Sepsis Transcriptomic Trajectory and Phenotypic Correlation Using Time Series Gene Expression from Childhood Meningococcal Infection

Asrar Rashid1,2, M. Toufiq1, Praveen Khilnani1, Zainab A. Malik4,6, Z. Hussain1, Hoda Alkhazaimi8, Javed Sharief2, Raziya Kadwa2, Berit S. Brusletto9, Amrita Sarpal3,10, Damien Chaussabel11, Rayaz Malik4,6, Nasir Quraishi15, Govind Benakatti2, Syed A. Zaki14, Rashid Nadeem16, Guftar Shaikh16, Ahmed Al-Dubai1, Amir Hussain1

1. School of Computing, Edinburgh Napier University. Edinburgh, UK.
2. NMC Royal Hospital. Abu Dhabi, UAE.
3. Sidra Medicine, Doha, Qatar.
4. Rainbow Children’s Hospital, Delhi, India.
5. College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences. Dubai, U.A.E.
6. Mediclinic City Hospital. Dubai, UAE.
7. Edinburgh Medical School, University of Edinburgh, UK
8. New York University Abu Dhabi, UAE
10. Weill Cornell Medicine-Qatar. Doha, Qatar.
11. The Jackson Laboratory, USA
12. Institute of Cardiovascular Science, University of Manchester. Manchester, UK.
13. Centre for Spinal Studies & Surgery, Queen’s Medical Centre. University of Nottingham. Nottingham, UK.
14. All India Institute of Medical Sciences. Hyderabad, India.
15. Dubai Hospital, Dubai, UAE.
17. Department of Cardiovascular Sciences, University Hospitals of Leicester NHS Trust. Leicester, UK.
18. NMC Provita, UAE.

*Corresponding author Dr Asrar Rashid asrar@mac.com

There are no financial relationships to disclose on behalf of the authors

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
ABSTRACT

Sepsis remains a leading cause of in-hospital morbidity and mortality. In the acute phase sepsis is described as a dysregulated process. The objective was to track temporal transcriptomic changes in infants with meningococcal disease (MSS). Therefore, two datasets underwent secondary analysis using temporal transcriptomic data. Methods applied to analysis included Principal Component Analysis (PCA), Transcript Time Course Analysis (TTCA) and Gene Set Expression Analysis (GSEA). Gene expression clustering algorithm for both datasets suggested three phases from the time of admission, and PCA plots indicated a gene-expression trajectory for both datasets. The data set from Kwan et al., showed that 410 genes differentiated survivors from the non-survivor, which included various cytokine, TNF, and apoptosis-associated pathways (normalized expression scores = -1.60, p = 0.02, and q = 0.15). Additionally, GSEA demonstrated gene sets significantly associated with disease severity. The genes for the cytokines CLC, HFE, HLA-F, NLRP3, and TNFRSF1B were significant. The Gene-sets elicited by GSEA in the Kwan and Emonts dataset had a high degree of molecular signature crossover (89.2% overlap), although a comparison of gene-expression trajectories across both datasets was difficult due to differing sample timing. Transcriptomic analysis demonstrated temporal trajectories in MSS-related gene expression. Larger studies using transcriptome expression are required could enhance the understanding of sepsis pathogenesis, improving prognostication and enable the development of precision therapies.
Introduction

Pediatric and neonatal sepsis is an underestimated major global health problem. In 2017 sepsis in children constituted half the global sepsis burden, with 20 million episodes of pediatric sepsis and 2.9 million deaths in children younger than five years {Rudd, 2020 #3205}{WHO, 2017 #3396}. Children with sepsis have a mortality rate approaching 50% {Fleischmann-Struzek, 2018 #3397} and survivors have significant morbidity. Among children with severe sepsis that are referred to the Pediatric Intensive Care Unit (PICU), the majority of deaths occur within the first 24hrs from multi-organ failure or refractory shock {Cvetkovic, 2015 #3400}.

Early identification and timely intervention are key to improving clinical outcomes in children with suspected sepsis. Evidence-based guidelines have been developed to improve early recognition of sepsis using clinical and laboratory data {Weiss, 2020 #3267}{Casserly, 2015 #3407} {Goldstein, 2005 #1803}. Although adherence to sepsis guidelines and initiatives has improved, it has failed to translate into an improvement in mortality {Paul, 2018 #3409}{Larsen, 2021 #3410}. Early endotype prediction, such as using a biomarker risk model to predict children developing sepsis-associated myocardial dysfunction, could be useful for the application of precision strategies{Lautz, 2022 #4019}. Also, the clinical utility of measuring sepsis metabolites and biomarkers showed an improvement in the translation of cellular pathophysiology to clinical outcomes {Wang, 2020 #3411}{Kim, 2020 #3412}. However, greater clarity is required to align the clinical features of the sepsis syndrome to altered immune function and physiological response, enabling more precise tailoring of treatment modalities. An enhanced understanding of sepsis dynamics is key to improving the prognostic capacity of biomarkers. Recently, a
microarray-based approach utilizing messenger RNA (mRNA) gene expression has been used in sepsis prognostication and to develop precision therapies (Stanski, 2020 #2526).

