www.sciencetranslationalmedicine.org/cgi/content/full/7/287/287ra71/DC1

Supplementary Materials for

A comprehensive time-course–based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set

Timothy E. Sweeney,* Aaditya Shidham, Hector R. Wong, Purvesh Khatri*

*Corresponding author. E-mail: tes17@stanford.edu (T.E.S.); pkhatri@stanford.edu (P.K.)

Published 13 May 2015, *Sci. Transl. Med.* **7**, 287ra71 (2015)
DOI: 10.1126/scitranslmed.aaa5993

This PDF file includes:

Materials and Methods
Fig. S1. Labeled PCA comparing healthy controls, SIRS/trauma patients, and sepsis patients.
Fig. S2. Violin plots for the data sets included in the discovery multicohort analysis.
Fig. S3. Neutrophil percentages for the Glue Grant patients with both complete blood count and microarray data.
Fig. S4. Performance of the infection *z* score in the sorted monocytes from the Glue Grant cohort.
Fig. S5. Performance of the infection *z* score in the sorted T cells from the Glue Grant cohort.
Fig. S6. Linear models of SIRS criteria and the infection *z* score.
Fig. S7. The infection *z* score in non–time-matched data sets.
Fig. S8. Comparison of the infection *z* scores in patients with acute infections to healthy controls and patients with autoimmune diseases.
Fig. S9. Ingenuity Pathway Analysis results for the 11-gene set.
Fig. S10. Schematic of the entire integrated multicohort analysis.
Legends for tables S1 to S4
Table S5. Linear models of infection score in the Glue Grant data.
Table S6. Comparison of infection *z* score across infection types.
Table S7. In silico transcription factor binding analyses for the 11-gene set.
Legend for table S8
Legend for movie S1
References (*55*–*79*)

Other Supplementary Material for this manuscript includes the following:

(available at www.sciencetranslationalmedicine.org/cgi/content/full/7/287/287ra71/DC1)

Table S1 (Microsoft Excel format). Summary spreadsheet of all data sets referenced in the manuscript.
Table S2 (.csv format). Design matrix of individual phenotypes for multicohort analysis.
Table S3 (.csv format). Summary statistics for the 82 genes that passed significance, heterogeneity, and effect size filtering after multicohort analysis.
Table S4 (Microsoft Excel format). Probe-level data for all 11 genes in the diagnostic set for all patients in the multicohort analysis.
Table S8 (.csv format). Design matrix for cell type enrichment analyses.
Movie S1 (.mp4 format). Rotation of a time-course–labeled PCA of trauma patients.

Supplemental Materials and Methods

Study Design

The purpose of this study was to use an integrated multicohort meta-analysis framework to analyze multiple gene expression data sets to identify a set of genes that can separate patients with sterile inflammation from patients with infectious inflammation. This framework has been described previously (*41, 42*).

Search

Two public gene expression microarray repositories (NIH GEO, ArrayExpress) were searched for all human data sets that matched any of the following search terms: sepsis, SIRS, trauma, shock, surgery, infection, pneumonia, critical, ICU, inflammatory, nosocomial. Data sets that compared either healthy controls or patients with non-infectious inflammation (SIRS, trauma, surgery, autoimmunity) to patients with acute infections and/or sepsis were kept for further study. Data sets that utilized endotoxin injection as a model for SIRS or sepsis were not included.

Data set Details

Six of the publicly available whole blood data sets were from the Genomics of Pediatric SIRS/Septic Shock Investigators (GPSSSI) (*15-20*). These data sets contain overlapping samples, for which Hector Wong provided a key of the unique patients. Those unique patients were then gcRMA normalized together and treated as a single data set (GPSSSI Unique).

In addition to the publicly-available data sets, we used the Inflammation and Host Response to Injury Program (Glue Grant) trauma data sets (*6, 12, 13*). The Glue Grant data sets consist of separate trauma patient cohorts sampled for either the entire buffy coat or sorted cells (neutrophils, monocytes, T-cells). Inclusion criteria are described elsewhere (*14*). Patients were sampled at the following days after admission: 0.5, 1, 4, 7, 14, 21, 28 days. The Glue Grant trauma cohort patients were classified as ‘infected’ if they had a nosocomial infection (pneumonia, urinary tract infection, catheter-related bloodstream infection, etc.), a surgical infection (excluding superficial wound infections), or underwent surgery for perforated viscus; infection definitions can be found at http://www.gluegrant.org/commonlyreferencedpubs.htm.

For meta-analyses, samples drawn within +/- 24 hours of the day of diagnosis of infection were included as infection cases. Time points with fewer than 20 patients were not included in the multicohort analysis. The Glue Grant also contains burn patients, but these were not included due to the difficulty of distinguishing clinically relevant infections from colonized burn wounds. Use of the Glue Grant was approved by both the Glue Grant Consortium and the Stanford University IRB (protocol 29798).

Gene Expression Normalization

All Affymetrix data sets were downloaded as CEL files and re-normalized using gcRMA (R package affy). Output from Agilent chips and custom arrays analyzed on GenePix scanners was background corrected, within-arrays loess normalized, and then between-arrays quantile normalized (R package limma). Illumina data sets were quantile normalized. The Glue Grant sorted-cell data sets were analyzed using custom arrays (GGH-1, GGH-2); these were normalized as previously described and used in their post-processed state (*55*). For all gene analyses, the mean of probes for common genes was set as the gene expression level. All probe-to-gene mappings were downloaded from GEO from the most current SOFT files on Dec 14, 2014.

