bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Meta-analysis reveals pathway signature of Septic Shock

Samanwoy Mukhopadhyay,[1] Abhaydeep Pandey,[1] Pravat K Thatoi[2], Bidyut K Das[2], Balachandran Ravindran [3], Samsiddhi Bhattacharjee[1], Saroj K Mohapatra[1*]

[1]National Institute of Biomedical Genomics, 741251 Kalyani, India
Sriram Chandra Bhanja Medical College and Hospital, Manglabag, 753007 Cuttack, India.
[2]Institute of Life Sciences, NALCO Square, 751023 Bhubaneswar, India.

*To whom correspondence should be addressed; E-mail: skm1@nibmg.ac.in

May 4, 2016

Septic shock is a major medical problem with high morbidity and mortality and incompletely understood biology. Availability of genome-wide expression data from different studies on septic shock empowers the quest for hitherto unidentified pathways by integration and meta-analysis of multiple data sets. Electronic search was performed on medical literature and gene expression databases. Selection of studies was based on the organism (human subjects), tissue of origin (circulating leukocytes) and the platform technology (gene expression microarray). Gene-level meta-analysis was conducted on the six selected studies to identify the genes consistently differentially expressed in septic shock. These genes were then subjected to pathway analysis. The identified up-regulated pathway $hsa04380$ **(Osteoclast Differentiation) was validated in an independent cohort of patients. A simplified model was generated showing the major gene-modules dysregulated in SS.**

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Septic shock (SS) is a serious medical condition that claims many lives every year worldwide. Approximately 2% of the patients admitted to the hospital are diagnosed with SS. Of these patients, half are treated in the intensive care unit (ICU), representing 10% of all ICU admissions (*2, 29*). Approximately 40-60% of the SS patients die within 30 days (*2*). The number of cases in the USA exceeds 750,000 per year (*2*), but the incidence of SS is largely unknown in those parts of the world with scarce ICU care. Extrapolating from treated incidence rates in the USA, Adhikari et al. estimated up to 19 million cases worldwide per year (*1*). However, the true incidence is expected to be far higher. Incomplete grasp of SS biology is compounded by the lack of specific drug for treating the condition. With a number of failed clinical trials, there is urgent need for new directions in research (*5*). Genome-wide expression profiling offers a complete picture of the condition and enables identification of genes and pathways of diagnostic, prognostic or therapeutic relevance (*37*).

The purpose of this study was to investigate genome-wide host response to SS by combining the power of multiple studies. Analysis was performed at two levels: genes and gene sets. A "gene set" or pathway consists of a set of functionally related genes, and provides higher-order information about gene expression and valuable insights into the biology of a disease. Accordingly, we have laid emphasis on discovery of pathway(s) enriched among the genes consistently differentially expressed among the multiple data sets from studies on SS.

A biological process involves a group of genes. The principle of enrichment analysis is that if a biological process is abnormal in a given condition, the co-functioning genes should have a higher (enriched) potential to be implicated as a relevant group. Because the analytic conclusion is based on a group of relevant and functionally related genes instead of on individual genes, it increases the likelihood of identifying the biological process(es) pertinent to the disease condition under study. A variety of methods are available for testing differential expression of gene sets. We started with the most popular method that takes a list of differentially expressed genes and tests whether the gene set is over-represented in this list. This was followed up and validated by other methods to discover a pathway that is consistently perturbed in SS.

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Search strategy and selection criteria

We searched the popular online database PubMed with the search string ("Systemic Inflammatory Response Syndrome"[MeSH] OR "septic shock"[MeSH] OR "Shock, Septic"[MeSH] OR "Endotoxemia"[MeSH]) AND ("gene expression profiling"[MeSH] OR "transcriptome"[MeSH] OR "microarray analysis"[MeSH] OR "Oligonucleotide Array Sequence Analysis"[MeSH]) with "Human" filter. Additionally, we searched the following gene expression databases: (1) National Centre for Biotechnology Information Gene Expression Omnibus (GEO) and (2) European Bioinformatics Institute ArrayExpress. All the queries were made on the 30th Aug, 2013. Entries from the gene expression databases were then cross-referenced with publications retrieved from PubMed. Selection of studies was based on the organism (human subjects), tissue of origin (circulating leukocytes) and the platform technology (gene expression microarray). The process of selection of the six studies of SS is described in Fig. 1. Only data sets published as full reports and meeting the previous three criteria were selected (Table 1). These data sets were subjected to systematic analysis, as described below.

Pre-processing of data

Series matrix files (containing normalized gene expression data) for all six studies were downloaded and read with the function *getGEO* of the bioconductor package "GEOquery" (*7*). All data were transformed to logarithmic scale (base 2). The function *nsFilter* of "genefilter" package (*8*) was applied to remove the probe sets with incomplete or duplicate gene annotation, and those carrying low expression signal. Each probe set ID was mapped to the corresponding Entrez gene ID. Genes common to all six studies were included in the analysis. These common genes were extracted and all further analyses were performed only on this list of common genes across all the studies. Some of the studies included samples from cases other than SS in the data set. However, these samples were excluded from analysis. For each study, control samples were retained for comparison with the cases of SS.

Identification of differentially expressed genes

For all genes within each study, we applied a Welch's 2-sample t-test (*36*) to compare expression values in samples of SS with that in samples of control subjects. A single overall p-value was then computed for each gene by meta-analyzing individual study level t-statistic using the following procedure.

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We first transformed each individual t-statistic to a Z-score Z_i retaining its sign by using the quantile transformation:

$$Z_i = \Phi^{-1}(F_{t_{n_i^{SS}+n_i^{CON}-2}}(t_i)) \tag{1}$$

We then followed a fixed effects meta-analysis (*11*) approach, where the Z-scores are combined using an optimal linear combination with weights equal to square root of the effective sample size of each study (i.e. the harmonic mean of case and control sample sizes). This gave the combined meta-analyzed Z-score Z_{meta}.

$$Z_{meta} = \frac{\sum Z_i \sqrt{n_i^{eff}}}{\sqrt{\sum n_i^{eff}}} \quad ; \quad n_i^{eff} = \frac{2 \times n_i^{SS} \times n_i^{CON}}{n_i^{SS} + n_i^{CON}} \tag{2}$$

Thus the final overall 2-sided p-value p_{meta} was obtained by:

$$p_{meta} = 2 \times \Phi(1 - |Z_{meta}|) \tag{3}$$

The p-values for all genes were then adjusted for multiple testing using the method of Benjamini and Hochberg (1995) (*3*) by a call to the function *p.adjust* of the R package "stats" (*28*). For each gene, mean log-fold change in gene expression was computed by averaging the log-fold changes over the six studies. Differential expression was detected by applying two thresholds: p-value < 0.01 and 2 fold or greater change in expression in SS compared to control.

Over-Representation Analysis (ORA)

The set of up-regulated genes was then subjected to over-representation analysis by applying hypergeometric test which is briefly described here. Let us consider two lists of genes: the first list being the set of up-regulated genes, and the second, genes that are part of a given KEGG pathway (*14*). The task is to find out if genes belonging to this pathway are also likely to be part of the list of up-regulated genes. This is captured in a 2×2 contingency table as shown in Table 2.

Over-representation analysis tries to capture whether n_{11} and n_{22} are larger than expected relative to the other two cell numbers; in other words do genes belonging to the KEGG pathway also "tend to be in the up-regulated category."

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Hypergeometric testing was applied to detect if a significant association exists between the two attributes of a gene: being up-regulated in SS and being member of a particular pathway. A significant association was detected at threshold of $p < 0.001$. Over-representation analysis was implemented by applying the function *hyperg* of the "Category" package (*9*).

Gene Set Enrichment Analysis (GSEA)

GSEA focuses on cumulative changes in the expression of multiple genes as a group, which shifts the focus from individual genes to groups of genes. By looking at several genes at once, GSEA identifies pathways with coordinated change in expression of the member genes. The following steps were taken. First, genes were mapped to pathways on an incidence matrix. Then *gseattperm* function from the "Category" (*9*) package was used for the permutation test. The function received the following as input: the gene expression data, sample phenotypic data, and the incidence matrix representing the gene sets of interest. Then it created a version of the data set with phenotype labels randomly scrambled, producing the corresponding ranked list, and computed the enrichment score (ES) of the gene set for this permuted data set. This process of scrambling the labels and re-computation of ES was repeated 10000 times to produce an empirical null distribution of the ES scores. The observed ES for each pathway was compared against the null distribution. Pathways with observed ES very different ($p = 0.0$) from the null distribution were considered significant. For each pathway a meta-analyzed GSEA p-value was obtained using Fisher's p-value product method (*38*).

Signaling Pathway Impact Analysis (SPIA)

Impact Factor analysis combines both over-representation analysis and functional class score (gene set enrichment analysis) approach, taking into account fold-change in gene expression, gene-gene interactions such as activation and inhibition, and the topology of the pathway (*33*). By combining two different evidences (one from the analysis using the hypergeometric model and the other from the probability of perturbation that takes the pathway topology into account), a pathway-level score was computed. For each gene, t-test between SS and control was performed. Log-fold changes of the differentially expressed ($p < 0.05$) genes were passed to the function *spia* of the package "SPIA" (*33*). This was repeated for each study separately, each time returning a list of pathways and associated p-values. For each pathway, Fisher product of the six p-values (one for each study) was calculated to return one p-value for pathway. The ten most significantly perturbed pathways were considered.

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Validation cohort

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, India and SCB Medical College Hospital, Cuttack, India. All the methods were carried out in accordance with the approved guidelines. Informed consent was obtained from all subjects who participated in the study. 12 healthy control and 9 cases of SS were included in the study. Whole blood was collected from the subjects in PAXgene Blood RNA tubes (BD/Preanalytics) and incubated at ambient temperature for at least 2 hours. RNA was isolated using PAXgene Blood RNA kit (Qiagen) as per manufacturer's instructions. Total RNA yield and quality were checked in Nanodrop 2000 spectrophotometer and Agilent Bioanalyzer RNA Nano 6000 chip. RNA samples with a RIN number ≥ 6 were converted to cRNA and hybridised onto either HumanHT-12 v4 BeadChip (Illumina) or Human Gene 2.0 ST Arrays (Affymetrix). Gene expression intensities were read using the standard protocol provided by the manufacturer.