We hypothesized that microarray analysis of mRNA gene expression could have utility in better understanding the pathogenesis of Neisseria meningitides septic shock (MSS). We have analyzed differential mRNA gene expression in children with severe sepsis to establish a transcriptomic trajectory in patients with MSS to identify key time points for intervention to improve clinical outcomes.
METHOD

Public transcriptome dataset selection

Public, online repositories were searched to find clinical studies involving temporal transcriptomic datasets involving Neisseria Meningococcal infection. The search term ‘Meningococcal Shock’ was parsed through the NCBI Gene Expression Omnibus (GEO) database (n= 64 datasets isolated) and Array Express EMBL-EBI (n=5 datasets isolated). Then the term ‘meningococcal sepsis’ was used in both Array Express (n=7 datasets isolated) and the NCBI GEO database (n=64 isolated). The term ‘meningococcal transcriptome’ was entered into Pubmed (n=8 datasets).

Two publicly available datasets were finalized. Dataset descriptive information are shown with the two patient groups compared (Table 1). The Kwan dataset {Kwan, 2013 #49} provides the most frequent blood sampling in the first 48 hours of meningococcal sepsis, followed by the Emonts temporal data-set, which covers a longer period (72 hours) {Emonts Marieke , 2008 #2615}. In the Kwan dataset patient, 1 to 5 are denoted by P1 to P5, and in the Emonts dataset patient, 1 to 6 are represented by P1 to P6. The time points used in the study are given an arbitrary value according to research design, starting from time zero (T=0).

Analytical Strategy

The datasets were first inspected for data processing methods indicated by the author(s), and appropriate normalization and log2 transformation was applied wherever needed using R-script for further downstream analysis.

Statistical and Gene Ontology analyses

Qlucore Omics Explorer (QOE) version 3.1 software (Qlucore AB, Lund, Sweden)
was used to analyze the differential expression of the genes (DEGs). Box plots were used to illustrate differences in gene expression in QOE. GSEA uses a computational method to determine whether an apriori-defined set of genes show statistically significant differences between two biological states {Subramanian, 2005 #2568}. For GSEA, the mean expression value of multiple probe IDs that matched an official gene symbol was calculated. This value was considered to represent the expression intensity of the corresponding gene symbol. Hierarchical clustering was based on Euclidean distance and average linkage clustering. Further, all gene expression data was centered to a mean equal to zero and scaled to a variance similar to one. A p-value of <0.05 was used to confirm statistical significance; a false discovery rate (FDR) or q value of < 0.25 was chosen for cut-off criteria. For GSEA, Gene ranks were downloaded from the broad institute database Molecular Signatures Database (MSigDB). In a ‘cross over’ GSEA method, we applied GSEA to the Kwan dataset to generate gene lists downloaded and checked for their presence in the Emonts dataset. ANOVA and t-testing was undertaken according to the design of the study sets. Further, GSEA-associated gene sets associated with cytokine production were used to analyze cell-signaling patterns in survivors versus the non-survivor.
Transcript Time Course Analysis

Transcript Time Course Analysis (TTCA) R software allowed temporal data exploration {Albrecht, 2017 #2534}. In the Kwan dataset, the last sample (48-hour collection) was simulated as a constant, or control, using the last sample for each dataset. Emonts had age-gender matched controls. The KEGG database was used to obtain curated genes for Vascular Endothelial Growth Factor (VEGF) associated signaling, apoptosis, complement, and coagulation. The Hugo database was used to curate the Heat Shock protein genes, including HSBAP1, Zinc finger, and Metallothionein genes. Simple Over-representation analyses (via hypergeometric-distribution-based testing) was performed using the TTCA results. This provided an analysis of significant genes related to ‘Consensus,’ ‘Early Response,’ ‘Middle Response,’ ‘Late Response,’ ‘Complete Response,’ ‘Dynamic,’ and ‘MaxDist.’ Similarly, pathfindR software {Ulgen, 2019 #3312} was used to analyze the genes from the TTCA. KEGG pathway and Gene Ontology Biological Process (GO-BP) enrichment analysis was undertaken using enrichR {Kuleshov, 2016 #3313}. 
Result

The Kwan dataset allowed a comparison of gene expression according to survival. The statistical filtering led to the isolation of 410 genes differentiating the survivors from the non-surviving patient. ICAM3 expression differed significantly according to survival ($p=1.895\times10^{-5}$). CRP did not show a similar differentiation. When GSEA was applied to differentiate according to disease severity, a gene signature for GOBP was identified for altered cytokine production. Of this 88 gene signature set, five genes (CLC, HFE, HLA-F, NLRP3, TNFRSF1B) were present in the Kwan dataset.

