viral infection from TP1 to TP2 ($p = 0.034$), while naïve B cell proportions increased more in BASIC Endotype 2 in comparison with Endotype 1 from TP1 to TP2 ($p = 0.047$; Supplementary Table 2).

Weighted gene co-expression network analysis (WGCNA)

Gene expression data were analysed using signed WGCNA[29,30] to identify gene co-expression (positive or inverse association) networks in participants between TP1 and TP2 (Fig. 3a and Supplementary Fig. 1). Ten gene co-expression modules (Fig. 3b, 3c), were identified from differential gene expression between TP1 and TP2, named by the hub gene (the gene with the highest connectivity for each module). To assess temporal change in module activity, module eigengene values were compared between paired TP1 and TP2 samples using Wilcoxon signed-rank tests, with Benjamini–Hochberg (BH) correction. Five gene co-expression modules showed significant temporal changes. *CMTM6*, *DYNLT1 and RAB2B* had higher eigengene values at TP1 than TP2, with negative median paired differences of -0.19, -0.15 and −0.077 respectively (BH-adjusted $p = 0.0019$, $p = 0.001$, $p = 0.033$). Conversely, *BCL2* and *RAN* had higher eigengene values at TP2 than TP1, with positive median paired differences of 0.18 and 0.089 respectively (BH-adjusted $p = 0.0019$, $p = 0.0088$).

Gene set enrichment analysis[31] (GSEA) of the *CMTM6* module, comprising 2805 genes, identified significant enrichment of innate immune function, myeloid cell differentiation and activation (all p <

0.001, Fig. 3c). GSEA of the *DYNLT1* module, comprising 143 genes, identified significant enrichment of cellular component pathways of the early endosome membrane and canonical inflammasome complex ($p < 0.05$, not shown). GSEA of the *RAB2B* module are described below. GSEA of the *BCL2* module, comprising 1943 member genes, identified significant enrichment of the adaptive immune response, including lymphocyte differentiation, B cell activation and proliferation and MHC class II assembly (all $p < 0.01$, Fig. 3d). GSEA of the *RAN* module, comprising 4324 genes, identified significant enrichment of transcription and translational pathways, including ribonucleoprotein complex biogenesis, ribosome biogenesis, rRNA processing and cytoplasmic and mitochondrial translation (all $p < 0.001$, Fig. 3e).

BASIC Endotype analysis

Shared patterns of immune responses – endotypes – had previously been described, and were associated with outcome, in a separate group of 382 children in the BASIC cohort using microarray analysis[14]. We therefore determined BASIC Endotype for each patient at each time point by using quantile normalised RNA-Seq data with a 20-gene signature of endotype generated by forward-selection partial least squares (FS-PLS)[32] and applied via a linear regression model trained and cross-validated within the normalised BASIC microarray dataset[14]. Endotype assignments were concordant with sample separation observed on PCA of normalised expression data across both microarray and RNA-Seq platforms (Fig. 4a). At TP1, 9 participants were classified as Endotype 1 and 10 as Endotype 2 (Fig. 1a). At TP2, 6 participants were classified as Endotype 1 and 13 as Endotype 2 following the transition of 3 participants from Endotype 1 to Endotype 2 (Fig. 4b).

WGCNA identified a single gene co-expression module (*RAB2B*) that was significantly associated with Endotype 1 compared with Endotype 2 at both TP1 and TP2 (median module eigenvalue differences −0.24 $p = 0.026$ and −0.2 $p = 0.047$ respectively; Fig. 4c). The *RAB2B* module comprised 479 genes and was significantly enriched for pathways related to myeloid differentiation and development, erythrocyte homeostasis, and proteasome-mediated ubiquitin-dependent catabolic processes (all $p < 0.01$; Fig. 4d).

RNA velocity analysis

We next modelled *future* gene expression of the significantly regulated gene co-expression modules identified from WGCNA. We used the RNA velocity analysis[33] algorithm VeloCD[34], specifically designed for bulk RNA-Seq data analysis, to create RNA velocity fate maps for each significantly regulated gene co-expression module for each participant. VeloCD models future expression levels for each gene based on the modelled relationship between the spliced and unspliced transcript expression. This enables projection of participants onto a low-dimensional transcriptomic space with RNA velocity arrows assigned to each participant-TP (Fig. 5a–d).

The multi-exon genes with spliced and unspliced transcript expression from the TP associated modules (*CMTM6*, Fig. 5a; *DYNLT1*, Fig. 5b; *BCL2*, Fig. 5c; *RAN*, Fig. 5d) and Endotype-associated module (*RAB2B*,

Fig. 5e) were separately run through this algorithm: with sample points colored by Endotype and shaped by TP. All fate maps showed some relationship between the future gene expression and TP or Endotype. TP1 participant-samples RNA velocity arrows pointing towards the TP2 samples was most clear for the *DYNLT1* module (Fig. 5b), which had strongest correlation with time from the WGCNA (Fig. 3b). Consistent with the directionality of the WGCNA correlation, the RNA velocity arrows of the *BCL2* and *RAN* modules show inverse directionality with time in comparison to *CMTM6* and *DYNLT1* (Fig. 5b-d).

The *RAB2B* module fate map (Fig. 5e) showed that the 6 participants who transition from Endotype 1 to Endotype 2 at TP2 ("transitioners") are placed between Endotype 1 "remainers" and Endotype 2 remainers. The Endotype 1 remainer RNA velocity arrows also point towards Endotype 2 suggesting that the immune processes captured by this gene co-expression module are moving all participants towards Endotype 2, with the Endotype 1 participants just much earlier on in this process.

DISCUSSION

We aimed to characterise the immune trajectory of childhood septic shock by studying the dynamic transcriptome. We identified changes in immune cell subsets and gene co-expression modules that were significantly regulated between TP1 and TP2: the *CMTM6* module, enriched for myeloid cell differentiation and activation, and the *DYNLT1* module, enriched for the canonical inflammasome complex, had higher expression at TP1. The *BCL2* module, enriched for lymphocyte differentiation, B cell activation and MHC Class II assembly, had higher expression at TP2. Our identification of gene co-expression modules is supported by a wide body of literature describing a transition from innate to adaptive immune function in early infection[3,11]. Our temporal data that estimate early high proportions of neutrophils and a late increase in CD4 and CD8 T cell proportions are concordant with this.