Raw data files from the platform HumanHT-12 v4 BeadChip (Illumina) were read with the function *read.ilmn* of the package "limma" (*32*). After reading the raw data the background correction and normalization using the control probes were done by the help of function *neqc* of the package "limma". For the samples processed on Human Gene 2.0 ST Arrays (Affymetrix), RMA-normalized data were read with the function *read.table*. Each probe ID was mapped to its corresponding Entrez gene ID from the gene symbol by the help of "org.Hs.eg.db" package. In case of duplicated/multiple probes for a single Entrez gene id, the probe with the highest variance was retained (and others dropped from further analysis using the function *nsFilter* of the package "genefilter"). Correction of batch effect was performed with the *ComBat* function of "sva" (*41*) package. Genes of the pathway hsa04380 were extracted and mean gene expression computed. Scatter plot of SS versus control was generated for visual inspection of the data.

Permutation test for enrichment

We performed a permutation-based enrichment test to provide evidence for overall up-regulation of the pathway hsa04380 in SS (validation cohort). For this, we used the function *permutationTest* of the package"resample" (*39*) to calculate the permutation-based p-value accounting for correlation among the pathway genes. First, we calculated the proportion of significantly up-regulated ($p < 0.05$) genes in the pathway by using a two-sample t-test to test for up-regulation of each pathway gene. This was observed

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

to be 0.447. Next, we reshuffled the sample groups (i.e. case control status) 100000 times and similarly calculated the proportion of up-regulated genes for each permutation replicate. Finally, the permutation-based p-value was obtained as the proportion of replicates where the simulated proportion was greater than the observed value.

Gene-level analysis: We then looked at the individual genes in the pathway that were significantly up-regulated after multiple-testing correction at an FDR level of 0.05. We further filtered the resulting gene list to include only genes that showed a high fold-change (2 or more) and were expressed in significant amounts (intensity of 100 or more).

Software used

All analyses except Ingenuity Pathway Analysis (described later) were done with R version 3.1.0 implemented under GNU/Linux operating system on a 32-bit i686-computer system. The analysis work flow is presented in Fig. 2.

Ingenuity Pathway Analysis of the up-regulated genes

Mean log-fold change in expression was calculated for the 200 genes up-regulated in SS, and submitted to the Ingenuity Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity). Core IPA analysis was performed without applying any threshold or cutoff. IPA performed a right-tailed Fisher's Exact Test and calculated a p-value for over-representation of *a priori* defined biological pathways in the data submitted. It also corrected the p-value to account for multiple testing by Benjamini-Hochberg method (FDR). IPA generated molecular networks with algorithmically generated pathways which were over-represented in the submitted data. Similarly it identified diseases and biological functions that were over-represented in the data. The three broad categories returned by IPA were interpreted.

Availability of data and materials

The dataset(s) and the R code supporting the conclusions of this article are available under the project **nibmgss** in the SourceForge repository
https://sourceforge.net/projects/nibmgss.

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Selection of studies and meta-analysis

By systematic search of literature, NCBI GEO and other gene expression databases, we selected six studies of SS (Fig. 1, Table 1). Normalized gene expression data of the studies were retrieved from GEO and analyzed to detect differentially expressed genes (Fig. 2). Differential expression was measured in terms of both log-fold change and p-value. For each gene, SS was compared with control and the six p-values were combined and adjusted for multiple testing (detailed meta-analysis procedure provided in the Methods section) to generate a single p-value per gene. For each gene, the six log-fold changes were averaged to produce a single log-fold change. Using stringent criteria (adjusted p-value < 0.01, fold change of 2 or more), we discovered 200 genes that were consistently up-regulated in SS. We did not discover any gene that was down-regulated at this level of significance, i.e., $p < 0.01$ and fold change of half or less (Supplementary Fig. S1).

The 200 genes up-regulated in SS

These included genes associated with inflammation and innate immunity (SERPINB1, S100A12, SLC11A1, TLR5, PADI4, CLEC5A, TLR8, LILRA5), critical illness (ANKRD22, CEACAM1, CLEC5A, G0S2, BCL-6, MMP8, TREM1, CCL4), anti-microbial activity (ADM, BP1, DEFA4, RNASE3), complement and coagulation systems (C1QB, F5, CR1, C3AR1 , CR1, C3AR1, CD59, SERPINA1), carbohydrate and lipid metabolism (PFKB2, PFKB3, SLC2A3, GYG1, PYGL, ACSL1, HGF, FAR2, DGAT2), and others. While many of these systems are known to be involved in SS, there were some surprises, such as, up-regulation of genes of bone metabolism (CPD, MAPK14, FOSL2, IL1R1, MAP2K6, SOCS3, OSCAR, SIRPA, ALPL, BST1, CA4, TNFAIP6) in SS (Supplementary Table S1). With strong evidence in favour of altered expression at gene level, we proceeded to analyse the data at pathway (gene set) level.

Pathway analysis: ORA, GSEA, SPIA, IPA

A pathway is a set of genes that participate in a single functional module. When a large number of genes in the same pathway are perturbed (or up-regulated, in this case), then the pathway itself may be inferred to be perturbed. Secondly, small changes in gene expression are usually not detectable with standard methods. However, if the changes happen in a co-ordinated fashion over a set of functionally related genes (i.e., a pathway), then the cumulative effect of the small gene-level changes shall result in a

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

large and detectable perturbation at the pathway-level. Lastly, the gene set in a pathway has a specific network structure or topology, such that, the nodes proximal to the point of origin (for example, the signal from the membrane), has a greater regulatory impact compared to the distal nodes. All these considerations were taken into account in the current analysis.

Kyoto Encyclopedia of Genes and Genomes (KEGG) (*14*) is a database of gene sets (pathways) arranged by function. Each pathway consists of multiple genes and a gene may belong to more than one pathway. We reasoned that, for a pathway to be up-regulated in SS, a large number of genes in the pathway need to be up-regulated. In other words, we sought to find the pathway(s) that is (are) significantly "enriched" with the genes in our list of 200. Using the method of "Over-Representation Analysis" (ORA; described in detail in the Methods section), three pathways were identified: hsa04610 - Complement and Coagulation Cascades, hsa04380 - Osteoclast Differentiation, and hsa05202 -Transcriptional Misregulation in Cancer (Supplementary Table S2). The complement system is a proteolytic cascade in blood plasma participating in innate immunity. When activated, it leads to recruitment of inflammatory cells and elimination of pathogens. Dysregulation of the complement system (especially activated C3 and C5) is known to occur in SS (*23*). Similarly, blood coagulation is an essential defense mechanism for blocking off the pathogen and preventing spread of inflammation, in the initial phase. However, the inflammatory state found in SS alters the hemostatic balance in favour of a procoagulant state, with disseminated intravascular coagulation (DIC) (*17*). The other two pathways are not commonly associated with sepsis or SS, although osteoclasts arise from differentiation of blood monocytes and cause bone resorption, which is markedly increased in critically ill patients (*31*). The third pathway hsa05202 consists of many genes that are transcriptionally altered in different malignancies.

We next applied two other tools, "Gene Set Enrichment Analysis" (GSEA) and "Signaling Pathway Impact Analysis" (SPIA), that are methodologically independent of ORA, and provide additional support to our finding (Fig. 2). GSEA calls upon a global (i.e., genome-wide, not limited to any pre-selected list) search strategy to detect the KEGG pathway (s) with significant up-regulation in SS compared to control. SPIA combines elements of ORA and GSEA, with attention to gene-gene interactions and pathway topology. There was a single pathway (hsa04380) that was common to the results from the three approaches (ORA, GSEA, SPIA). Further, we performed Ingenuity Pathway Analysis (IPA) with the top 200 up-regulated genes, that returned a list of pathways. Two pathways among the top ten were related to bone loss in Rheumatoid Arthritis. This and the presence of other pathways, such as, p38 MAPK

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

signaling, complement system and TLR signaling, were similar to the results from KEGG analysis. IPA discovered the following diseases and functions enriched among the up-regulated genes: Infectious Disease, Connective Tissue Disorder, Inflammatory Diseases and Skeletal and Muscular Disorders. The network with significant biological function termed "infectious disease, inflammatory response, connective tissue disorders" was discovered to be differentially regulated in SS (Supplementary Fig. S2). Over all, IPA results were similar to the pathway analysis with KEGG, and generally consistent with the finding of altered skeletal function in SS. Thus, multiple analytic approaches, ORA, GSEA, SPIA and IPA, together revealed hsa04380 (Osteoclast Differentiation) with unequivocal transcriptional up-regulation in SS. We then sought to validate the pathway up-regulation in an independent set of SS cases.

Validation of hsa04380 in an independent cohort

Batch-corrected whole blood gene expression data (Supplementary Fig. S3) of the pathway hsa04380 was obtained from an independent set of human subjects (Supplementary Table S3). Mean expression values (for the two groups: control and SS) were calculated and a scatter-plot generated (Fig. 3). In this plot, each point corresponds to a single gene. The points near the identity line (the diagonal in the Figure), correspond to genes with similar expression level in control and SS groups. The genes that are up-regulated in SS are expected to be significantly deviated from the diagonal toward the SS axis. Indeed, for most of the genes, there is much higher expression in SS, as shown in Fig. 3. Additional evidence for up-regulation of the pathway hsa04380 was obtained from the permutation-based enrichment test, where we resampled (reshuffled) the case/control labels of the samples and calculated a permutation-based p value for the proportion of up-regulated genes in the pathway. We observed that up-regulation of the pathway was highly significant (permutation $p = 0.00028$; 100000 replicates). On looking at individual genes, we found that 60 genes were FDR significant and 25 genes passed the expression filters. Supplementary Fig. S4 shows the boxplots of these highly significant genes in the validation cohort.