Emonts Data Analysis across different cell sub-types

The Emonts dataset includes data from peripheral mononuclear blood cells (PMBCs), lymphocytes, and monocytes and included control samples (Figure 1C).

Time Series Analysis

ANOVA PCA analysis of gene expression datasets from infants with MSS in the Kwan and Emonts studies identified a trajectorial pattern (Figures 1A and 1D). This was supported by heat maps (Figures 1B and 1E) demonstrating a starting state, Kwan data set according to survival

PCA was used to compare gene expression in relation to survival (Figure 2A), and the fatal case (P1) showed differing gene expression compared to the survivors (P2-P5). Gene sets related to various cytokines, TNF, and apoptosis were used for GSEA (Figures 2 E & F).

A. The cytokine gene set differentiating P1 from P2-P5 included the ‘ALTEMEIER_RESPONSE_TO_LPS_WITH_MECHANICAL_VENTILATION’
pathway (NES=-1.68, p=0.004 and q = 0.14) and
‘GO_BP_CYTOKINE_PRODUCTION_INVOLVED_IN_IMMUNE_RESPONSE’
pathway (NEW=-1.53, p=0.022, q= 0.24)

Negative enrichment here implies genes from the Kwan et al. dataset were
downregulated against the MSigDB geneset.

B. Apoptosis.gmt for P1 versus P2-P5 gave a gene set related to WP_APOPTOSIS
(NES=-1.61, p=0.02 and q=0.24). P2 versus P1,P3-P5 showed
WU_APOPTOSIS_BY_CDKN1A_VIA_TP53 (NES=-1.89, p=0.02 and q= 0.06).

C. TNF related gene sets BIOCARTA_NFKB_PATHWAY (NES =-
1.60,p=0.02,q=0.15) & ST_TUMOR_NECROSIS_FACTOR_PATHWAY (NES=-1.49,
p=0.006 and q=0.08)

The gene set the GSEA (see above) for
GO_REGULATION_OF_CYTOKINE_PRODUCTION_INVOLVED_IN_IMMUNE_RE
SPONSE gene-list was overlapped into the Kwan Data (Figure 1E).

Comparing Kwan and Emonts datasets using GSEA

GSEA was also used to compare the Emonts and Kwan datasets (Table 2). First,
GSEA was applied to the Kwan dataset eliciting significant gene-lists using ANOVA,
‘Time (hours),’ using Hallmark genes (H.aal.v7.2 symbols GMT). Gene overlap was
assessed according to gene symbol by performing a cross-over comparison of
significant gene lists and showed a high percentage overlap between the two
independent datasets. Herein gene lists were highly conserved across the two
datasets. GSEA was also applied in the Kwan dataset to compare the deceased
patient with survivors (Table 3).
**TTCA PathfindR-results**

Data from the Kwan study was processed by TTCA and showed a decrease in pro-inflammatory IL1A and C3 compared to controls (Figures 3 A&G). TTCA also showed a slight increase then decrease in NFKB1 gene-expression, with an increase in TNF compared to controls (Figures 3 C&E). In the Emonts study TNF, IL1A and NFKB1 were increased compared to controls. Further, the Emonts dataset showed an initial fall of C3 gene expression between 0 to 24 hours, followed by a gradual rise (Figure 3H). Similarly, for the Emonts study, there was a fall in NFKB1 gene expression over the initial 8 hours, followed by a sustained rise.

Comparison across the datasets was difficult due to the differing time periods of both studies. IL1A gene expression decreased from 0 to 8 hours and then showed a positive linear gradient from 12 to 24 hours in the Kwan dataset, whereas a consistent increase occurred from 0 to 72 hours in the Emonts study. There was a similar rise and fall of NFKB1 in both datasets. However, the intensity scales (y-axis) differ. The shape of the TNF intensity and fall of C3 gene expression in one dataset by TTCA was similar to the other datasets. The Kwan and Emonts datasets have a similar shape for C3 gene expression from 0-24 hours, with a continuing increase in the Emonts study from 24 to 72 hours. A zig-zag change in intensity occurred for PLAT gene expression with a fall between 12 to 24 hours for Kwan and 24 to 72 hours for Emonts datasets.