Labelled PCA Method

The labelled principal components analysis (PCA) method is an implementation of the constrained optimization described in equations 6 and 7 in section 4.1 of Koren and Carmel (*56*). This optimization computes a linear transformation of the data that maximizes the pairwise distance between points in different labelled classes of the data while maintaining the constraint that the transformed data are orthogonal to each other. This orthogonal constraint is slightly different to the constraint employed by PCA, which demands that the transformed basis is mutually orthogonal, not the transformed data itself. While PCA is a projection scheme, labelled PCA is a general form of a linear transformation due to this difference in constraint.

Call X the original data set, with m rows (data points) and n columns (each data point has n elements). Y is an m by 1 matrix that has a different listing for each class. In other words, Y (i) equals Y (j) if and only if elements i and j are part of the same labelled class. L is a symmetrical m by m matrix whose (i,j) entry is -1, unless Y (i) equals Y (j). In this latter case, the entry is 0.

Finally, all diagonal entries (where i equals j) are filled so that row i sums to 0. Koren and Carmel prove in Lemma 3.2 that the eigenvectors of transpose (X)*L*X provide a mapping that maximizes the pairwise distance between points in different labelled classes of the data. However, this transformation remains a projection scheme, which means that these eigenvectors are orthogonal to each other. This latter result limits the utility of the transformed data, but is more generalizable. The general linear projection used in this paper instead finds the vectors v that solve the equation $Av=\lambda Bv$, where A is transpose (X)*L*X, and B is transpose (X)*X. Although more expressive, this method is not as robust as the labelled PCA projection scheme, however, since solutions to this generalized form require that B is not singular. Since B is not the identity matrix, the old orthogonal constraint used in projections does not have to hold. Instead, solutions to this form require that the basis is mutually orthogonal with respect to the covariance basis of the original data.

Labelled PCA Applications

All data sets that contain a comparison of non-infectious SIRS, ICU, or trauma patients to sepsis patients were converted from probes to genes, and then bound into a single large matrix and quantile normalized. Genes not present in all data sets were thrown out. Patients with sepsis at any time (either on admission or hospital-acquired) were grouped in a single class, and Lasso-penalized regression was applied to separate sterile SIRS patients from sepsis patients (R package glmnet). Labelled PCA was carried out using the genes selected by the penalized regression, on the classes of sterile SIRS versus sepsis. The same graph was then re-labelled to show which samples are from hospital-acquired (or late) sterile SIRS or sepsis patients. The same set of genes from the penalized regression was then used in labelled PCA to compare healthy, sterile SIRS, and sepsis patients. This same graph was then re-labelled to show which samples are from hospital-acquired or late sterile SIRS or sepsis patients.

To examine the effects of time on gene expression in SIRS/trauma and infection, all data sets that include serial measurements over time were selected. From the Glue Grant data sets, only buffy coat arrays were included, so as not to overwhelm the signal from the other data sets. The selected data sets were converted from probes to genes, and then bound into a single large matrix and quantile normalized. Genes not present in all data sets were thrown out. To reduce the gene set in an unbiased manner, CUR matrix decomposition was used to select the top 100 genes

with the greatest orthogonality in the combined data sets (R package rCUR) (*57*). Labelled PCA was then carried out with each time point used as a different class (split at 1, 2, 3, 4, 5, 6, 10, 20, and 40 days). The resulting PCA was graphed in 3D, colored by time point, and a short video of rotations of the 3D space was captured using R package rgl.

Multicohort Analysis

We performed a multicohort analysis (*41, 42*) comparing gene expression in non-infected SIRS/trauma patients versus patients with infections or sepsis. All data sets with comparisons of SIRS/trauma patients to septic/infected patients at the same time point were selected for inclusion in the multicohort analysis; thus, comparisons of patients at admission to those with sepsis at a later time-point were excluded (see justification for this model in the Results). The admission data sets were limited to samples from patients within 48 hours of admission. The Glue Grant trauma data sets were split into time bins of days since injury, excluding the initial 24 hours after admission (see Supplemental Materials and Methods). Each of these time bins was treated as a separate data set in the multicohort analysis, where time-matched never-infected patients were compared to patients within +/- 24 hours of diagnosis of infection (infection as defined above). Patients who were diagnosed with infection more than 24 hours earlier are thus censored in this comparison. This method allows for detection of deviation due to infection from the 'standard' changes in gene expression over time due to recovery from trauma. A design matrix for the multicohort analysis is shown in Supplemental Table 2.

After selecting the input data sets, we applied two meta-analysis methods: one combining effect sizes using Hedges' g, the other using Fisher's sum of logs method combining p-values (see schematic in Supplemental Figure 10). Given n data sets, this method is applied n times in a leave-one-data set-out fashion. A false discovery rate (FDR) threshold was set (0.01), and genes with a q-value below the FDR threshold in both the effect size and the Fischer's sum of log analyses at every round of the leave-one-out analysis were selected. The genes were then subjected to a data set heterogeneity test, such that in a test for heterogeneity across all input data sets, a p-value greater than 0.01 was required for each gene; this removes genes that show significantly different effects across different data sets. Next, all genes with a summary effect size <1.5 fold were removed. Finally, all genes found to be statistically significant in the multicohort analysis according to all three of the above criteria (Supplemental Table 3) were subjected to a greedy forward search model, where, starting with the most significant gene, all

remaining genes are added to the gene score one at a time, and the gene with the greatest increase in discriminatory ability is added to the final gene list. The probe-level data for the genes remaining after forward search are shown in Supplemental Table 4.