The 25 highly significant genes of the pathway are well-connected in the pathway map, and known to be expressed in neutrophils and monocytes, which are the two abundant sub-types of leukocytes. These 25 genes are not only sgnificantly up-regulated in the validation cohort but also significantly ($p < 0.05$; in gene level metaanalysis) up-regulated in the Discovery cohort (6 SS datasets).

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Significant genes of the pathway in SS

Of these 25 key genes of this pathway (Supplementary Table S4), there were genes whose encoded proteins form a signal transduction cascade from the plasma membrane to the nucleus. There were genes coding for proteins situated on the plasma membrane (IL1R, TNFα, TGFβ), or belonging to MAP Kinase group (Rac1, MKK6 and p38), AP1 transcription factor complexes (JUNB, FOSL2). This suggested that a signal starting from the membrane may be transduced through the MAP kinase subnetwork to the nucleus culminating in transcriptional regulation of osteoclast differentiation-specific genes. The list also included other genes (OSCAR, TREM2, SIRPB1, NFATC1) that suggested a parallel path from the membrane leading to up-regulation of the transcription factor PU.1 and expression of genes for osteoclast differentiation. Based on these 25 genes, a simple model of osteoclast differentiation in SS was proposed (Fig. 4).

Since our analysis was performed on whole blood expression data, we were also interested in the cellular and plasma distribution of the gene products. A literature search for tissue specific expression revealed that many of these genes are expressed in neutrophils and monocytes, two important leukocyte sub-types. Additionally, almost all of the gene products have been reported to be detectable in plasma (Supplementary Table S4).

Discussion

In contrast with gene-level analysis, pathway analysis strives to identify a perturbed pathway (a set of functionally related genes) whose transcriptional alteration is significantly associated with the given condition. There exist several approaches to this end, such as ORA, GSEA, or SPIA, each with its own advantages (*15*). ORA considers the number of genes differentially expressed, but ignores any value associated (such as fold-change). GSEA considers all genes in a pathway and is more likely to detect coordinated changes in gene expression. SPIA, in addition, takes the topology of the pathway into account. Here, combination of three different methods gives leverage in finding the pathway that is unequivocally altered in SS.

In order to discover the stable blood transcriptional changes in SS, we conducted three different analyses (ORA, GSEA, SPIA) and intersected the resulting lists to obtain a single KEGG pathway (hsa04380 - Osteoclast Differentiation Pathway) as a stable expression signature of SS in circulating

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

leukocytes. We confirmed this result in an independent validation cohort and demonstrated that up-regulation of the pathway is distinct in SS compared to controls. Further, we subjected the list of up-regulated genes to Ingenuity Pathway Analysis and observed broad agreement with this finding. To our knowledge, this is the first time that osteoclast differentiation pathway has been shown to be significantly perturbed in SS.

The skeletal and the immune systems are known to interact with each other. For example, in immune disorders, such as rheumatoid arthritis, the ratio of bone resorption to bone formation is skewed toward the former. This ratio is balanced by the activity of two groups of cells: osteoclasts and osteoblasts. Osteoclasts are specialised bone cells of monocyte-macrophage lineage (*10, 34*) responsible for bone resorption whereas osteoblasts are responsible for bone formation. It was earlier shown that there is increased bone resorption in critically ill patients (*31*). However, our finding suggests that the role of this pathway may be more specific in SS.

We propose a simplified model of transcriptional changes in hsa04380 during SS, as described below and shown in Fig. 4. The Osteoclast differentiation pathway may be conceptualised in terms of interactions among a few key processes or subnetworks. Firstly, pro-inflammatory cytokine-receptor complexes (such as, TNFα-TNFR and IL1-IL1R) can cause osteoclast differentiation with the help of MITF and MAPK, NFκB signaling (*16, 18*). In parallel, IL-1 and TNFα activate NFκB to up-regulate expression of osteoclast specific genes CD40LG (also known as TRAP), CTSK (cathepsin K) and OSCAR (osteoclast associated receptor), three important regulators of bone mass. It is worth mentioning that IL1-mediated osteoclast differentiation is distinct from the well-known RANKL-RANK signaling mediated activation (*16*). We observed significant up-regulation of CD40LG, cathepsin K and OSCAR, but not RANKL or RANK. The association of TNFα, IL-1, IL-6 and IL-7 has been detected in a variety of chronic inflammatory bone diseases such as rheumatoid arthritis, osteoarthritis, and periodontal diseases. These cytokines are produced by macrophages, lymphocytes, osteoblasts and bone marrow stromal cells under the regulation of NFκB, and also stimulate NFκB signaling in the target cells by positive feedback (*20*). These genes are significantly up-regulated in SS.

The second subnetwork starts with a set of closely related and functionally similar immunoglobulin-like receptors among which the most prominent one is OSCAR (*6*). Each receptor molecule bears ITAM transmembrane domains that transduce signal from plasma membrane to the nucleus through second messenger molecules, such as Syk (a tyrosine kinase) and BLNK, which carry the signal to PLCγ, resulting

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

in intracellular calcium release, that is known to activate transcription factors. Two transcription factors PU.1 (a well-known biomarker for osteoclast differentiation) (*30*) and AP1 (JUN, FOSL) (*35*) form a complex that results in transcription of osteoclast-specific genes in the precursor cells. Both PU.1 and AP1 are significantly up-regulated in SS.

It is also interesting, that AP1 is involved in cellular proliferation, transformation and death. There is also a report that apoptotic cell death (mostly of lymphocytes) is observed in patients suffering from SS (*42*). An alternative interpritation would suggest prognosis towards severe diease phenotype (i.e., SS) by AP1-induced excessive lymphocyte apoptosis.

Thirdly, the MAPK module (includes genes such as MAP2K6, JNK, p38 along with Reactive Oxygen Species (ROS) genes NCF2, NCF4) plays an important role in activating AP1 transcription factor complex and regulates genes responsible for the differentiation process (*22*). Generation of reactive oxygen species and MAP kinase pathway signaling play a critical role in this pathway. IL1R1, a prominent mediator involved in many cytokine-induced immune and inflammatory responses, and genes of MAP kinase pathway (MAPK14, MAPK26, FOSL2, IL1R1, CA4), all associated with inflammation, are up-regulated in SS. The signalling molecule p38, known to be essential for osteoclast differentiation (*24*), is sensitive to activation by ROS in bone marrow monocyte-macrophage lineage (BMM) cells (*21*). Mitochondrial dysfunction in SS causes generation of various reactive oxygen (and nitrogen) species, contributing to activation of MAP kinase pathway and subsequent osteoclastogenesis (*21*). Although little attention has been paid to the role of ROS in differentiation of macrophages and monocytes into osteoclasts, we observed that MAP kinase pathway and genes associated with ROS generation are significantly up-regulated in SS. Additionally, both MKK6 and p38 (up-regulated in SS) cause transdifferentiation of neutrophils to monocyte or dendritic cells, which is of relevance during inflammation (*13, 19*).

In most studies, discharge from the ICU is considered the survival end-point. However, the survivors themselves are prone to many long term complications including musculoskeletal ones. By rigorous analysis, we show that patients of SS have increased osteoclast differentiation. Long term follow-up of the critically ill patients after discharge might reveal new insights into the biology of SS and provide fresh directions for therapeutic innovation. Since excessive osteoclast activity contributes to enhanced bone resorption, patients of SS (especially those that survive the acute episode) are likely at risk for bone loss and fracture. Therefore, this pathway is a potential target for therapeutic intervention. There already exists a molecule Simvastatin (*25*), which inhibits osteoclast differentiation by scavenging reactive

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

oxygen species. Zinc is known to inhibit osteoclast differentiation in some other model systems (*12, 27*). It is interesting that prophylactic zinc supplementation improves treatment outcome in human infants aged 7–120 days with probable serious bacterial infection (*4*). We think that zinc supplementation might also play a role in lowering the osteoclastogenic bone loss in SS patients, which might contribute to patient survival and long term recovery (*12*). Antioxidants could also be useful in limiting osteoclast differentiation. However, the value of any therapeutic intervention on the pathway needs to be ascertained through independent testing. Lastly, we have shown in our study that there are 25 genes of the pathway which are significantly differentially expressed in SS (Supplementary Table S4) compared to controls. Since their protein products are detectable in peripheral circulation, it is plausible to suggest these genes, singly or in combination, as plasma biomarkers of SS.

Conclusion

Firstly, systematic analysis of multiple data sets enabled identification of the core set of genes that are consistently up-regulated in SS. Secondly, we applied three different methods (Over-Representation Analysis, Gene Set Enrichment Analysis and Signaling Pathway Impact Analysis) to arrive at the pathway hsa04380 (Osteoclast Differentiation) with consistent and significant up-regulation in SS. We validated this result with additional bioinformatic analysis and pathway gene expression assay in an independent cohort of SS. Thirdly, we established that the expression signature is distinct in SS compared to control. The 25 genes short-listed, either singly or in combination, may serve as an expression signature of SS. It should be noted that, we do not claim causality, but simply an association of up-regulation of hsa04380 with SS. However, in the light of this finding, altered osteoclast differentiation in septic shock deserves greater attention.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

S.K.M. conceived the study; S.K.M., S.B., S.M. and A.B.P. conducted the data analysis; S.M. conducted the experiment; B.K.D. and P.T. recruited the patients in the validation study; S.K.M, B.R., S.M., B.K.D., P.T. interpreted the results; S.M., S.B. and S.K.M. wrote the manuscript. All authors reviewed

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the manuscript.

Acknowledgements

We are grateful to the patients of SCB Medical College Hospital for contributing blood samples for the study. We thank the Director, NIBMG, India for facilitating this work. This study is supported by an intramural grant from NIBMG, Kalyani and an extramural grant from the Department of Biotechnology, Govt of India. S.M. acknowledges research fellowship provided by the University Grants Commission of India. We thank our colleagues at the NIBMG core facility and the ILS laboratory for timely processing of the samples, often at short notice. S.K.M. acknowledges very helpful advice received from Prof. Gagandeep Kang, Christian Medical College, Vellore during data analysis.