Curated genes from TTCA analysis were enriched using PathfindR with enrichment charts eliciting significant pathways (Figure 4). Further, for both datasets, enrichment plots showed significance for the hematopoietic cell lineage (3 J&K), and GO_BP
enrichment plots showed significance for the neutrophil degradation pathway (Figure 3 L&M). KEGG gene enrichment showed the presence of pathways related to Coronavirus infection, Ribosome, and Th17 cell differentiation in both the Kwan and Emonts datasets. Table 4 shows TTCA gene enrichment comparison of Kwan versus Emonts datasets.
Discussion

Severe sepsis has been regarded as a dysregulated immune process. Thereby the primary aim of this study was to enhance our understanding of the immune repertoire during severe sepsis by utilizing mRNA-based analysis. Two independent clinical temporal datasets in infants with acute invasive meningococcal disease were explored regarding gene expression. Data was filtered to differentiate genes and discern their relationship to time, allowing a dynamic representation of altered gene expression. PCA and heat map plot analysis depicted three phases in the trajectorial patterns. Also, PCA of the Kwan data separated gene expression according to survival. Therefore a gene-expression trajectory suggests a degree of temporal immune organization in severe sepsis, at least at the level of transcript signaling.

GSEA aids the interpretation of differences in molecular signatures from gene expression datasets. Thus using GSEA, Kwan et al. showed a temporal fall in oxidative phosphorylation gene expression {Raman, 2015 #42}. Further GSEA-based modular cluster analysis has shown value in identifying early predictive biomarkers of sepsis {Li, 2021 #3099}{Zeng, 2021 #3503}, relating changes to outcomes by Kaplan-Meier analysis {Groeneveld, 2019 #3501}. A novel application of GSEA in the current study was to identify differences between survivors and non-survivors with MSS. Furthermore, on labeling the Kwan data according to gene symbol, Hallmark gene sets showed molecular signatures of significance. Moreover we showed that these gene sets were also identifiable in the Emonts dataset. Suggesting a high degree of conservation for molecular signatures across MSS datasets. Such GSEA methodology may enhance the understanding of dynamic alterations in molecular signatures and could aid in prognostication.
The Kwan et al. dataset contained a list of genes potentially useful to differentiate survivors from non-survivors, including ICAM-3, NLRP3 and TNFRSF1B. Differential host gene expression analysis showed ICAM3 gene up-regulation after 6 hours. Of relevance, ICAM3 is expressed in a human meningeal-based model incubated with meningococci, indicating the importance of ICAM-3 in invasive meningococcal disease (Robinson, 2004 #3492). ICAM-3 may have a role in early antigen-independent T-cell interactions with antigen-presenting cells, which may be necessary for pathogen surveillance (Montoya, 2002 #3493). Its role in severe sepsis and up-regulation associated with mortality has not been previously documented. Although ICAM3 has been shown to bind to the dengue virus envelope protein and given the endothelialopathy in Dengue, the predilection for endothelial damage in MSS could be mediated via ICAM3 (Kerrigan, 2019 #3494). Garnacho-Montero et al. showed elevated NLRP3 gene expression in critically ill patients, particularly those with sepsis, and this was sustained today 7 in sepsis survivors (Garnacho-Montero, 2020 #3496). Our study indicates an initial rise in NLRP3 in the fatal case compared to the survivors in the first 48 hours. NLRP3 is an inflammasome that activates macrophages, epithelial cells, and endothelial cells to produce pro-inflammatory cytokines IL-18 and IL-1β, coined the ‘cytokine storm’ in response to the host-pathogen (Fricke-Galindo, 2021 #3499) and was first popularized with H5N1 influenza (Avian) virus infection (Tisoncik, 2012 #3498). The NLRP3 inflammasome pathway also has significance for SARS-CoV2 disease and is responsible for the leukopenia, neutrophilia, and inflammation associated with COVID-19 (Fricke-Galindo, 2021 #3499)(Farahani, 2021 #3500). TNFRSF1B gene expression was increased in the non-survivor. This gene is a part of the gene-expression profile in a murine model of sepsis-associated pulmonary dysfunction (Wang, 2018 #3082). TNFRSF1B has also
been expressed in vivo and in vitro following influenza virus infection with a pulmonary cytokine storm {Wang, 2018 #3082}. Also, both datasets confirmed proof of common mechanistic pathways with coronavirus through TTCA analysis. TTCA allowed a temporal illustration of gene expression and was combined with Pearson correlation and pathR enrichment analysis. While there were potential similarities in trajectories on TTCA, this was difficult to quantify due to the differing design of the two independent studies. Further, identifying KEGG gene enrichment pathways relating to coronavirus infection suggests an overlap in disease pathogenesis between MSS and Severe Acute Respiratory Syndrome (SARS).

