Immunosuppression, rather than inflammation, is a salient feature of sepsis in an Indian cohort

Samanwoy Mukhopadhyay[1], Pravat K Thatoi[2], Bidyut K Das[2], Saroj K Mohapatra[1*]

1 National Institute of Biomedical Genomics, Kalyani, West Bengal, India.
2 Department of Medicine, Sriram Chandra Bhanja Medical College and Hospital, Cuttack, Odisha, India.

* Corresponding author
Email: skm1@nibmg.ac.in (SKM)

1 Abstract

Sepsis remains a lethal ailment with imprecise treatment and ill-understood biology. A clinical transcriptomic analysis of sepsis patients was performed for the first time in India and revealed large-scale change in blood gene expression in patients of severe sepsis and septic shock admitted to ICU. Three biological processes were quantified using scores derived from the corresponding transcriptional modules. Comparison of the module scores revealed that genes associated with immune response were more suppressed compared to the inflammation-associated genes. These findings will have great implication in the treatment and prognosis of severe sepsis/septic shock if it can be translated into a bedside tool.

2 Introduction

Sepsis is a condition with severe systemic inflammation accompanied by dysregualted host response to infection. It is one of the leading causes of hospital stay, death and economic burden in worldwide health-care [1–3]. Sepsis is generally considered a disease continuum from bacterial infection through systemic inflammation, progressing to severe sepsis with onset of organ failure, ultimately developing into the most lethal septic shock [1]. Pneumonia, urinary tract infection and intra-abdominal infection are the chief causes leading to sepsis [1]. Bacteria form the majority of etiologic microorganisms but the bacterial classes differ geographically. Gram-negative bacteria in India, such as, *Escherichia coli*, *Klebsiella species* and *Pseudomonas aeruginosa* [4], dominate in contrast to the gram-positives in the West. While there are many pathogens causing sepsis, human host response to the infection is very complex. A large body of literature has provided insights into the genome-scale biology of sepsis [5–8]. However, the clinical predictions from these discoveries remain to be tested in Indian patients. We set out to perform unbiased blood transcriptomics in patient with severe sepsis or septic shock (SS)

from an Indian ICU. One of the major focus of our study was to find the genes and gene sets associated with poor outcome. Accordingly analysis was performed at multiple levels (genes, gene sets and networks) converging on key molecular processes (transcriptional modules) associated with outcome.

3 Results

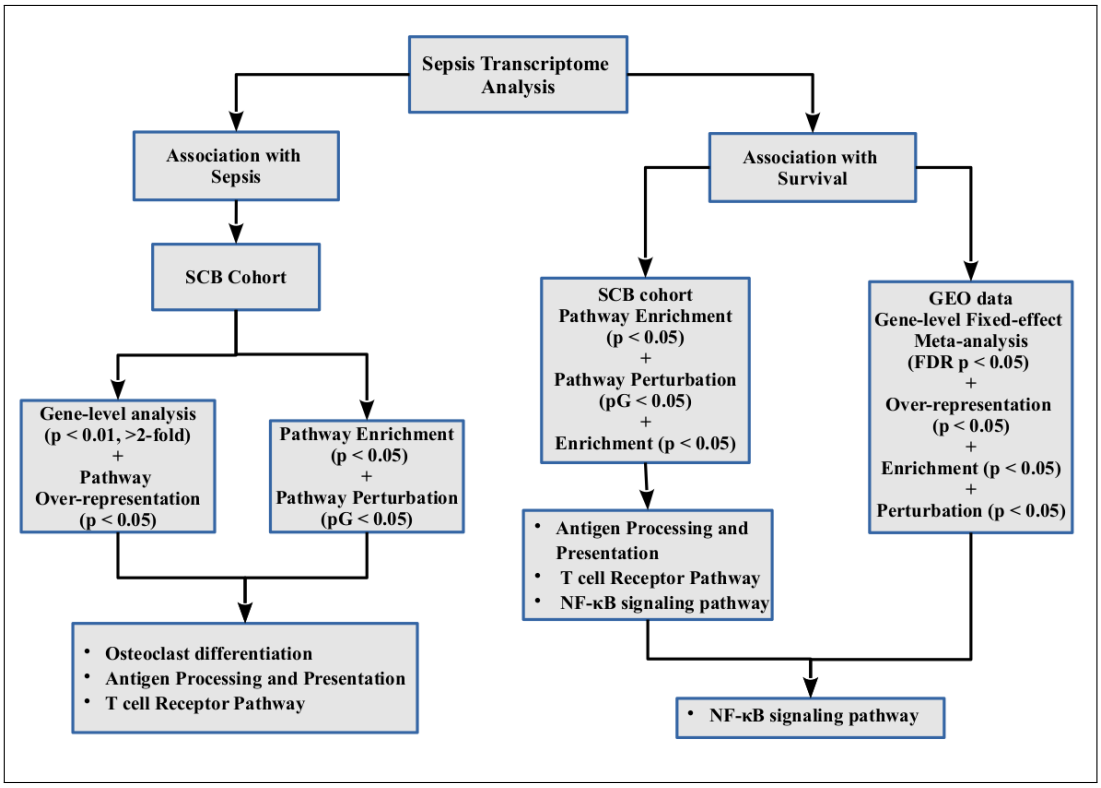

Figure 1. Analysis flow with results: The left arm describes the steps for case-control analysis and the right arm describes the analysis for association with survival.

The analysis plan for our study contained two arms (Figure 1)–firstly detection of transcriptomic changes associated with sepsis and secondly to identify pathways associated with survival.

3.1 Genome-level changes in gene expression

Differential gene expression analysis revealed 4221 genes (24% of the total number of 17513 genes assayed) to be significantly altered in sepsis patients compared to age- and gender-matched healthy controls (Figure 2). This established large scale change in gene expression in sepsis, impacting multiple pathways.

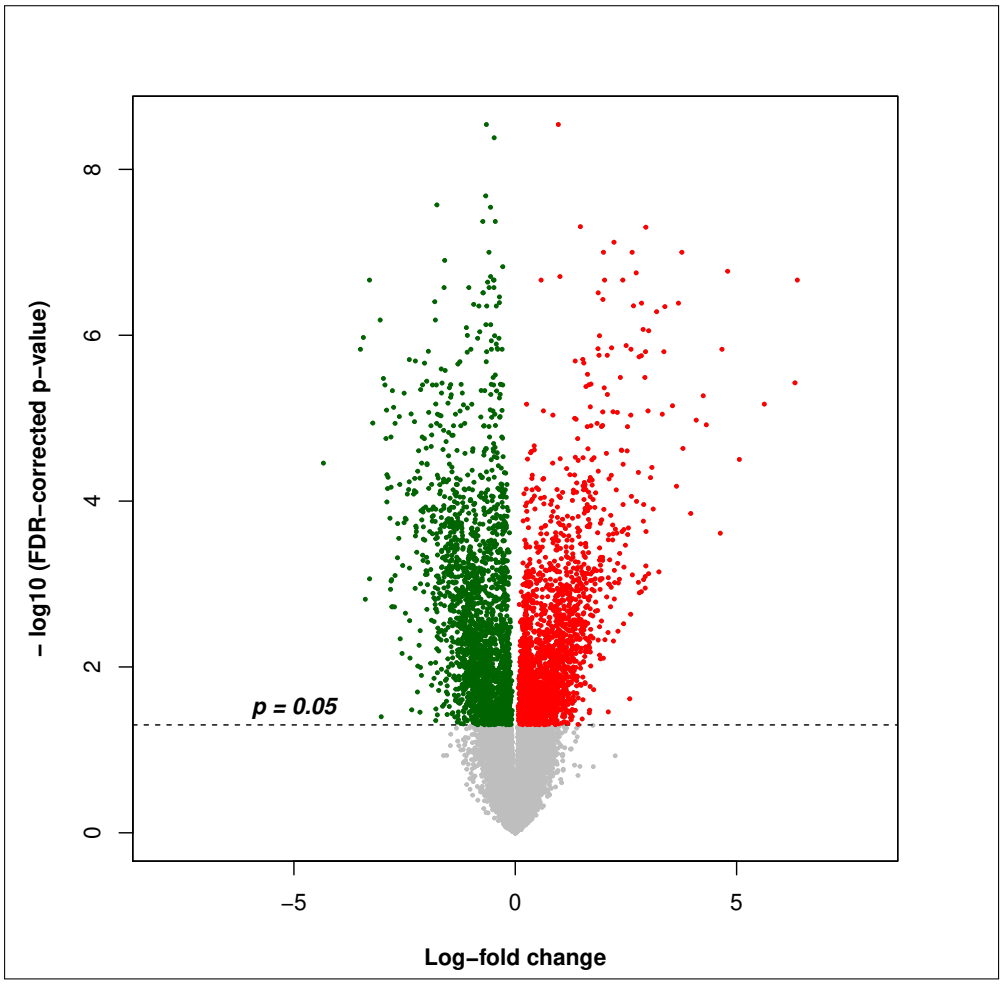

Figure 2. Volcano plot showing genome-wide change in gene expression in Sepsis: The Y-axis is in the negative logarithmic scale, showing more significant genes (smaller p-values) at the top. Significant genes (FDR $p < 0.05$) are shown above the dashed line. Green and red dots represent down- and up-regulated genes respectively. Approximately 24% of the genome (4221 out of 17513 genes) was observed to be differentially expressed.

3.2 Temporal change in gene expression

Temporal analysis of differentially expressed (DE) genes (FDR $p < 0.05$, 2 fold or greater change) revealed a non-random trend toward the baseline with time (Figure 3). Additionally, we observed a difference between survivors and non-survivors (Figure S4) suggesting delayed recovery of the transcriptome in non-survivors compared to survivors. This is consistent with the trends observed in critically ill subjects [9].

3.3 Pathways associated with disease

To extract higher-order signal, we conducted analysis at the level of "gene sets" (pathways). Three complementary approaches were pursued: (over-representation) analysis based on

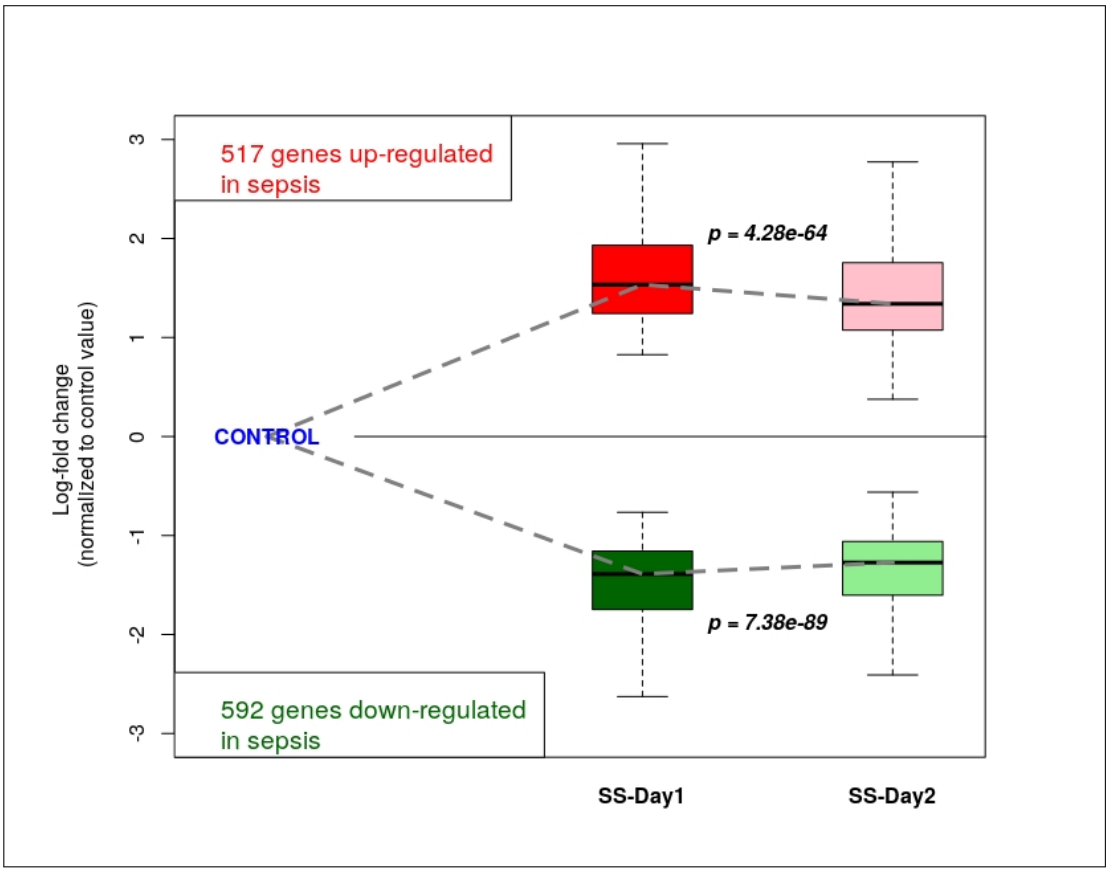

Figure 3. Temporal change of DE genes: Temporal change of DE genes (FDR p < 0.05, 2 fold-change or more), there is a non-random trend toward the baseline with time (p-values from paired t-tests are provided). This is consistent with earlier findings from patients with trauma [9].

hypergeometric test (ORA), permutation-based gene set enrichment analysis (GSEA) and topology based perturbation analysis. The intersection from the three analyses consisted of the pathways Osteoclast Differentiation, Antigen Processing and Presentation (AgPP) and T-cell Receptor (TCR) signalling (Table 1, Figure 4).

KEGG ID	KEGG pathway name	Direction of change in SS compared to Control	p value
hsa04612	Antigen processing and presentation (AgPP)	Down-regulated	3.51e-05
hsa04660	T-cell receptor signaling (TCR)	Down-regulated	5.82e-07
hsa04380	Osteoclast differentiation	Up-regulated	2.91e-06

Table 1. Key pathways observed to be perturbed in SS compared to control.

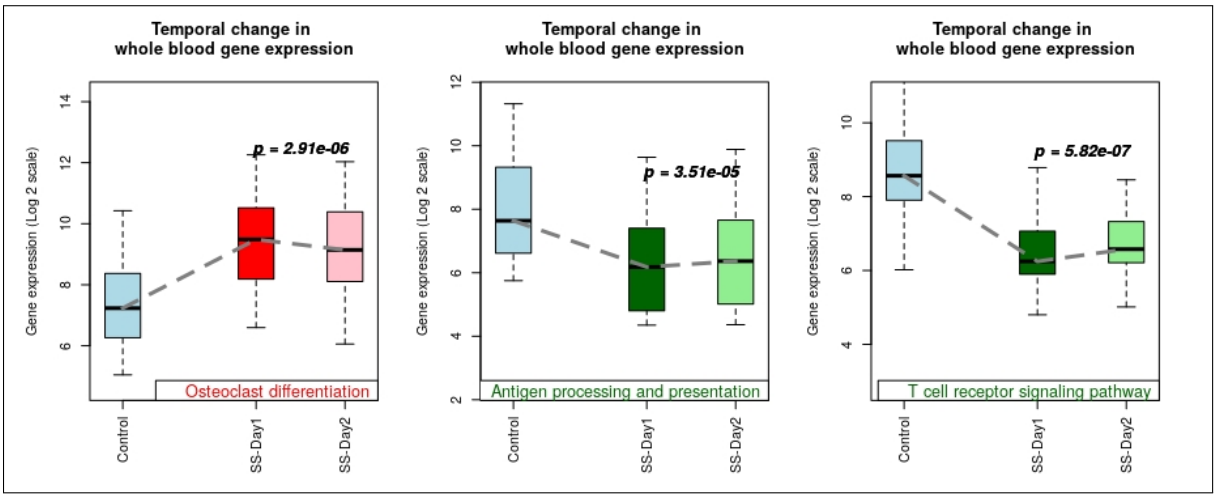

Figure 4. Pathways perturbed in disease: Each box represents median expression of genes in the given pathway. Data are plotted for three time points. "Control" (Healthy individuals), 1st day of sepsis (SS-Day1), 2nd day of sepsis (SS-Day2). The dotted line connects the median values of the three boxes and represents the trajectory of the pathway gene expression during sepsis.

3.4 Pathways associated with poor outcome in sepsis

Testing for differential expression in non-survivors compared to survivors revealed 1333 genes (7.6% of the genome) to be associated with survival ($p < 0.05$). Compared to the extent of transcriptional alteration in disease process, difference between survivors and non-survivors of sepsis was small, in terms of both the number of genes significantly affected (7.6% with survival compared to the 24% with sepsis) and magnitude of fold-change. Combined over-representation and perturbation analysis [16] revealed three pathways (Table 2) associated with survival. Each pathway was interrogated further as described below.

KEGG ID	KEGG pathway name	Direction of change in Non-survivors	p value
hsa04064	NF-kB signalling	Down-regulated	3.5e-03
hsa04612	Antigen processing and presentation (AgPP)	Down-regulated	3.6e-09
hsa04660	T-cell receptor signalling (TCR)	Down-regulated	9.33-05

Table 2. Key pathways observed to be down-regulated in non-survivors in SCB cohort compared to survivors.

3.5 NF-κB signalling pathway

NF-κB signalling pathway is known to be associated with cell survival, immunity and inflammation. This pathway was observed to be significantly down-regulated in non-survivors compared to survivors ($p = 0.0012$) in SCB cohort as well as in other cohorts (Figure 5, Figure S5). Next, we considered the role of NF-κB, a transcription factor in regulation of several genes of this pathway. We observed that the target genes of NF-κB (involved in antigen presentation, T-cell activity and macrophage activation) were significantly ($p = 0.004$) down-regulated (Figure 8) in non-survivors. The differential impact of NF-κB on host immunity was tested through relative difference in expression of genes associated with different macrophage types i.e. M1-specific pro-inflammatory genes and M2-specific hypo-inflammatory genes. Downregulation of M2-specific gene expression ($p = 0.07$) was observed in non-survivors of SCB cohort while M1-specific gene expression was observed to be up-regulated in survivors.