We extended temporal analysis by examining the relationship between previously identified BASIC Endotypes, and temporal changes in gene expression. This identified an additional gene co-expression module, *RAB2B*, that was significantly positively associated with BASIC Endotype 1 and was enriched for myeloid differentiation and development, erythrocyte homeostasis, and proteasome-mediated ubiquitin-dependent catabolic processes. The *RAB2B* module was broadly consistent with previous work, which identified myeloid response as key drivers of BASIC Endotype – and of endotypes or sepsis response syndromes in sepsis in adults[5,8,9]. Previous WGCNAs of sepsis at single TPs have identified endothelial cell dysfunction and reduced T cell proliferation (hub gene *SLC51A*) and myeloid leukocyte activation (hub gene *C3orf86*) to be elevated in sepsis, and the adaptive immune response (hub gene *DDX24*) to be decreased in sepsis in comparison to controls[35]. A previous WCGNA in adult sepsis at a single TP, with the temporal course of analysed in a murine model of sepsis led to similar conclusions[36]. In addition, clinical studies of sepsis in adults have previously identified an immunosuppressive neutrophil population – probably released through emergency granulopoiesis[37] – that is associated with reduced CD4 T cell function and worse sepsis outcome in adults[38]. These analyses, taken as a whole, support a recognition of neutrophils to be key mediators of immune trajectory in child (and adult) sepsis.

Establishing a causal relationship between gene co-expression modules and outcome in sepsis (most notably *CMTM6* and similar emergency granulopoiesis/myeloid activation modules identified elsewhere) would require either a Mendelian randomisation approach or a (quasi)randomised blockade of emergency granulopoiesis, linked to clinical outcomes in large data sets. To our knowledge, such studies have yet to be completed but are feasible. Such an approach would inform the repurposing or discovery of immune modulator therapies for immune dysfunction in sepsis, and an analysis across the life-course may have significant implications for our understanding of immune ageing.

Studies of sepsis aspire to discriminate pathological immune dysfunction from physiological immune function. This is hampered by practical and ethical challenges in sampling a single participant from pathogen colonisation/inoculation to transition to infection and transition to sepsis, even in carefully controlled challenge studies. As a result, temporal changes, identified through repeated sampling, may need to be used to identify patients that have a pathological immune trajectory after the diagnosis of sepsis is made (and treatment and support commenced). Children in the BASIC cohort were recruited exceptionally early (before or at PICU admission) in comparison with many other adult and child cohorts[8,39]. This early TP1 sample, combined with TP2, enabled us to identify temporal changes that may be underestimated in other cohorts. We extended this by using RNA velocity analysis[33,34] (VeloCD algorithm) to model the immune trajectory of participants from TP1 to TP2. We hypothesize that RNA velocity may improve the identification of patients with a trajectory towards worsening immune dysfunction by characterising changing gene expression from a single sample[40]. This approach, based on the relationship between the spliced and unspliced transcript expression of the genes in each co-expression module, could be accelerated by limiting sequencing and analysis of smaller gene signatures from each module; with a potential test-to-result time making clinical implementation feasible for critically unwell patients.

This study had a sample size of 19 participants (38 samples) and is underpowered to discover more subtle gene co-expression modules that may differ between TPs or clusters of patients. In addition, all participants survived, and all participants had an immune trajectory that was relatively stereotyped across PCA, digital cytometry, Endotype transition and RNA velocity fate maps. This prevented a comparison between the immune trajectories of patients who did, and did not, survive sepsis that may have been informative in identifying gene co-expression modules that were associated with poor outcome (and are therefore more likely to represented pathological immune dysfunction). Although the wider BASIC cohort was developed with a deferred consent model for sampling[23], consent was subsequently required for retention and analysis of samples. Consent from parents for children who subsequently died was rarely achieved, leading to a survivor bias in the wider BASIC cohort. Providing a compassionate and ethically-sound approach to gain consent from parents faced with this tragedy in future would enhance the ability to learn from the samples and data of children who do die early within a PICU.

Clinical insight would also benefit from biological analyses of single cells, using cytometry (including immune state markers), single cell RNA-Seq, epigenomic studies and stimulation or co-culture

experiments to examine post-transcriptional and post-translational immunology. A limitation in many studies of immune trajectory in sepsis in both children and adults has been the exclusion of neutrophils due to methodological and logistical challenges in handling these fragile cells[41,42]. Our data suggest that this is a significant omission, and the exclusion of neutrophils from analyses may miss important age-related effects. Given the increasing realisation of post-sepsis morbidity and mortality[11,43], future clinical studies should also incorporate medium- and long-term follow-up of participants beyond hospital discharge, and possibly through routinely collected healthcare data.

In conclusion, this study provides a model for the analysis of temporal changes in gene expression in sepsis, including the estimation of immune cell proportions, derivation of gene co-expression modules (that are associated with clinically important features), endotype membership, and the modelling of future immune trajectory based on RNA velocity fate maps. The use of these approaches in larger data sets – particularly in combination with other -omic analyses – would provide further support for incorporating immune trajectory into clinical decision-making in sepsis in children and adults.

MATERIALS AND METHODS

Clinical recruitment and data

We analysed data collected as part of the Biomarkers of Acute Serious Illness in Children (BASIC) study that recruited participants between 2014–2016[23]. BASIC was a prospective cohort study that enrolled critically ill children admitted to four PICUs in London and Eastern England, UK, during emergency transport by the Children's Acute Transport Service (CATS), the regional paediatric critical care retrieval service. Children aged 0–16 years with an indwelling central venous or arterial catheter were eligible for the BASIC cohort. Infants < 36 weeks of corrected gestational age and those with agreed end-of-life care plans were not recruited. Recruitment followed deferred consent processes, with written informed consent obtained by the study research team from parents/legal guardians within 24–48 hours[25]. According to UK best practice, the study was approved by both the Health Research Authority (HRA, reference: 136866) and ethical approval was provided by the National Research Ethics Committee East Midlands—Nottingham 2 (reference: 13-EM-0399) independently of the authors' academic and clinical affiliations. Procedures followed were in accordance with the Helsinki Declaration of 1975. The study reporting follows the Standards of Reporting of Diagnostic Accuracy Studies 2015 Update (https://www.equator-network.org/reporting-guidelines/stard/). The clinical trials registry for Biomarkers of Acute Serious Illness in Children (BASIC) study is at: https://clinicaltrials.gov/study/NCT03238040, trial number: NCT03238040. Data presented in this paper represents a retrospective analysis of data collected during the BASIC study.

Of the 674 children recruited to the BASIC study, Timepoint (TP) 1 samples have been previously analysed and presented from 382 (57%) children using Illumina microarray technology[14] and these children and data were not analysed again here. Of the remaining 292 children in the cohort, 12 (4.1%)

children died, 5 of whom with infection/sepsis; none of these children had Timepoint 2 (TP2) samples available for analysis. We therefore selected 19 children with septic shock (defined operationally as children with suspected or confirmed infection who were invasively ventilated and receiving vasoactive infusions via a central venous catheter) from whom TP1 and TP2 RNA-stabilised samples were available (see Fig. 1 and Supplementary Table 1).