This analysis has limitations such as the fact that the Kwan cohort lacked controls which would have been useful in mitigating confounding factors. A work around in TTCA analysis was to label the last Kwan sample in the time series as the control, assuming that the child was then the most stable. In contrast, the Emonts dataset had age and gender-matched controls, which possibly resulted in the different significant gene sets generated by TTCA compared to the Kwan data. Also, the small sample numbers in both studies is a limitation as well as the differences in the duration of study between the two datasets. The Kwan dataset was derived over 48 hours compared to 72 hours in the Emonts study. However, we believe that in the scientific literature no other clinical sepsis dataset currently surpasses the Kwan or Emonts datasets in terms of the number of transcript samples. We suggest that future studies should incorporate sequential sampling exploiting the temporal connectivity of data. However the optimal sample timing to define the transition of immunological function in acute sepsis remains unresolved. Kamaleswaran et al. described an Artificial Intelligence (AI) physiomarker multi-time point strategy for outcome prediction in sepsis {Kamaleswaran, 2018 #2671}. Such an AI-supervised approach combined with
the temporal gene expression sampling could result in novel perspectives with respect to transcriptomic pathophysiology in acute severe sepsis.
Conclusion

Transcriptomic analysis outlined in this paper has demonstrated three-phase temporal trajectories in MSS-related gene expression, with specific patterns related to survival. Future studies are warranted to validate the findings in this paper. Moreover, understanding how mRNA changes associated with MSS overlap with the cytokine storm described in SARS-CoV-2, requires further research.
Figure 1. Gene Expression Trajectory from 2 data series of infants with Meningococcal Sepsis

**Figure 1A**: Kwan dataset Multi-group analysis (p≤0.01, 728 genes) showing the expression levels of 728 genes define differences between time points. The PCA plot shows the time as being the single most important component in differences between samples (principal component, 52%).

**Figure 1B**: Heat map with red indicating up-regulation of gene expression, green indicating low expression. At...
time 0, the majority of these genes exhibit low expression, by 48h, the majority are increased (see yellow boxes). The heap map shows that gene expression transitions from an initial state to a final state. The vertical and horizontal hierarchical clustering heat-map indicated that time points 0 and 48 hours were the most different in terms of gene expression and demonstrated the greatest transition from one pattern to another. Emonts Data PCA is shown according to different cell types (Blood, Lymphocyte and Monocytes), ANOVA undertaken (p=0.01 and q=0.012, giving 4543 genes of significance (Figure 1C: ANOVA according to cell type p=0.01 and q=0.012). Controls are included in this dataset (Figure 1D). PCA is from a Multi-group ANOVA analysis with respect to Time (eliminating by ANOVA the 'Cell type'), (p=0.002, q=0.14, axis 41%, 13%, and 8%), 776 genes were differentially up-regulated across time points (Figure 1E) with the associated heatmap (Figure 1F) which again shows the transition from an initial state to a final state.
Figure 2. Kwan, non-survivor versus survivors, differential gene expression, heat plot, and GSEA

Figure 2: Figure 2A Kwan survival analysis. 2 group (t-test) by ‘Patient ID,’ P1 (non-survivor) versus the surviving patients (P2-P5) elicited 410 genes that differentiated the two group PCA (p=1.0e-04, q=0.008), in the non-collapsed gene symbol mode (Figure 2A). The differential gene expression between the two groups is shown in the form of the heat-map, the vertical axis
representing the 410 genes (Figure 2B). ICAM3 was included in the list of 410 genes and shown here with a heat map, showing up-regulation in the surviving patients (Figure 2C). GSEA was undertaken, t-Test ‘Patient ID,’ P1 (non-survivor) versus the surviving patients (P2-P5). Figure 2E: GSEA was undertaken using cytokine.gmt, GSEA significant P1 versus the other patients for GO_BP_CYTOKINE_PRODUCTION_INVOLVED_IN_IMMUNE_RESPONSE and WP_APOPTOSIS Figure 2F. The Gene-set GO_BP_CYTOKINE_PRODUCTION_INVOLVED_IN_IMMUNE_RESPONSE contained 83 genes and was checked in the original 410 gene list. Five genes (genes CLC, HFE, HLA-F, NLRP3, TNFRSF1B) overlapped both this molecular gene set and the original 410 genes (Figure 2G), differentiating the non-survivor (P1) against the surviving patients as one gene probe occurred twice in the gene list, the gene-symbols were collapsed according to averaging. Gene expression for C-reactive protein was not seen to differentiate from a survival point-of-view (Figure 2H) (Pro-Calcitonin gene was not available on this microarray panel).
Figure 3. TTCA Analysis for the Kwan and Emonts Datasets
Figure 3. TTCA analysis for the Kwan dataset (Figure 3 A, C, E, G, and I) and the Emonts dataset (Figure 3 B, D, F, and H). Using data from TTCA, data bar plots were generated for Kwan (Figure 3 K and M) and Emonts (Figure 3 L and N). Using TTCA, IL1A shows a fall for Kwan (Figure 3A) but a rise in Emonts (Figure 3B). NFKB1 a fall and climb over the 24 hours (Figure 3C), whereas a rise is noted in Kwan (Figure 3D). TNF a rise is noted in both datasets (Figure 3E and F), C3 a fall in both datasets (Figure 3G and H), and PLAT a rise and fall noted in both (Figure 3I and J). KEGG enrichment shows the coronavirus pathway to be enriched for both datasets (K and L), antigen presentation, and hematopoietic cell lineage.
Figure 4. From the TTCA data, PathindR enrichment charts were generated for both the Kwan (Figure 4A and B) and Emonts datasets (Figure 4C and D). For both datasets, Ribosome gene enrichment was significant for complete and dynamic responses. Also, T cell receptor signaling pathways were noted to be significant in the complete response for both datasets. Th17 cell differentiation was noted to be significant in Kwan (complete response) and Emonts (complete and dynamic response).
## TABLE 1: Demographic Comparison for Kwan versus Emonts

