**Epigenetic Profiling in Severe Sepsis (EPSIS): a pilot study of DNA methylation profiles in critical illness**

Alexandra Binnie\* 1, Christopher J Walsh\* 2,3\*, Pingzhao Hu 4, Dhruva J. Dwivedi 5, Alison Fox-Robichaud 5,6, Patricia C. Liaw 5,6, Jennifer LY Tsang 6,7, Jane Batt 2,3, Gabriela Carrasqueiro 8,9, Sahil Gupta2,3, John C. Marshall2,3, Pedro Castelo-Branco# 8,9,10, Claudia C. dos Santos# 2,3 *For the EPSIS study (Epigenetic Profiling in Severe Sepsis) of the Canadian Critical Care Translational Biology Group (CCCTBG).*

**Supplementary Material**

**Patient inclusion and exclusion criteria:**

The inclusion criteria for sepsis were a modification of those defined by Bernard et al. (S1). Patients were eligible for inclusion into the septic group of this study if they had a confirmed or suspected infection on the basis of clinical data at the time of screening, at least one dysfunctional organ system, 3 or more signs of systemic inflammatory response syndrome (SIRS), and were expected to remain in the ICU for ≥ 72 hours. The presence of organ dysfunction was defined as: (1) SBP ≤90 mm Hg or MAP ≤70 mm Hg or SBP < 40 mm Hg for at least 1 hour despite fluid resuscitation, adequate intravascular volume status, or use of vasopressor in attempt to maintain systolic BP ≥90 or MAP ≥ 70 mm Hg, (2) P/F Ratio ≤ 250 in the presence of other dysfunctional organs or systems, or ≤ 200 if lung is the only dysfunctional organ, (3) acute rise in creatinine > 171 mM or urine output 1.5 times the upper limit of normal, and (4) platelet count < 50,000 or a 50% drop over the 3 days prior to ICU admission. The inclusion criteria for septic shock were the same as those for sepsis except that the patient had to have been on vasopressors within the previous 24 hours. Patients were excluded if they were < 18 years old, were pregnant or breastfeeding, or were receiving palliative care only. The ‘parent study’ DYNAMICS (https://clinicaltrials.gov/ct2/show/ NCT01355042

(S2)) took place in 6 hospitals (Foothills Hospital, Calgary, AL; St. Paul’s Hospital, Vancouver, BC; Hamilton General, Hamilton and St. Joseph’s Health Center, in Hamilton, ON; Ottawa Hospital, in Ottawa, ON; and Centre Hospitalier de l’Universite de Laval, Quebec, QC). Whole blood samples were collected only at Hamilton Health Sciences Centre and St Joseph's Health Care in Hamilton, ON.

**DNA Collection**: DNA samples on “day 1” of ICU admission were obtained from 486 patients at 2 ICU sites (Hamilton Health Sciences Centre and St. Joseph’s Healthcare Hamilton). Whole blood was collected in DNA PaxGENE tubes or as a buffy coat, as per the Canadian Critical Care Translational Biology Group standard operating procedures (<http://www.ccctbg.ca/sopmenu.html>).

**Sample Preparation**: Genomic DNA was extracted from buffy coat or PaxGene tubes using the QIAquick DNA purification kit (Qiagen, Hilden, Germany). Bisulfite conversion was performed using the EZ-96 DNA Methylation Kit (Zymo Research, Irvine, CA), according to the manufacturer’s instructions.

**Methylation data normalization and processing:** Unsupervised intra-array normalization was performed with a normal-exponential fit using out-of-band probes (Noob) for background correction and dye-bias equalization (S3), followed by beta-mixture quantile normalization; this combination has been shown to consistently increase signal sensitivity (S4). Unsupervised between-array normalization was performed using Functional Normalization (FunNorm), an extension of quantile normalization that uses summarized control probes as surrogates for unwanted variation (S5); the number of principal components was set to 2. Singular value decomposition (SVD) analysis (S6) was performed on technical variables (microarray slide, array position, and control probes) and phenotypic variables (age and sex) to identify confounding factors. Technical confounding related to batches and DNA processing methods was corrected using Combat (S7) and confirmed by SVD analysis. Replicate samples from the same patient were processed in 3 separate batches; correlation between replicates was 0.998 after correction. All analyses were performed using methylation *β*-values, calculated as SignalMeth/SignalMeth + SignalUnmeth, with the exception of Combat where *β*-values were logit transformed into *M*-values. White blood cell type proportions were estimated from methylation data using the reference-based model proposed by Houseman (S8).