We used patient demographic data (sex at birth, age, self-reported ethnicity), clinical characteristics (reason for PICU admission), details of infection status (including phenotyping for bacterial, viral, unknown and non-infectious illnesses using standardized published criteria[27]), clinical severity markers such as retrospectively applied Phoenix Sepsis Score (PSS)[2] at admission to PICU as appropriate, interventions performed (invasive ventilation, vasoactive drugs) during transport and PICU discharge outcome (vital status and duration of organ support).

RNA sampling and sequencing

Whole blood was sampled into PAXgene (Qiagen, UK) tubes and stored appropriately at -80ºC. RNA was extracted according to manufacturer's instructions using the QIAsymphony PAXgene blood miRNA extraction kit (Qiagen, UK). Extracted samples of RNA were DNA depleted (RNA Clean and Concentrator-5 with DNase, Zymo Research, CA, USA) and globin depleted (GLOBINclear Human Kit, ThermoFisherScientific) and sequenced on an Illumina NextSeq2000 at a core genomics facility (Imperial Genomics Facility, Imperial NIHR Biomedical Research Centre) to generate 50 bp paired-end reads. Raw fastq files were aligned to the human genome (GRCh38) using Salmon 1.10.2 (https://salmon.readthedocs.io/en/latest/salmon.html)[45] and gene counts quantified prior to differential gene expression analyses.

Differential gene expression analysis

Statistical analysis of gene expression was undertaken in R: A Language and Environment for Statistical Computing (version 4.4.1)[46]. Differential gene expression analysis was performed using DESeq2[47] (v1.42) in R. Transcript-level abundance estimates from Salmon were imported using tximport, and genes with fewer than 10 total reads across all samples were excluded prior to analysis. To identify genes differentially expressed between TPs while accounting for variability between participants, patient ID was included as a covariate in the DESeq2 design formula. The DESeq function was run with default parameters, and normalised counts were extracted for downstream use. Differential expression between TP2 and TP1 was assessed, with genes considered significant if they met a Benjamini-Hochberg-adjusted p-value < 0.05 and an absolute $\log_2$ fold change ≥ 0.5. Results were visualised using a volcano plot, with the top 20 most significant genes labelled.

Digital cytometry to impute immune cell populations was undertaken with Cibersort (https://cibersortx.stanford.edu; quantile normalisation was disabled, run mode was relative and 100 permutations for significance testing were used)[44] using the reference data set LM22 and used to infer the differential proportions of immune cells between participants at TP1 and TP2.

Weighted gene co-expression network analysis

Weighted gene co-expression network analysis was performed using the WGCNA R package[30] (v1.73) to identify modules of co-expressed genes and relate their activity to clinical response. A soft-thresholding power of β = 5 was selected based on the criterion of approximate scale-free topology (R^2>0.8), as determined by the pickSoftThreshold function. A signed gene co-expression network was constructed using blockwiseModules with the following parameters: minModuleSize = 30, deepSplit = 2, and mergeCutHeight = 0.25. The signed network type was chosen to retain the directionality of gene-gene correlations, distinguishing positively from negatively co-expressed gene pairs. The grey module, comprising genes not assignable to any co-expression module, was excluded from downstream analyses.

Module eigengenes (MEs), defined as the first principal component of each module's expression matrix, were compared between paired TP1 and TP2 samples using Wilcoxon signed-rank tests. Differences in ME values between assigned endotypes were assessed at each timepoint using Wilcoxon rank-sum tests. Modules were labelled by their hub gene, defined as the intramodular gene with the highest module membership. P values were adjusted for multiple testing using the Benjamini–Hochberg method. Gene ontology (GO) enrichment analysis of selected modules was performed using clusterProfiler[48] v4.14.4, with a q-value threshold of 0.05.

BASIC Endotypes

Microarray transcriptomic data from the previously presented 382 children at point of PICU admission alongside BASIC Endotype assignment determined by unsupervised k-means clustering[14] were analysed alongside paired TP RNA-Seq data from the separate 19 children from the same BASIC cohort[23]. Genes common to both datasets were retained (n = 2373). Both datasets were $\log_2$ transformed and quantile normalised to the microarray derived quartile distribution using preprocessCore[49]. PCA confirmed appropriate distributional alignment between platforms[50]. A 20-gene signature, discriminatory between BASIC Endotype 1 and 2, was derived from the microarray cohort using forward selection partial least squares (FS-PLS)[32,51]. A penalised logistic regression classifier (glmnet, 10-fold cross-validation) was trained on the microarray data and applied to RNA-Seq samples to assign endotype membership. A continuous BASICq score was calculated using diffusion mapping and scaled to the interval [0,1] to validate endotype membership. Endotype assignments were verified as concordant with sample separation observed on principal component analysis (PCA) of quantile normalised expression data across both microarray and RNA-Seq platforms.

RNA velocity and transcriptomic trajectory methods

The raw RNA-Seq fastq files were re-mapped and re-quantified using STAR[52], featureCounts[53] and the GRCh38 genome assembly fasta and GTF files from Ensembl. Counts were then summed across sequencing lanes and those mapped to introns and exons for each gene (meta features) were used to

derive library size normalised spliced and unspliced expression values, as previously described[34]. Genes without both spliced and unspliced transcript expression were then removed as well as those containing any introns labelled with the biotype “retained intron” or overlap with an annotated exon of another gene. It is important to note that this may filter out key genes from the following modules, including the hub genes.

Of the 2805 genes from the *CMTM6* module, 1579 had spliced and unspliced transcript expression. This was 1079/1943 (55.5%), 86/143 (60.1%), 2589/4076 (63.5%) and 300/479 (62.6%) from the *DYNLT1, BCL2*, *RAN* and *RAB2B* and modules respectively. We then put the genes from each module into RNA velocity analysis (VeloCD algorithm[34]) using all 38 samples (from 19 participants) and examined the resulting fate maps. This was run using a number of neighbour range of 2 to 37 and transition probability number of neighbour range of 2 to 36.

DECLARATIONS

COMPETING INTERESTS

JCK reports a grant to his institution from the Danaher Beacon Programme for work on RNA biomarker point-of-care test development in sepsis for endotype assignment in adult patients. Danaher had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. No other authors declare competing interests.