References and Notes

1. Neill K J. Adhikari, Robert A. Fowler, Satish Bhagwanjee, and Gordon D. Rubenfeld. Critical care and the global burden of critical illness in adults. *Lancet*, 376(9749):1339–1346, Oct 2010.

2. Angus, Linde-Zwirble DC, Lidicker WT, Clermont J, Carcillo G, Pinsky J, and MR. Epidemiology of severe sepsis in the unit-ed states: analysis of incidence, outcome,and associated costs of care. *Crit Care Med*, 29:1303–1310, 2001.

3. Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300, 1995.

4. Shinjini Bhatnagar, Nitya Wadhwa, Satinder Aneja, Rakesh Lodha, Sushil Kumar Kabra, Uma Chandra Mouli Natchu, Halvor Sommerfelt, Ashok Kumar Dutta, Jagdish Chandra, Bimbadhar Rath, Mamta Sharma, Vinod Kumar Sharma, Mohini Kumari, and Tor A. Strand. Zinc as adjunct treatment in infants aged between 7 and 120 days with probable serious bacterial infection: a randomised, double-blind, placebo-controlled trial. *Lancet*, 379(9831):2072–2078, Jun 2012.

5. Jonathan Cohen, Steven Opal, and Thierry Calandra. Sepsis studies need new direction. *Lancet Infect Dis*, 12(7):503–505, Jul 2012.

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

6. Tania N Crotti, Anak ASSK Dharmapatni, Ekram Alias, and David R Haynes. Osteoimmunology: Major and costimulatory pathway expression associated with chronic inflammatory induced bone loss. *Journal of Immunology Research*, 2015.

7. Sean Davis and Paul S Meltzer. Geoquery: a bridge between the gene expression omnibus (geo) and bioconductor. *Bioinformatics*, 23(14):1846–1847, Jul 2007.

8. R. Gentleman, V. Carey, W. Huber, and F. Hahne. *genefilter: methods for filtering genes from microarray experiments*. R package version 1.46.1.

9. R. Gentleman, S. Falcon, and D. Sarkar. *Category Analysis*. R package version 2.30.0.

10. Siamon Gordon and Philip R Taylor. Monocyte and macrophage heterogeneity. *Nature Reviews Immunology*, 5(12):953–964, 2005.

11. Larry V. Hedges and Jack L. Vevea. Fixed- and random-effects models in meta-analysis. *Psychological Methods*, 3:486–504, 1998.

12. Mamiko Hie and Ikuyo Tsukamoto. Administration of zinc inhibits osteoclastogenesis through the suppression of rank expression in bone. *Eur J Pharmacol*, 668(1-2):140–146, Oct 2011.

13. Christof Iking-Konert, Benedikt Ostendorf, Oliver Sander, Markus Jost, Christof Wagner, Leo Joosten, Matthias Schneider, and GM Hänsch. Transdifferentiation of polymorphonuclear neutrophils to dendritic-like cells at the site of inflammation in rheumatoid arthritis: evidence for activation by t cells. *Annals of the rheumatic diseases*, 64(10):1436–1442, 2005.

14. M. Kanehisa and S. Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28(1):27–30, Jan 2000.

15. Purvesh Khatri, Marina Sirota, and Atul J. Butte. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol*, 8(2):e1002375, 2012.

16. Jung Ha Kim, Hye Mi Jin, Kabsun Kim, Insun Song, Bang Ung Youn, Koichi Matsuo, and Nacksung Kim. The mechanism of osteoclast differentiation induced by il-1. *The Journal of Immunology*, 183(3):1862–1870, 2009.

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

17. Elizabeth G. King, Gustavo J. Bauzá, Juan R. Mella, and Daniel G. Remick. Pathophysiologic mechanisms in septic shock. *Lab Invest*, 94(1):4–12, Jan 2014.

18. Kanichiro Kobayashi, Naoyuki Takahashi, Eijiro Jimi, Nobuyuki Udagawa, Masamichi Takami, Shigeru Kotake, Nobuaki Nakagawa, Masahiko Kinosaki, Kyoji Yamaguchi, Nobuyuki Shima, et al. Tumor necrosis factor α stimulates osteoclast differentiation by a mechanism independent of the odf/rankl–rank interaction. *The Journal of experimental medicine*, 191(2):275–286, 2000.

19. René Köffel, Anastasia Meshcheryakova, Joanna Warszawska, Annika Hennig, Karin Wagner, Almut Jörgl, Daniela Gubi, Doris Moser, Anastasiya Hladik, Ulrike Hoffmann, Michael B. Fischer, Wim van den Berg, Marije Koenders, Clemens Scheinecker, Bernhard Gesslbauer, Sylvia Knapp, and Herbert Strobl. Monocytic cell differentiation from band-stage neutrophils under inflammatory conditions via mkk6 activation. *Blood*, 124(17):2713–2724, Oct 2014.

20. S A Krum, J Chang, G Miranda-Carboni, and C Y Wang. Novel functions for NFκB: inhibition of bone formation. *Nat Rev Rheumatol*, 6:607–611, 2010.

21. Lee, Choi NK(1), Baik YG, Han JY, Jeong SY, Bae DW, Kim YS, Lee N, and SY. A crucial role for reactive oxygen species in rankl-induced osteoclast differentiation. *Blood*, 106:852–859, 2005.

22. Xiaotong Li, Nobuyuki Udagawa, Kanami Itoh, Koji Suda, Yoshiyuki Murase, Tatsuji Nishihara, Tatsuo Suda, and Naoyuki Takahashi. p38 mapk-mediated signals are required for inducing osteoclast differentiation but not for osteoclast function. *Endocrinology*, 143(8):3105–3113, 2002.

23. John D. Lambris Maciej M. Markiewski, Robert A. DeAngelis. Complexity of complement activation in sepsis. *J Cell Mol Med*, 12(6A):2245–2254, 2008.

24. Masahito Matsumoto, Masakazu Kogawa, Seiki Wada, Hiroshi Takayanagi, Masafumi Tsujimoto, Shigehiro Katayama, Koji Hisatake, and Yasuhisa Nogi. Essential role of p38 mitogen-activated protein kinase in cathepsin k gene expression during osteoclastogenesis through association of nfatc1 and pu.1. *J Biol Chem*, 279(44):45969–45979, Oct 2004.

25. Ho-Jin Moon, Sung Eun Kim, Young Pil Yun, Yu-Shik Hwang, Jae Beum Bang, Jae-Hong Park, and Il Keun Kwon. Simvastatin inhibits osteoclast differentiation by scavenging reactive oxygen species. *Experimental & molecular medicine*, 43(11):605–612, 2011.

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

26. Jeffrey E Nowak, Kelli Harmon, Charles C Caldwell, and Hector R Wong. Prophylactic zinc supplementation reduces bacterial load and improves survival in a murine model of sepsis. *Pediatric critical care medicine: a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*, 13(5):e323, 2012.

27. Kwang Hwan Park, Boryung Park, Dong Suk Yoon, Seung-Hyun Kwon, Dong Min Shin, Jin Woo Lee, Hyun Gyu Lee, Jae-Hyuck Shim, Jeon Han Park, and Jae Myun Lee. Zinc inhibits osteoclast differentiation by suppression of ca2+-calcineurin-nfatc1 signaling pathway. *Cell Commun Signal*, 11:74, 2013.

28. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

29. Rangel-Frausto, Pittet MS, Costigan D, Hwang M, Davis T, Wenzel CS, and RP. The natural history of the systemic inflammatory response syndrome (sirs): a prospective study. *JAMA*, 273:117–123, 1995.

30. Sudarshana M Sharma, Agnieszka Bronisz, Rong Hu, Krupen Patel, Kim C Mansky, Said Sif, and Michael C Ostrowski. Mitf and pu. 1 recruit p38 mapk and nfatc1 to target genes during osteoclast differentiation. *Journal of Biological Chemistry*, 282(21):15921–15929, 2007.

31. Lorna M. Smith, Brian Cuthbertson, John Harvie, Nigel Webster, Simon Robins, and Stuart H. Ralston. Increased bone resorption in the critically ill: association with sepsis and increased nitric oxide production. *Crit Care Med*, 30(4):837–840, Apr 2002.

32. Gordon K Smyth. Limma: linear models for microarray data. In R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber, editors, *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, pages 397–420. Springer, New York, 2005.

33. Adi Laurentiu Tarca, Purvesh Kathri, and Sorin Draghici. *SPIA: Signaling Pathway Impact Analysis (SPIA) using combined evidence of pathway over-representation and unusual signaling perturbations*, 2013. R package version 2.16.0.

34. Nobuyuki Udagawa, Naoyuki Takahashi, Takuhiko Akatsu, Hirofumi Tanaka, Takahisa Sasaki, Tatsuji Nishihara, Toshihiko Koga, T John Martin, and Tatsuo Suda. Origin of osteoclasts: mature

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

monocytes and macrophages are capable of differentiating into osteoclasts under a suitable microenvironment prepared by bone marrow-derived stromal cells. *Proceedings of the National Academy of Sciences*, 87(18):7260–7264, 1990.

35. EF Wagner and K Matsuo. Signalling in osteoclasts and the role of fos/ap1 proteins. *Annals of the rheumatic diseases*, 62(suppl 2):ii83–ii85, 2003.

36. B. L. Welch. On the comparison of several mean values: an alternative approach. *Biometrika*, 38:330–336, 1951.

37. Hector R. Wong. Clinical review: sepsis and septic shock–the potential of gene arrays. *Crit Care*, 16(1):204, 2012.

38. R. A. Fisher. Statistical Methods for Research Workers. Oliver and Boyd (Edinburgh), New York, 1925. ISBN 0-05-002170-2.

39. T. Hesterberg. *resample: Resampling Functions*. R package version 0.4. 2015.

40. J. T. Leek and W. E. Johnson and H. S. Parker and E. J. Fertig and A. E. Jaffe and J. D. Storey. *sva: Surrogate Variable Analysis*. R package version 3.18.0. 2014.