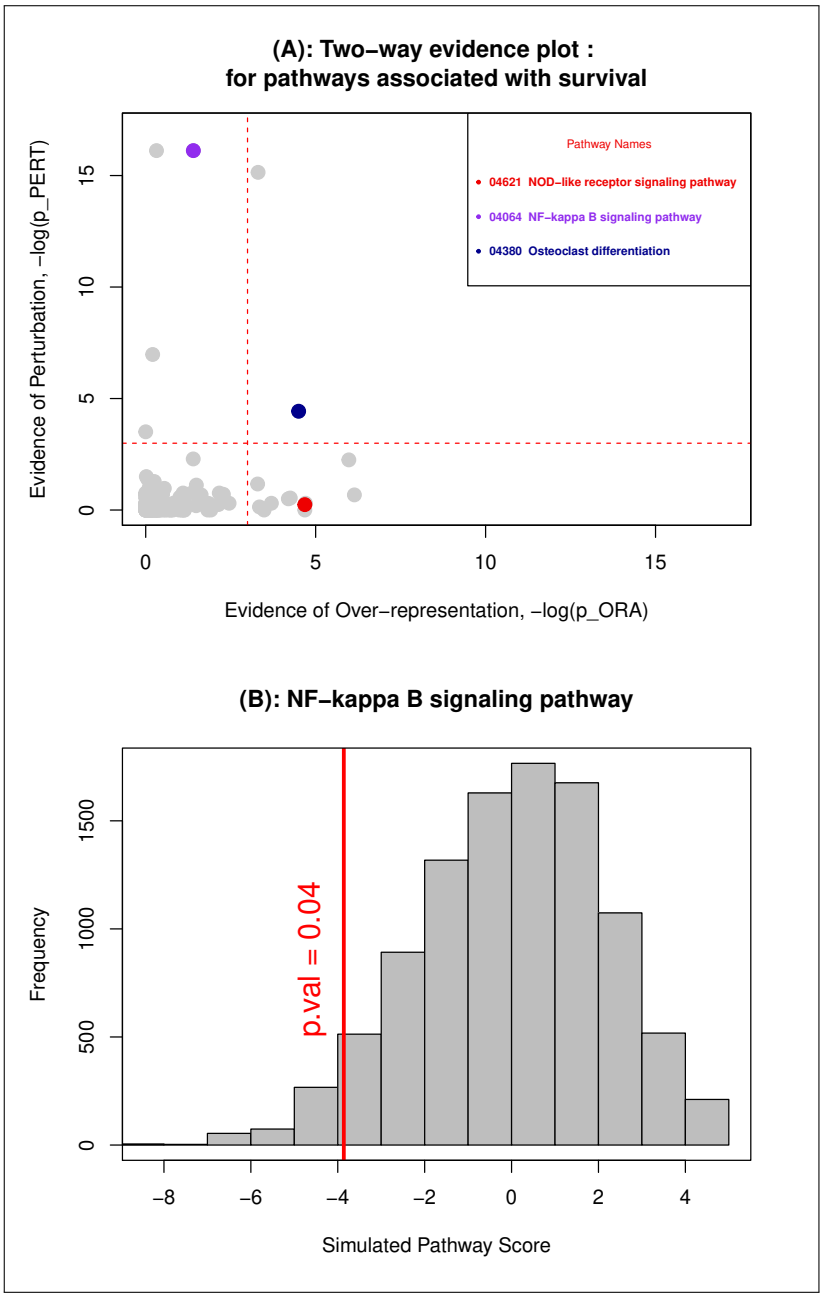

Figure 5. Evidence of perturbation of NF-κB signalling pathway: NF-κB signalling pathway is associated with survival. **(A)** NF-κB signalling pathway, Osteoclast differentiation and NOD-like receptor signalling pathway. Each point in this plot represents a single KEGG pathway. The horizontal axis records over-representation (negative log p-value) evidence and the vertical axis records the perturbation (negative log p-value) evidence. The three pathways mentioned here are all significant when both the evidences are combined (i.e., Fisher's product of the p-values). This plot is obtained from analysing data from multiple cohorts published earlier. **(B)** NF-κB pathway was also observed to be significantly perturbed in the SCB cohort by permutation based GSEA method with significant difference between survivors and non-survivors ($p = 0.04$).

3.6 Antigen processing and presentation (AgPP)

This pathway was observed to be down-regulated in non-survivors compared to survivors (Figure S6). MHC class II genes (Table S3) of this pathway were found to be down-regulated in non-survivors compared to survivors, suggesting a role in impaired adaptive immunity functions (potentially mediated by T helper cells) in non-survival.

3.7 T-cell receptor signalling (TCR)

The T-cell receptor (TCR) signalling pathway was observed to be down-regulated in survivors compared to control, and it was further down-regulated in non-survivors (Table 1; Figure S7). Down-regulation of this pathway (along with antigen presentation) in non-survivors is consistent with the association of diminished adaptive immune response with poor outcome.

3.8 Transcriptional modules underlying survival

Following the principle of monotonic change across disease severity, KEGG pathways [10] were reconstructed into three transcriptional modules associated with SS. Each of the three modules was observed to be significantly ($p < 0.05$) associated with outcome (Figure 6) and immunosuppression was observed to be the module with the highest magnitude of change (Figure S8). A detailed description of modules is provided in Supplementary Text 2.

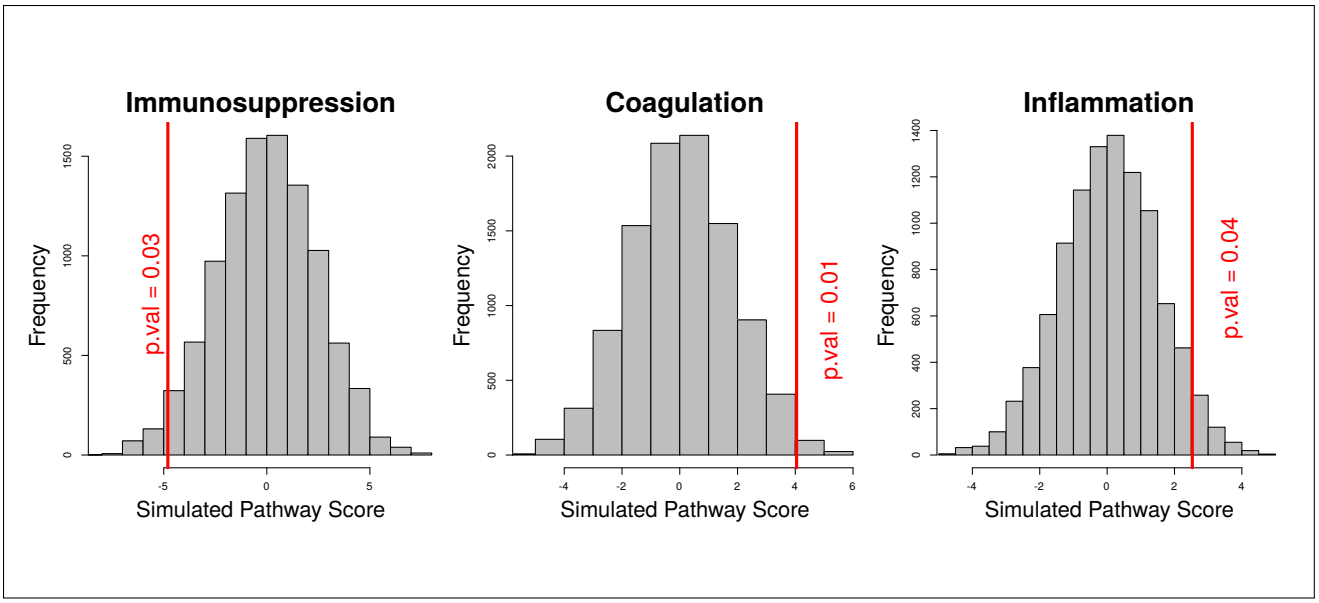

Figure 6. Significance of 3 biological modules: Histogram of simulated scores (GSEA permutation, 10000 times) and observed score (in red straight line) for three key biological modules. The histogram represents the null distribution of the gene set perturbation score. The deviation of the observed score for this gene set from the histogram suggests association between the gene sets and the disease outcome (non-survival). Significance of the module perturbation is denoted by the p-value.

3.8.1 Immunosuppression:

This module, is derived from the KEGG pathways: Antigen processing and presentation; T-cell receptor signalling and NK cell mediated cytotoxicity. Each of these pathways was observed to be significantly down-regulated in survivors compared to control, and further down in non-survivors (monotonic decrease). Antigen processing and presentation and T cell receptor mediated pathway genes were observed to be significantly down-regulated (GSEA p value 0.02 and 0.09 respectively) in non-survivors compared to survivors (Figure S6, S7). Both MHC class I and MHC class II genes were observed to be down-regulated (Table S3).

3.8.2 Coagulation:

A well known key pathophysiological characteristics of sepsis is coagulopathy. A pro-coagulant state is observed in septic patients due to the interaction of pro-inflammatory cytokines and tissue factor [11]. We observed Factor III, V, VII, XI to be up-regulated monotonically from healthy to survivors and further up-regulated in non-survivors, suggesting a more severe coagulopathy for those patients that did not survive.

3.8.3 Inflammation:

We observed several genes associated with inflammation to be monotonically up-regulated from control through survivors to non-survivors. Up-regulation of many chemokines (TGF-β, IL-13, IL-4, IL-6, and IL-21) and key signalling molecules (PI3K/AKT, PKC, WASP, ROCK) are associated with regulation of actin cytoskeleton (GSEA, p value = 0.04) (which helps in trans-endothelial migration of leukocytes).

4 Discussion

Unbiased and stringent testing revealed about one-fourth of the genome to be transcriptionally altered in day 1 of SS compared to control (Figure 2). This profound change is induced by the disease state, and is comparable with the scale of change generally observed in critical illness or sepsis (Table S1). With passage of time in the ICU, the transcriptome slowly returns to the baseline, with a small but significant difference between day 1 and day 2 of SS (Figure 3). Of note, there is a deviation in the trajectories of non-survivors compared to survivors, with the magnitude of differential expression being greater in non-survivors at both the time points. In general, delayed genomic trajectory in non-survivors suggests an increased disease severity that resists restoration of baseline expression. This trend is part of the generic host response seen in critically ill humans [9]. We have leveraged this observation in the quest for transcriptional modules with stable association with outcome.

Functionally, there is evidence of up-regulation of osteoclast differentiation and down-regulation of antigen processing and presentation, T-cell receptor signaling (Table 2). While immuno-suppression continues to be a recurring theme in sepsis biology, osteoclast differentiation has only recently been reported in our previous work in the context of septic shock with some evidence of its association with survival [12].

Immunosuppression in our patients is a result of down-regulation of NF-κB signalling in non-survivors as established by stringent testing (Figure 5). Monotonic decrease in non-survivors, i.e., positive correlation between disease severity and differential expression, suggests that this dysregulation is a stable consequence of SS. Targets of NF-κB include molecules of the immune system, for example, those associated with T-cell receptor signaling, antigen presentation and alternative macrophage activation. These are significantly down-regulated in non-survivors (Figure 8-9; Table 2). Similarly, NF-κB-mediated down-regulation of M2-type genes (Figure 9) was consistent with an immunosuppressed state in the non-survivors. In spite of the tremendous interest in the role of NF-κB in

inflammation in general and sepsis in particular [13, 14], NF-κB has resisted being a successful target of sepsis therapy [15–18]. Further investigation is required to improve our understanding of the precise role of NF-κB in determining the outcome of SS.

Monotonic change across disease severity spectrum (healthy control, survivors, non-survivors) has the potential to better define molecular underpinning of SS outcome. This is true not only at the individual gene level but also at the level of functionally connected gene sets, i.e., transcriptional modules (Figure 7). The modules correspond to the three major dysregulated pathophysiologic processes considered part of the hallmark of sepsis: immunosuppression, coagulopathy and inflammation. These transcriptional modules capture progressive dysfunctional processes in SS, culminating in poor outcome (Figure 6). Although each of the three modules is significantly different between survivors and non-survivors, the magnitude of immunosuppression is greater than that of coagulation, with least perturbation for the inflammation module (Figure S8).

Our finding raises questions for management of patients in this region. If validated on a larger set of patients, it appears that immune-enhancement e.g., immune adjuvant therapy [19]) shall be a better strategy than inhibition of inflammation.

One caveat is that this study included a small cohort of samples (n=40). Therefore, the results must be considered preliminary in nature. Secondly all our patients were recruited from a single clinical center. In order to adequately capture the demographic diversity of the region, a future multi-centric study shall need to include patients from different geographical locations across the country.

5 Conclusion

Profound transcriptional reprogramming in SS was interrogated by complementary bioinformatic strategies revealing down-regulation of immune pathways. NF-κB appears to be at the center of this change. Data-driven integration of KEGG annotation and sepsis pathophysiology led to identification of three key transcriptional modules associated with survival – immunosuppression, coagulopathy and inflammation. Quantitatively, the magnitude of immunosuppression is much more than that of inflammation, with potential role in patient stratification. The three modules (containing 288 genes) can be tested (with the help of quick gene expression profiling technologies at bedside) to determine which category each patient belongs. Specific therapy can then be catered to each patient based on that categorisation.

6 Materials and methods

6.1 Differential gene expression analysis

Welch t-test [20] was performed to detect genes DE in SS compared to gender and age matched healthy controls. Correction for multiple testing was performed to reduce False Discovery Rate (FDR) [21]. Threshold of FDR p-value of < 0.01 and 2-fold or greater change; was applied to detect DE genes.

6.2 Pathway analysis methods

6.2.1 Over-Representation Analysis (ORA)

Significantly up-regulated genes were subjected to over-representation analysis by applying hypergeometric test [22]. Correction for multiple testing was performed to reduce FDR. A significant association was detected at threshold of FDR $p < 0.001$.

6.2.2 Gene Set Enrichment Analysis (GSEA)

Permutation test was performed by randomly scrambling the sample labels (case/control) and computing the enrichment score (t-statistics) of the gene set for this permuted data set [23]. Multiple rounds of this process generated the null-distribution of the enrichment score. Pathways with observed enrichment score significantly deviated (FDR $p < 0.05$) from the null distribution were considered significant [23].

6.2.3 Perturbation Analysis of Signalling Pathways

Analysis based on pathway topology (derived from [24]) revealed the probability of perturbation of the pathway. In this process the positions of the differentially expressed genes are scrambled in the pathway and for each combination a perturbation score is calculated. Ultimately a total perturbation score is calculated by summing up the cumulative perturbation score from each iteration of gene swapping. In our analysis perturbation score was applied along with over-representation score visualise pathways (Figure 5) . By combining two different evidences (one from the hypergeometric test model and the other from the probability of perturbation that takes the pathway topology into account), a two way evidence visualisation was made possible [24].

6.3 Construction of Biological Modules

We reconstructed three modules (immunosuppression, coagulation and inflammation) from KEGG pathways associated with SS (Figure 7). First, we selected genes (nodes) from significantly perturbed pathways, that were also monotonically dysregulated from baseline in survivors and further in non-survivors. We reasoned that the extent of perturbation of biological process shall increase with severity of disease, and therefore, the genes that are monotonically changing are key nodes of outcome-associated modules. Then we extended the network of genes based on the connectivity as described in a well-curated pathway database (KEGG [10]). Wherever possible, we made the connectivity parsimonious with minimal number of non-significant genes (included only to preserve the continuity of signal flow, in keeping with the directed nature of curated KEGG graph). Three modules were constructed to describe three aspects of sepsis pathophysiology – inflammation, coagulation and immunosuppression. Each module was then considered an independent gene set and tested for significance of transcriptional difference between non-survivors and survivors by permutation-based testing of pathway score.

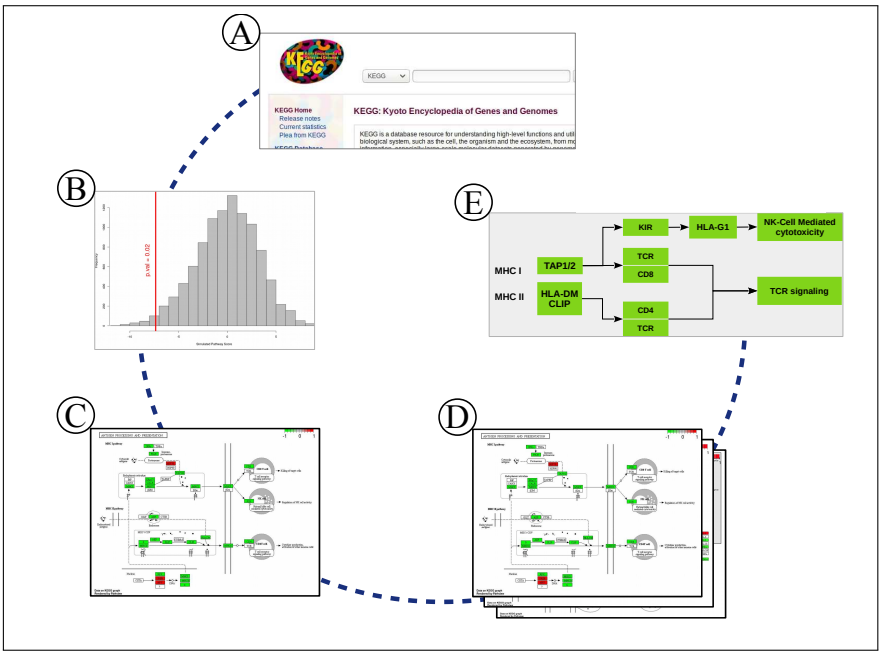

Figure 7. Steps of module construction: The sequence of steps leading to the construction of the three biological modules. (A) Selection of key biological processes (e. g. Immune pathways, Signalling pathways, Metabolic pathways) from KEGG database. (B) Testing for significant perturbation of the pathway in non-survivors. (C) Selection of pathways with genes differentially expressed and monotonic change from control through survivors to non-survivors. Finding the pathways which are monotonically changed. (D) Merging the monotonic pathways into a set of essentially connected nodes. (E) Construction of three key transcriptional modules (one is displayed here).

6.4 Selection of data sets from GEO

Electronic search was performed in NCBI GEO (Gene Expression Omnibus) on 06th December 2017, with the search string: "sepsis or septic shock or survivor or non-survivor". Application of human filter (species-human) resulted in 9 transcriptome data sets (Figure S9; Table S4). One study (GSE78929) was rejected as it contained data from a tissue (muscle tissue) and not blood. Blood was chosen as it captures the systemic cellular response in systemic inflammatory disease such as sepsis [25]. Hierarchical clustering of the gene expression data led to clear age-specific segregation of sepsis transcriptome (Figure S10). The data from our SCB cohort consisted only of adult sepsis cases. Accordingly, all subsequent analysis for validation was performed on adult data sets only.

6.5 SCB (Srirama Chandra Bhanja Medical College) cohort

Transcriptomic profiling was performed on sepsis cases and on matched (by age and gender) healthy controls, who were not suffering from any inflammatory diseases, and were not related to the patients. Samples were collected at two different time points in the cases. A

comparison between survivors and non-survivors was planned to identify genomic changes that are specifically associated with survival. Patients suspected of sepsis were recruited into the study with defined inclusion and exclusion criteria mentioned in Table S2. Two blood samples from the sepsis cases were collected, the first at the time of diagnosis (D1), and a second sample after 24 hours (D2) . Representative single blood sample was collected from each of the healthy control individuals. Blood samples were collected from the patients of SS and healthy subjects after obtaining approval from the Institutional Ethical Committees of the National Institute of Biomedical Genomics, Kalyani and SCB Medical College, Cuttack. All the methods were carried out in accordance with the approved guidelines. Informed written consent was obtained from all subjects who participated in the study. A total of 27 patients (23 with paired transcriptome data for two time points) and 12 healthy control subjects (for the single time point) were recruited in this study. For each of the cases and matched control samples quality of RNA (isolated from whole blood) was assessed by the following criteria: ratio of absorbance of light at two wavelengths (A_{260}/A_{280}) should be between 1.8 - 2 and RNA Integrity Number (RIN) of 6 or higher. Another criterion for selection of subjects was based on temporal analysis, which requires that the two temporal samples from each subject should pass the quality assessment mentioned above.