AUTHOR CONTRIBUTION

SS, CD, ER, JHa, ST, AC, MJP, PR and MJC contributed to the conceptualization and design of the study. SS, CD, ER, YF, JHe, AG, VW, SN, NP, NE, ML, MK, SN, MLP, PR and MJC. MJC, JHa, DHC, KB, MK and PR were responsible for acquisition, analysis or interpretation of data. YF, SY, JHe, VW, SN, NP, NE, ML, MJP, PR were responsible for clinical investigation. SS, CD, ER, KB and MJC, were responsible for visualization of data. PR, ML, JHe, MK and MJP were responsible for funding acquisition. PR, MJC, MK, KB, JCK, AC, ML were responsible for supervision. Writing of the original draft was led by MJC, SS, CD, ER and PR. Review and editing of subsequent drafts was led by MJC, SS, CD, ER, PR, MJP and AC. MJC, SS, CD, and ER have directly accessed and verified the underlying data in the manuscript. All authors have had full access to the data and approve of submission.

ACKNOWLEDGEMENTS

We thank all the parents and children participating in this study and the medical, nursing, and research teams at the participating sites for their help in study setup, recruitment, data collection, and monitoring of study data for their invaluable support. Prof Ramnarayan, Prof Herberg, Prof Cunnington, Prof Levin

and Dr Kaforou, and Prof Peters receive funding from the National Institute for Health and Care Research (NIHR), in part through the Imperial-NIHR (PR, JH, AC, ML, CD, MK) and UCL-NIHR Biomedical Research Centres (MJP), respectively. Dr Dunican is supported by the NIH. Dr Robinson is supported by an NIHR Academic Clinical Fellowship. Dr Hageman, Prof Herberg, Prof Peters and Prof Ramnarayan received support from the Great Ormond Street Hospital Children's Charity. Dr Carter, Dr Timouma and Prof Knight receive funding via the National Health Service Genomic Network of Excellence Award for Severe Presentations of Infectious and Immune Disorders. KB is a member of the Wellcome Sanger Institute which is funded by the Wellcome Trust 220540/Z/20/A. For Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript arising from this submission.

DATA AVAILABILITY

The datasets generated and/or analysed during the current study are available in the ArrayExpress repository, https://www.ebi.ac.uk/biostudies/ArrayExpress/studies/E-MTAB-17576, including links to the raw RNA-Seq data (fastq files) held in the European Nucleotide Archive. Pseudo-anonymised clinical data will be made available on reasonable request to the corresponding author.

All code for the analysis is available at: https://github.com/michaeljamescarter/basic_trajectory and https://github.com/DrClaireDunican/VeloCD.

REFERENCES

1. Singer M et al (2016) The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 315:801–810. 10.1001/jama.2016.0287
2. Schlapbach LJ et al (2024) International Consensus Criteria for Pediatric Sepsis and Septic Shock. JAMA 331:665–674. 10.1001/jama.2024.0179
3. Carter MJ et al (2024) Susceptibility to childhood sepsis, contemporary management, and future directions. Lancet Child Adolesc Health 8:682–694. 10.1016/S2352-4642(24)00141-X
4. Wong HR et al (2009) Identification of pediatric septic shock subclasses based on genome-wide expression profiling. BMC Med 7:34. 10.1186/1741-7015-7-34
5. Cano-Gamez E et al (2022) An immune dysfunction score for stratification of patients with acute infection based on whole-blood gene expression. Sci Transl Med 14:eabq4433. 10.1126/scitranslmed.abq4433
6. Sweeney TE et al (2018) A community approach to mortality prediction in sepsis via gene expression analysis. Nat Commun 9. 10.1038/s41467-018-03078-2
7. Scicluna BP et al (2017) Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study. Lancet Resp Med 5:816–826. 10.1016/s2213-2600(17)30294-1

8. Davenport EE et al (2016) Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study. Lancet Resp Med 4:259–271. 10.1016/s2213-2600(16)00046-1
9. Scicluna BP et al (2025) A consensus blood transcriptomic framework for sepsis. Nat Med 31:4119–4130. 10.1038/s41591-025-03964-5
10. Antcliffe DB et al (2019) Transcriptomic Signatures in Sepsis and a Differential Response to Steroids. From the VANISH Randomized Trial. Am J Respir Crit Care Med 199:980–986. 10.1164/rccm.201807-1419OC
11. Shankar-Hari M et al (2024) Reframing sepsis immunobiology for translation: towards informative subtyping and targeted immunomodulatory therapies. Lancet Respir Med. 10.1016/S2213-2600(23)00468-X
12. Wong HR et al (2009) Identification of pediatric septic shock subclasses based on genome-wide expression profiling. BMC Med 7:34
13. Atreya MR et al (2024) Identification and transcriptomic assessment of latent profile pediatric septic shock phenotypes. Crit Care 28:246. 10.1186/s13054-024-05020-z
14. Carter MJ et al (2025) Pediatric critical illness endotypes reveal distinct outcomes and immune pathways shared across cause of illness. *iScience* 28, 114210. 10.1016/j.isci.2025.114210
15. Wong HR, Hart KW, Lindsell CJ, Sweeney TE (2021) External Corroboration That Corticosteroids May Be Harmful to Septic Shock Endotype A Patients. Crit Care Med 49:e98–e101. 10.1097/ccm.0000000000004709
16. Schlapbach LJ et al (2024) Host gene expression signatures to identify infection type and organ dysfunction in children evaluated for sepsis: a multicentre cohort study. Lancet Child Adolesc Health. 10.1016/s2352-4642(24)00017-8
17. Herberg JA et al (2016) Diagnostic test accuracy of a 2-transcript host RNA signature for discriminating bacterial vs viral infection in febrile children. JAMA 316:835–845. 10.1001/jama.2016.11236
18. Jackson HR et al (2023) Diagnosis of multi-system inflammatory syndrome in children by a whole-blood transcriptional signature. J Ped Inf Dis Soc
19. Habgood-Coote D et al (2023) Diagnosis of childhood febrile illness using a multi-class blood RNA molecular signature. *Med* 4, 635–654 e635. 10.1016/j.medj.2023.06.007
20. Kwan A, Hubank M, Rashid A, Klein N, Peters MJ (2013) Transcriptional instability during evolving sepsis may limit biomarker based risk stratification. PLoS ONE 8:e60501. 10.1371/journal.pone.0060501
21. Cvetkovic M et al (2015) Timing of death in children referred for intensive care with severe sepsis: implications for interventional studies. Pediatr Crit Care Med 16:410–417. 10.1097/pcc.0000000000000385
22. Wedgwood M, Randle E, Honsel M, Ramnarayan P, Peters MJ (2025) Timing of Death in Children Referred for Intensive Care With Sepsis: Comparison of Two Cohorts in the United Kingdom, 2005–2011 vs. 2018–2023. Pediatr Crit Care Med 26:e1323–e1329. 10.1097/pcc.0000000000003825