41. J. T. Leek and W. E. Johnson and H. S. Parker and E. J. Fertig and A. E. Jaffe and J. D. Storey. *sva: Surrogate Variable Analysis*. R package version 3.18.0. 2014.

42. R. Hotchkiss and E. F. Paul and D. T. Bradley and W. C. Kevin and J. P. Matuschak and M. B. George and G. K. Timothy and E. Irene. Apoptotic cell death in patients with sepsis, shock, and multiple organ dysfunction. *Critical care medicine*, 27:1230-1251, 1999.

Figures

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

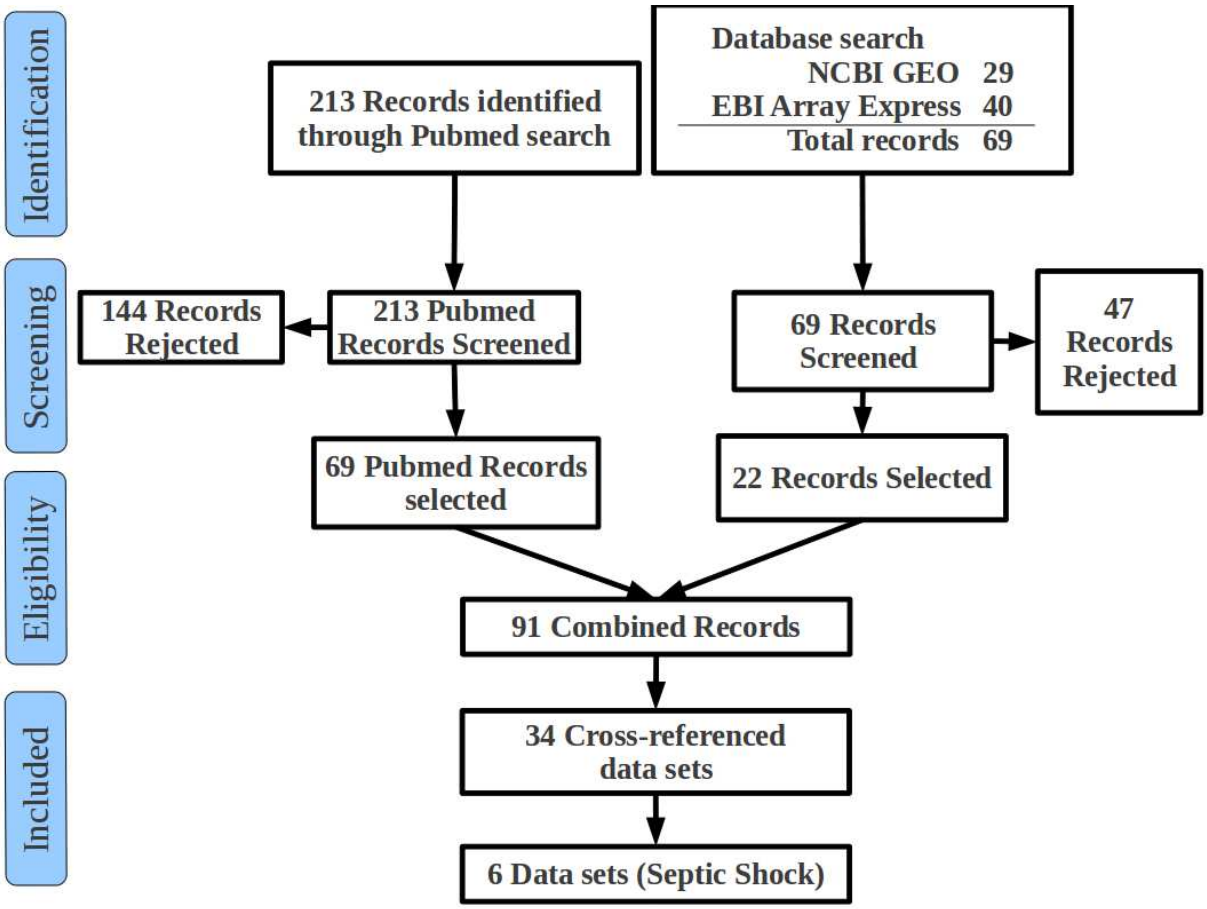


Figure 1: **Selection of studies.** Electronic search on the Gene expression and Literature databases was performed to identify the 213 Pubmed records and 69 gene expression data sets. These records were manually screened to select the studies that were done in human subjects, on circulating leukocytes and gene expression profiling done using gene expression microarray platform, resulting in 91 combined (Pubmed and expression data) records. Duplicate records were removed, and records without data from human sepsis or SS, were deleted. One record was added from a source outside those mentioned above. Of the resulting 34 data sets (cross-referenced with Pubmed articles), six were derived from studies on SS. These six data sets were selected for meta-analysis.

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

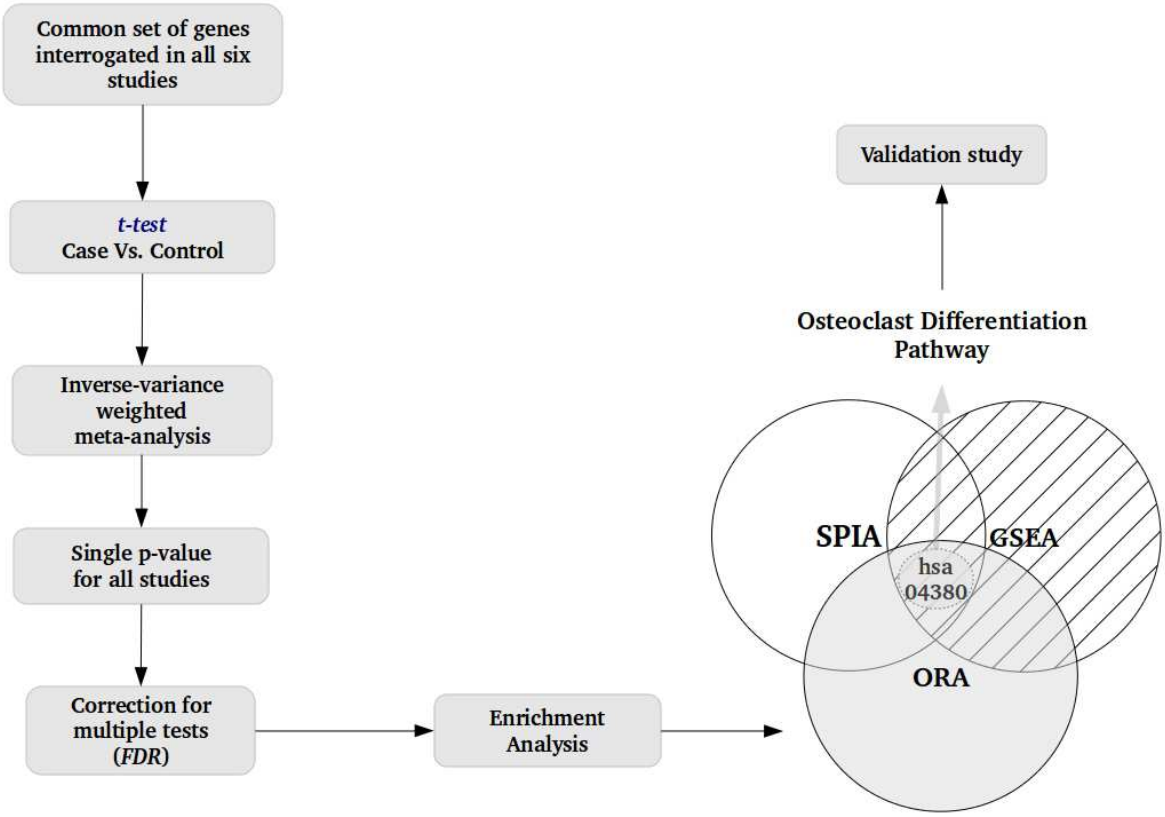


Figure 2: **Overview of the analysis work-flow.** First, gene-level analysis was performed by comparing SS with control. The p-values from multiple studies were combined to compute a single p-value per genes. The p-values were adjusted for multiple testing (FDR). 200 genes with low p-value ($p < 0.001$) and large fold change (2 or greater) were considered significantly differentially expressed (up-regulated) in SS. The 200 up-regulated genes were then subjected to three pathway analysis methods [Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA) and Signaling Pathway Impact Analysis (SPIA)]. A single pathway (hsa04380 - Osteoclast Differentiation) was returned as the common result of the three approaches. Up-regulated gene expression of hsa04380 in SS was validated in an independent patient cohort.