Infection *z* score

Genes that were found to be significant after multicohort analysis were separated according to whether their effects were positive or negative (where 'positive' means a positive effect size in sepsis as compared to SIRS/trauma, and 'negative' means a negative effect size in sepsis as compared to SIRS/trauma). The class discrimination power of these gene sets was then tested using a single gene score. The gene score used is the geometric mean of the gene expression level for all positive genes minus the geometric mean of the gene expression level of all negative genes multiplied by the ratio of counts of positive to negative genes. This was calculated for each sample in a data set, and the scores for each data set were then standardized to yield a *z* score ('infection *z* score'). Genes not present in an entire data set were excluded; genes missing for individual samples were set to 1. To obtain an infection *z* score for data sets with negative gene expression values (two-channel arrays), the entire data set was scaled by the minimum value present in the data set, to ensure all values were positive (since the geometric mean yields imaginary values for negative input).

Class discriminatory power was examined comparing the infection *z* scores for classes of interest in each examined data set. The infection *z* score ranges were examined with violin plots, and, since they cannot be assumed to have normal distributions, are shown with interquartile range and compared using Wilcoxon rank-sum test. ROC curves of the infection *z* score were constructed within data sets compared to sepsis, and the area under the curve (AUC) was calculated using the trapezoidal method.

Forward Search

To obtain a parsimonious gene set that discriminates SIRS/trauma patients and septic/infected patients, all genes found to be statistically significant in the multicohort analysis were subjected to a greedy forward search model, where, starting with the most significant gene in the data set, all remaining genes are added to the gene score one at a time, and the gene with the greatest increase in discriminatory ability is added to the final gene list. Here, discriminatory

ability was defined as a weighted ROC AUC, wherein the infection *z* score is tested in each discovery data set, and the resulting AUC is multiplied by the total number of samples in the data set. The function then maximizes the sum of weighted AUCs across all discovery data sets for each step. In this way, excellent class discrimination in a small data set does not outweigh modest gains in class discrimination in a very large data set. The function stops at an arbitrarily defined threshold; we used a stopping threshold of 0.5 (such that when the function cannot find a gene that will increase the total discovery weight AUCs of the current infection *z* score by more than 0.5, it will terminate). This final resulting gene set is thus maximized for discriminatory power in the discovery cohorts, though is not optimized as a global maximum. The probe-level data for the genes remaining after forward search are shown in Supplemental Table 3.

Discovery Cohort Examinations

The final gene score was used to compute infection *z* scores in each discovery data set. The admission data sets were analyzed separately and separate ROC plots plotted. For the hospital-acquired (Glue Grant) data sets, infection scores were standardized (converted into *z* scores) once for the whole cohort as opposed to normalizing the different time-bins separately to show changes over time in the same patients. The infection *z* scores were then analyzed for significance using repeated-measures analysis of variance. ROC curves were plotted for the individual time-bins treated as separate data sets in the multicohort analysis.

For the Glue Grant data sets, two time-course analyses of infection *z* score were carried out for both the buffy coat and neutrophil data sets. First, the average infection *z* score was compared over time using linear regression for patients within +/- 24 hours of infection and for non-infected patients. Repeated-measures analysis of variance was used to compare infected and non-infected groups to each other and to test for the significance of changes over time. Next, boxplots were constructed for each time window, such that the infection *z* scores for the patients in that time window who were never infected were compared to patients at >5 days prior to their day of diagnosis with infection, 5-1 days prior to diagnosis, or +/- 24 hours of diagnosis. For each time point (except for the 0-1 day window), the trend in infection *z* score across the different groups was tested with the Jonckheere trend (JT) test. The infection *z* scores at the admission time point ([0,1)) were tested as the outcomes variable in multiple linear regression, examining the contributory effects of both injury severity score and time to infection.

Validation

The final gene set was tested in several validation cohorts completely separate from the discovery cohorts. The sorted-cells cohort of the Glue Grant was broken into time bins, and AUCs were calculated separately for each time bin. Note that no infections within +/- 1 day of diagnosis were captured in this cohort after 18 days from the time of injury, so the [18,24) day bin is never shown.

The validation cohorts included three data sets that examined trauma patients over time (GSE6377, GSE12838, and EMEXP3001), all of whom developed infections (mostly ventilator-associated pneumonia). These data sets do not include controls, and so they were compared to the Glue Grant non-infected patients as a baseline. These three validation data sets and the Glue Grant buffy coat non-infected samples were first linearly scaled by a factor of the geometric mean of four housekeeping genes (*GAPDH, ACTN1, RPL9, KARS*) (*58*). The data sets were then joined on overlapping genes, and batch-corrected between data sets using the ComBat empiric Bayes batch-correction tool, with parametric priors (R Package sva) (*59*). The ComBat correction was controlled for day after injury (so that relative differences between days stay relatively different). The infection *z* score was then calculated for the joined data sets, and the validation data sets were plotted against the loess curve from the non-infected Glue Grant cohort. Patients within +/- 24 hours of their diagnosis of infection in the validation data sets were then compared to day-matched ComBat-co-normalized non-infected Glue Grant buffy coat patients, and ROC curves were constructed.