23. Feinstein Y et al (2018) Cohort profile of the Biomarkers of Acute Serious Illness in Children (BASIC) study: a prospective multicentre cohort study in critically ill children. BMJ Open 8:e024729. 10.1136/bmjopen-2018-024729
24. Dunican C Predicting trajectories of acute illness using RNA velocity of whole blood. (Under review).
25. Woolfall K et al (2015) How parents and practitioners experience research without prior consent (deferred consent) for emergency research involving children with life threatening conditions: a mixed method study. BMJ Open 5:e008522. 10.1136/bmjopen-2015-008522
26. Carter MJ et al (2025) Evaluation of Phoenix Sepsis Score Criteria: Exploratory Analysis of Characteristics and Outcomes in an Emergency Transport PICU Cohort From the United Kingdom, 2014–2016. Pediatr Crit Care Med. 10.1097/pcc.0000000000003682
27. Nijman RG et al (2021) A Novel Framework for Phenotyping Children With Suspected or Confirmed Infection for Future Biomarker Studies. Front Pediatr 9:688272. 10.3389/fped.2021.688272
28. Newman AM et al (2015) Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 12:453–457. 10.1038/nmeth.3337
29. Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 4., Article17, 10.2202/1544-6115.1128
30. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559. 10.1186/1471-2105-9-559
31. Subramanian A et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102:15545–15550. 10.1073/pnas.0506580102
32. Rawlinson D, Zhou C, Kaforou M, Cao KL, Coin L (2025) J. M. A flexible framework for minimal biomarker signature discovery from clinical omics studies without library size normalisation. PLOS Digit Health 4:e0000780. 10.1371/journal.pdig.0000780
33. La Manno G et al (2018) RNA velocity of single cells. Nature 560:494–498. 10.1038/s41586-018-0414-6
34. Dunican C (2025) e. a. Predicting trajectories of acute illness using RNA velocity of whole blood. https://doi.org/10.21203/rs.3.rs-5764288/v1
35. Li Q et al (2023) A gene network database for the identification of key genes for diagnosis, prognosis, and treatment in sepsis. Sci Rep 13:21815. 10.1038/s41598-023-49311-x
36. Zhang R et al (2026) Identification and validation of hub genes associated with neutrophil immunosuppression during Sepsis. Cell Immunol 422:105083. 10.1016/j.cellimm.2026.105083
37. Manz MG, Boettcher S (2014) Emergency granulopoiesis. Nat Rev Immunol 14:302–314. 10.1038/nri3660
38. Kwok AJ et al (2023) Neutrophils and emergency granulopoiesis drive immune suppression and an extreme response endotype during sepsis. Nat Immunol 24:767–779. 10.1038/s41590-023-01490-5

39. Boeddha NP et al (2018) Mortality and morbidity in community-acquired sepsis in European pediatric intensive care units: a prospective cohort study from the European Childhood Life-threatening Infectious Disease Study (EUCLIDS). Crit Care 22. 10.1186/s13054-018-2052-7
40. Raman S et al (2015) Oxidative phosphorylation gene expression falls at onset and throughout the development of meningococcal sepsis-induced multi-organ failure in children. Intensive Care Med 41:1489–1490. 10.1007/s00134-015-3817-y
41. Brodin P, Duffy D, Quintana-Murci L (2019) A Call for Blood. Hum Immunol Immunity 50:1335–1336. 10.1016/j.immuni.2019.05.012
42. Li C et al (2024) The effect of whole blood logistics on neutrophil non-specific activation and kinetics ex vivo. Sci Rep 14:2543. 10.1038/s41598-023-50813-x
43. Prescott HC, Osterholzer JJ, Langa KM, Angus DC, Iwashyna TJ (2016) Late mortality after sepsis: propensity matched cohort study. BMJ 353:2975. 10.1136/bmj.i2375
44. Newman AM et al (2019) Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 37:773–782. 10.1038/s41587-019-0114-2
45. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14:417–419. 10.1038/nmeth.4197
46. R: A Language and Environment for Statistical Computing (2023)
47. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. 10.1186/s13059-014-0550-8
48. Yu G, Wang LG, Han Y, He QY (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. Omics 16:284–287. 10.1089/omi.2011.0118
49. Bolstad B, preprocessCore (2025) A collection of pre-processing functions (R package version 1.72.0)
50. Foltz SM, Greene CS, Taroni JN (2023) Cross-platform normalization enables machine learning model training on microarray and RNA-seq data simultaneously. Commun Biology 6:222. 10.1038/s42003-023-04588-6
51. The mt R package: mt: Metabolomics Data Analysis Toolbox (v. 2.0-1.2) (2025)
52. Dobin A et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. 10.1093/bioinformatics/bts635
53. Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930. 10.1093/bioinformatics/btt656

Figures

a

Clinical characteristic	Patients	Endotype 1 at TP1	Endotype 2 at TP1
n	19	9 (47%)	10 (43%)
Sex at birth (%)			
Female	6 (32%)	3 (33%)	3 (30%)
Male	13 (68%)	6 (67%)	7 (70%)
Patient age group (%)			
0 to <1 month	1 (5%)	0	1 (5%)
1 to 11 months	4 (21%)	3 (16%)	1 (5%)
12 to 23 months	2 (11%)	1 (5%)	1 (5%)
24 months to 4 years	6 (32%)	2 (11%)	4 (21%)
5 years to 12 years	4 (21%)	2 (11%)	2 (11%)
>12 years	2 (11%)	1 (5%)	1 (5%)
Clinical syndrome			
Undifferentiated septic shock	7 (37%)	3 (33%)	4 (40%)
Pneumonia	6 (32%)	5 (56%)	2 (20%)
Meningitis/encephalitis	3 (16%)	1 (11%)	2 (20%)
Cellulitis/soft tissue infection	1 (5%)	0	1 (10%)
Pleural effusion/empyema	1 (5%)	0	1 (10%)
Phoenix sepsis score (median, IQR, max)	3 (2–5, 7)	3 (2–5, 7)	3 (1–4, 5)
Treatment			
Duration mechanical ventilation (h; median, IQR, max)	133 (91–224, 330)	122 (88–266, 330)	125 (87–150, 259)
Duration vasoactive infusions (h; median, IQR, max)	45 (16–102, 251)	53 (36–148, 251)	32 (15–56, 216)
Outcomes			
Ventilator-free days at 30 days (median, IQR)	25 (22–26)	25 (19–26)	25 (24–26)
Vasoactive infusion-free days at 30 days (median, IQR)	28 (26–29)	28 (24–29)	29 (28–29)
Mortality (%)	0	0	0

Abbreviations: TP, Timepoint; h, hours; Endotype refers to BASIC endotypes.