<table>
<thead>
<tr>
<th>Patient Identifier</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>&quot;p-Value Comparison&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neisseria Meningococcal Serotype</td>
<td>Negative culture (presumes meningococcal sepsis)</td>
<td>GpB</td>
<td>GpB</td>
<td>GpB</td>
<td>GpB</td>
<td>GpB</td>
<td>GpB</td>
<td>PCR negative</td>
<td>GpB</td>
<td>GpB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>1N</td>
<td>2N</td>
<td>3N</td>
<td>4N</td>
<td>5N</td>
<td>1R</td>
<td>2R</td>
<td>3R</td>
<td>4R</td>
<td>5R</td>
<td>6R</td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Mechanical Ventilation</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Female</td>
<td>Female</td>
<td>Female</td>
<td>Male</td>
<td>Male</td>
<td>Male</td>
<td>Male</td>
<td>Male</td>
<td>Male</td>
<td>Male</td>
<td>Male</td>
<td></td>
</tr>
<tr>
<td>Protein C</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Hospital</td>
<td>Kwan</td>
<td>Kwan</td>
<td>Kwan</td>
<td>Kwan</td>
<td>Emonts</td>
<td>Emonts</td>
<td>Emonts</td>
<td>Emonts</td>
<td>Emonts</td>
<td>Emonts</td>
<td>Emonts</td>
<td></td>
</tr>
<tr>
<td>Number of samples taken</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>1.08</td>
<td>0.83</td>
<td>1.83</td>
<td>2</td>
<td>0.75</td>
<td>5.06</td>
<td>1.37</td>
<td>1.79</td>
<td>2.09</td>
<td>1.51</td>
<td>8.04</td>
<td>0.13</td>
</tr>
<tr>
<td>Duration of PICU admission (DAYS)</td>
<td>9</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>60</td>
<td>4</td>
<td>0.46</td>
</tr>
<tr>
<td>Median PRISM Score</td>
<td>15</td>
<td>7</td>
<td>15</td>
<td>13</td>
<td>4</td>
<td>26</td>
<td>23</td>
<td>21</td>
<td>25</td>
<td>28</td>
<td>14</td>
<td>0.0035</td>
</tr>
<tr>
<td>Leucocyte count (*10⁻⁹/L)</td>
<td>6.8</td>
<td>18.4</td>
<td>18.8</td>
<td>3.7</td>
<td>15.1</td>
<td>16.1</td>
<td>5.1</td>
<td>35.5</td>
<td>18.1</td>
<td>8.3</td>
<td>9.9</td>
<td>0.6</td>
</tr>
<tr>
<td>Platelet count (*10⁻⁹/L)</td>
<td>129</td>
<td>126</td>
<td>123</td>
<td>88</td>
<td>82</td>
<td>111</td>
<td>26</td>
<td>184</td>
<td>91</td>
<td>116</td>
<td>124</td>
<td>0.97</td>
</tr>
<tr>
<td>CRP (mmol/L)</td>
<td>52</td>
<td>81</td>
<td>60</td>
<td>138</td>
<td>96</td>
<td>51</td>
<td>78</td>
<td>176</td>
<td>258</td>
<td>105</td>
<td>31</td>
<td>0.44</td>
</tr>
<tr>
<td>Lactate (mmol/L)</td>
<td>6.75</td>
<td>0.99</td>
<td>1.71</td>
<td>6.48</td>
<td>0.6</td>
<td>9</td>
<td>5.7</td>
<td>1.7</td>
<td>1.8</td>
<td>9.5</td>
<td>2.2</td>
<td>0.43</td>
</tr>
<tr>
<td>Base excess (mmol/L)</td>
<td>1.2</td>
<td>2.4</td>
<td>8</td>
<td>-15.1</td>
<td>-6.9</td>
<td>-9</td>
<td>-11</td>
<td>-7</td>
<td>-6</td>
<td>-12</td>
<td>-9</td>
<td>0.32</td>
</tr>
<tr>
<td>Urea (mmol/L)</td>
<td>11.7</td>
<td>4.9</td>
<td>8</td>
<td>6</td>
<td>3.3</td>
<td>7</td>
<td>7.4</td>
<td>4.4</td>
<td>11.4</td>
<td>7.6</td>
<td>6</td>
<td>0.77</td>
</tr>
<tr>
<td>Potassium (mmol/L)</td>
<td>4.5</td>
<td>4.5</td>
<td>4</td>
<td>3.4</td>
<td>3</td>
<td>3</td>
<td>3.7</td>
<td>3.2</td>
<td>3.4</td>
<td>3.2</td>
<td>3.4</td>
<td>0.14</td>
</tr>
<tr>
<td>Bilirubin (micromol/L)</td>
<td>12</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>9</td>
<td>3</td>
<td>3</td>
<td>12</td>
<td>8</td>
<td>0.78</td>
</tr>
</tbody>
</table>