6.5.1 Sample processing and quality assessment

Venous blood was collected from patients in PAXgene Blood collection tubes from BD, Franklin Lakes, New Jersey, USA (cat. no. 762165), containing stabilising reagent that keeps the blood cells fixed and preserves cellular RNA before further use. PAXgene Blood RNA isolation Kit from Qiagen, Hilden, Germany (cat. no. 762164) was used to isolate the RNA from the whole blood according to manufacturer's instructions. Spectroscopy was done in Nanodrop-2000 (ND-2000, from Thermo Fisher Scientific, Waltham, Massachusetts, USA) for checking RNA quality and concentration. Agilent Bioanalyzer RNA nano kit (Agilent RNA 6000 Nano) was used for checking RNA quality. Illumina TotalPrep RNA Amplification Kit was used to convert the total RNA to biotinylated cRNA for hybridisation to microarray chip. Transcriptome profiling was done on the following microarray platforms: Affymetrix (GeneChip^{TM} Human Gene 2.0 ST Array, cat. no. 902113) and Illumina Microarray Chips (HumanHT-12 v4 Expression BeadChip Kit, cat. no. BD-103-0204)

6.5.2 Software:

All analysis were done in R [26], a language and environment for statistical computing in Linux operating system.

6.5.3 Data availability:

A vignette with reproducible code chunks for analysis is provided in the supplementary data. All data and code are available at (https://figshare.com/) under project **("ssnibmgsurv")**.

7 Author contributions

SM: contributed to Data Curation, Microarray Hybridization, Investigation, Analysis, Methodology, Software, Validation, Visualization, Writing, Reviewing and Editing of the original draft.

PT: contributed to Data Curation, Investigation, Reviewing the original draft.

BD: contributed to Data Curation, Investigation, Funding Acquisition, Project Administration, Resources, Supervision, Reviewing the original draft.

SKM: conceptualized the study and contributed to Funding Acquisition, Project Administration, Resources, Supervision, Analysis, Methodology, Software, Reviewing and editing of the original draft.

All authors contributed to Manuscript Reviewing and approved of the final manuscript.

8 Acknowledgments

This work was supported by an intramural grant from NIBMG, Kalyani and an extramural grant (No. BT/PR5548/MED/29/571/2012, duration of 3 years, sanctioned on 27-05-2013) from the Department of Biotechnology, Government of India. S.M. acknowledges research fellowship (No. F.2-45/2011(SA-I)) provided by the University Grants Commission of India. We are grateful to the patients of SCB Medical College Hospital for contributing blood samples for the study. We acknowledge the kind suggestion given by Prof. Partha P Majumder during the analysis of this data. We acknowledge the kind suggestions of Dr. Samsiddhi Bhattacharjee regarding the metaanalysis of published data. We acknowledge the kind suggestions by Prof. B. Ravindran for his kind discussion on many topics of sepsis. We thank our colleagues at the NIBMG core facility and the ILS laboratory for timely processing of the samples, often at short notice.

References

1. Angus DC, Van der Poll T. Severe sepsis and septic shock. New England Journal of Medicine. 2013;369(9):840–851.

2. Mellhammar L, Wullt S, Lindberg Å, Lanbeck P, Christensson B, Linder A. Sepsis incidence: a population-based study. In: Open Forum Infectious Diseases. vol. 3. Oxford University Press; 2016.

3. Vincent JL, Marshall JC, Ñamendys-Silva SA, François B, Martin-Loeches I, Lipman J, et al. Assessment of the worldwide burden of critical illness: the intensive care over nations (ICON) audit. The Lancet Respiratory Medicine. 2014;2(5):380–386.

4. Chatterjee S, Bhattacharya M, Todi SK. Epidemiology of adult-population sepsis in India: a single center 5 year experience. Indian journal of critical care medicine: peer-reviewed, official publication of Indian Society of Critical Care Medicine. 2017;21(9):573.

5. Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, Cho RJ, et al. A network-based analysis of systemic inflammation in humans. Nature. 2005;437(7061):1032.

6. Wong HR, Shanley TP, Sakthivel B, Cvijanovich N, Lin R, Allen GL, et al. Genome-level expression profiles in pediatric septic shock indicate a role for altered zinc homeostasis in poor outcome. Physiological Genomics. 2007;30(2):146–155.

7. Wong HR, Cvijanovich N, Allen GL, Lin R, Anas N, Meyer K, et al. Genomic expression profiling across the pediatric systemic inflammatory response syndrome, sepsis, and septic shock spectrum. Critical Care Medicine. 2009;37(5):1558.

8. Wong HR, Freishtat RJ, Monaco M, Odoms K, Shanley TP. Leukocyte subset-derived genome-wide expression profiles in pediatric septic shock. Pediatric critical care medicine: a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies. 2010;11(3):349.

9. Xiao W, Mindrinos MN, Seok J, Cuschieri J, Cuenca AG, Gao H, et al. A genomic storm in critically injured humans. Journal of Experimental Medicine. 2011;208(13):2581–2590.

10. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Research. 2000;28(1):27–30.

11. Simmons J, Pittet JF. The coagulopathy of acute sepsis. Current Opinion in Anaesthesiology. 2015;28(2):227.

12. Mukhopadhyay S, Thatoi PK, Pandey AD, Das BK, Ravindran B, Bhattacharjee S, et al. Transcriptomic meta-analysis reveals up-regulation of gene expression functional in osteoclast differentiation in human septic shock. PLoS ONE. 2017;12(2):e0171689.

13. Abraham E. Nuclear factor—κB and its role in sepsis-associated organ failure. The Journal of Infectious Diseases. 2003;187(Supplement_2):S364–9.

14. Liu SF, Malik AB. NF-κB activation as a pathological mechanism of septic shock and inflammation. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2006;290(4):L622–L645.

15. Li X, Cui X, Li Y, Fitz Y, Hsu L, Eichacker PQ. Parthenolide has limited effects on nuclear factor-$\kappa\beta$ increases and worsens survival in lipopolysaccharide-challenged C57BL/6J mice. Cytokine. 2006;33(6):299–308.

16. Li X, Su J, Cui X, Li Y, Barochia A, Eichacker PQ. Can we predict the effects of NF-κB inhibition in sepsis? Studies with parthenolide and ethyl pyruvate. Expert Opinion on Investigational Drugs. 2009;18(8):1047–1060.

17. Uwe S. Anti-inflammatory interventions of NF-κB signaling: potential applications and risks. Biochemical Pharmacology. 2008;75(8):1567–1579.

18. Su J, Li X, Cui X, Li Y, Fitz Y, Hsu L, et al. Ethyl pyruvate decreased early nuclear factor-κB levels but worsened survival in lipopolysaccharide-challenged mice. Critical Care Medicine. 2008;36(4):1059–1067.

19. Ono S, Tsujimoto H, Hiraki S, Aosasa S. Mechanisms of sepsis-induced immunosuppression and immunological modification therapies for sepsis. Annals of Gastroenterological Surgery. 2018;2(5):351–358.

20. Welch BL. On the comparison of several mean values: an alternative approach. Biometrika. 1951;38(3-4):330–336.

21. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological). 1995; p. 289–300.

22. Fisher R. Book :Statistical methods for research workers. V; 1925. Available from: http://psychclassics.yorku.ca/Fisher/Methods.

23. Oron AP, Jiang Z, Gentleman R. Gene set enrichment analysis using linear models and diagnostics. Bioinformatics. 2008;24(22):2586–2591.

24. Tarca AL, Draghici S, Khatri P, Hassan SS, Mittal P, Kim Js, et al. A novel signaling pathway impact analysis. Bioinformatics. 2008;25(1):75–82.

25. Kohane IS, Valtchinov VI. Quantifying the white blood cell transcriptome as an accessible window to the multiorgan transcriptome. Bioinformatics. 2012;28(4):538–545.

26. R Core Team. R: A Language and Environment for Statistical Computing; 2013. Available from: http://www.R-project.org/.

27. Flohé SB, Agrawal H, Flohé S, Rani M, Bangen JM, Schade FU. Diversity of interferon γ and granulocyte-macrophage colony-stimulating factor in restoring immune dysfunction of dendritic cells and macrophages during polymicrobial sepsis. Molecular Medicine. 2008;14(5-6):247–256.

28. Bo L, Wang F, Zhu J, Li J, Deng X. Granulocyte-colony stimulating factor (G-CSF) and granulocyte-macrophage colony stimulating factor (GM-CSF) for sepsis: a meta-analysis. Critical Care. 2011;15(1):R58.

29. Venet F, Foray AP, Villars-Méchin A, Malcus C, Poitevin-Later F, Lepape A, et al. IL-7 restores lymphocyte functions in septic patients. The Journal of Immunology. 2012;189(10):5073–5081.

30. Inoue S, Unsinger J, Davis CG, Muenzer JT, Ferguson TA, Chang K, et al. IL-15 prevents apoptosis, reverses innate and adaptive immune dysfunction, and improves survival in sepsis. The Journal of Immunology. 2010;184(3):1401–1409.

31. Brahmamdam P, Inoue S, Unsinger J, Chang KC, McDunn JE, Hotchkiss RS. Delayed administration of anti-PD-1 antibody reverses immune dysfunction and improves survival during sepsis. Journal of Leukocyte Biology. 2010;88(2):233–240.

32. Zhang Y, Zhou Y, Lou J, Li J, Bo L, Zhu K, et al. PD-L1 blockade improves survival in experimental sepsis by inhibiting lymphocyte apoptosis and reversing monocyte dysfunction. Critical Care. 2010;14(6):R220.

33. Midwood KS, Williams LV, Schwarzbauer JE. Tissue repair and the dynamics of the extracellular matrix. The International Journal of Biochemistry & Cell Biology. 2004;36(6):1031–1037.

34. Awad F, Assrawi E, Jumeau C, Georgin-Lavialle S, Cobret L, Duquesnoy P, et al. Impact of human monocyte and macrophage polarization on NLR expression and NLRP3 inflammasome activation. PLoS ONE. 2017;doi:10.1371/journal.pone.0175336.

9 Figures and Tables

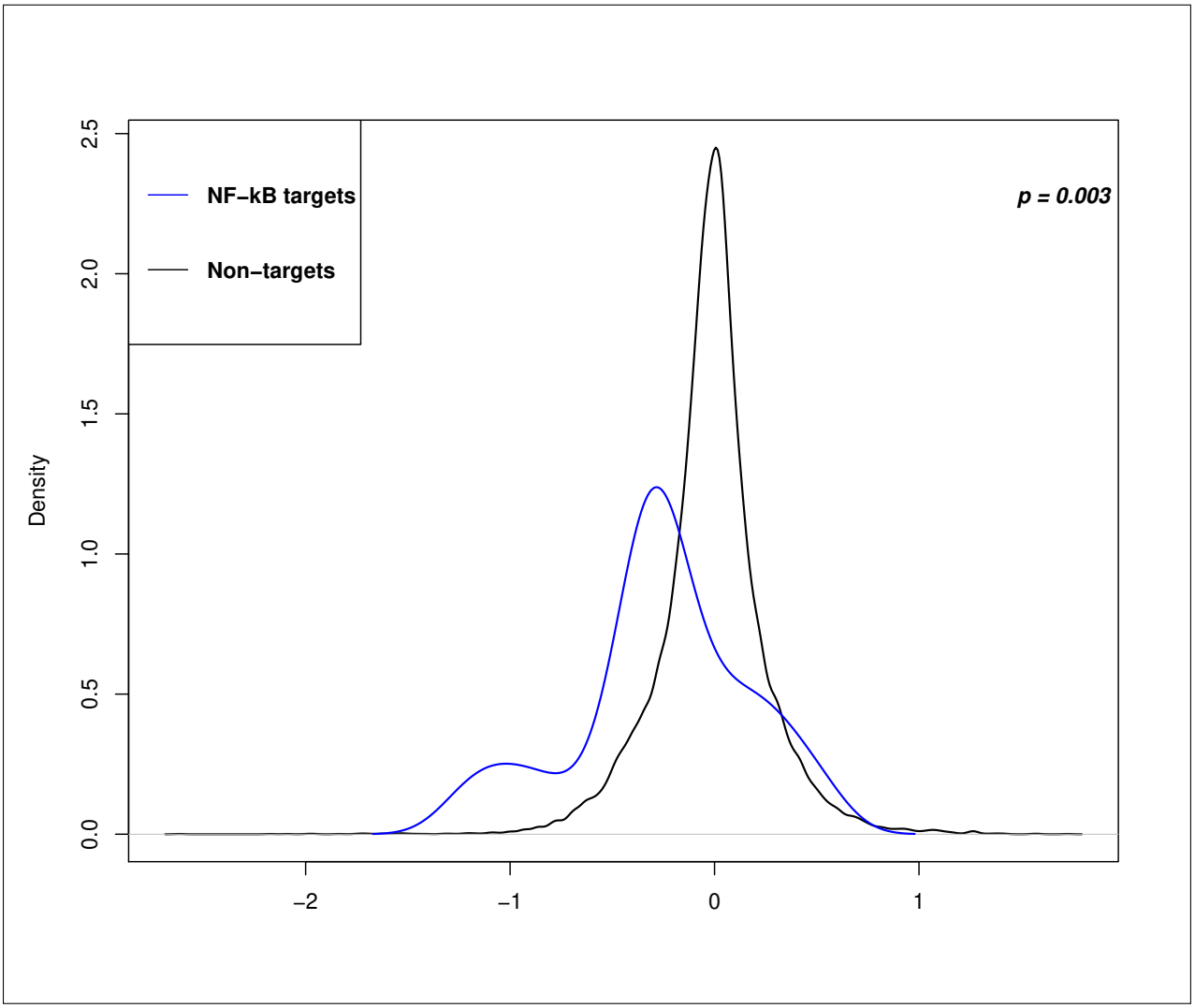

Figure 8. NF-κB target gene expression density plot: A density plot of targets of NF-κB between non-survivors and survivors that include antigen processing and presentation genes and various immune receptor genes. The gray peak in the background represents all the non-target genes in the genome and the blue peak represents the NF-κB targets. There is a shift in the target gene expression to the left (p= 0.005), suggesting NF-κB-induced down-regulation of its targets.

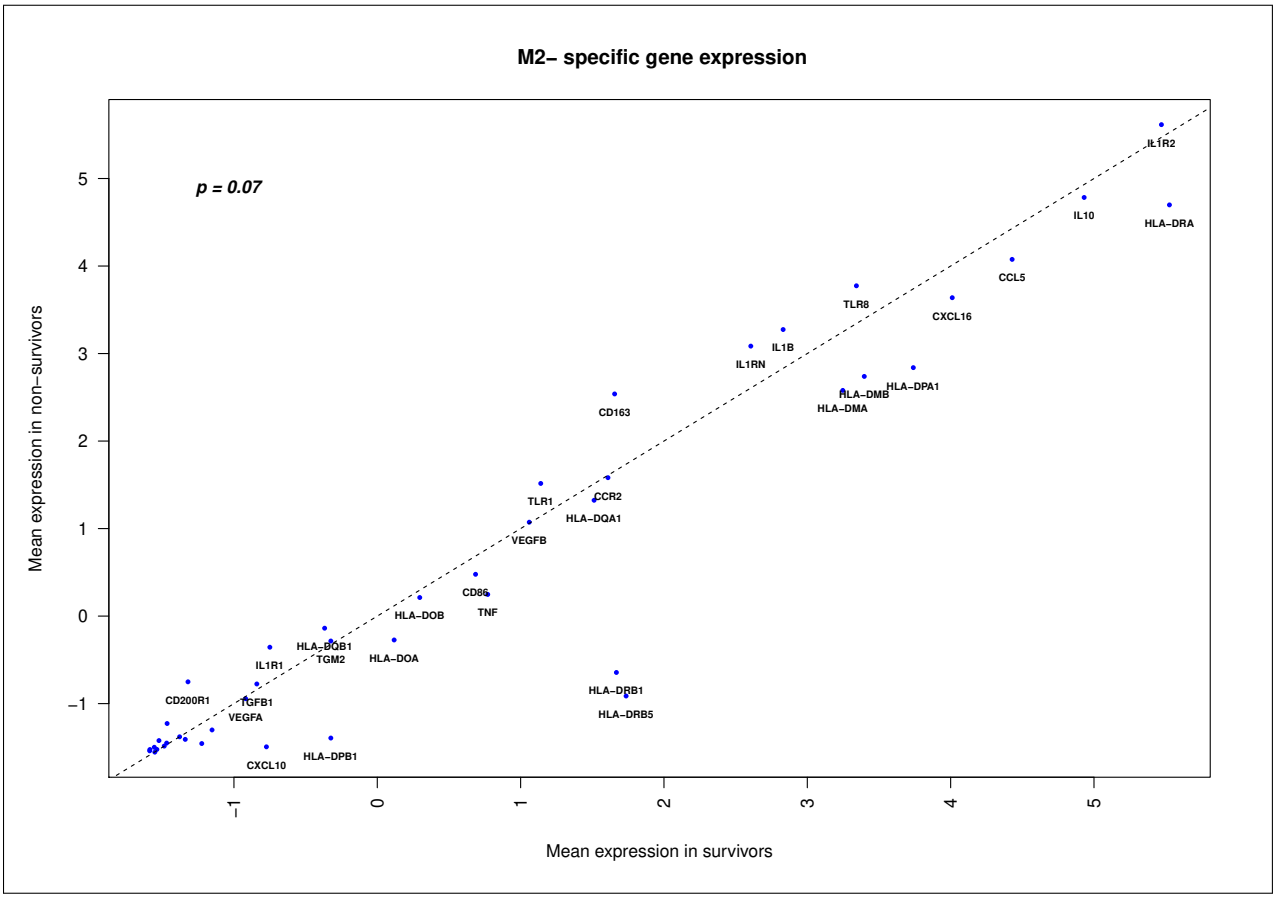

Figure 9. M2-specific gene expression: Scatterplot of M2 target gene expression in non-survivors compared to survivors. Each point represents a single gene with mean expression in the two groups of patients: survivors (x-axis) and non-survivors (y-axis). For most of the genes there is higher mean expression in survivors, suggesting significant M2-specific (p = 0.07) under-expression in non-survivors.

10 Supplementary Text 1

10.1 Immunostimulatory therapy in sepsis

Our result is supportive of Ono *et al.* [19] who argued in favour of immunostimulation over targeting inflammation as a possible therapeutic startegy to treat patients suffering from sepsis. As the immunosuppression involves chiefly the adaptive immune system, there have been attempts to boost adaptive immunity through IFNγ, granulocyte-macrophage colony-stimulating factor (GMCSF), or granulocyte colony-stimulating factor GCSF [27] . However these attempts have failed to demonstrate a clear survival benefit to those who received these therapies. A meta-analysis [28] reports that GCSF and GMCSF failed to show survival benefit in sepsis patients. Two Interleukins IL-7 [29] and IL-15 [30] have been shown to have adaptive immunostimulatory function in sepsis patients. IL7 treatment to sepsis patients restored IFN-gamma secretion and T-cell proliferation [29] in them. PD-1 is another interesting target molecule for immunostimulatory therapy, as it causes immunosuppression through IL-10 expression [19]. Anti PD-1 [31] and PDL-1 [32] therapy in bacterial and fungal murine sepsis showed increased survival. Disruption of the PD-1/PDL-1 axis seems to be a novel approach for restoring immune function in sepsis patients [19] .