b

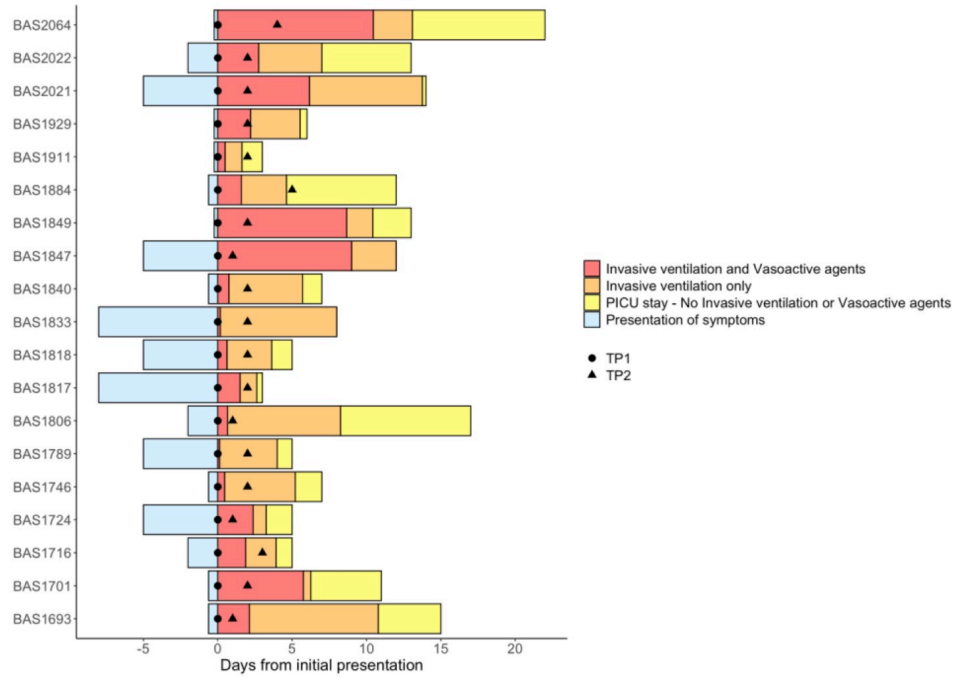


Figure 1

Clinical characteristics and sampling. a, Clinical characteristics of the BASIC study sub-cohort (n=19 patients) and characteristics of these participants by BASIC endotype at Timepoint 1 (retrieval or admission to PICU). **b,** Time course of preceding illness, PICU admission and sampling at Timepoint 1 and Timepoint 2 for study participants.

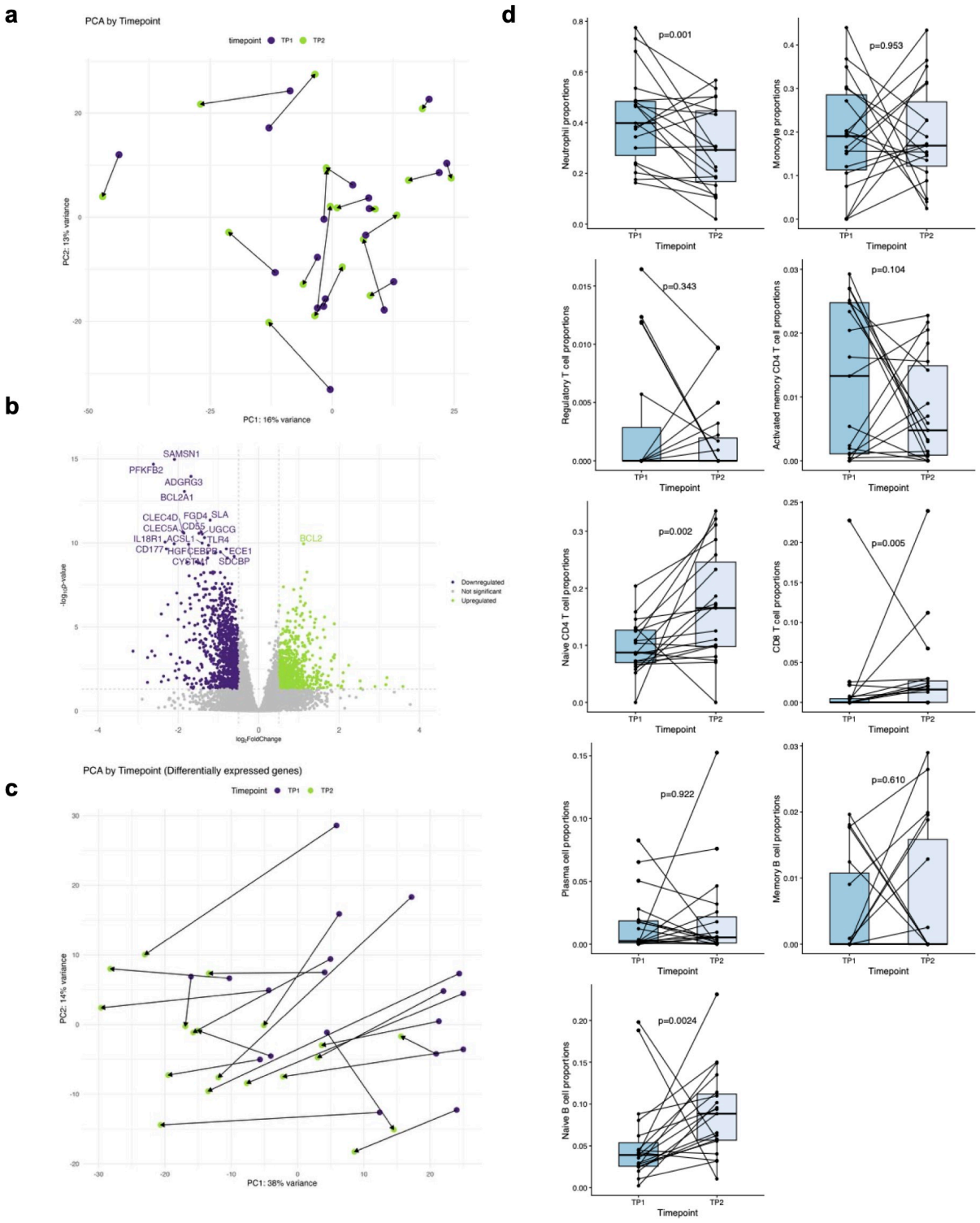


Figure 2

Differential expression of genes by time point to show immune trajectories. a, Principal component analysis (PCA) plot of participants (n=19, n=38 samples), with arrows showing the difference between a patients transcriptome from Timepoint 1 (red) to Timepoint 2 (blue) for all genes. **b,** Volcano plot of gene expression, with differentially expressed genes (Benjamini-Hochberg adjusted p<0.05, and absolute $\log_2$ fold-change ≥1.5, red points (individual genes) indicate when this value is 1.5 or above, blue when

this is -1.5 and below). **c,** PCA plot of participants (including trajectory of patient transcriptomes) using only differentially expressed genes. Red indicates timepoint 1 and blue timepoint 2 samples. **d,** Imputed cell populations as proportions of leukocytes at Timepoint 1 and Timepoint 2 (Cibersort[44]).