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

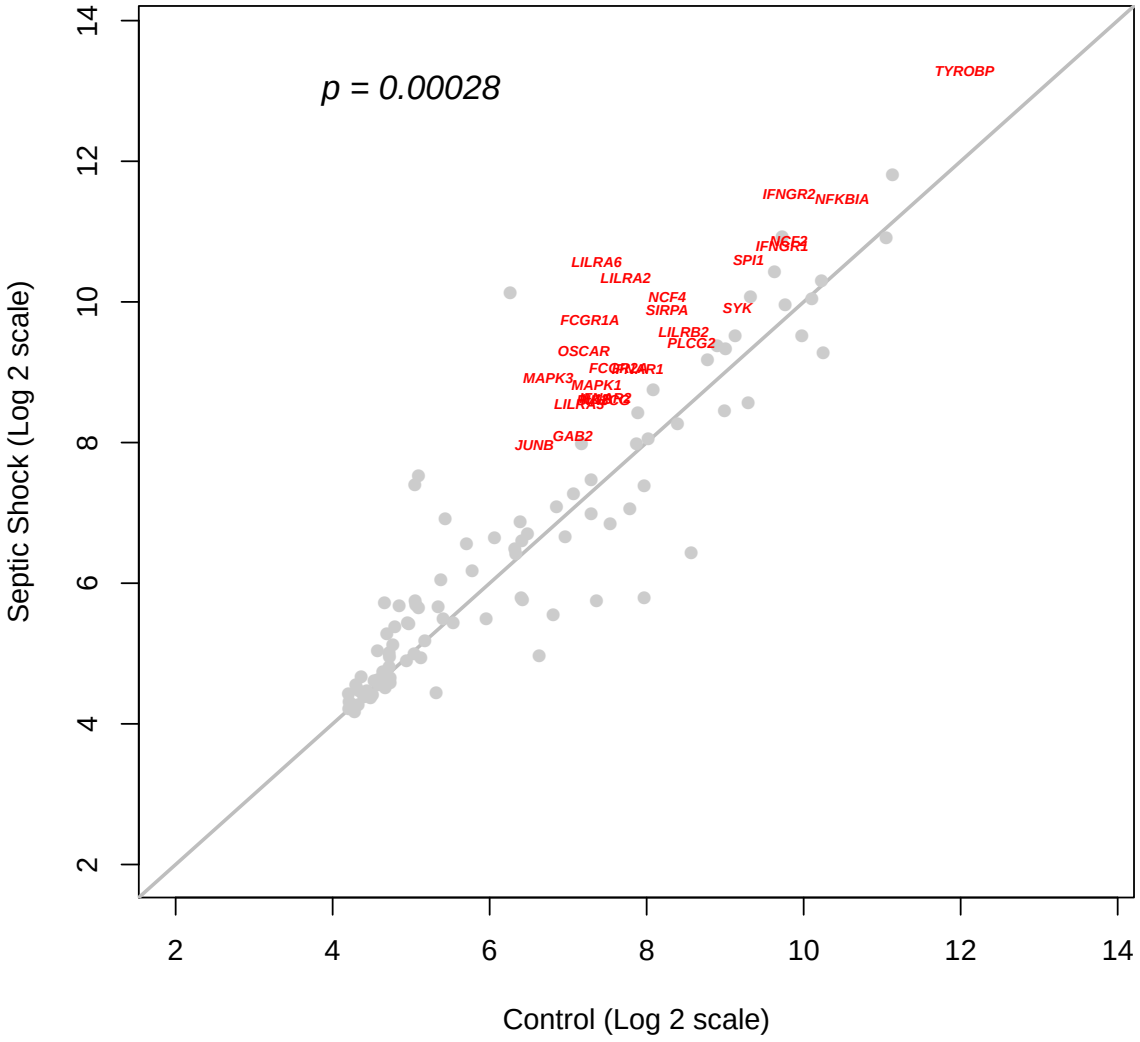


Figure 3: **Up-regulation of Osteoclast Differentiation pathway in septic shock.** Scatter plot of hsa04380 (Osteoclast Differentiation) pathway genes from expression data generated in an independent validation cohort. Each point corresponds to a single gene, and its coordinates reflect mean gene expression in healthy controls (x-axis) and SS (y-axis). The points close to the identity line (the diagonal in the figure), correspond to genes with similar expression level in the control and the SS groups. Points away from the diagonal represent genes that are differentially expressed in SS (up-regulated if the points are in the upper half of the plot area; down-regulated otherwise). Many of the genes are up-regulated in SS as shown by a large number of points deviated from the diagonal toward the SS-axis. The up-regulation of these genes in SS is statistically significant ($p = 0.00028$, permutation test for enrichment). Additional testing was performed for each gene (unpaired t-test between control and SS) leading to selection of individual genes in the pathway that were significantly up-regulated after multiple-testing correction at an FDR level of 0.05. An expression filter was applied to identify the genes that showed a high fold-change (2 or more) and were expressed in significant amounts (intensity of 100 or more). These genes have been shown in red on the plot.

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

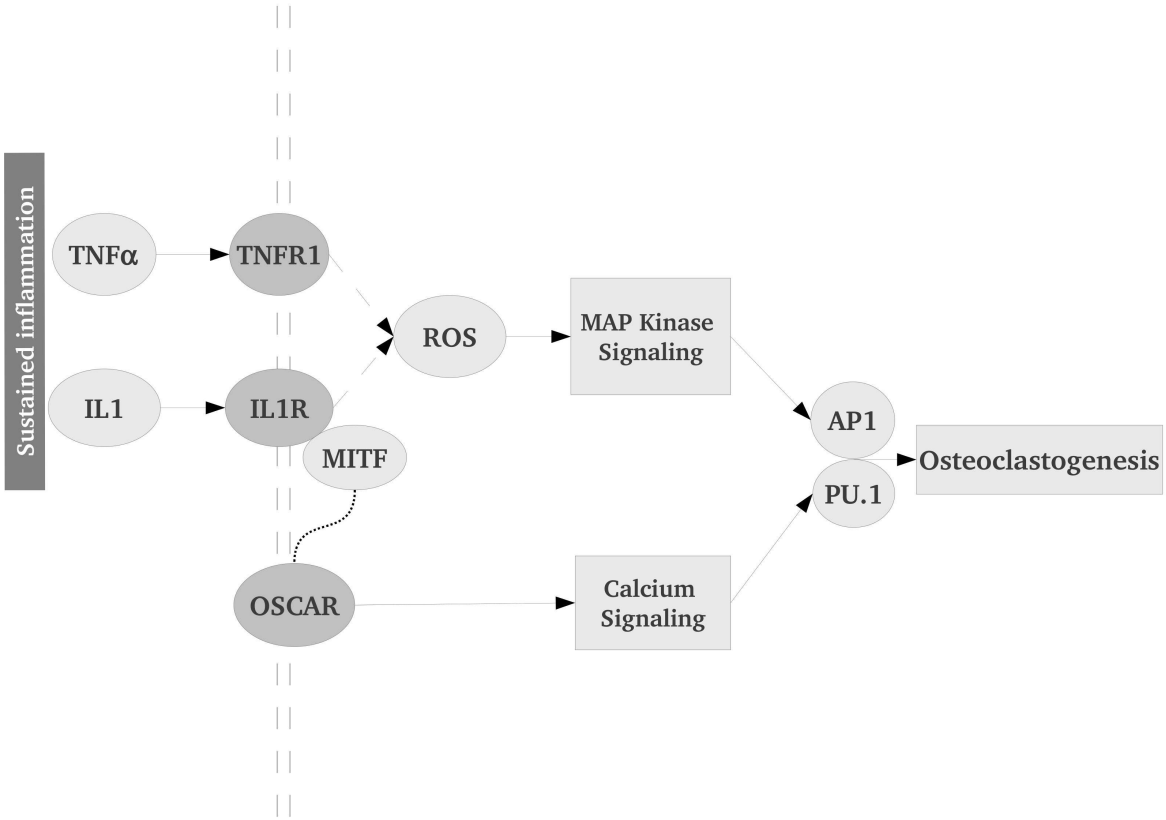


Figure 4: **A simplified model of the pathway.** A simplified model of the Osteoclast Differentiation Pathway up-regulation in SS. Sustained inflammation leads to activation of osteoclast differentiation through proinflammatory cytokines TNFα and IL1, followed by MAP Kinase and Calcium signaling. Reactive Oxygen Species (ROS) play a major role in activation of MAP Kinase. OSCAR is an important receptor on the membrane that causes signal transduction via calcium-signaling. AP1 and PU.1, two important transcription factors, integrate multiple signals and induce expression of osteoclastogenesis-specific genes.

Study ID	Author-Year	PMID	Study Design	Total No. of Samples	Mean Age (yr)	Clinical Setting	Inclusion Criteria	Control Group	Tissue	Sample collection and RNA isolation method	Platform	Data normalization
GSE4607	Wong2007 Cvijanovich2008 Wong2012	17374846 18460642 23107287	Observational study	84	Children <10 years of age	PICU	Septic Shock	Healthy controls	Whole blood	PAXgene	Affymetrix Human Genome U133 Plus 2.0 Array	RMA
GSE8121	Shanley2007	17932561	Longitudinal	75	Children <10 years of age	PICU	Septic Shock	Healthy controls	Whole blood	PAXgene	Affymetrix Human Genome U133 Plus 2.0 Array	RMA
GSE9692	Cvijanovich2008	18460642	Retrospective observational study	45	Children <10 years of age	PICU	Septic Shock	Healthy controls	Whole blood	PAXgene	Affymetrix Human Genome U133 Plus 2.0 Array	RMA
GSE13904	Wong2009	19325468	Prospective observational study	124	Children <10 years of age	Paediatric wards	Septic Shock	Healthy controls	Whole blood	PAXgene	Affymetrix Human Genome U133 Plus 2.0 Array	RMA
GSE26378	Wynn2011	21738952	Prospective observational study	103	Children <10 years of age	Paediatric wards	Septic Shock	Healthy controls	Whole blood	PAXgene	Affymetrix Human Genome U133 Plus 2.0 Array	RMA
GSE26440	Wong2009	19624809	Prospective observational study	130	Children <10 years of age	Paediatric wards	Septic Shock	Healthy controls	Whole blood	PAXgene	Affymetrix Human Genome U133 Plus 2.0 Array	RMA

Table 1: **SS Study Characteristics.** The table shows characteristics (such as sample size, study design, clinical parameters for inclusion criteria and details about the platform technology used to generate the data) of the six selected gene expression studies of septic shock

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

	Number of genes in the given KEGG pathway	Number of genes NOT in the given KEGG pathway
Number of genes up-regulated in SS	n_{11}	n_{12}
Number of genes NOT up-regulated in SS	n_{21}	n_{22}

Table 2: **A** 2×2 **contingency table.** The table shows the four quantities of interest while estimating if a pathway is over-represented (enriched) among the set of differentially expressed genes.