All other data sets found in the initial search that allow for comparison between healthy or SIRS/trauma and sepsis patients were used for simple class discrimination validation. All data sets obtained from whole blood or neutrophils are shown. Studies carried out in PBMCs were selected for only those that examined SIRS/trauma and sepsis patients. Data sets using PBMC samples that did not include both a sterile SIRS group and a sepsis group were excluded. All peripheral blood healthy vs sepsis patient data sets were grouped into a single violin plot and tested jointly for separation (Wilcoxon rank-sum) since they were all being used to make the same comparison. ROC curves were carried out on each individual data set separately to show the discriminatory capability of the infection *z* scores within each data set.

Glue Grant SIRS Evaluation

To evaluate the effectiveness of SIRS as screening criteria for infection in the Glue Grant cohort, all patients were classified as either non-infected or within +/- 24 hours of infection, with infection as defined above. Patients were censored >24 hours after infection diagnosis. SIRS criteria were defined according to standard international guidelines (Temperature <36C or >38C, respiratory rate >20 or PaCO2<32, total WBC <4,000 or >12,000, and HR >90). Patients missing any criteria were excluded. Each criterion was stored as a binary variable for each patient for each day. Logistic regression was run on the data both with and without inclusion of the *z* score, and ROC AUC was calculated for both models. The two models were then compared using the continuous net reclassification index (R package PredictABEL).

Gene Set Evaluation

The final gene set was evaluated for transcription factor binding sites using two online tools, EncodeQT (*60*) and PASTAA (*61*). Positive and negative genes were evaluated separately, since they are hypothesized to be under separate regulatory control. The EncodeQT tool was used with 5000 upstream and 5000 downstream base pairs from transcription start site. A similar analysis was carried out with PASTAA, examining the region -200 base pairs from transcription start site, only for those factors which were conserved for both mouse and human. The top ten significant transcription factors were recorded for both analyses.

Cell-Type Enrichment Tests

GEO was searched for gene expression profiles of clinical samples of relevant immune cell types. The search was limited to only samples run on Affymetrix platforms, to ensure platform effect homogeneity. All data sets used were downloaded in RAW format and gcRMA normalized separately. For each sample, the mean of multiple probes mapping to the same gene was taken as the gene value. Genes not present in all samples were thrown out. For multiple samples all corresponding to the same cell type, the mean of the samples was taken as the final value, thus creating a single vector for each cell type. To obtain a *z* score for a gene set in each cell type vector, we took the geometric mean of the ‘positive’ genes’ expression, and subtracted from it the geometric mean of the ‘negative’ genes’ expression, times the ratio of negative genes to positive genes (same procedure as for the infection *z* score). These scores are then

standardized across all cell types, such that the score represents the number of standard deviations away from the group mean. This thus represents how enriched a given gene set is in a given cell type, relative to other tested cell types.

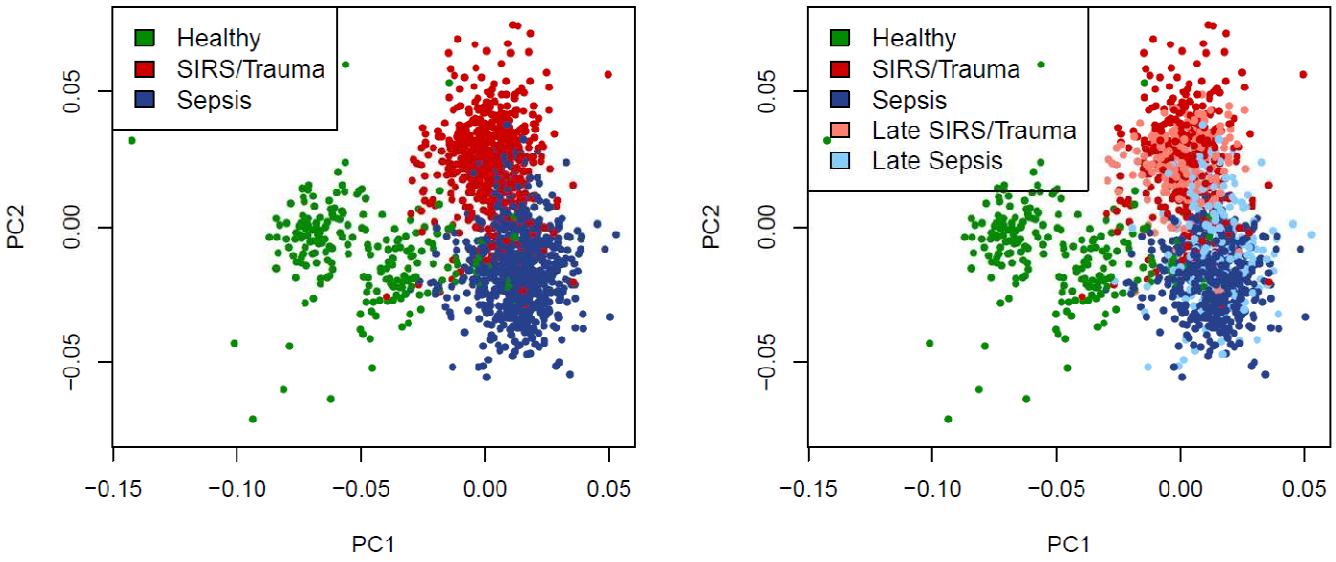
A total of 18 GEO data sets that matched criteria were used: GSE3982 (*62*), GSE5099 (*63*), GSE8668 (*64*), GSE11292 (*65*), GSE12453 (*66*), GSE13987 (*67*), GSE14879 (*68*), GSE15743 (*69*), GSE16020 (*70*), GSE16836 (*71*), GSE24759 (*72*), GSE28490 (*73*), GSE28491 (*73*), GSE31773 (*74*), GSE34515 (*75*), GSE38043 (*76*), GSE39889 (*77*), GSE42519 (*78*), GSE49910 (*79*); see Supplemental Table 8 for cell type mapping design matrix.