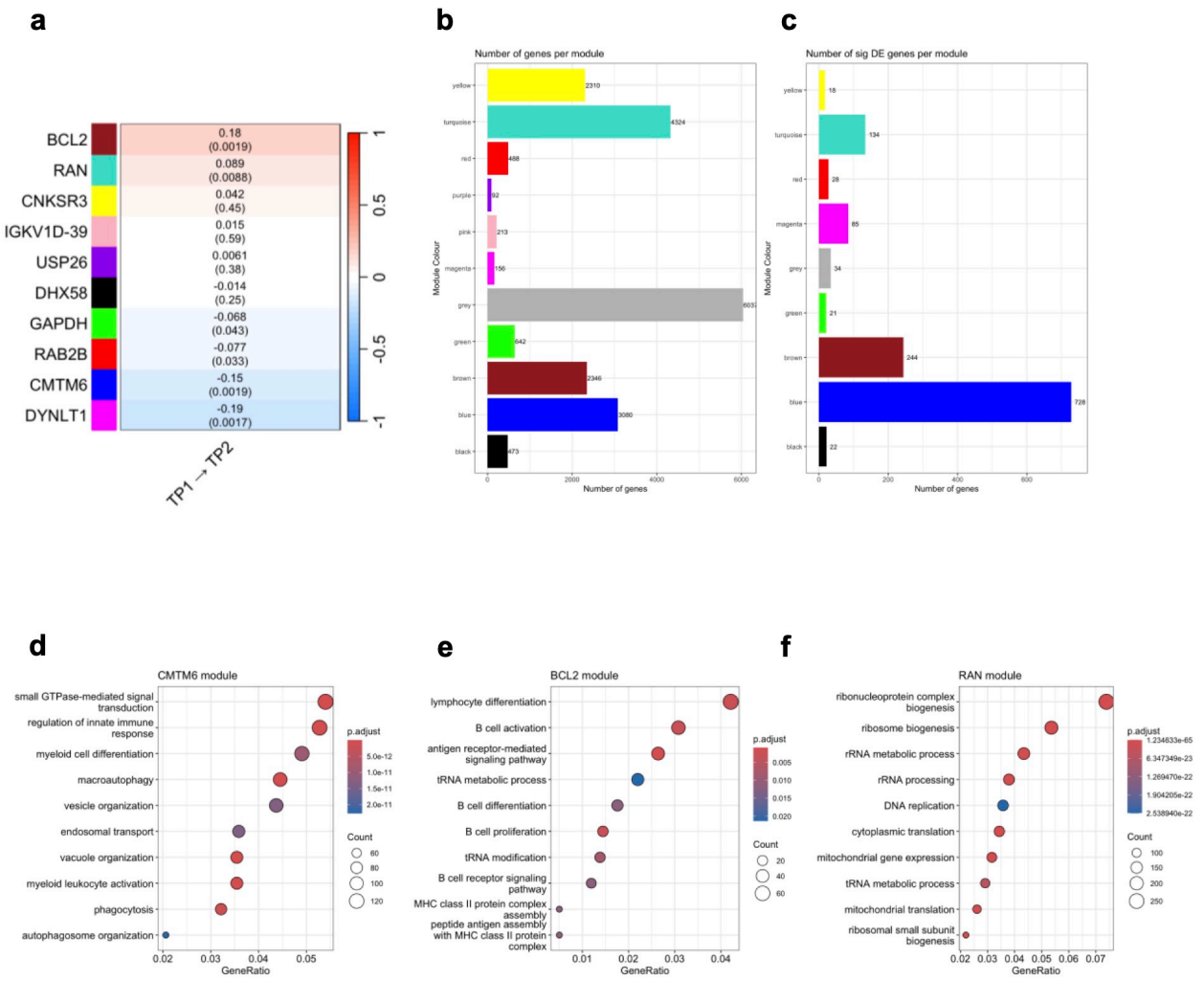


Figure 3

Biological mechanisms associated with immune trajectory. a, Weighted gene co-expression network analysis (WGCNA) to identify key biological mechanisms between Timepoint 1 and Timepoint 2. **b - c,** All genes, and all differentially expressed genes (DEGs) for modules identified in WGCNA. **d,** Gene set expression analysis (GSEA) for the *CMTM6* module. **e,** GSEA for the *BCL2* module. **f,** GSEA for the *RAN* module.