11 Supplementary Text 2

Immunosuppression: This module (Figure S1) consists of MHC molecules with co-stimulators (CD4, CD8), transporters (TAP1/2), and receptors (KIR, TCR) resulting in NK-cell mediated cytotoxicity and T-cell receptor signaling. TAP1, TAP2 help in transport of the processed cytosolic pathogenic antigen with the help of MHC class I. NK cells are the part of innate immunity that detects the pathogen infected cell by MHC class I and interact through Killer-cell immunoglobulin-like receptors (KIRs) located in the plasma membrane of NK cells. The NK cells mostly play an instrumental role in removal of virus infected cells. MHC class II mediated antigen presentation, observed to be down-regulated in non-survivors leading to impaired TCR signalling with MHC class II and CD4, CD8 all were significantly down-regulated in non-survivors. This is consistent with general and perhaps reversible immune paralysis/ suppression in non-survivors.

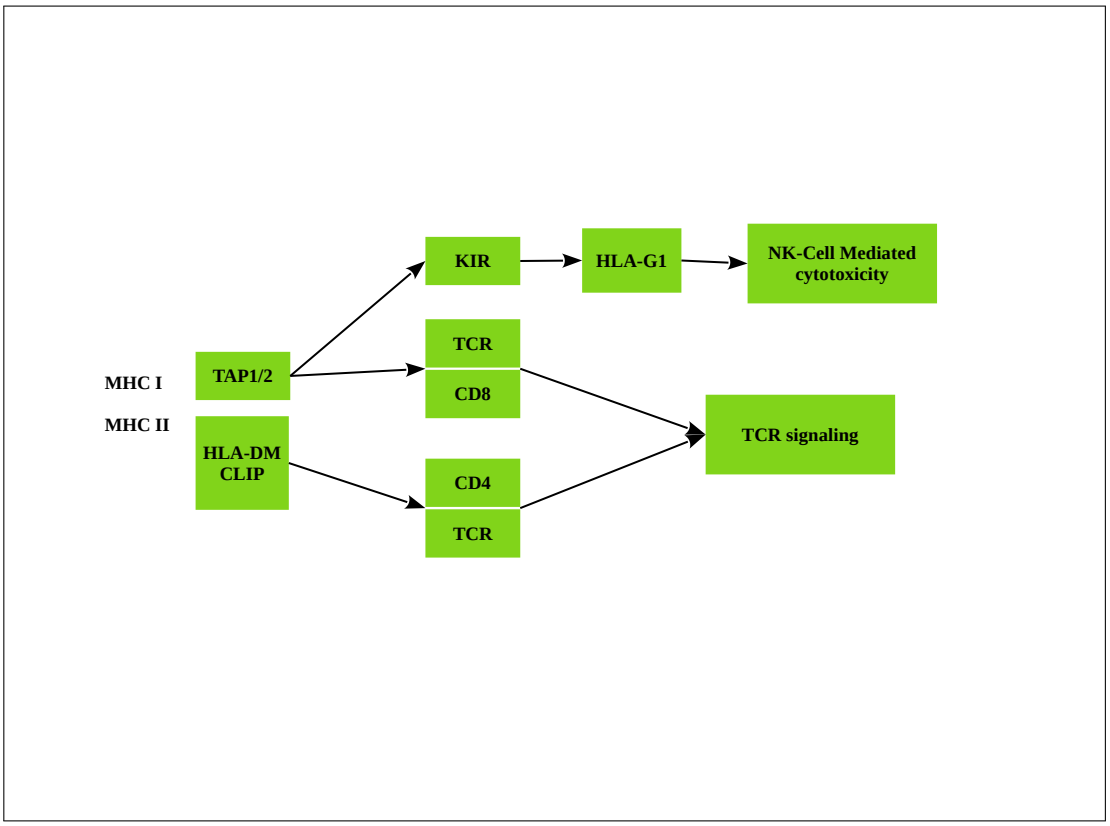

Figure S1. Immunosuppression module A schematic representation of the connected nodes in the immunosuppression module.

Coagulation: Platelets can detect vascular injury through collagen (up-regulated in non-survivors and promoting clot formation by platelet activation [33]. This module (Figure S2) consists of collagen-induced platelet activation and clotting factors both causing formation of fibrin and a persistent pro-coagulant phenotype. GPVI (up-regulated in non-survivors) is a key receptor protein that plays an important role in platelet activation after the collagen is bound to it. Up-regulation (GSEA $p = 0.03$) of extracellular matrix (ECM) pathways suggests damage to the ECM. This activates platelets to the site of injury that contribute to repair of the damaged tissue at that site [33].

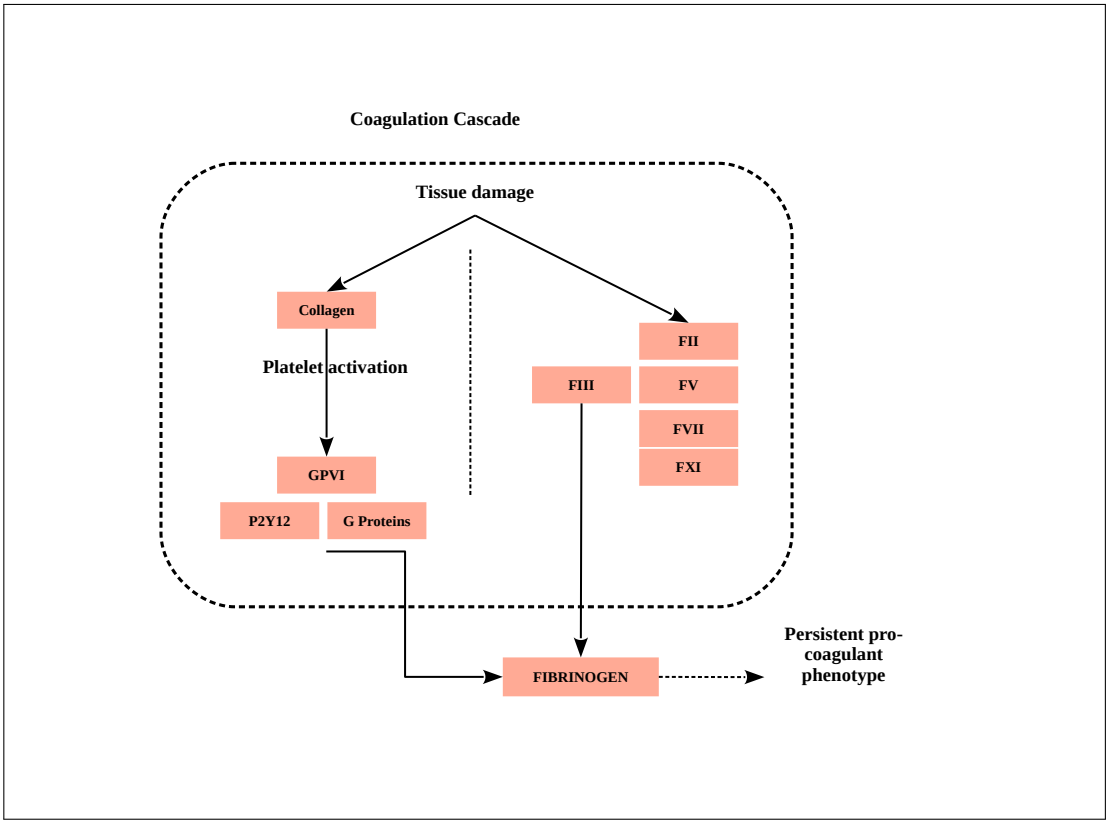

Figure S2. Coagulation Module: A schematic representation of the connected nodes in the coagulation module.

Inflammation: This module (Figure S3) consists of pattern recognition receptors, chemokines, signaling molecules (PI3K/AKT, MAP Kinases), leading to regulation of actin cytoskeleton and leukocyte migration. As expected, there is up-regulation of genes associated with inflammation.

The main input signal of the inflammation emanates from not only the PAMPS (that recognise microbial antigens), but also the damage-associated molecular patterns (DAMPs); that indicates tissue damage due to host inflammation. Toll like receptors recognize pathogen-associated molecular patterns (PAMPs) [34] as well as DAMPs.

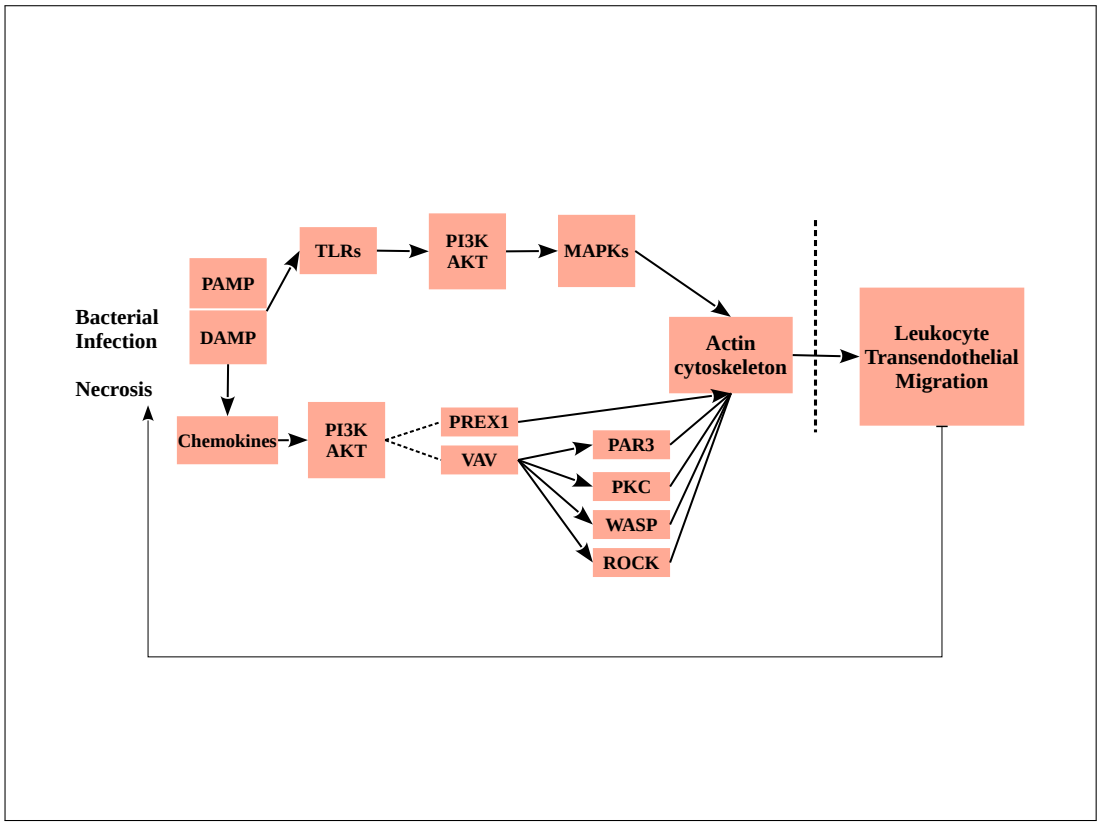

Figure S3. Inflammation Module: A schematic representation of the connected nodes of the inflammation module.

12 Supplementary Figures and Tables: Immunosuppression, rather than inflammation, is a salient feature of sepsis in an Indian cohort.

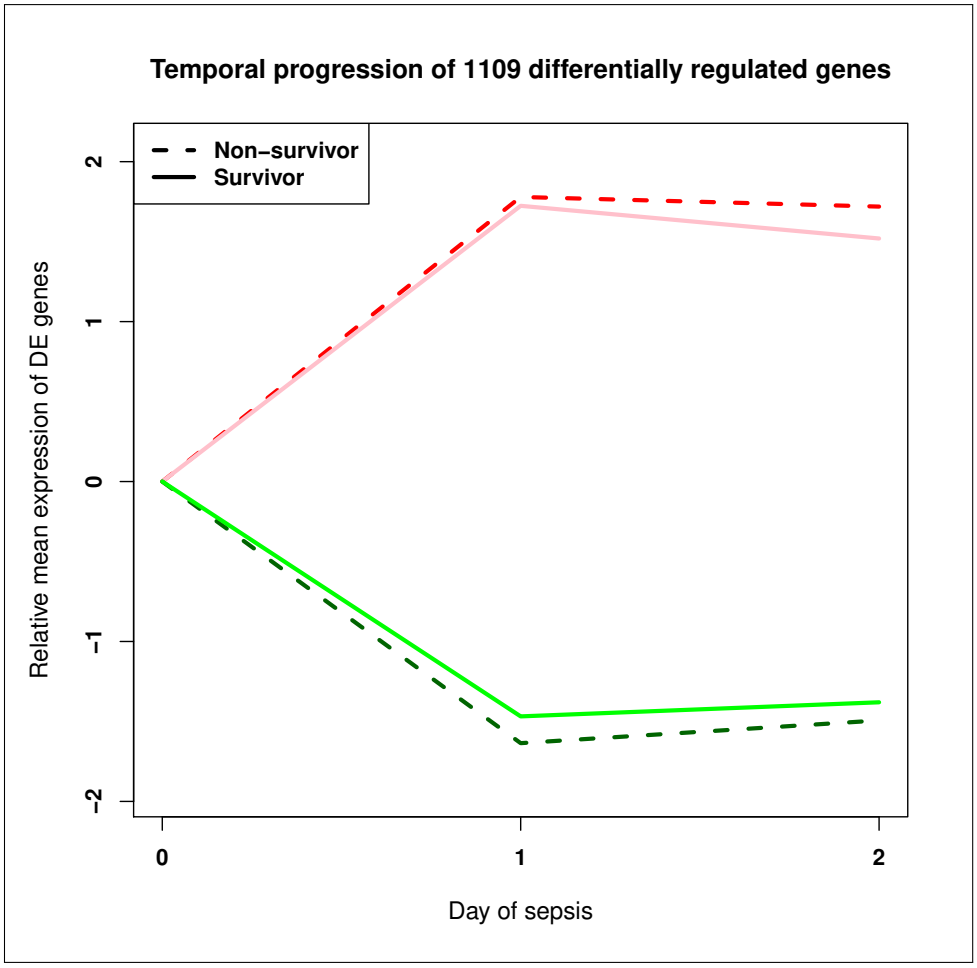

Figure S4. Trajectory of DE genes: This is a line plot of mean expression of DE genes in SS patients compared to control. Compared to the survivor group, generally the non-survivor gene expression is more deviated from controls and continues to diverge with time.

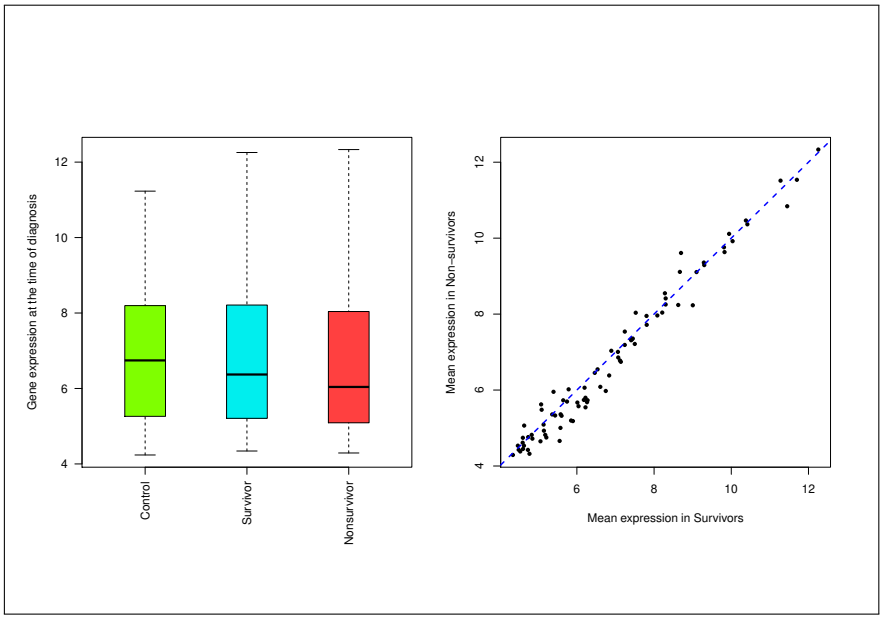

Figure S5. NF-κB pathway: Boxplot and scatterplot of NF-κB pathway in SCB cohort. The pathway is monotonically down-regulated from control through survivors to non-survivors on Day-1.

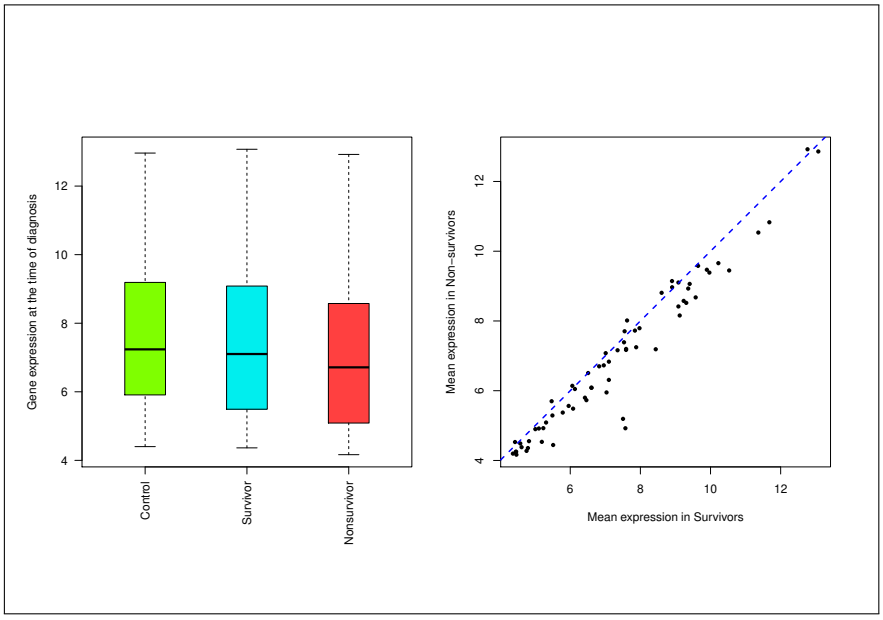

Figure S6. Antigen Processing and Presentation Pathway: Boxplot and scatterplot of Antigen Processing and Presentation Pathway pathway in SCB cohort. The pathway is monotonically down-regulated from control through survivors to non-survivors on Day-1.