* A p-value comparison is shown comparing the Kwan group against the Emonts group, indicating that the two groups have similar demographic characteristics apart from the PRISM score.
Table 2: Comparison of the Emonts versus Kwan Dataset using GSEA

<table>
<thead>
<tr>
<th>Name of Gene Set</th>
<th>Size</th>
<th>Matches to the Kwan dataset</th>
<th>Percentage</th>
<th>ES</th>
<th>p value</th>
<th>p value</th>
<th>Matches to the Emonts dataset</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>HALLMARK_MYC_TARGETS_V1</td>
<td>200</td>
<td>185</td>
<td>92.5%</td>
<td>0.53</td>
<td>1.98</td>
<td>0.002</td>
<td>191</td>
<td>95.5%</td>
</tr>
<tr>
<td>HALLMARK_BILE_ACID_METABOLISM</td>
<td>112</td>
<td>111</td>
<td>99.1%</td>
<td>0.40</td>
<td>1.37</td>
<td>0.003</td>
<td>112</td>
<td>89.2%</td>
</tr>
<tr>
<td>HALLMARK_COAGULATION</td>
<td>138</td>
<td>132</td>
<td>95.7%</td>
<td>0.39</td>
<td>1.40</td>
<td>0.007</td>
<td>138</td>
<td>100%</td>
</tr>
<tr>
<td>HALLMARK_XENOBOTIC_METABOLISM</td>
<td>200</td>
<td>190</td>
<td>95.0%</td>
<td>0.40</td>
<td>1.41</td>
<td>0.007</td>
<td>195</td>
<td>97.5%</td>
</tr>
<tr>
<td>HALLMARK_COLESTEROL_HOMEOSIS</td>
<td>74</td>
<td>69</td>
<td>93.2%</td>
<td>0.50</td>
<td>1.62</td>
<td>0.007</td>
<td>71</td>
<td>95.9%</td>
</tr>
<tr>
<td>HALLMARK_MTORC1_SIGNALING</td>
<td>200</td>
<td>189</td>
<td>94.5%</td>
<td>0.46</td>
<td>1.67</td>
<td>0.01</td>
<td>195</td>
<td>97.5%</td>
</tr>
<tr>
<td>HALLMARK_HYPOXIA</td>
<td>200</td>
<td>193</td>
<td>96.5%</td>
<td>0.39</td>
<td>1.40</td>
<td>0.015</td>
<td>194</td>
<td>97%</td>
</tr>
<tr>
<td>HALLMARK_REACTIVE_OXYGEN_SPECIAL</td>
<td>49</td>
<td>44</td>
<td>89.8%</td>
<td>0.50</td>
<td>1.57</td>
<td>0.03</td>
<td>47</td>
<td>95.9%</td>
</tr>
</tbody>
</table>

Table 2. Gene set generated from collapsing (average) the Kwan dataset and then applying Hallmark Gene Set for GSEA (h.all.v7.2.symbols.gmt), according to time. This led to column a. The data was then ported back into Emonts dataset (including controls and patients with MSS) to see the number of matches b. The percentage that matched in relation to the size of the original dataset are shown for Kwan (columns c) and for Emonts (column d). Enrichment score (ES), Normalised Enrichment score (NES), p and q values pertain to the GSEA undertaken on the Kwan dataset.
Table 3: t-Test Kwan dataset with one patient compared against all others in the group using GSEA