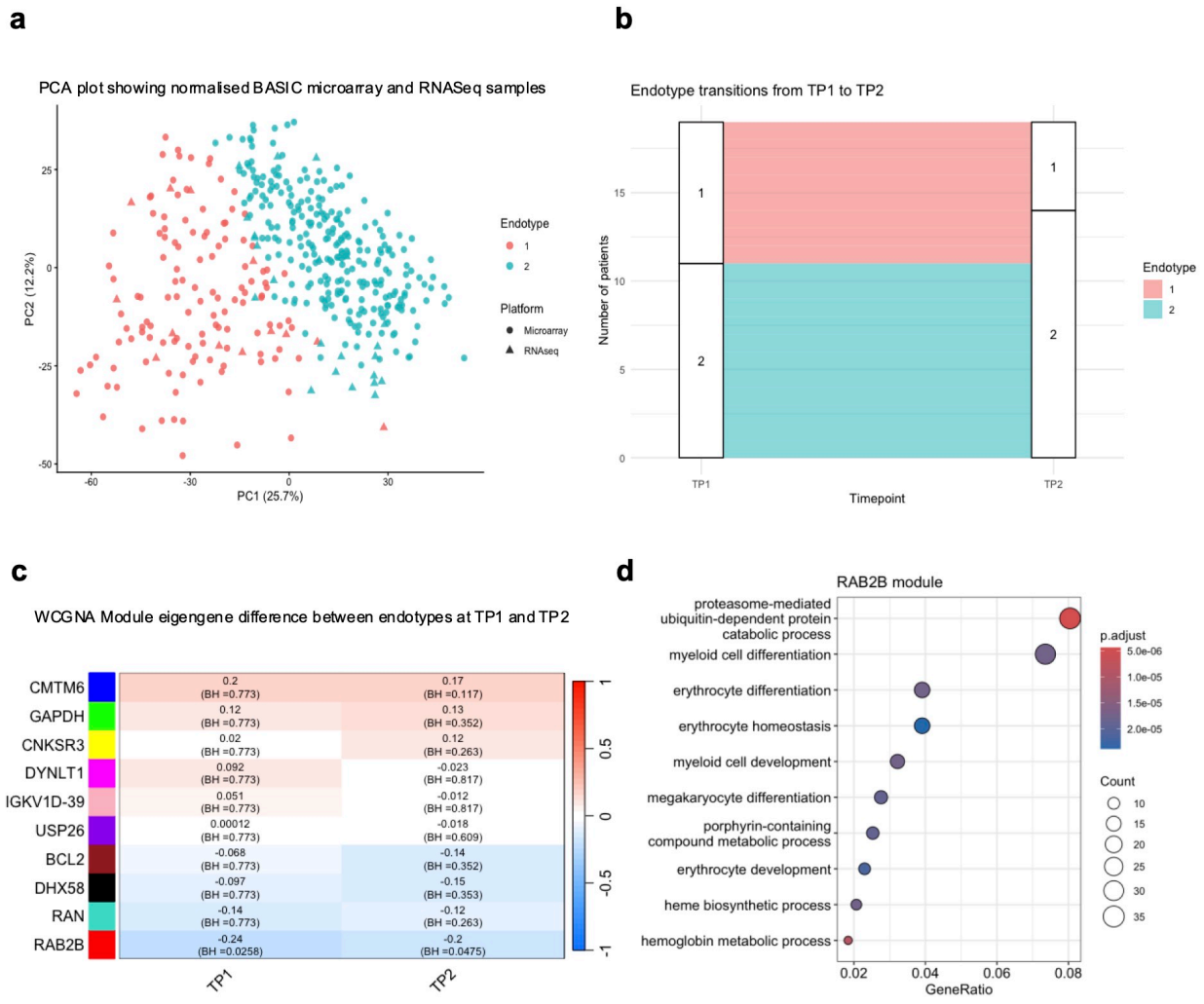


Figure 4

Associations of BASIC endotype with immune trajectory. Across all figure components, red indicates endotype 1 and blue endotype 2. **a,** Allocation of RNA-Seq patient-time points to BASIC endotypes following quantile normalisation and application of 20 gene signature, alongside BASIC microarray allocated endotypes by k-means clustering[14]. **b,** Sankey plot showing changes in BASIC endotype for participants over time, colored by Timepoint (TP) 1 endotype. **c,** Correlation of BASIC endotype membership and weighted gene co-expression network analysis (WGCNA) modules identified in the cohort at Timepoint 1 and Timepoint 2. **d,** Gene set expression analysis (GSEA) for the *RAB2B* module.

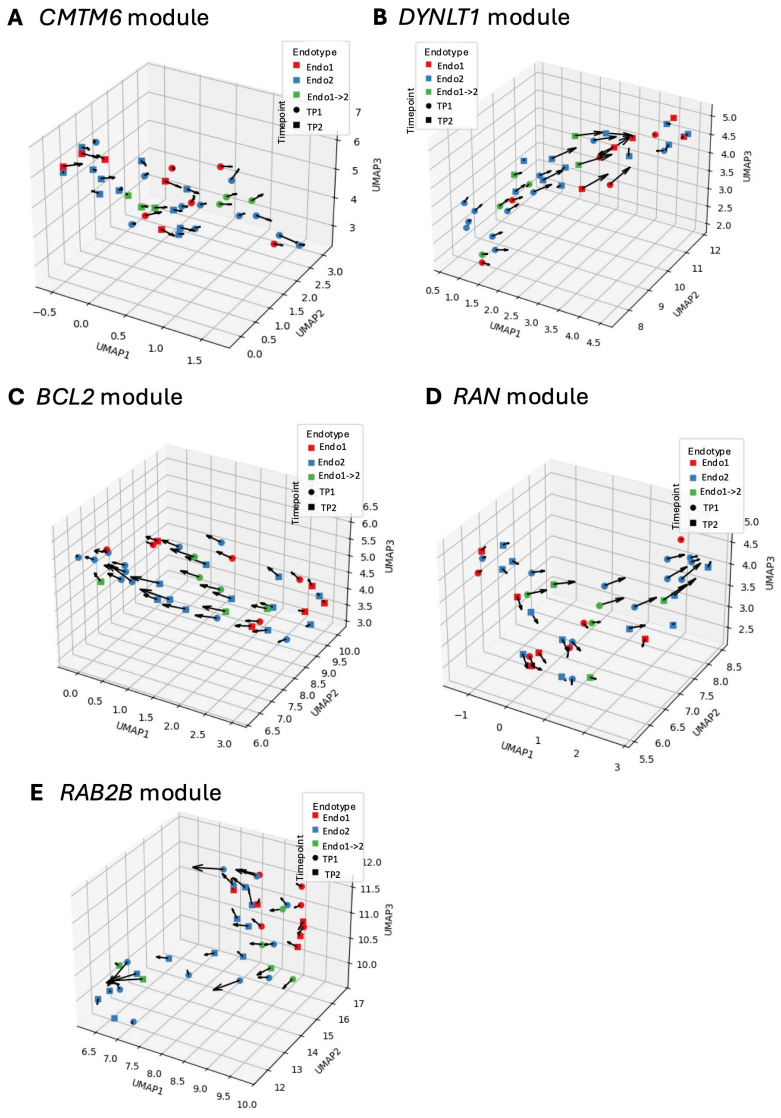


Figure 5

RNA velocity analysis. **a-d**, Uniform Manifold Approximation and Projection (UMAP)-based RNA velocity fate maps of the samples in three dimensions. Arrows represent RNA velocities, the future transcriptomic trajectories of each sample (n=38 samples, n=19 participants). The Transition Probability Number of Neighbors (TPNN) value used for all fate maps was 36. The fate maps show samples points colored by endotype (red: endotype 1, blue: endotype and green: individuals who transitioned from endotype 1 to 2) and shaped by timepoint (circles: TP1, squares: TP2). **a,** fate map constructed from 1579 genes from the Weighted Gene Co-expression Network Analysis (WGCNA) by timepoint, *CMTM6* module and a Number of Neighbors (NN) value of 25. **b,** Was constructed using 86 genes from the *DYNLT1* module that had spliced and unspliced transcript expression (NN: 16). **c,** Was constructed using 1079 genes from the *BCL2* module that had spliced and unspliced transcript expression (NN: 23). **d,** Was constructed using 2,589 genes from the *RAN* module that has spliced and unspliced transcript expression (NN: 11). **e,** Was constructed using 300 genes from the *RAB2B* module that has spliced and unspliced transcript expression (NN: 26).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- BASICtimepointssupplementary23062026.pdf
- REPORTINGFORSPECIFICMATERIALS.docx