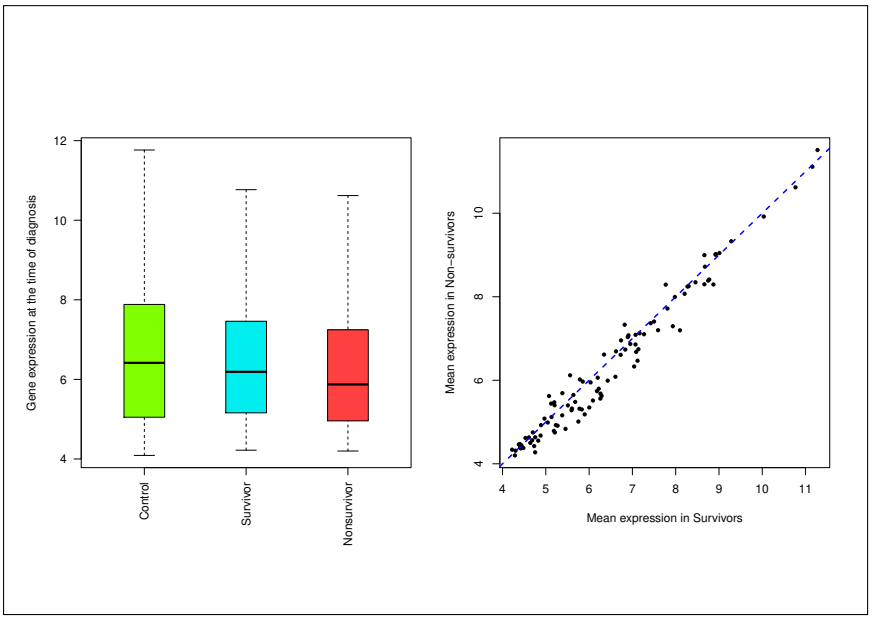

Figure S7. TCR signaling Pathway: Boxplot and scatterplot of TCR signaling pathway in SCB cohort. The pathway is monotonically down-regulated from control through survivors to non-survivors on Day-1.

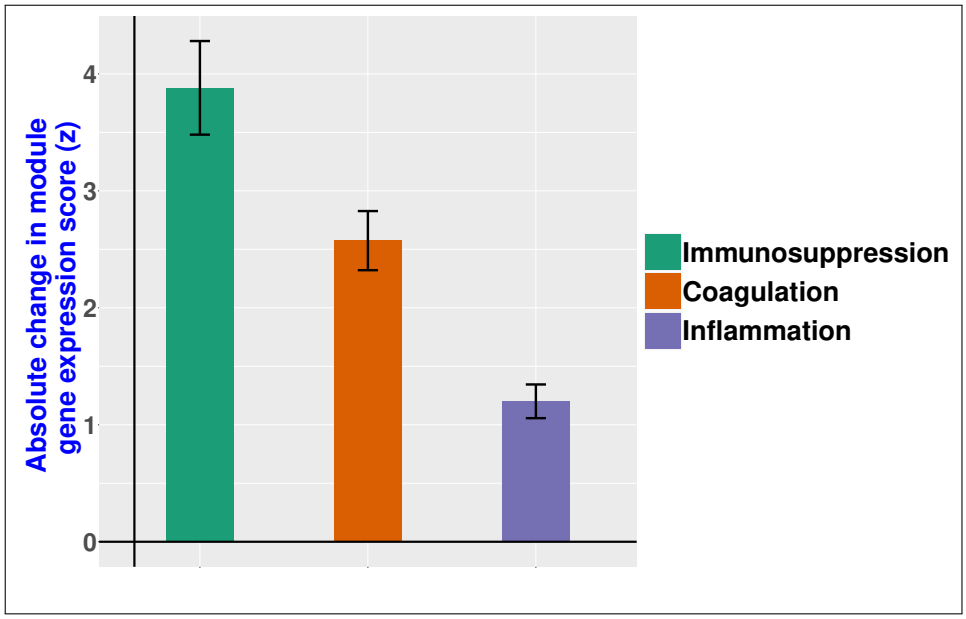

Figure S8. Mean module score in SCB cohort: Barplot shows the magnitude of gene expression of 3 key modules in each sepsis patients. For each module a gene expression score (z) was calculated to capture the relative expression of the genes of the module in the sepsis patients compared to the healthy subjects. Quantitatively, magnitude of change in expression is highest for the immunosuppresion module.

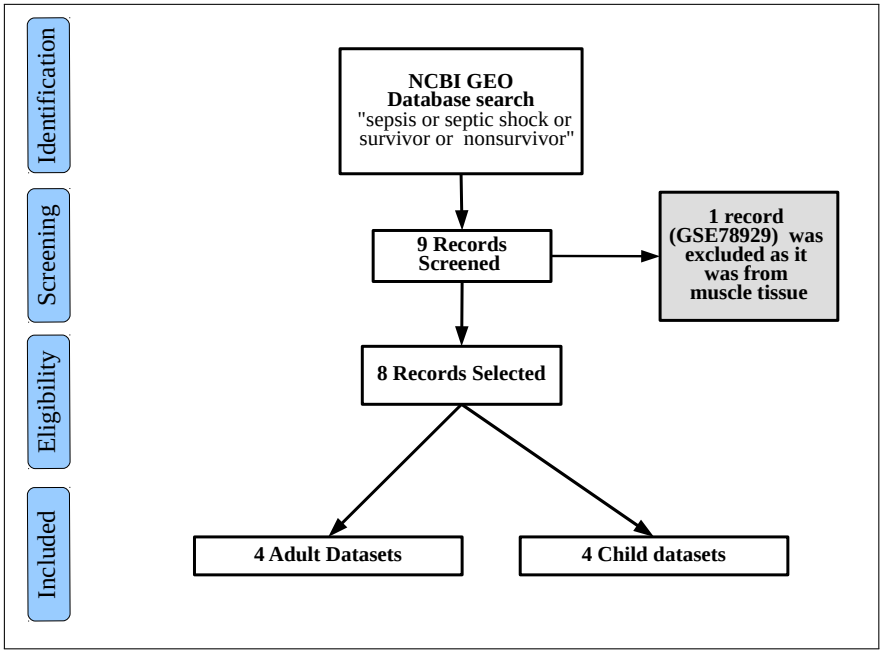

Figure S9. PRISMA flowchart: For selection and screening of the data sets from NCBI GEO.

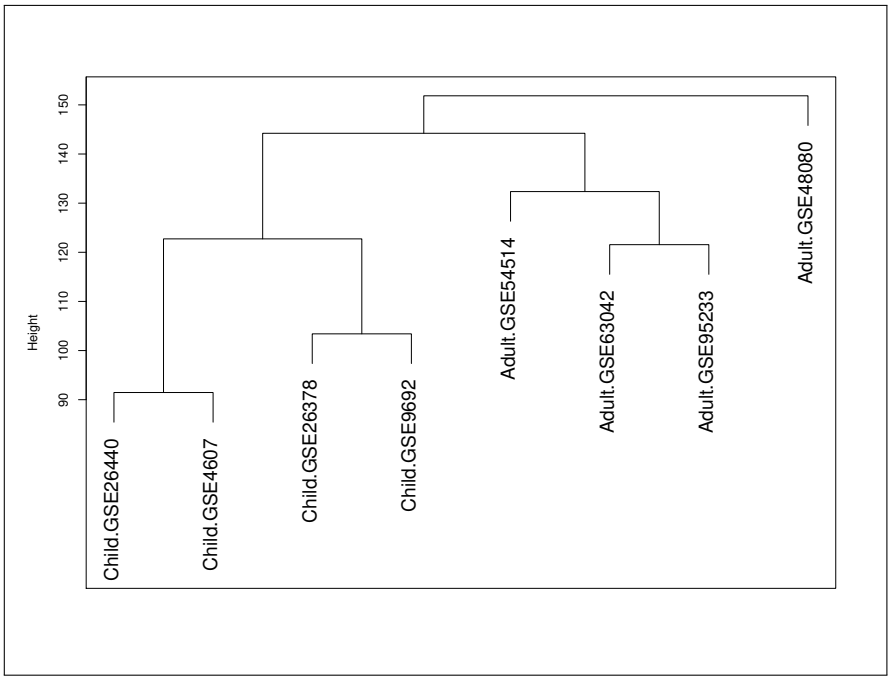

Figure S10. Hierarchical Clustering: Using expression of the common genes in the 8 data sets from GEO. The paediatric data (annotated with prefix "Child") are clearly segregated from the "Adult" studies.

Author_year	Data Source	Disease	Country	# Probes assayed	# DE probes	% change	FDR p threshold	Fold change threshold	Pubmed ID
Parnell_2013	GSE54514	Sepsis	Australia	48804	3677	7.53	0.05	NA	23807251
Severino_2014	GSE48080	Sepsis	Brazil	19596	151	0.77	0.05	1.7	24667684
Wong_2007	GSE4607	Septic shock	USA	54675	2482	4.54	0.05	NA	17374846
Wong_2009	GSE13904	Septic shock	USA	54675	1867	3.41	0.001	2	17374846
Wong_2010	NA	Septic shock	USA	54675	2355	4.31	0.05	1.5	20009785
Calvano_2005	GlueGrant	Endotoxin Shock	USA	44000	5093	11.58	0.1	NA	16136080
Xiao_2011	GSE11375	Critical Injury	USA	47000	5136	10.93	0.001	2	22110166
This manuscript	NA	Severe sepsis / septic shock	India	17513	4221	24.10	0.05	NA	NA
Re-analysis *	GSE4607	Septic shock	USA	22017	6179	28.1 %	0.05	NA	NA
	GSE13904	Septic shock	USA	22017	7420	33.7 %	0.05	NA	NA
	GSE26378	Septic shock	USA	22017	8143	37 %	0.05	NA	NA
	GSE26440	Septic shock	USA	22017	10014	45.5 %	0.05	NA	NA
	GSE8121	Septic shock	USA	22017	6517	29.6 %	0.05	NA	NA
	GSE9692	Septic shock	USA	22024	12202	55.4 %	0.05	NA	NA
	GSE95233	Septic shock	France	22017	11580	52.6 %	0.05	NA	NA

Table S1. Genomic change in available SS reports: This table contains genomic change in available reports in SS and our results of re-analysis of data downloaded from NCBI GEO. Details are available in Materials and Methods.

	Clinical features
Inclusion criteria	**Any two of the following clinical parameters is present in patient** – Body temperature >38 °C or <36 °C – Heart rate >90 beats per minute – Respiratory rate >20 breaths/min or arterial $PaCO_2$ <32 mmHg – WBC count >12 x 10^9 per microlitre or <4 x 10^9 per microlitre.
Exclusion criteria	Viral or fugal infection Chronic inflammatory diseases such as asthama, rheumatoid arthritis etc. Nosocomial (hospital acquired) infections

Table S2. Inclusion and exclusion criteria : Clinical and laboratory features for recruitment of sepsis patients for the study

Genes	Function
HLA-DPB1	Forms a part of MHC class II beta chain
HLA-DRB1	Forms a part of MHC class II beta chain
HLA-DRB5	Forms a part of MHC class II beta chain
TAP1	Involved in transport of MHC class II protein
TAP2	Involved in transport of MHC class II protein
CD74	Involved in the formation and transport of MHC class II protein

Table S3. Annoatated table for MHCII genes with their function.

GSE ID	PMID	Data Normalization	Group	Cell type	Study Design	Country	Total(n)	Survivor+Non-survivor (first time-point)	Number of Survivors	Number of Nonsurvivors	Mean Age(yr)	Clinical setting	Age Group
GSE26378	21738952	RMA	Septic shock	Whole blood	Prospective	USA	103	82	70	12	<10 years	PICU	Child
GSE26440	19624809, 21738952	RMA	Septic shock	Whole blood	Prospective	USA	130	98	81	17	<10 years	PICU	Child
GSE4607	17374846, 18460642	RMA	Septic shock	Whole blood	Prospective	USA	123	42	33	9	<10 years	PICU	Child
GSE9692	18460642	RMA	Septic shock	Whole blood	Prospective	USA	45	10	24	6	<10 years	PICU	Child
GSE48080	24667684, 26484123	RMA	Septic shock	peripheral mono-nuclear cells	Prospective	Brazil	10	10	5	5	>18 years	ICU	Adult
GSE54514	23807251	RMA	Sepsis	Whole blood	Prospective	Australia	53	35	26	9	>18 years	ICU	Adult
GSE63042	25538794	Reads Per Kilobase of transcript, per Million mapped reads (RPKM)	Septic shock, Severe sepsis, Sepsis	Whole blood	Prospective	USA	106	82	54	28	>18 years	ICU	Adult
GSE95233	28341250	RMA	Septic shock	Whole blood	Prospective	France	73	51	34	17	>18 years	ICU	Adult

Table S4. Study characteristics of all the datasets selected for survival meta-analysis.

Vignette: Immunosuppression, rather than inflammation, is a salient feature of sepsis in an Indian cohort

Samanwoy Mukhopadhyay[*] *and Saroj Kant Mohapatra*[†]

21 August, 2019

Contents

1 General summary

Sepsis is one of the leading diseases associated with high mortality and morbidity, due to the systemic nature of this illness, transcriptomic technology is particularly suited to investigation of molecular underpinning of survival from sepsis episodes. We adopted an analysis approach that combined published transcriptome data and data generated in our laboratory from Indian sepsis patients leading to the discovery of key immune pathways to be altered in non-survivors compared to survivors. This is the first clinical transcriptomic study on sepsis from India, showing that non-survival is associated with down-regulated adaptive immune pathways and significant M2-specific immune-suppression, possibly regulated by NF-κB signalling. Three Biological processes related to sepsis were observed to be significantly altered in non-survivors. A patient-specific analysis reveals up-regulation of coagulation and inflammation but a strong down-regulation of immunosuppression modules in Indian sepsis patients.

[*]sm3@nibmg.ac.in
[†]skm1@nibmg.ac.in

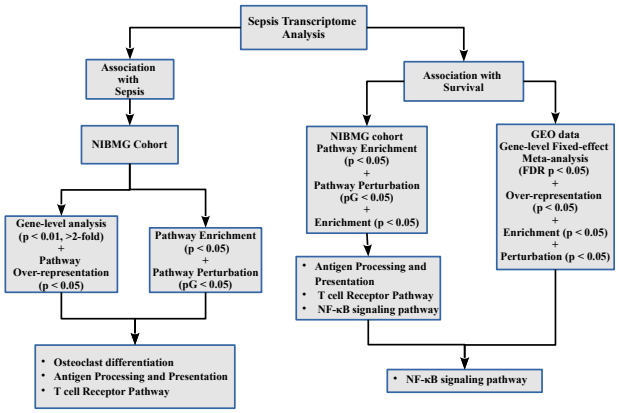

Figure 1: Analysis flow with results: The left arm describes the steps for case-vs-control analysis whereas the right arm describes the steps for identification of pathways associated with survival

2 Getting the data

Availability of data and materials Data and R code are available at the following link (https://figshare.com/) (search for the project **ssnibmgsurv**). The data are accessed through the two data packages listed below. The code can be downloaded as a single zip file (**ssnibmgsurvdoc.zip**). Upon uncomperssing the zip, install the two data packages as described below and run the subsequent code to generate the appropriate output.

2.1 Installation of the data package ssnibmgsurv

Dowload the file **ssnibmgsurv_1.0.tar.gz** from https://figshare.com/ (search for **ssnibmgsurv**). Change the directory to where you saved the file. Start R. At the R prompt, issue the following command:

```
install.packages(pkgs="ssnibmgsurv_1.0.tar.gz", repos=NULL)
# Now the data package ssnibmgsurv is installed on your computer.
# Check with the following command:
library("ssnibmgsurv")
```

2.2 Installation of the data package ssgeosurv

Dowload the file **ssgeosurv_1.0.tar.gz** from https://figshare.com/ (search for **ssgeosurv**). Change the directory to where you saved the file. Start R. At the R prompt, issue the following command:

```
install.packages(pkgs="ssgeosurv_1.0.tar.gz", repos=NULL)
# Now the data package ssgeosurv is installed on your computer.
# Check with the following command:
library("ssgeosurv")
```

Now both the data packages are installed in your computer; let's focus on starting the analysis.

2.3 Running the analysis: Setting working directory

It is assumed that you have access to a folder **ssnibmgsurvdoc**. Start R and set the working directory to **ssnibmgsurvdoc**. Run the code chunks as follow.

3 Case control analysis: Identifying genes and pathways DE in sepsis

```
# Clearing the workspace and close any graphics window if open
rm(list=ls())
graphics.off()
# Loading the preliminary libraries including data packages
source("Rcode/prelim.R")
```

```
source("Rcode/getData.R")
# Using age- and gender-matched controls
matched.12 <- c("C11","C8","C1","C7","C17","C10","C21","C18","C20","C4","C9",
                "C12","42D1","1D1","8D1","50D1","60D1", "90D1","62D1","70D1",
                "19D1","32D1","14D1","61D1")
esetm <- eset[, matched.12]
rttm <- rowttests(esetm, "Group")
lfcdm<-rttm$dm
pdm <- p.adjust(rttm$p.value, method="BH")
names(pdm) <- rownames(rttm)
names(lfcdm) <- rownames(rttm)
egs.all <- featureNames(esetm)
# Removing some variables that are not to be used for further analysis
rm(snames, ptids)
```

```
# Draw a volcano plot to show that 24% of the genome
# are perturbed in sepsis (FDR p < 0.05)
# sepsis and matched control samples are used here
source("Rcode/makeplot_1_volcano_for_genomic_storm.R")
```

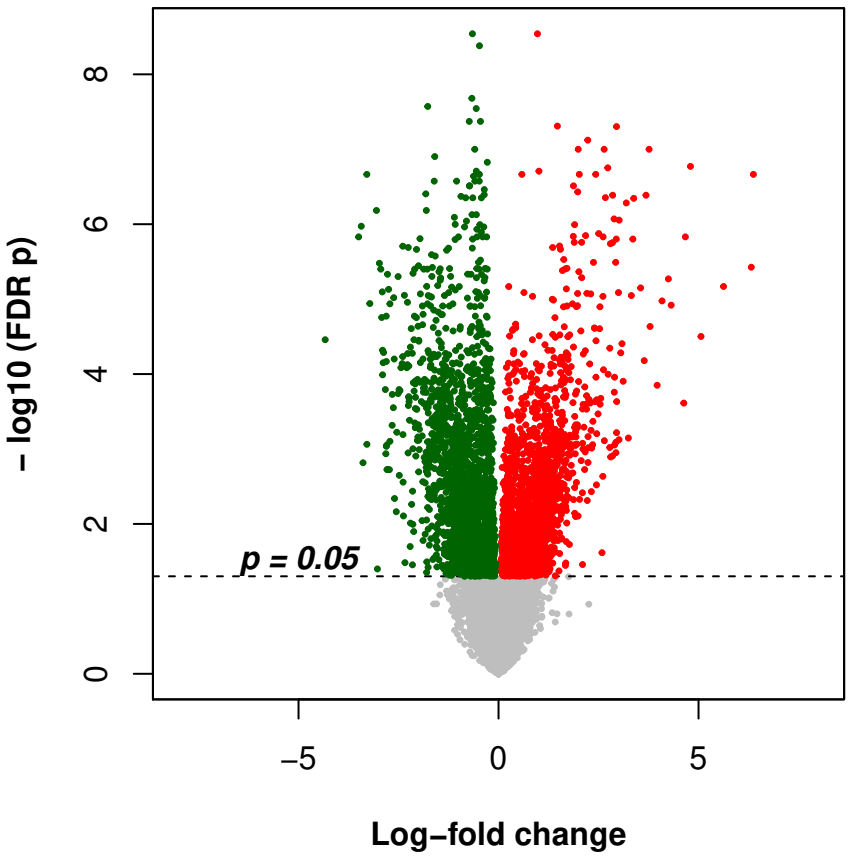

Figure 2: Volcano plot showing 24 percent of the genome perturbed in sepsis compared to healthy control (FDR p < 0.05). This establishes large scale change in gene expression in sepsis, and possible multiple pathways being perturbed.