**Differential Methylation (DE) Detection:** The 414,818 quality-filtered CpG probes were analysed using the adjacent site clustering (A-clustering) algorithm

(S9) to identify clusters of ≥ 2 CpGs with correlated methylation levels. The association of the methylation clusters and patient status (septic or non-septic) was tested by generalized estimating equations (GEE) to identify differentially methylated regions (DMRs) correlating with sepsis status. Estimated cell proportions derived from the Houseman cell mixture deconvolution are presented in Table 1. The analysis was adjusted for sex, age, DNA processing method, estimated percent neutrophils (neutrophils accounted for 93.4% of the variance in the cell count), and 28-day mortality. A random effects model was used to cluster MODS (high versus low*)* in the GEE model to account forcorrelations between probes based on MODS.

Significant methylation clusters and individual CpG sites within clusters were defined according to corrected p-value using Benjamini and Hochberg false discovery rate (FDR) < 0.05 and difference of methylation *β*-values >= 0.02. A-clustering parameters were used in Spearman correlation for calculating the distance between adjacent CpG probes (*average* clustering type with mean distance between sites at least 0.25, 1000bp distance restriction). Annotation provided by the Illumina Human Methylation 450k set was used to determine the genomic locations of probes. CoMET (S10) was used to assign probes to their gene targets using the human reference genome 19

(S11).

**Cross-referencing with transcriptomic datasets**

Seven datasets containing whole blood gene expression data from septic and non-septic patients were identified in GEO (Table S1)

(S12). Individual samples were designated as septic or non-septic according to documentation available in GEO or information from contributors (communication). In datasets containing multiple time points, only data from the first time point (usually Day 1 of hospital admission) was included. Differentially-expressed genes were identified using the GEO2R functionality (http://www.ncbi.nlm.nih.gov/geo/geo2r/) with correction for multiple comparisons (FDR < 0.05).

**Weighted Gene Co-Expression Network Analysis:** We sought to identify clusters of highly correlated CpGs (co-methylation modules) that also correlated with clinical phenotypes. The 207,408 CpG probes with the highest median absolute deviation (top 50% of the quality filtered probes) across samples (ignoring sepsis status) were selected for further analysis. Linear regression was used to correct CpG probe M-values for sex, age, DNA processing method, and estimated percent neutrophils. The residual values from the linear regression were used for further analysis. Sample data was separated into septic (n=66) and non-septic (n=68) and input separately into WGCNA. Signed co-methylation networks were detected using the blockwiseModules function. We chose signed networks as they have been shown to detect modules with more significant enrichment of functional classifications

(S13). Pairwise Tukey’s biweight correlations for the set of CpGs was calculated using bicor function. Adjacency transformation was calculated by raising the correlation matrix to the power of 14, which was chosen using the scale-free topology criterion (S14). For each pair of CpG probes the topological overlap measure was calculated based on the adjacency matrix and the topological overlap dissimilarity measure was used as input for average linkage hierarchical clustering. Dynamic Hybrid tree cutting algorithm was used to cut branches of the dendrogram because it produces robustly defined modules. To obtain moderately large and distinct modules we set the minimum module size to 50 probes. Probes that did not fulfill these criteria were assigned to the predefined “grey module”. Each module was summarized by its first principle component of the CpG methylation values (module eigengenes [MEs]). We then tested the association of each ME with clinical variables using the polycor package to perform correlations between order clinical variables and continuous ME values.

**Module preservation analysis**: We applied the module preservation statistic Zsummary in the *modulePreservation* function in WGCNA (S15) to determine the degree of conservation of network topology between the gene modules in the Septic and non-septic groups. The Zsummary statistic integrates the overlap in module membership with the connectivity (sum of connections) and density (mean connectivity) patterns of the each module. The recommended significance thresholds, Zsummary < 2 implies no evidence for module preservation, 2 < Zsummary < 10 implies weak to moderate evidence, and Zsummary > 10 implies strong evidence for module preservation.

Table S1 (*related to Figure 4*): Sepsis gene expression datasets from Gene Expression Omnibus

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Abbreviation | GEO name | Contributor(s) | Year | Septic cohort | N | Non-septic cohort | N | Ref |
| Parnell  | GSE54514 | Parnell GP and Tang BM | 2014 | Septic ICU  | 35 | Healthy controls | 18 | (S16) |
| Tang | GSE6535 | Tang BM | 2006 | Septic ICU  | 55 | Non-septic ICU | 17 | (S17) |
| Wong derivation | GSE26440 | Wong HR | 2011 | Pediatric septic ICU  | 98 | Pediatric healthy controls | 32 | (S18) |
| Wong validation | GSE26378 | Wong HR | 2010 | Pediatric septic ICU  | 82 | Pediatric healthy controls | 21 | (S19) |
| Scicluna | GSE65682 | Scicluna BP and van der Poll T | 2015 | Septic ICU (pneumonia) | 192 | Healthy controls | 42 | (S20) |
| Scicluna ICU | GSE65682 | Scicluna BP and van der Poll T | 2015 | Septic ICU (pneumonia) | 108 | Non-septic ICU  | 33 | (S20) |
| McHugh | GSE74224 | Leo M and Thomas Y | 2015 | Septic ICU  | 74 | Non-septic ICU  | 31 | (S21) |
| Dolinay | GSE32707 | Dolinay T et al. | 2011 | Septic ICU  | 42 | Non-septic ICU  | 21 | (S22) |