Two gene sets were tested in this manner: both the entire set of genes found to be significant after the initial multicohort analysis, and the subset of genes found to be most diagnostic after forward search. Figure 7 shows the *z* score (enrichment for the given gene set) in each cell subtype (black dots), as well as a box plot for the overall distribution of *z* scores (shown in red).

Statistics and R

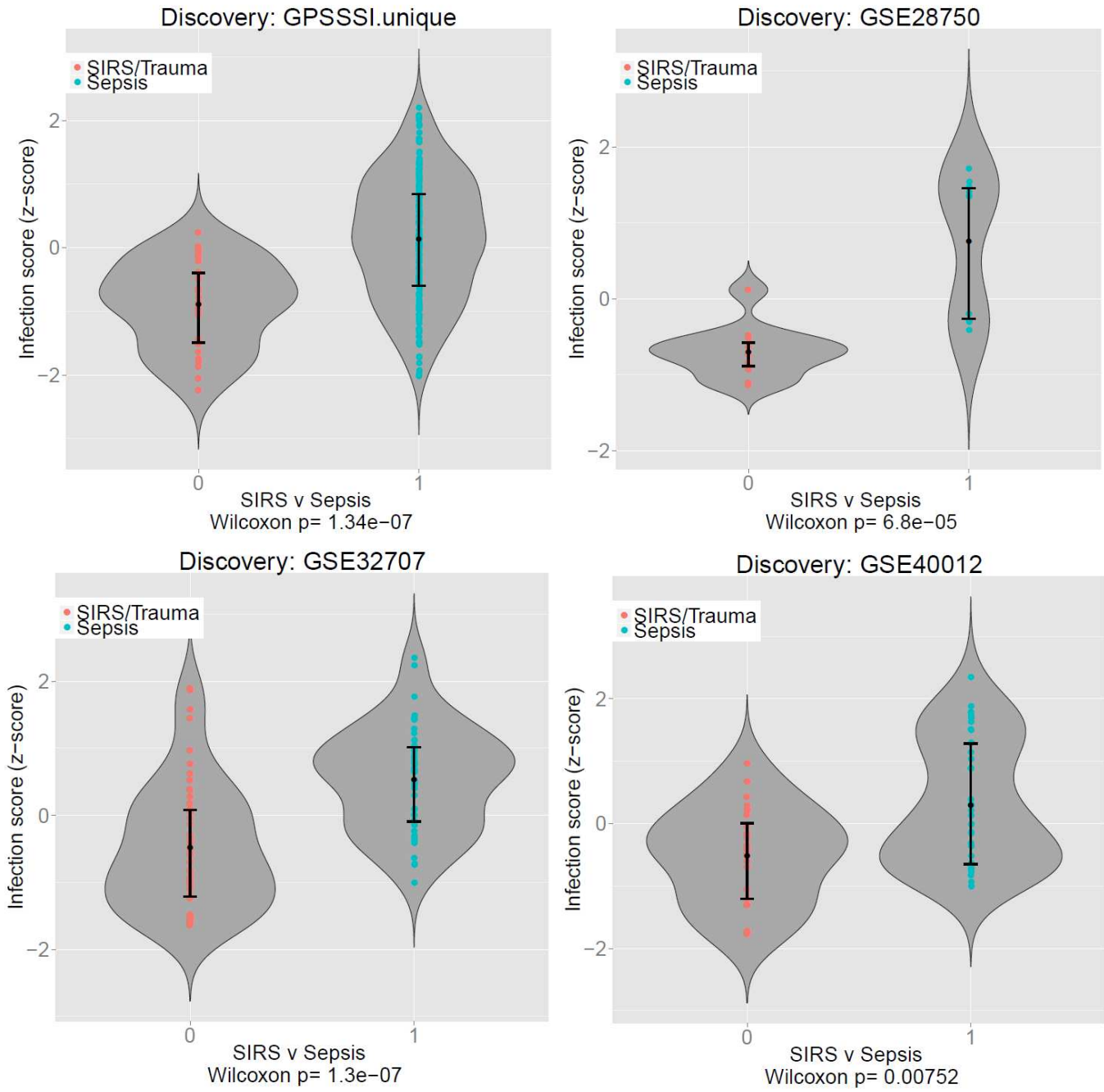
All computation and calculations were carried out in the R language for statistical computing (version 3.0.2). Significance levels for p-values were set at 0.05, and analyses were two-tailed, unless specified otherwise.