3.1 Genome-level changes in gene expression

Differential gene expression analysis revealed 1109 genes to be altered in sepsis patients compared to age- and gender-matched healthy controls. Volcano plot (Figure 2) showed 24% of the genome perturbed in sepsis compared to healthy control (FDR p < 0.05). This establishes large scale change in gene expression in sepsis, and possible multiple pathways being perturbed.

```
# Detect the highly significant DE genes of sepsis - FDR p < 0.01; 2-fold
# Draw box-plot to show temporal changes in control vs cases
upg <- egs.all[which(pdm<0.01 & lfcdm>1)]
downg <- egs.all[which(pdm<0.01 & lfcdm<(-1))]
deg.d1<-union(upg,downg)

# Line plot showing slow return or non-survivors to baseline gene expression
source("Rcode/makeplot_2_temporal_plot_for_DEgns.R")
```

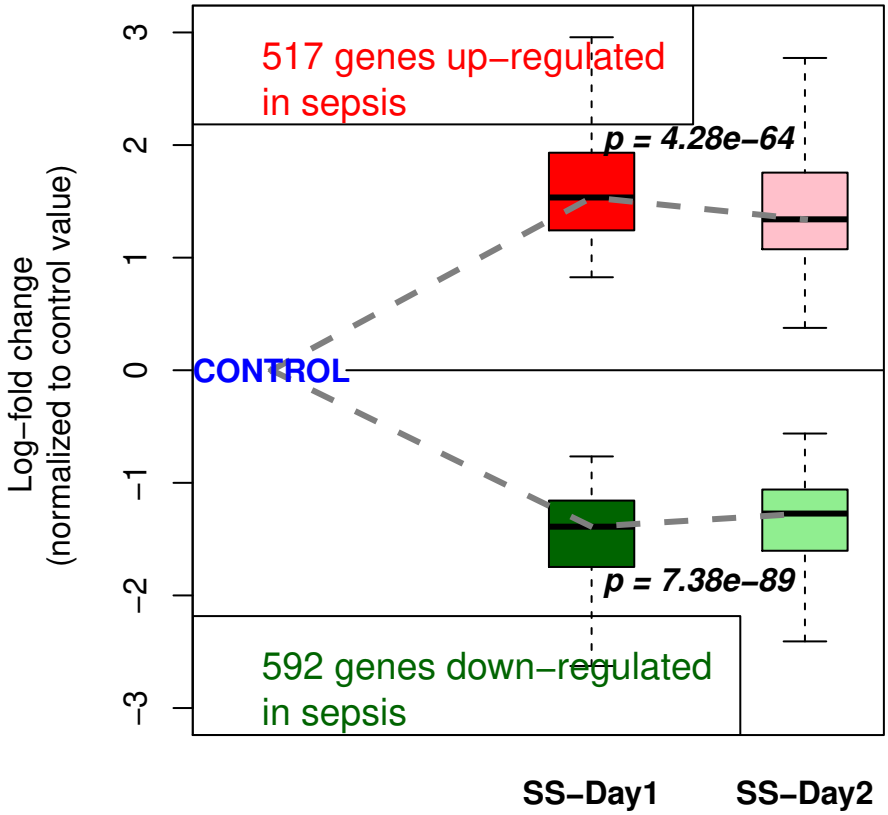

Figure 3: Temporal change of DE genes (FDR $p < 0.05$, 2 fold-change or more), there is a non-random trend toward the baseline with time (p-values from paired t-tests are provided in the legend). This is consistent with earlier findings from patients with trauma.

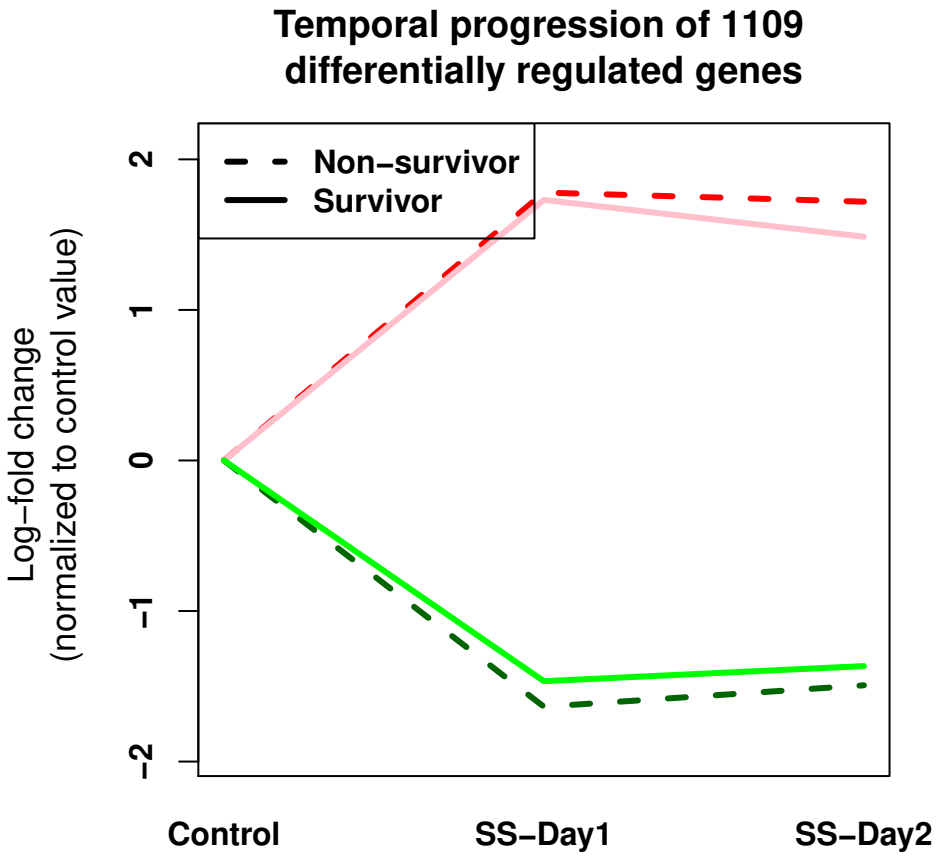

Figure 4: Temporal change of DE genes (FDR $p < 0.05$, 2 fold-change or more), there is a non-random trend toward the baseline with time (p-values from paired t-tests are provided in the legend). The delayed return to baseline is associated with non-recovery from sepsis.

```
# draw trajectory of DE genes survivor versus non-survivor
source("Rcode/drawDEtrajectorySurvival.R")
```

3.2 Pathway analysis results

```
# Pathway Analysis: ORA, GSEA, SPIA
# ORA - Over representation analysis
pORAup <- getORApvals(upg, egs.all)[,"p"]
pORAup <- p.adjust(pORAup, method="BH")
pORAdown <- getORApvals(downg, egs.all)[,"p"]
pORAdown <- p.adjust(pORAdown, method="BH")

NIBMG.ORA.disease <- cbind(pORAup, pORAdown)
colnames(NIBMG.ORA.disease) <- paste(colnames(NIBMG.ORA.disease),
                                     "NIBMG.disease", sep="_")

# GSEA - Gene Set Enrichment Analysis
# SPIA - Signaling Pathway Impact Analysis
# Warning running this code will take long time (approx 10 minutes each)
source("Rcode/run_GSEA_sepsis_vs_control.R") # GSEA
```

```
##
##  Loading GSEA permutation t.test result from file ...
##   done!
```

```
source("Rcode/run_SPIA_sepsis_vs_control.R") # SPIA
```

```
##  Loading SPIA
##  result from file ...  done!
```

```
# Combine result from 3 pathway analyses and print the Down and Up pathways#
pathsDown = intersect(intersect(names(which(pGSEAdown < 0.01)),
                                names(which(pORAdown < 0.01))), names(which(pG< 0.01)))
pathways.list[ paste0("path:", pathsDown)]
```

```
##                                                 path:hsa03013
##                       "RNA transport - Homo sapiens (human)"
##                                                 path:hsa04612
## "Antigen processing and presentation - Homo sapiens (human)"
##                                                 path:hsa04660
##   "T cell receptor signaling pathway - Homo sapiens (human)"
##                                                 path:hsa05332
##           "Graft-versus-host disease - Homo sapiens (human)"
```

```
# Removing some variables that are not to be used for further analysis
rm(pathsDown)

pathsUp = intersect(intersect(names(which(pGSEAup < 0.01)),
                              names(which(pORAup < 0.01))), names(which(pG< 0.01)))
pathways.list[ paste0("path:", pathsUp)]
```

```
##                                                   path:hsa04380
##             "Osteoclast differentiation - Homo sapiens (human)"
##                                                   path:hsa05133
##                               "Pertussis - Homo sapiens (human)"
##                                                   path:hsa05150
##         "Staphylococcus aureus infection - Homo sapiens (human)"
##                                                   path:hsa05202
## "Transcriptional misregulation in cancer - Homo sapiens (human)"
```

```
##                                                    path:hsa05322
##            "Systemic lupus erythematosus - Homo sapiens (human)"
```

```
# Removing some variables that are not to be used for further analysis
rm(pathsUp)
```

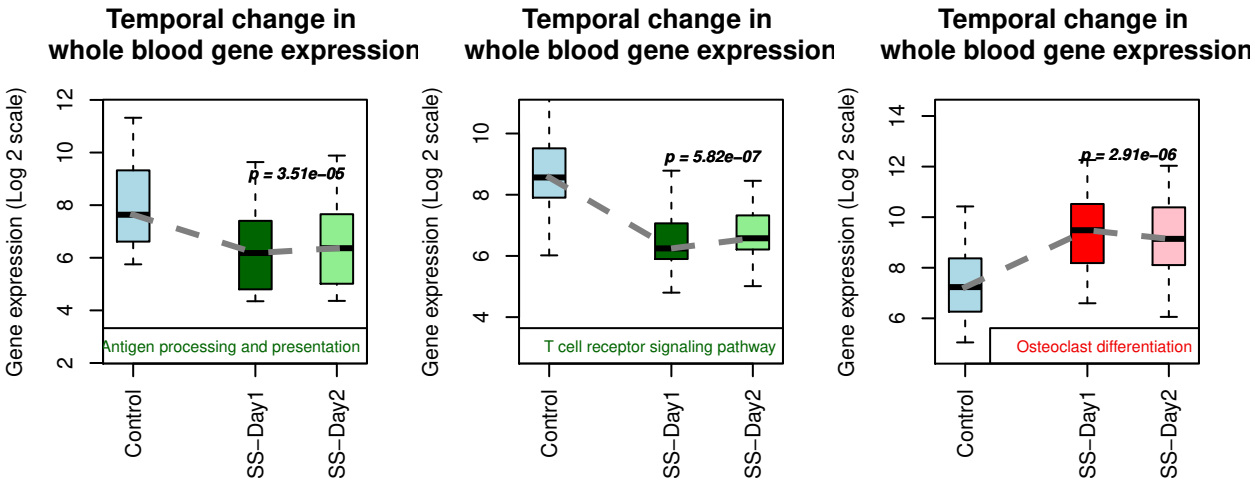

Figure 5: Temporal boxplot of 3 key pathwayes altered in Sepsis by combined pathway analysis.

```
par(mfrow=c(1,3))
# Box plot of two pathways down-regulated in sepsis
box.plot.KEGG(id="hsa04612", direction="down") # Antigen processing and presentation
box.plot.KEGG(id="hsa04660", direction="down") # T cell receptor signaling
# Box plot up-regulated pathway: Osteoclast Differentiation
box.plot.KEGG(id="hsa04380", direction="up")
```

4 Survival analysis

4.1 Survival analysis using published transcriptome data (from NCBI GEO)

```
# Section B: Survival analysis
# Getting genes and pathways associated with survival
# uses NIBMG and published data sets
#####################################################
rm(list=ls())
graphics.off()

#############################################################
# Preliminries
source("Rcode/prelim.R")
```

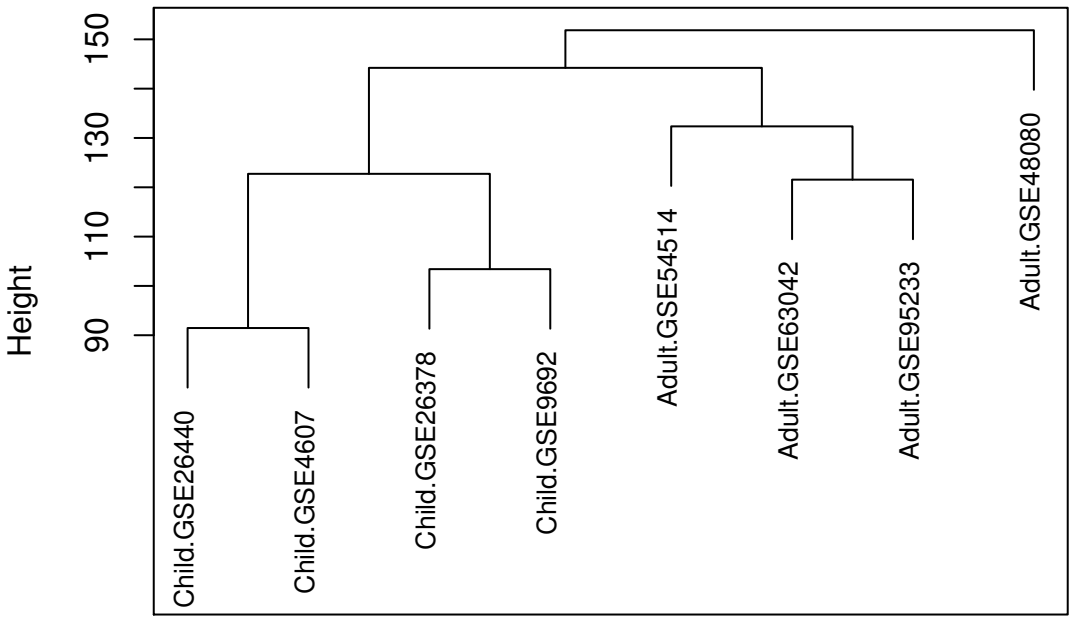

Figure 6: Survival analysis with eight published data sets: human adults and children with sepsis; hierarchical clustering with log-fold change in gene expression led to evidence of developmental age-specific differential perturbation; i.e., separate clusters for adult and child data sets. In view of this difference, further analysis was confined to to Adult data sets when combined with NIBMG data.

```
############################
# Get the expression set
source("Rcode/getData.R")
library(ssgeosurv)
data(ss.list) # eight data sets = 4 adult + 4 child
data(ss.surv.list) # 8 datasets with survivors and non-survivors
studies = read.table(file="metadata/studies.txt", header=TRUE, sep="\t")
study.ids = as.character(studies$study.id)
study.type = as.character(studies$age)
names(study.type) = study.ids

########################################
# Day 1 non-survivor vs survivor with FDR p cutoff 0.01
rtt1 <-rowttests(eset.s, factor(eset.s$Outcome))
sel1 <- rownames(rtt1)[which(rtt1$p.value < 0.01)]
sel1.nibmg.lfc <- rtt1[sel1,]

# Hierarchical clustering of log-fold change of 8 studies
source("Rcode/run_hclust.R")
box()
```

```
# Pathway analysis in GEO data
# ORA: over-representation Analysis
source("Rcode/run_ORA_surv_nonsurv_analysis.R")
# Perform permutation-based GSEA analysis
source("Rcode/run_GSEA_surv_nonsurv_analysis.R")
```

```
## Loading gsea_child data from file ... done!
## File exists ...
## Done!...
## Loading gsea_adult data from file ... done!
## File exists ...
## Done!...
```

```
##############################
# Two-way evidence plot adult
source("Rcode/run_SPIA_surv_nonsurv_analysis_adult.R")
```

```
res <- spia.res[[1]]
resall.adult <-  data.frame(rep(names(ss.adult)[1],  nrow(res)),
                            res$Name, res$ID, res$NDE,
                            res$pNDE, res$tA, res$pPERT,
                            res$pG, res$pGFdr, res$Status)

col.nm <- as.character(sapply(strsplit(colnames(resall.adult), "res."), "[[", 2))
col.nm[1] <- "Study"

colnames(resall.adult)<- col.nm

for(i in 2:length(ss.adult)) {
  res <- spia.res[[i]]
  resedited <- data.frame(rep(names(ss.adult)[i],
                          nrow(res)), res$Name, res$ID, res$NDE, res$pNDE,
                          res$tA, res$pPERT, res$pG, res$pGFdr, res$Status)

  colnames(resedited)<- col.nm
  resall.adult <- rbind(resall.adult, resedited, deparse.level=0)
}

############################################################
# Calculate Fisher's product of p-values of pertabation for all pathways
keggs <- as.character(unique(resall.adult$ID))

# Create empty vector for capturing fisher product of pG
pPERT.Fp.adult <- vector(mode="numeric", length=length(keggs))
names(pPERT.Fp.adult) <- keggs

for(id in keggs) {
  pvec <- resall.adult[resall.adult$ID==id, "pPERT"]
  pPERT.Fp.adult[id] <- Fisher.test(pvec)["p.value"]

}
rm(pvec)

# Combine by Fisher product the two p values
# for perturbation (pb) and hypergeometric test (ph)
pb <- pPERT.Fp.adult
pb <- p.adjust(pb, "fdr")
ph <- as.numeric(pNDE.paths.ad[paste("hsa",names(pb),sep=""),"p"])
names(ph) <- names(pb)

# Use a floor value for p
ph[ph < 1e-07] <- 1e-07
pb[pb < 1e-07] <- 1e-07

pGmeta.adult <- combfunc(pb,ph, "fisher")

# Capture the fisher product of pPERT pG Meta p values into a dataframe
fisher.prod.spia.adult <- as.data.frame(cbind(paste("hsa",
                                                    names(pPERT.Fp.adult), sep=""),
                                              as.numeric(ph), as.numeric(pPERT.Fp.adult),
```

```
                                                as.numeric(pGmeta.adult)))
colnames(fisher.prod.spia.adult) <- c("paths", "pNDE.adult",
                                      "pPERT.adult", "pG.meta.adult")

######################
# Following code is derived from SPIA::plotP
# The pG threshold is the p-value 0.05 corrected for the number of
# pathways being considered
tr= 0.05
#tr<-  0.05/length(pb)

# plot neg.log.p_PERT against neg.log.p_NDE
plot(-log(ph), -log(pb), col="gray80",
     xlim = c(0, max(c(-log(ph), -log(pb)) +1, na.rm = TRUE)),
     ylim = c(0, max(c(-log(ph), -log(pb) +1), na.rm = TRUE)),
     pch = 19, main = "Two-way evidence plot : Adult Sepsis", cex = 1.5,
     xlab = "Evidence of Over-representation, -log(p_ORA)",
     ylab = "Evidence of Perturbation, -log(p_PERT)")

# For selected  pathways for visualisation: NLR, NFkB, Osteoclast
###################################
sel.paths.ad <- c("04621", "04064", "04380")
col.vec <- c("red2", "purple2", "darkblue")
points(-log(ph)[sel.paths.ad ], -log(pb)[sel.paths.ad ], pch = 19, col = col.vec,
       cex = 1.5)
abline(v = -log(tr), lwd = 1, col = "red", lty = 2)
abline(h = -log(tr), lwd = 1, col = "red", lty = 2)
path.nms <- as.character(sapply(strsplit(
  pathways.list[paste("path:hsa", sel.paths.ad, sep="")],
  " -"), "[[", 1))
# Add a legend to the plot
legend("topright",  title="Pathway Names",
       legend= paste(sel.paths.ad , path.nms, sep="  "),
       text.font=2, text.col = col.vec, horiz=F,
       cex=0.8, pch=20, col= col.vec)
```