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Data

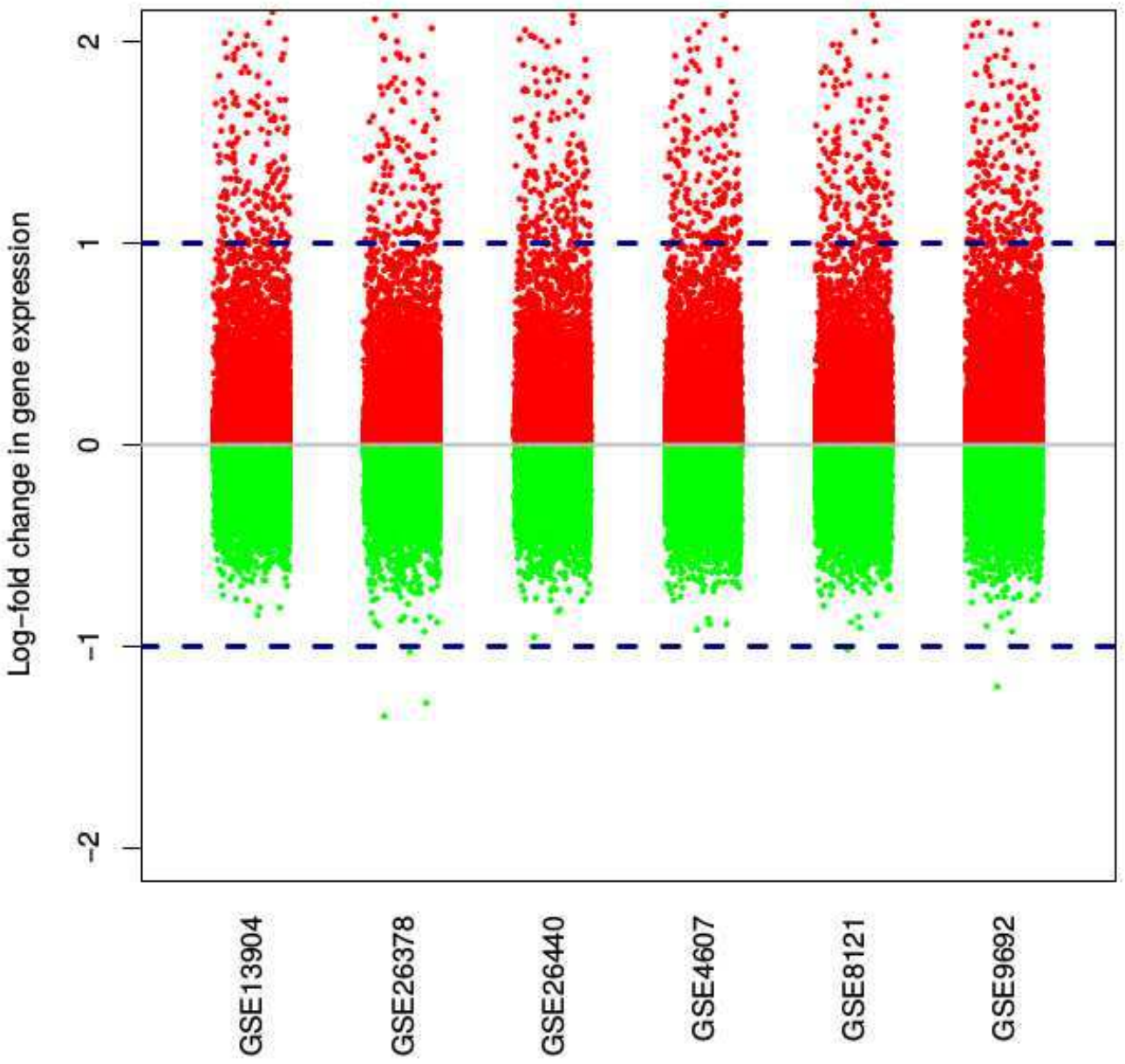
bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Meta-analysis reveals pathway signature of Septic Shock

Samanwoy Mukhopadhyay[1], Abhay Deep Pandey[1], Pravat K Thatoi[2], Bidyut K Das[2], Balachandran Ravindran[3], Samsiddhi Bhattacharjee[1] and Saroj Kant Mohapatra[1*]

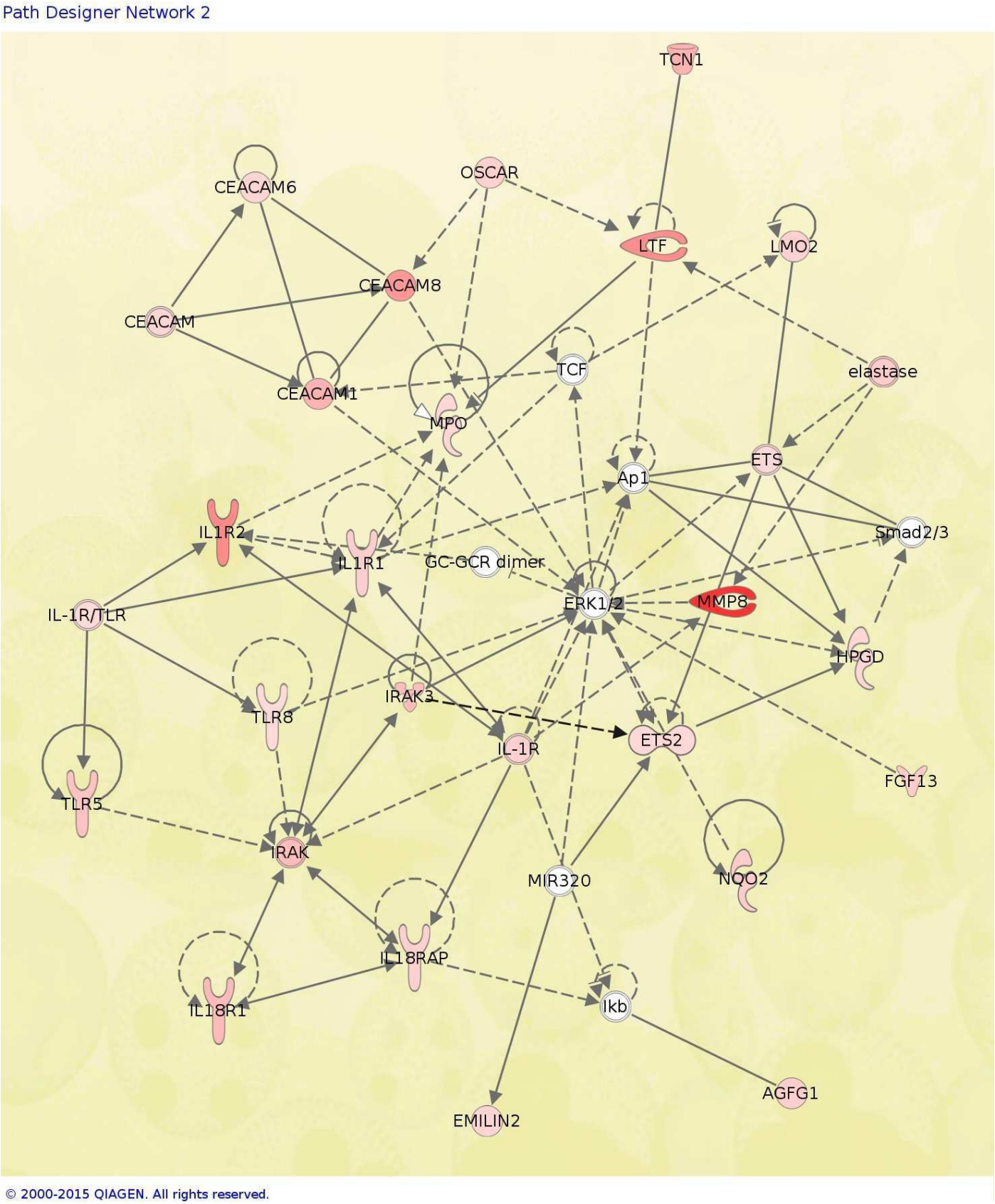
[1]National Institute of Biomedical Genomics, 741251, Kalyani, India
[2]Sriram Chandra Bhanja Medical College and Hospital, 753007,Cuttack, India
[3]Institute of Life Sciences,NALCO Square, 751023,Bhubaneswar, India
[*]skm1@nibmg.ac.in

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



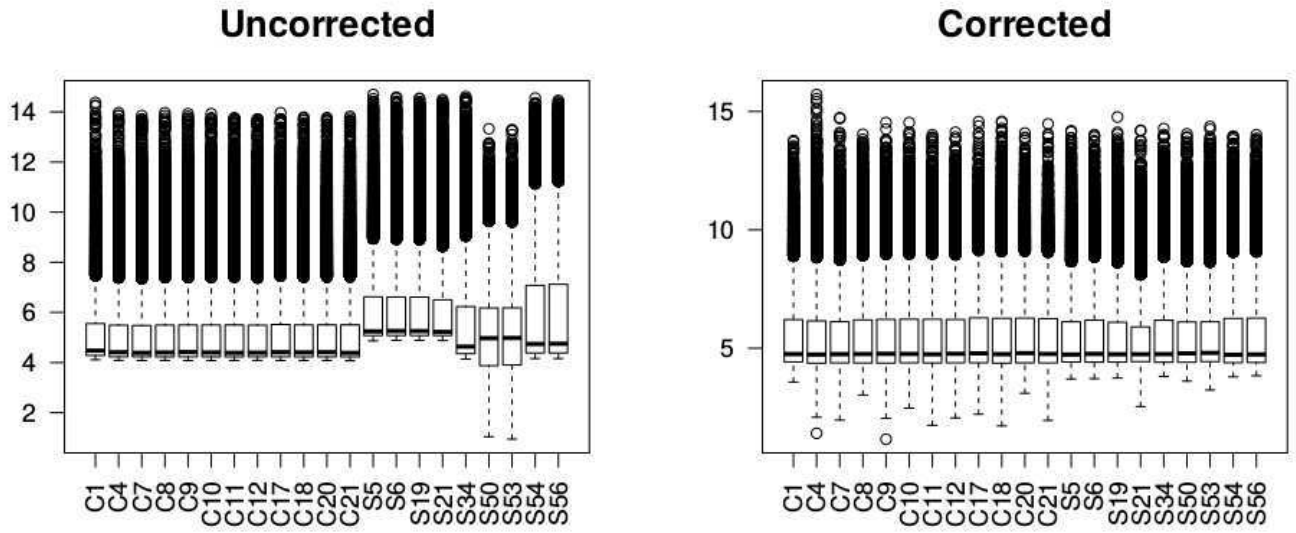
Supplementary Fig.S1: This is a dot plot representing log fold change in gene expression for all the filtered and quantile-normalised genes of each study. Green and Red dots represent down-regulated and up-regulated genes respectively. Any up-regulated gene, i.e., with expression higher in SS compared to healthy controls is shown as a red point, and a down-regulated gene is shown as a green point. The base line at 0 corresponds to "no change" in gene expression. The 2 dotted lines correspond to the threshold of 2-fold change (i.e., log-fold change of 1) in expression. Any point above the top dotted line represents a gene whose expression level is twice or more in SS compared with healthy controls. Similarly, any point below the bottom dotted line represents a gene whose expression level is at least half (or lower) in SS compared with healthy controls. Note that a change of 2 in the linear scale corresponds to a change of 1 in the log-2 scale. There are far more up-regulated genes (red points) than down-regulated genes (green points). Additionally there are many up-regulated genes with 2-fold or greater change in gene expression. This figure provides compelling visual evidence of high-intensity gene up-regulation in SS.