Supplemental Figures

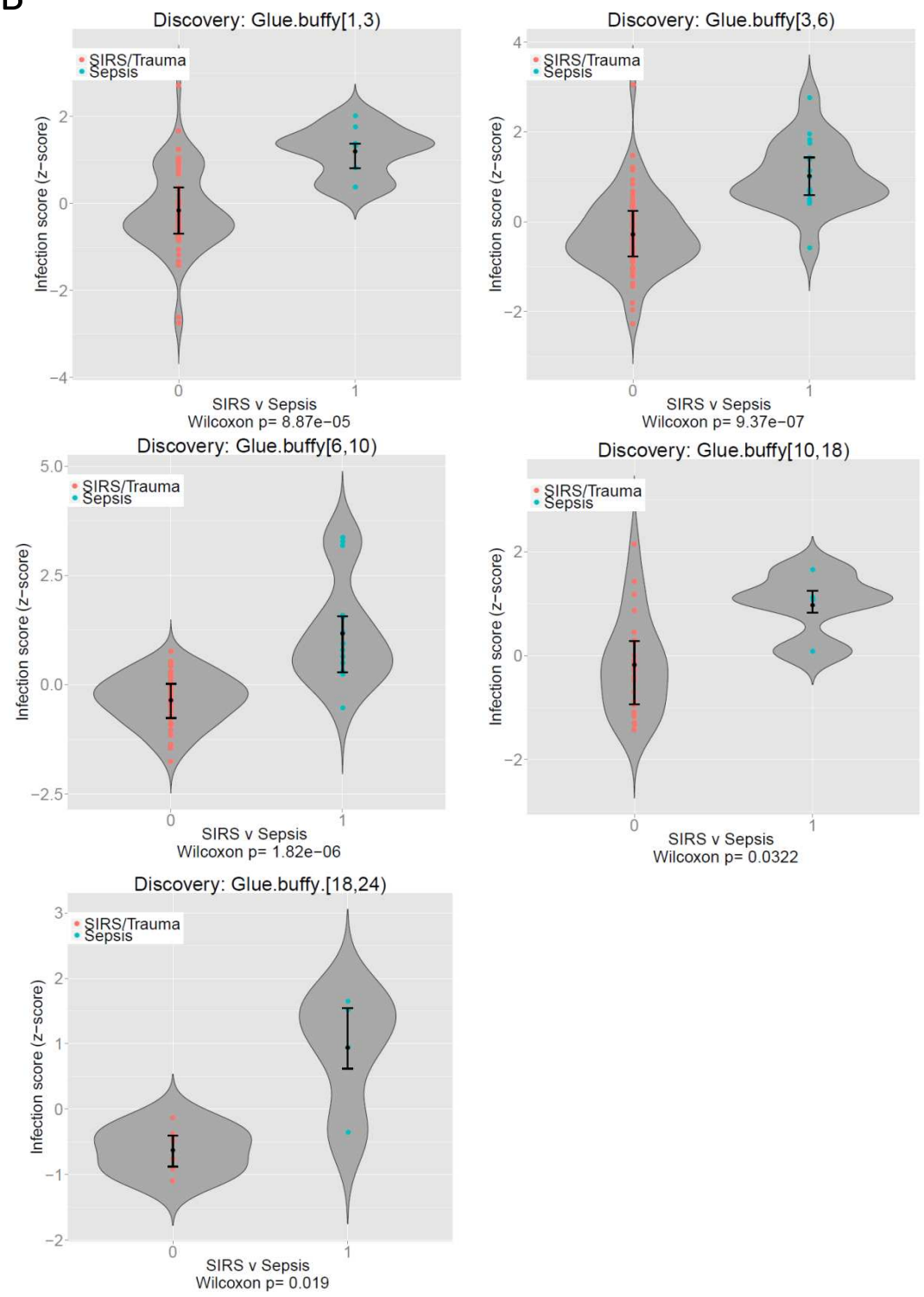


Supplemental Figure 1. Labelled PCA comparing healthy controls, SIRS/trauma patients, and sepsis patients. A. Healthy controls and SIRS/trauma and sepsis patients appear to be largely separable in the transcriptomic space, with only a minimal non-separable set. B. The same labelled PCA is shown, with labels updated to reflect patients in recovery from non-infectious SIRS/trauma and patients with hospital-acquired sepsis; the 'late' group (>48 hours after hospital admission) is much harder to separate. N= 1316 combined from 15 studies.