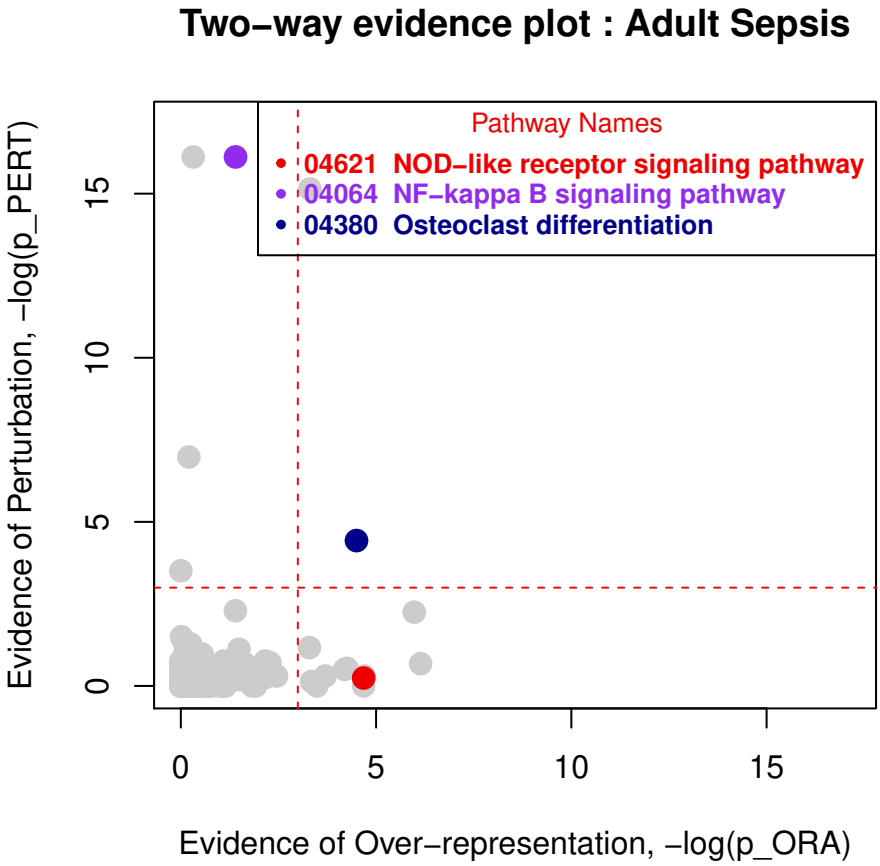

Figure 7: KEGG pathways associated with survival in GEO data; NF-kappaB signalling pathway, Osteoclast differentiation and NOD-like receptor signalling pathway.

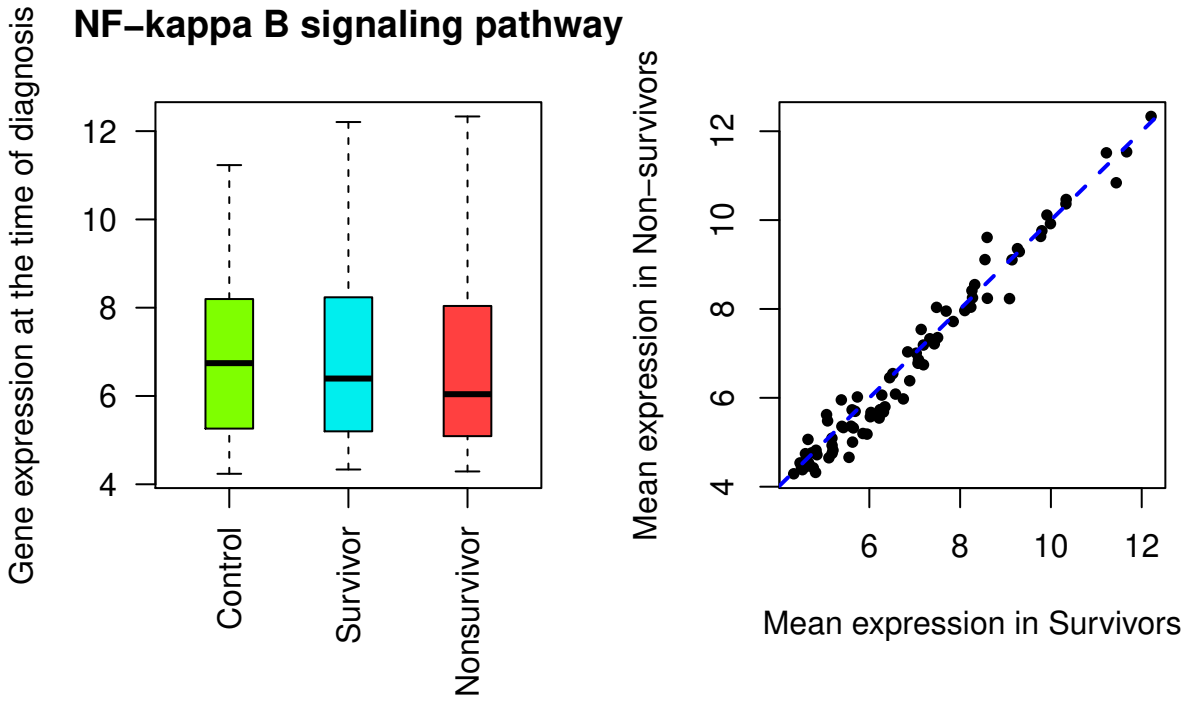

Figure 8: NF-kappa B signalling pathway boxplot and scatterplot

4.2 Down-regulation of NF-κ B signalling pathway genes in non-survivors

```
rm(list=ls())
#############################################################
# Preliminries
source("Rcode/prelim.R")
```

```
# Get the expression set
source("Rcode/getData.R")

# show gene expression trend for NF-kB signaling pathway
getPval(keggid="hsa04064", drawPlot=TRUE, getSigGenes=TRUE)
```

```
##  [1] "TNFRSF13C" "CSNK2A2"   "ICAM1"
##  [4] "LCK"       "BCL2A1"    "RELB"
##  [7] "BCL2L1"    "TRAF5"     "BCL10"
## [10] "CD40"
```

```
# Draw the permutation histogram of GSEA for NF-kB signaling pathway
drawPermutHist(keggid="hsa04064", eset=eset.s, fac=factor(as.character(eset.s$Outcome)))
```

```
##   p.val    zobs
##  0.0441 -3.8605
```

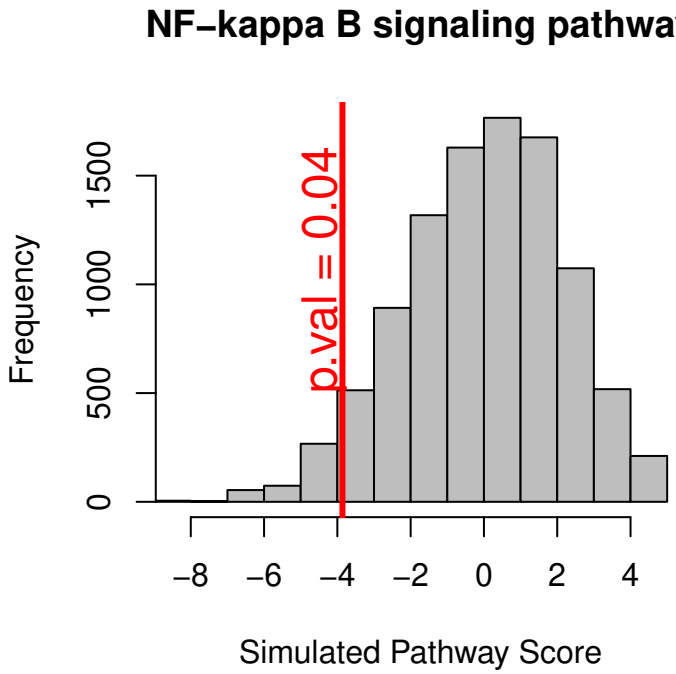

Figure 9: NF-kappa B signalling pathway histogram. Permutation bases Gene set enrichment analysis creats a histogram of simulated pathway scores (in gray bars). The red line shows the observed pathway score in nonsurvivors when compared to survivor.

4.3 Relative gene expression of the targets of NF-κB

```
##############################################################################
# drawing a histogram for NFkB targets
##############################################################################
source("Rcode/plotDensityNFkbTargets.R")
```

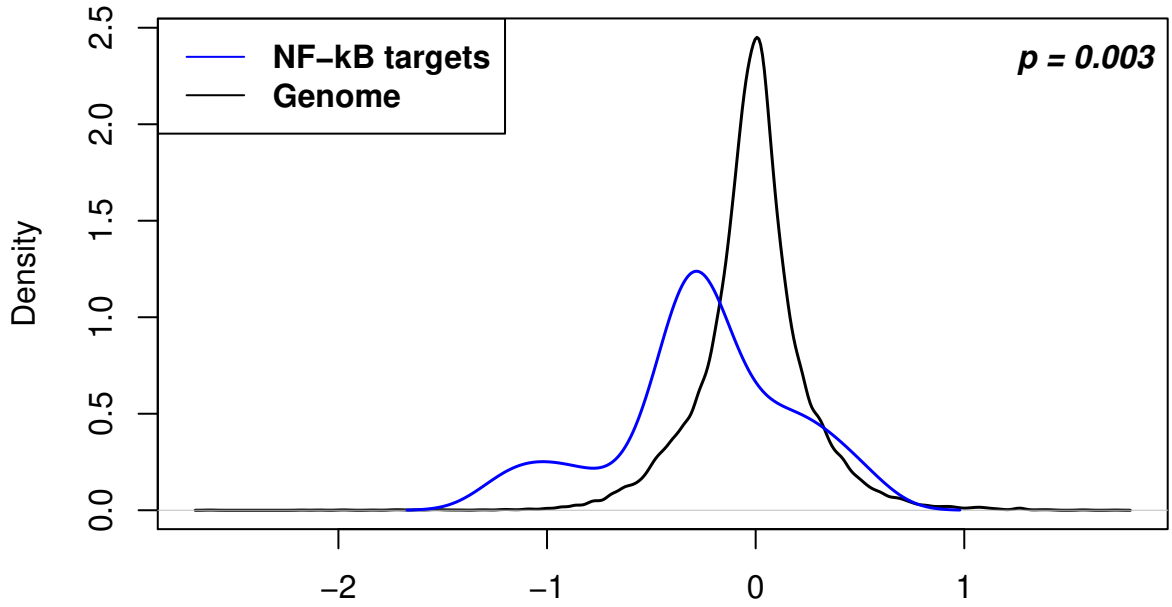

Figure 10: Relative gene expression of the targets of NFkB (i.e., Antigen processing and presentation genes and Immune receptor Genes). The gray peak in the background representsdistribution of all genes in the genome. There is significant down-regulation of the targets in the non-survivors (blue line).

4.4 M2 macrophage-specific down-regulation of gene expression in non-survivors

```
# check the expression of M1 vs M2 markers in data from SCB cohort
# read the file containing gene IDs
###############
sel.gns.dat <- read.table("metadata/M1_M2_markers.txt", sep="\t", header=T)
m1.gns <- sel.gns.dat[,1]
m1.gns <- intersect(m1.gns, featureNames(eset))
m2.gns <- sel.gns.dat[,2]
m2.gns <- intersect(m2.gns, featureNames(eset))

# function for plotting macrophage-specific gene expression
plotMgexp = function(type="M1", normalizeByControl=FALSE) {
  if(type=="M1") {
    egs = m1.gns
  } else {
    egs = m2.gns
  }

  # M1 gene expression
  gexp.s = rowMeans(exprs(eset[egs, eset$Outcome=="Surv" & eset$Group=="D1"]))
  gexp.ns = rowMeans(exprs(eset[egs, eset$Outcome=="Nonsurv" & eset$Group=="D1"]))

  if(normalizeByControl==TRUE) {
    gexp.c = rowMeans(exprs(eset[egs, which.ctrl]))
    gexp.s  = gexp.s-gexp.c
    gexp.ns = gexp.ns-gexp.c
  }
```

```
  plot(x = gexp.s, y=gexp.ns, las=1, cex=0.62, col= "blue", pch=16,
       ylab = "Mean expression in non-survivors", xlab = "Mean expression in survivors",
       main=paste0(type, "-specific gene expression"))
  abline(0,1, lty=2, xlim= c(-3, 10), ylim=c(-3, 10))
  gsyms <- as.character(unlist(mget(m1.gns, org.Hs.egSYMBOL)))
  #gsyms[intersect(which(gexp.1[,1] < -1), which(gexp.1[,2] < -1) )] <- ""
  text(x = gexp.s, y=gexp.ns, labels=gsyms, cex= 0.58, pos=2, offset = 0.75, font=2)

  pval <- t.test(gexp.s, gexp.ns, paired = T)$p.value
  legend.str <- paste("p = ", formatC(pval, digits=1), sep="")
  legend("topleft", legend.str, bty="n", text.font=4)
}
```

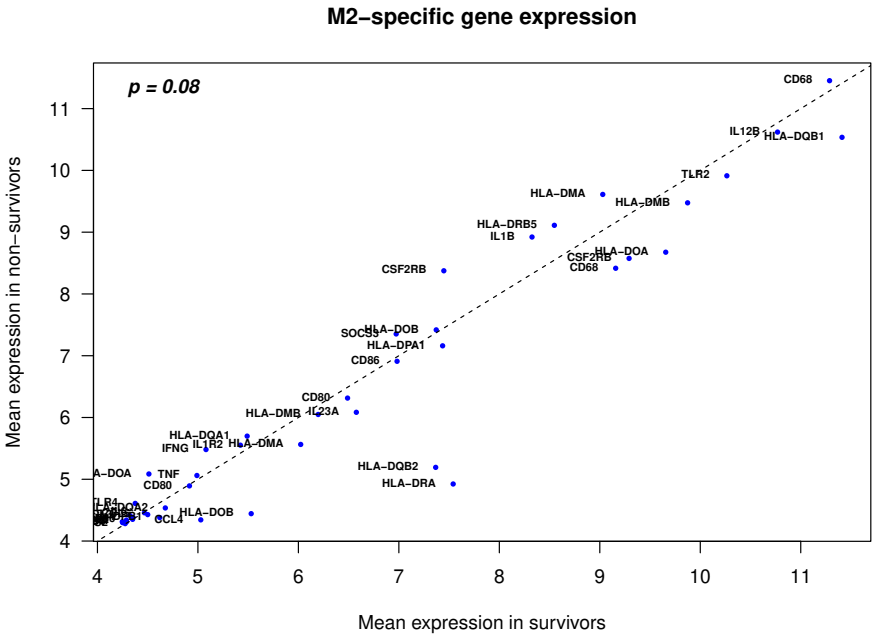

Figure 11: M1 macrophages (classically activated macrophages) are pro-inflammatory, important in host defence against the pathogens, phagocytosis, secretion of pro-inflammatory cytokines and microbicidal molecules. M2 macrophages (alternatively activated macrophages) participate in regulation of resolution of inflammation and repair of damaged tissues. M2-specific under-expression is observed in non-survivors (p = 0.02).

```
plotMgexp("M2")
```

4.5 Down-regulation of Antigen processing and presentation signalling pathway genes in non-survivors

```
#######################################################
# Survivor versus non-survivor
#######################################################
getPval("hsa04612", drawPlot=TRUE, getSigGenes=TRUE) # AgPP
```

```
##  [1] "HLA-DPA1" "HLA-DPB1" "HLA-DRB1"
##  [4] "HLA-DRB5" "HSPA5"    "KIR2DL3"
##  [7] "KIR3DL2"  "CIITA"    "PSME1"
## [10] "TAP1"     "TAP2"     "CD74"
```

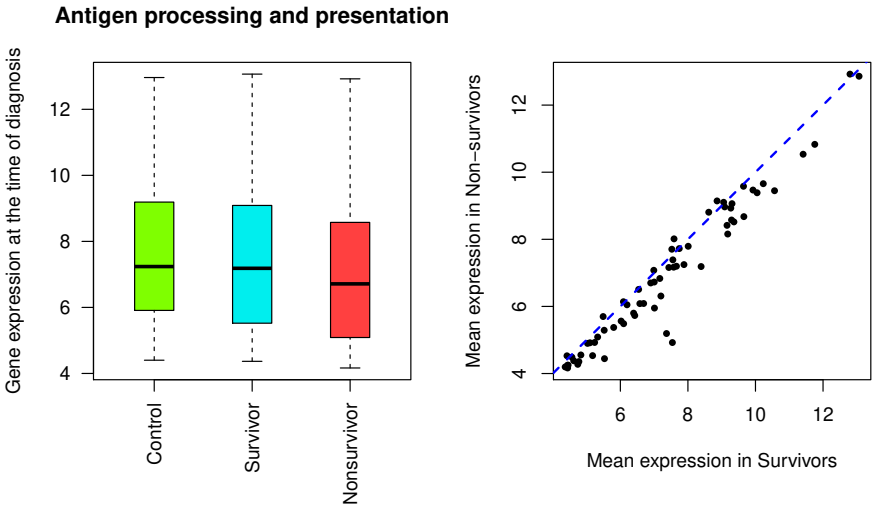

Figure 12: Status of AgPP pathway in NIBMG data with Box/Scatterplot

```
drawPermutHist("hsa04612", eset=eset.s, fac=factor(as.character(eset.s$Outcome)))# AgPP
```

```
##    p.val    zobs
##   0.0154 -7.3492
```

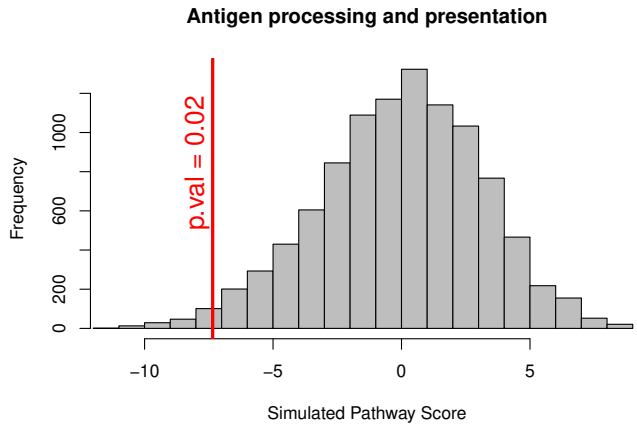

Figure 13: Status of AgPP pathway in NIBMG data, permutation based Gene set enrichment histogram.