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

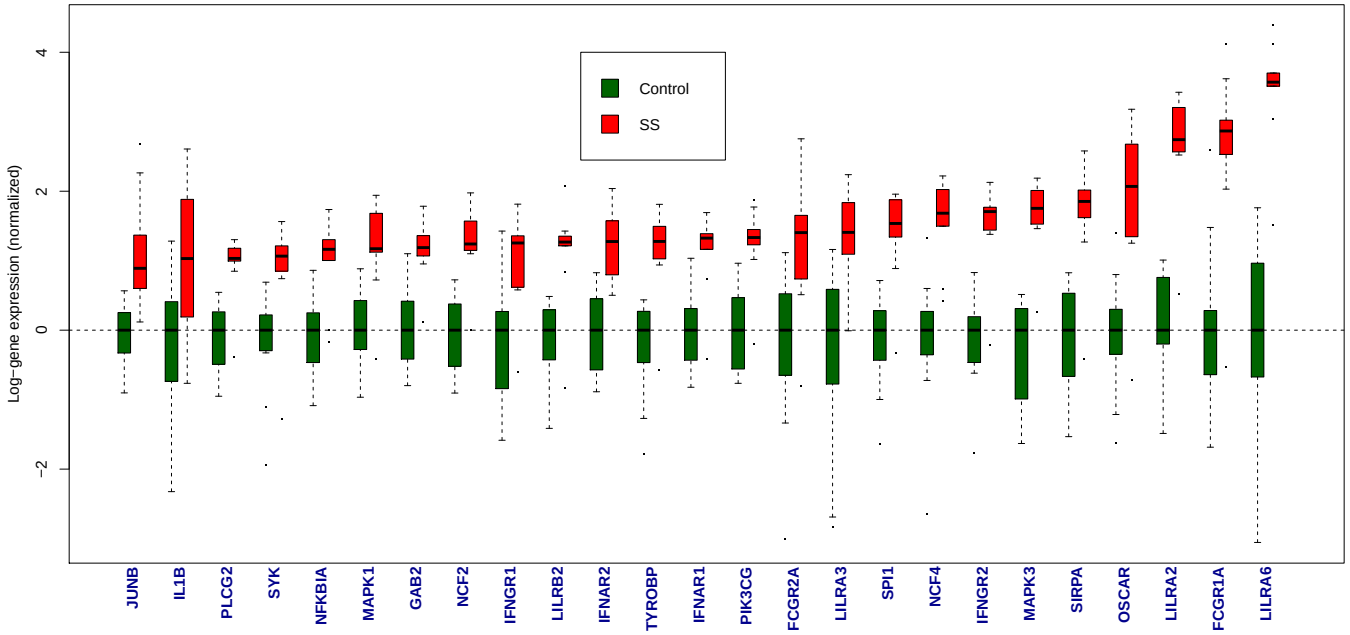


Supplementary Fig.S2: Network "Infectious Disease, Inflammatory Response, Connective Tissue Disorders", generated by Ingenuity Pathway Analysis with top 200 up-regulated genes with mean log-fold change in gene expression.

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Supplementary Fig.S3: ComBat function (of sva package) was used to remove the batch effect from validation cohort samples as shown in the box plots above, from left(Uncorrected) to right (Corrected).



Supplementary Fig.S4: Box plots of the highly significant 25 genes of the pathway hsa04380 up-regulated in SS. Green color corresponds to the control subjects while the red color corresponds to the cases of SS. Gene symbols are shown at the bottom. For each gene, log-intensity of gene expression has been normalized to the median expression of the control group.

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Biological function	Genes
Septic Shock[1],[2]	ANKRD22, CEACAM1, CLEC5A, G0S2, BCL-6, MMP8, TREM1, CCL4
Inflammation & Innate Immunity [3]	SERPINB1, S100A12, SLC11A1, TLR5, PADI4, CLEC5A, TLR8, LILRA5
Anti Microbial Activity [3]	ADM, BP1, DEFA4, RNASE3
Complement System [3]	C1QB, F5, CR1, C3AR1
Coagulation System [3]	CR1, C3AR1, CD59, SERPINA1
Carbohydrate Metabolism [4]	PFKB2, PFKB3, SLC2A3, GYG1, PYGL
Lipid Metabolism [5]	ACSL1, HGF, FAR2, DGAT2
Bone Metabolism [3]	CPD, MAPK14, FOSL2, IL1R1, MAP2K6, SOCS3, OSCAR, SIRPA, ALPL, BST1, CA4, TNFAIP6
Calcium Signalling [3]	S100A12, DYSF, MRVI1, LILRA5
Cytoskeletal And Cell Adhesion [3]	LIMk2, DSC2
Cancer-associated Genes [3]	BCL2A1, BCL6, CEACAM1, CD63, DACH1, DSC2, ELANE, ETS2, HPGD, IL1RN, LMO2, MMP8, MMP9, MPO, NQO2, IL1R2, PLXNC1, OLFM4, RAB20

Supplementary Table S1: Annotation of top 200 genes identified after ORA, as significantly up-regulated in septic shock into associated principal biological functions.

1 Mizock, B. A. (2001). Alterations in fuel metabolism in critical illness: hyperglycaemia., *Best Pract Res Clin Endocrinol Metab* 15 : 533-551.
2 Rival, T.; Cinq-Frais, C.; Silva-Sifontes, S.; Garcia, J.; Riu, B.; Salvayre, R.; Genestal, M. and Caspar-Bauguil, S. (2013). Alteration of plasma phospholipid fatty acid profile in patients with septic shock., *Biochimie* 95 : 2177-2181.
3 http://www.ncbi.nlm.nih.gov/gene/
4 Wong, H. R. (2013). Genome-wide expression profiling in pediatric septic shock., *Pediatr Res* 73 : 564-569.
5 Wong, H. R.; Shanley, T. P.; Sakthivel, B.; Cvijanovich, N.; *et al.* (2007). Genome-level expression profiles in pediatric septic shock indicate a role for altered zinc homeostasis in poor outcome., *Physiol Genomics* 30 : 146-155.

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

KEGG ID	p	odds ratio	Pathway Name
hsa04610	0.00001	11.5	Complement and coagulation cascades
hsa04380	0.00006	6.59	Osteoclast differentiation
hsa05202	0.00046	4.77	Transcriptional misregulation in cancer

Supplementary Table S2: Top 3 enriched KEGG pathway selected after Over Representation Analysis with top 200 genes.

ID	Age	SEX	Primary Diagnosis	Vasopressors	Infective focus
5	70	M	Septic Shock	Dopamine	NA
6	55	F	Septic Shock	Dopamine	Right middle lobe pneumonia
19	14	F	Septic Shock	NA	Right lower lobe pneumonia
21	20	M	Septic Shock	NA	NA
34	32	M	Septic Shock	Dopamine	Left lower lobe pneumonia
50	38	M	Septic Shock	Noradrenaline	Right leg abscess
53	75	M	Septic shock	Dopamine	Right lower lobe pneumonia
54	38	F	Septic Shock	Noradrenaline	Right lower lobe pneumonia
56	63	F	Septic shock	Dopamine	NA

Supplementary Table S3: Clinical characteristics of SS patients in Validation Cohort.

bioRxiv preprint doi: https://doi.org/10.1101/051706; this version posted May 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Entrez Ids	Gene Symbols	Gene expressed in Tissues[1,2]	Protein levels in plasma [2,3]
2209	FCGR1A	NK Cells	Detected
2534	FCGR2A	Monocyte, Whole Blood	Detected
9846	GAB2	CD14+ Monocyte, Whole Blood	Detected
3454	IFNAR1	NK Cells	Detected
3455	IFNAR2	Dendritic Cells	Detected
3459	IFNGR1	Whole Blood, Monocyte	Detected
3460	IFNGR2	NK cells , Monocyte, Whole Blood	Detected
3552	IL1A	Monocyte	Detected
3726	JUNB	Myeloid	Detected
11027	LILRA2	Monocyte	Detected
11026	LILRA3	Monocyte	Detected
79168	LILRA6	Monocyte	Detected
10288	LILRB2	Monocyte	Detected
5594	MAPK1	CD56+ NK cells, Monocyte, Whole Blood	Detected
5595	MAPK3	Whole Blood, Monocyte	Detected
4688	NCF2	Whole Blood, Monocyte	Detected
4689	NCF4	Whole Blood, Monocyte	Detected
4792	NFKBIA	CD33+ Myeloid	Detected
126014	OSCAR	Neutrophil	Detected
5294	PIK3CG	NK cells, Monocyte, Whole Blood	Detected
5336	PLCG2	NK Cells	Detected
140885	SIRPA	Whole Blood	Detected
55423	SYK	Whole Blood, Monocyte	Detected
6688	SPI1	Whole Blood, Monocyte	Detected
7305	TYROBP	Monocyte	Detected

Supplementary Table S4: Annotation of 25 genes of the pathway hsa04380 that are significantly up-regulated in SS.

[1] http://www.biogps.org
[2] http://www.genecards.org/
[3] http://www.plasmaproteomedatabase.org/