Table S1: Characteristics of seven datasets downloaded from the Gene Expression Omnibus comprising gene expression data from septic and non-septic patients. Patients were allocated to septic or non-septic cohorts based on accompanying documentation and information from authors (*private communication, B.Scicluna*).

Supplementary references:

S1. Bernard GR, Vincent JL, Laterre PF, et al.: Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med* 2001; 344:699–709

S2. Jackson Chornenki NL, Coke R, Kwong AC, et al.: Comparison of the source and prognostic utility of cfDNA in trauma and sepsis. *Intensive Care Med Exp* 2019; 7:29

S3. Triche TJ Jr, Weisenberger DJ, Van Den Berg D, et al.: Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Research* 2013; 41:e90–e90

S4. Liu J, Siegmund KD: An evaluation of processing methods for HumanMethylation450 BeadChip data. *BMC Genomics* 2016; 1–11

S5. Fortin J-P, Labbe A, Lemire M, et al.: Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biology* 2014; 15:503

S6. Teschendorff AE, Menon U, Gentry-Maharaj A, et al.: An Epigenetic Signature in Peripheral Blood Predicts Active Ovarian Cancer. *PLOS OnE* 2009; 4:e8274

S7. Johnson WE, Li C, Rabinovic A: Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; 8:118–127

S8. Houseman EA, Accomando WP, Koestler DC, et al.: DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012; 13:86

S9. Sofer T, Schifano ED, Hoppin JA, et al.: A-clustering: a novel method for the detection of co-regulated methylation regions, and regions associated with exposure. *Bioinformatics* 2013; 29:2884–2891

S10. Martin TC, Yet I, Tsai P-C, et al.: coMET: visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns. *BMC Bioinformatics* 2015; 16:721

S11. Lander ES, Linton LM, Birren B, et al.: Initial sequencing and analysis of the human genome. *Nature* 2001; 409:860–921

S12. Edgar R, Domrachev M, Lash AE: Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research* 2002; 30:207–210

S13. Mason MJ, Fan G, Plath K, et al.: Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics* 2009; 10:327

S14. Langfelder P, Horvath S: WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; 9:559

S15. Langfelder P, Luo R, Oldham MC, et al.: Is my network module preserved and reproducible? *PLoS Computational Biology* 2011; 7:e1001057

S16. Parnell GP, Tang BM, Nalos M, et al.: Identifying Key Regulatory Genes in the Whole Blood of Septic Patients to Monitor Underlying Immune Dysfunctions. *Shock* 2013; 40:166–174

S17. Tang BMP, McLean AS, Dawes IW, et al.: Gene-expression profiling of Gram-positive and Gram-negative sepsis in critically ill patients\*. *Critical Care Medicine* 2008; 36:1125–1128

S18. Wong HR, Cvijanovich N, Allen GL, et al.: Genomic expression profiling across the pediatric systemic inflammatory response syndrome, sepsis, and septic shock spectrum\*. *Critical Care Medicine* 2009; 37:1558–1566

S19. Wynn JL, Cvijanovich NZ, Allen GL, et al.: The influence of developmental age on the early transcriptomic response of children with septic shock. *Mol Med* 2011; 17:1146–1156

S20. Scicluna BP, Klein Klouwenberg PMC, van Vught LA, et al.: A Molecular Biomarker to Diagnose Community-acquired Pneumonia on Intensive Care Unit Admission. *Am J Respir Crit Care Med* 2015; 192:826–835

S21. McHugh L, Seldon TA, Brandon RA, et al.: A Molecular Host Response Assay to Discriminate Between Sepsis and Infection-Negative Systemic Inflammation in Critically Ill Patients: Discovery and Validation in Independent Cohorts. *PLoS Med* 2015; 12:e1001916

S22. Dolinay T, Kim YS, Howrylak J, et al.: Inflammasome-regulated Cytokines Are Critical Mediators of Acute Lung Injury. *Am J Respir Crit Care Med* 2012; 185:1225–1234