<table>
<thead>
<tr>
<th>Patient Against Rest</th>
<th>Cytokine.gmt</th>
<th>Negative or Positive NES scores</th>
<th>TNF.gmt</th>
<th>Negative or Positive NES scores</th>
<th>Apoptosis</th>
<th>Negative or Positive NES scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 versus Rest</td>
<td>45</td>
<td>All negative</td>
<td>132</td>
<td>All negative</td>
<td>26</td>
<td>All negative</td>
</tr>
<tr>
<td>P2 versus Rest</td>
<td>6</td>
<td>1 negative, 5 Positive</td>
<td>5</td>
<td>All positive</td>
<td>11</td>
<td>10 negative</td>
</tr>
<tr>
<td>P3 versus Rest</td>
<td>2</td>
<td>All positive</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>P4 versus Rest</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>P5 versus Rest</td>
<td>86</td>
<td>1 Negative, 85 Positive</td>
<td>74</td>
<td>All positive</td>
<td>33</td>
<td>All positive</td>
</tr>
<tr>
<td>GSEA Label</td>
<td>CYTOKINE</td>
<td>CYTOKINE</td>
<td>TNF</td>
<td>TNF</td>
<td>APOPTOSIS</td>
<td>APOPTOSIS</td>
</tr>
</tbody>
</table>

Table 3. From the Kwan datasets Patients (P1 to P5) are compared against the collective using three sets of gmt files (Cytokine, TNF and Apoptosis). To right of each gmt classifier, the net enrichment scores (NES) breakdown according to sign is shown. A negative NES implies genes are down-regulated in the patient mentioned against the collective, whereas a positive sign suggests up-regulation of gene expression. On GSEA when the analysis does not generate genes there is no NES score i.e. Not Applicable (NA).
Table 4: Curated genes from TTCA gene enriched according to the ‘Complete Response’ category

<table>
<thead>
<tr>
<th>Gene Set</th>
<th>Kwan Dataset</th>
<th>Emonts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fold Enrichment</td>
<td>P-value</td>
</tr>
<tr>
<td>KEGG_APOPTOSIS</td>
<td>1.82</td>
<td>0.004</td>
</tr>
<tr>
<td>KEGG_VEGF_SIGNALING_PATHWAY</td>
<td>0.47</td>
<td>0.931</td>
</tr>
<tr>
<td>KEGG_COMPLEMENT_AND_COAGULATION_CASCADES</td>
<td>1.53</td>
<td>0.196</td>
</tr>
<tr>
<td>Heat Shock Proteins (HUGO)</td>
<td>1.48</td>
<td>0.173</td>
</tr>
<tr>
<td>Metallothionein Genes (HUGO)</td>
<td>0.97</td>
<td>0.656</td>
</tr>
<tr>
<td>Zinc Finger Genes</td>
<td>1.29</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Table 4. TTCA analysis and enrichment generated the ‘Complete Response’ category for both the Emonts and the Kwan Datasets. The pathway gene sets are compared for each dataset. Genes are enriched for apoptosis in both datasets according to the KEGG apoptosis pathway. For Kwan the Zinc Finger Genes are enriched, whereas not for Emonts. For Emonts VEGF, complement and coagulation cascades and Heat Shock Proteins are also noted.
Data Availability
The secondary analysis of datasets undertaken in this paper were from gene expression microarray data available from public repositories. Readers can access the data from the online ArrayExpress database (EMBL-EBI ArrayExpress accession number E-MEXP-3850; https://www.ebi.ac.uk/arrayexpress/experiments/E-MEXP-3850) and from the National Library of Medicine GEO Dataset Repository ID GSE11755 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11755. There are no restrictions to data-access from the aforementioned repositories

Acknowledgments
Dr David Inwald, PICU Addenbrookes Hospital, Cambridge, UK & To Professor Hector Wong for his great contribution to sepsis genomics

Research and Ethics Committee
The paper was waived from review by the hospital Central Scientific Committee and the Research and Ethics Committee given the secondary analysis of data from primary studies.

Author Contributions
Conceived and designed the experiments: AR
Performed the experiments: AR
Analyzed the data: AR
Contributed analysis, methods, and tools: AR wrote the first draft of the paper: AR
Revised critically for importance and intellectual content: MT, PK, ZAM, ZH, JS, RK, BSB, AS, DC, RM, NQ, GB, SAZ, RN, GS, AA,AH

Funding Statement
Financial support was not required for this study