4.6 Down-regulation of T cell receptor signalling pathway genes in non-survivors

```
#######################################################
# Survivor versus non-survivor
#######################################################
getPval("hsa04660", drawPlot=TRUE, getSigGenes=TRUE) # TCR
```

```
## [1] "RASGRP1" "ITK"     "LCK"
## [4] "NFKBIE"  "SOS2"    "BCL10"
## [7] "CD3D"    "CD3E"    "CD3G"
```

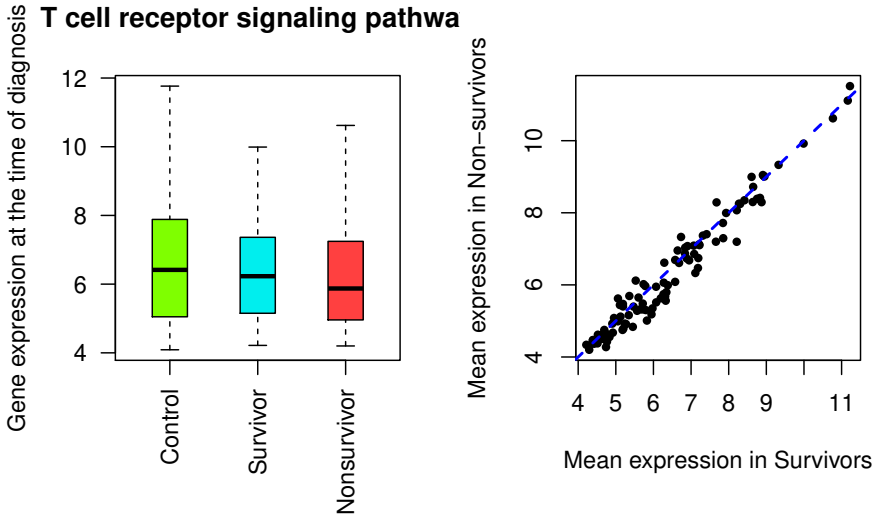

Figure 14: Box plot and scatter plot showing down-regulation of TCR pathway in non-survivors

```
drawPermutHist("hsa04660", eset=eset.s, fac=factor(as.character(eset.s$Outcome)))# TCR
```

```
##  p.val   zobs
##  0.114 -2.671
```

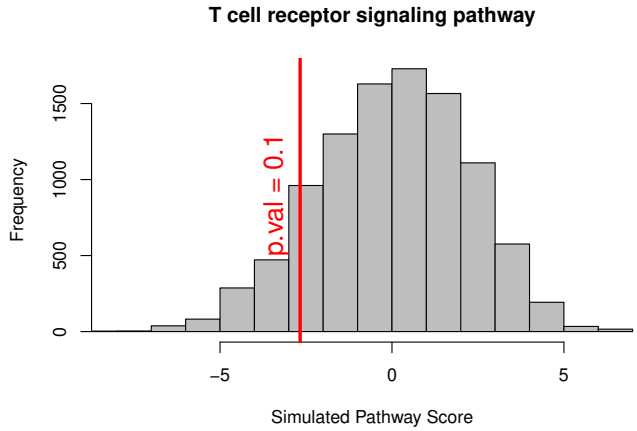

Figure 15: Down-regulation of TCR pathway in SCB cohort, with histogram from permutation based Gene set enrichment test.

```
net.dat <- read.table("metadata//network_input.csv", sep="\t", header=T)
net.list = with(net.dat, split(x=Egid, f=Process))

par(mfrow=c(3, 1), mar=c(2, 16, 2, 16))
for(i in 1:3) {
  drawPermutHist(geneids=net.list[[i]],
                 titlestr=names(net.list)[i],
                 fac=eset.s$Outcome,
                 eset=eset.s, cex= 0.80)
  box()
}
```

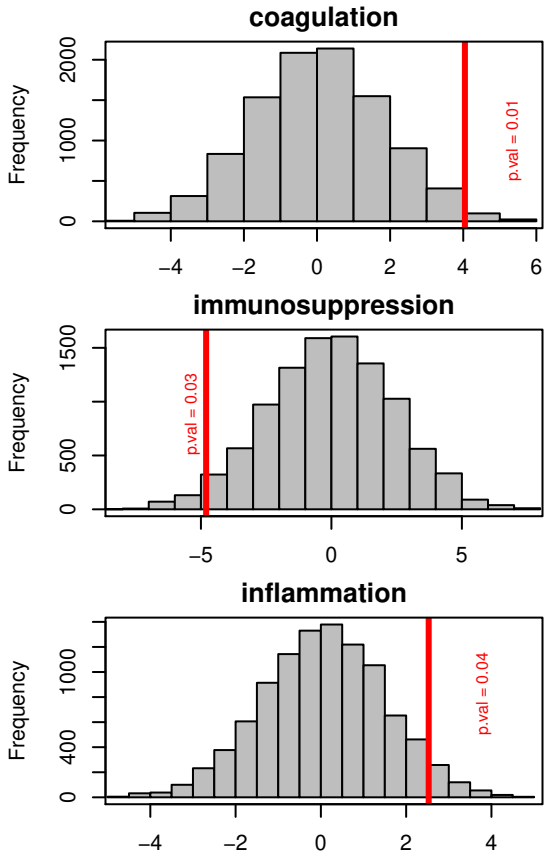

Figure 16: Three Biological processes found to be differentially enriched in Nonsurvivors. Refer to the main manuscript for more details.

5 Module scores in SCB cohort

```
# get scores for the three modules from NIBMG data
fn = "Results/nibmgModuleScores.rda"
if(file.exists(fn)) {
  cat("Reading NIBMG module scores from file ...")
  load(fn)
  cat(" done!\n")
} else {
  sids = sampleNames(eset)[-c(which.ctrl)]
  ids.con = sampleNames(eset)[which.ctrl]
  modScores = sapply(names(net.list), function(id) {
    egs = as.character(net.list[[id]])
    sapply(sids, function(sid) {
      eset.curr = eset[,c(sid, ids.con)]
      eset.curr$Group = factor(as.character(eset.curr$Group))
      rtt = rowttests(eset.curr,"Group")
      rttstats = rtt[egs, "statistic"]
      z = sum(rttstats)/sqrt(length(egs))
      return(z)
    })
  })
  save(modScores, file=fn)
}
```

```
## Reading NIBMG module scores from file ... done!
```

```
# format the data
sids = rownames(modScores)
sids.df = do.call(rbind,strsplit(sids, split="D"))
colnames(sids.df) = c("Pt","Day")
rm(sids)
dat = data.frame(sids.df, sids=rownames(modScores), modScores)

dat.split = with(dat, split(sids, Pt))
dat.split = sapply(dat.split, as.character)

par(mar=c(3,5,2,1))
dat = sapply(dat.split, function(x)
      as.matrix(dat[unlist(x),
      c("coagulation","immunosuppression","inflammation")]))
dat1 = dat
for(i in 1:length(dat)) {dat1[[i]] = rbind(dat[[i]], c(0,0,0))}
o = order(sapply(dat, function(x) min(x)))
dat1 = dat1[o]
plotdat = t(do.call(rbind, dat1))
mycols = c("blue", "darkgreen", "red")
b = barplot(plotdat, width=1, ylim=c(-15, 15), beside = T,
        col=mycols, , ylab="Module score",
        border=mycols, las=2, cex.names=0.9)
legend("top", c("Coagulation", "Immunosuppression", "Inflammation"), horiz=T,
       col=mycols,
       border=mycols,
       pch=15, pt.cex= 1.7, bty="n", inset=c(-0.05, 0))
```

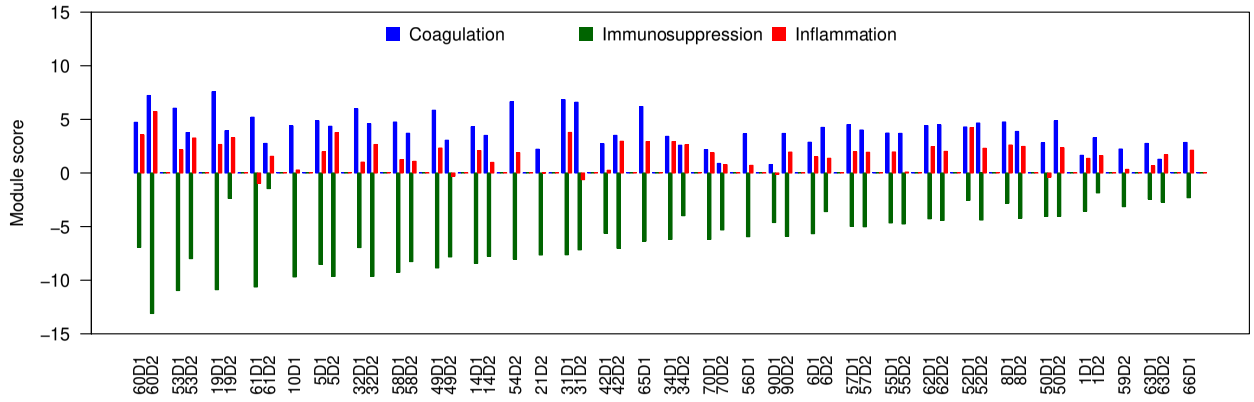

Figure 17: Barplot showing the score of 3 key modules in each sepsis patient

```
box()
```

```
require(ggplot2)
dat <- t(abs(plotdat))
apply(t(abs(dat)), 1, median)
```

```
##       coagulation immunosuppression
##             2.807             3.983
##      inflammation
##             0.768
```

```
df <- dat
sem <- function(x){
  sd(x)/sqrt(length(x))
}
my_mean <- apply(df, 2, mean)
my_sem <- apply(df, 2, sem)
# new data frame for storing the mean and sem
mean_sem_old <- data.frame(means=my_mean, sems=my_sem, group=colnames(df))
mean_sem <- rbind(mean_sem_old[2,], mean_sem_old[1,], mean_sem_old[3,])
rownames(mean_sem) < c("A", "B", "C")
```

```
## [1] FALSE FALSE FALSE
```

```
# larger font
theme_set(theme_gray(base_size = 2))
# factorize the variable for legend
mean_sem$group <- factor(mean_sem$group, levels = mean_sem$group)
ggplot(mean_sem, aes(x=group, y=means, fill=group)) + theme(legend.position="right") +
   geom_bar(stat='identity', width=0.4) +
  geom_errorbar(aes(ymin=means-sems, ymax=means+sems), width=.12) +
    scale_fill_brewer(palette="Dark2")+
    xlab('')  + theme(legend.text=element_text(size=12, face="bold")) +
    theme(legend.title = element_blank())+
    ylab('Absolute module score (z)') +
    theme(plot.title = element_text(color="red", size=18, face="bold"),
          axis.title.x = element_text(color="blue", size=12, face="bold"),
          axis.title.y= element_text(color="blue", size=12, face="bold"))+
  theme(axis.text.x = element_text(size=0.0012)) +
  theme(axis.text.y = element_text(size=11, face="bold")) +
  geom_hline(yintercept=0, color="black", size= 0.46) +
  geom_vline(xintercept=0.61, color="black", size= 0.46)
```

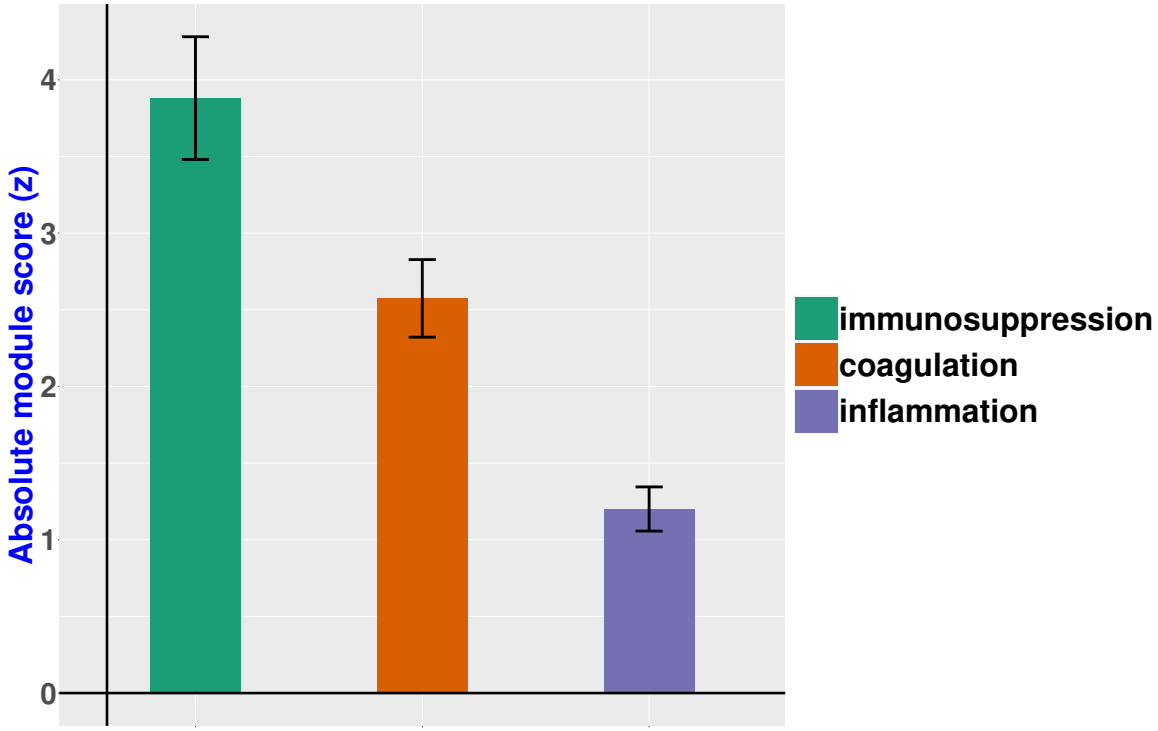

Figure 18: Barplot showing the magnitude of 3 key modules in each sepsis patients. Each bar represents mean of patient-level module score with SEM as the error bar. For all three modules, absolute z-scores have been used. It is clear that there is much greater immunnosuppression compared to inflammtion.

6 Session Information

```
## R version 3.4.4 (2018-03-15)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 18.04.1 LTS
##
## Matrix products: default
## BLAS: /usr/lib/x86_64-linux-gnu/openblas/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/libopenblasp-r0.2.20.so
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8
##  [2] LC_NUMERIC=C
##  [3] LC_TIME=en_IN.UTF-8
##  [4] LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_IN.UTF-8
##  [6] LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_IN.UTF-8
##  [8] LC_NAME=C
##  [9] LC_ADDRESS=C
## [10] LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_IN.UTF-8
## [12] LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats4    parallel  stats
## [4] graphics  grDevices utils
## [7] datasets  methods   base
##
## other attached packages:
##  [1] ssgeosurv_1.0
##  [2] ssnibmgsurv_1.0
##  [3] ggplot2_3.1.0
##  [4] KEGG.db_3.2.3
##  [5] pathview_1.18.2
##  [6] stringr_1.4.0
##  [7] gplots_3.0.1
##  [8] pca3d_0.10
##  [9] rgl_0.99.16
## [10] KEGGREST_1.18.1
## [11] SPIA_2.30.0
## [12] KEGGgraph_1.38.0
## [13] Category_2.44.0
## [14] Matrix_1.2-15
## [15] GSEABase_1.40.1
## [16] graph_1.56.0
## [17] annotate_1.56.2
## [18] XML_3.98-1.19
## [19] illuminaHumanv2.db_1.26.0
## [20] hgu133plus2.db_3.2.3
## [21] org.Hs.eg.db_3.5.0
## [22] AnnotationDbi_1.40.0
## [23] IRanges_2.12.0
## [24] S4Vectors_0.16.0
```

```
## [25] genefilter_1.60.0
## [26] limma_3.34.9
## [27] GEOquery_2.46.15
## [28] Biobase_2.38.0
## [29] BiocGenerics_0.24.0
##
## loaded via a namespace (and not attached):
##  [1] bitops_1.0-6
##  [2] bit64_0.9-7
##  [3] RColorBrewer_1.1-2
##  [4] webshot_0.5.1
##  [5] httr_1.4.0
##  [6] Rgraphviz_2.22.0
##  [7] tools_3.4.4
##  [8] R6_2.4.0
##  [9] KernSmooth_2.23-15
## [10] lazyeval_0.2.1
## [11] colorspace_1.4-0
## [12] DBI_1.0.0
## [13] manipulateWidget_0.10.0
## [14] withr_2.1.2
## [15] tidyselect_0.2.5
## [16] bit_1.1-14
## [17] compiler_3.4.4
## [18] xml2_1.2.0
## [19] labeling_0.3
## [20] caTools_1.17.1.1
## [21] scales_1.0.0
## [22] readr_1.3.1
## [23] RBGL_1.54.0
## [24] digest_0.6.18
## [25] rmarkdown_1.13
## [26] XVector_0.18.0
## [27] pkgconfig_2.0.2
## [28] htmltools_0.3.6
## [29] htmlwidgets_1.3
## [30] rlang_0.3.1
## [31] RSQLite_2.1.1
## [32] shiny_1.2.0
## [33] bindr_0.1.1
## [34] jsonlite_1.6
## [35] crosstalk_1.0.0
## [36] gtools_3.8.1
## [37] dplyr_0.7.8
## [38] RCurl_1.95-4.12
## [39] magrittr_1.5
## [40] Rcpp_1.0.0
## [41] munsell_0.5.0
## [42] stringi_1.4.3
## [43] yaml_2.2.0
## [44] zlibbioc_1.24.0
## [45] plyr_1.8.4
## [46] grid_3.4.4
## [47] blob_1.1.1
```

```
## [48] gdata_2.18.0
## [49] promises_1.0.1
## [50] crayon_1.3.4
## [51] miniUI_0.1.1.1
## [52] lattice_0.20-38
## [53] Biostrings_2.46.0
## [54] splines_3.4.4
## [55] hms_0.4.2
## [56] knitr_1.23
## [57] pillar_1.3.1
## [58] codetools_0.2-16
## [59] glue_1.3.1
## [60] evaluate_0.14
## [61] png_0.1-7
## [62] httpuv_1.4.5.1
## [63] gtable_0.2.0
## [64] purrr_0.2.5
## [65] tidyr_0.8.2
## [66] assertthat_0.2.0
## [67] xfun_0.8
## [68] mime_0.6
## [69] xtable_1.8-3
## [70] later_0.7.5
## [71] survival_2.43-3
## [72] tibble_2.0.1
## [73] memoise_1.1.0
## [74] ellipse_0.4.1
## [75] bindrcpp_0.2.2
```