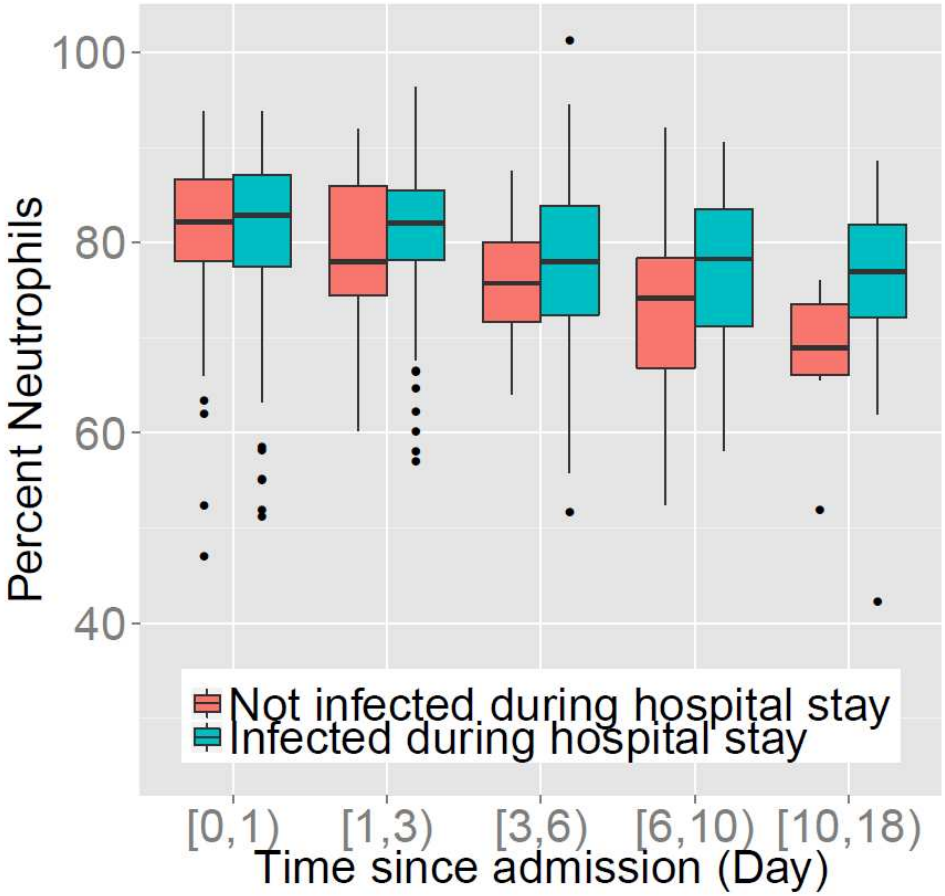
A


Discovery: GPSSSI.unique
SIRS/Trauma
Sepsis
Infection score (z−score)
2
0
−2
0
1
SIRS v Sepsis
Wilcoxon p= 1.34e−07
Discovery: GSE28750
SIRS/Trauma
Sepsis
Infection score (z−score)
2
0
−2
0
1
SIRS v Sepsis
Wilcoxon p= 6.8e−05
Discovery: GSE32707
SIRS/Trauma
Sepsis
Infection score (z−score)
2
0
−2
0
1
SIRS v Sepsis
Wilcoxon p= 1.3e−07
Discovery: GSE40012
SIRS/Trauma
Sepsis
Infection score (z−score)
2
0
−2
0
1
SIRS v Sepsis
Wilcoxon p= 0.00752

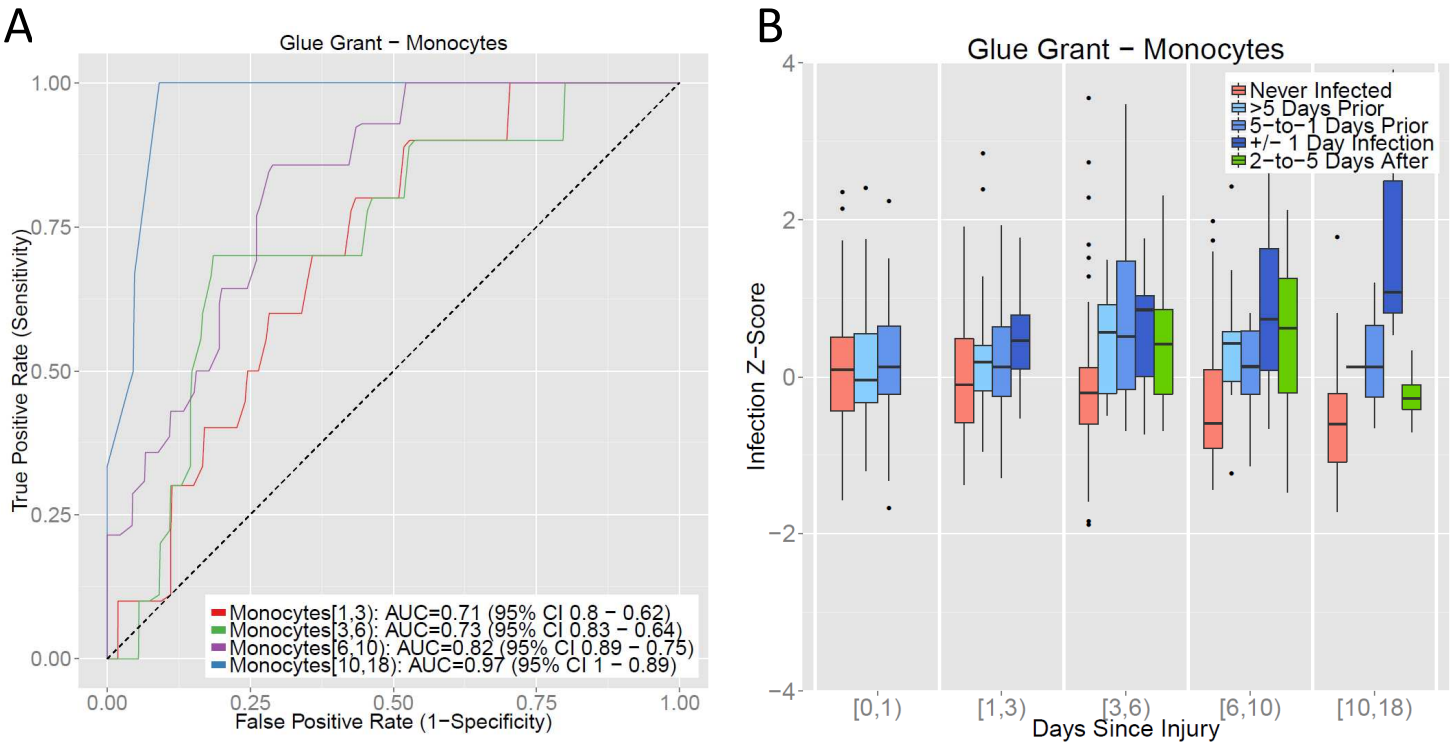
B



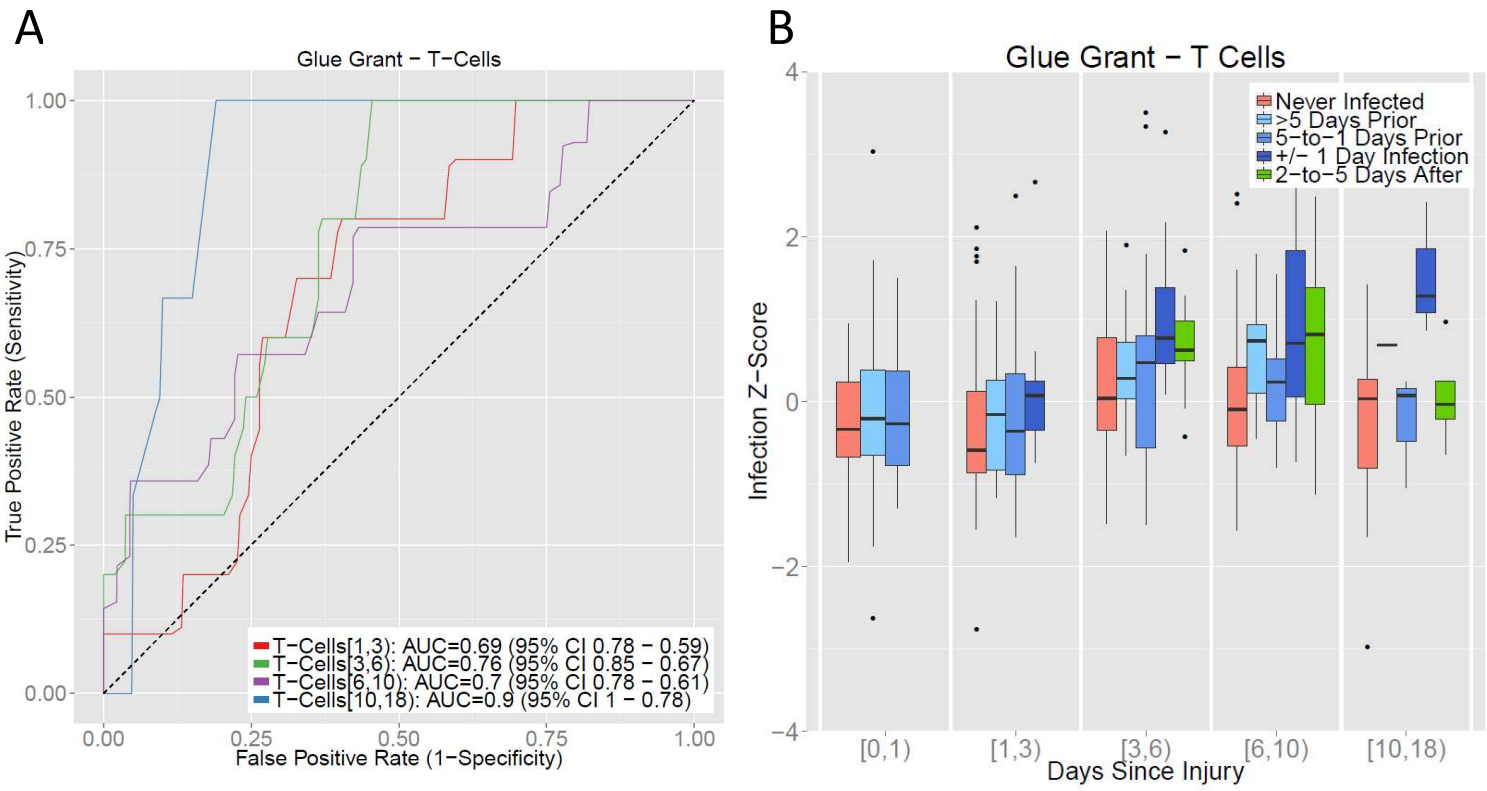
Supplemental Figure 2. Violin plots for the data sets included in the discovery multicohort analysis. (A) Data sets comparing SIRS/ICU/trauma to sepsis patients at admission. (B) Glue Grant Buffy Coat cohorts, comparing non-infected trauma patients to sepsis patients at matched time points. Error bars show middle quartiles. P-values from Wilcoxon rank-sum test.



Supplemental Figure 3. Neutrophil percentages for the Glue Grant patients with both complete blood count and microarray data. Median neutrophil percentage is between 75-85% for all time points. Patients who were ever infected during their hospital stay are compared to patients never infected during their hospital stay.



Supplemental Figure 4. Performance of the infection *z* score in the sorted monocytes from the Glue Grant cohort. These are the same patients as the neutrophils validation cohort in Figure 4(B,D,F). (A) ROC curves for each of the four sampled time bins. (B) Boxplots of infection *z* score by time since injury. Patients never infected are compared to patients >5 days prior to infection, 5-to-1 days prior to infection, within +/- 1 day of diagnosis (cases), and 2-to-5 days after infection.



Supplemental Figure 5. Performance of the infection *z* score in the sorted T cells from the Glue Grant cohort. These are the same patients as the neutrophils validation cohort in Figure 4(B,D,F). (A) ROC curves for each of the four sampled time bins. (B) Boxplots of infection *z* score by time since injury. Patients never infected are compared to patients >5 days prior to infection, 5-to-1 days prior to infection, within +/- 1 day of diagnosis (cases), and 2-to-5 days after infection.

A.

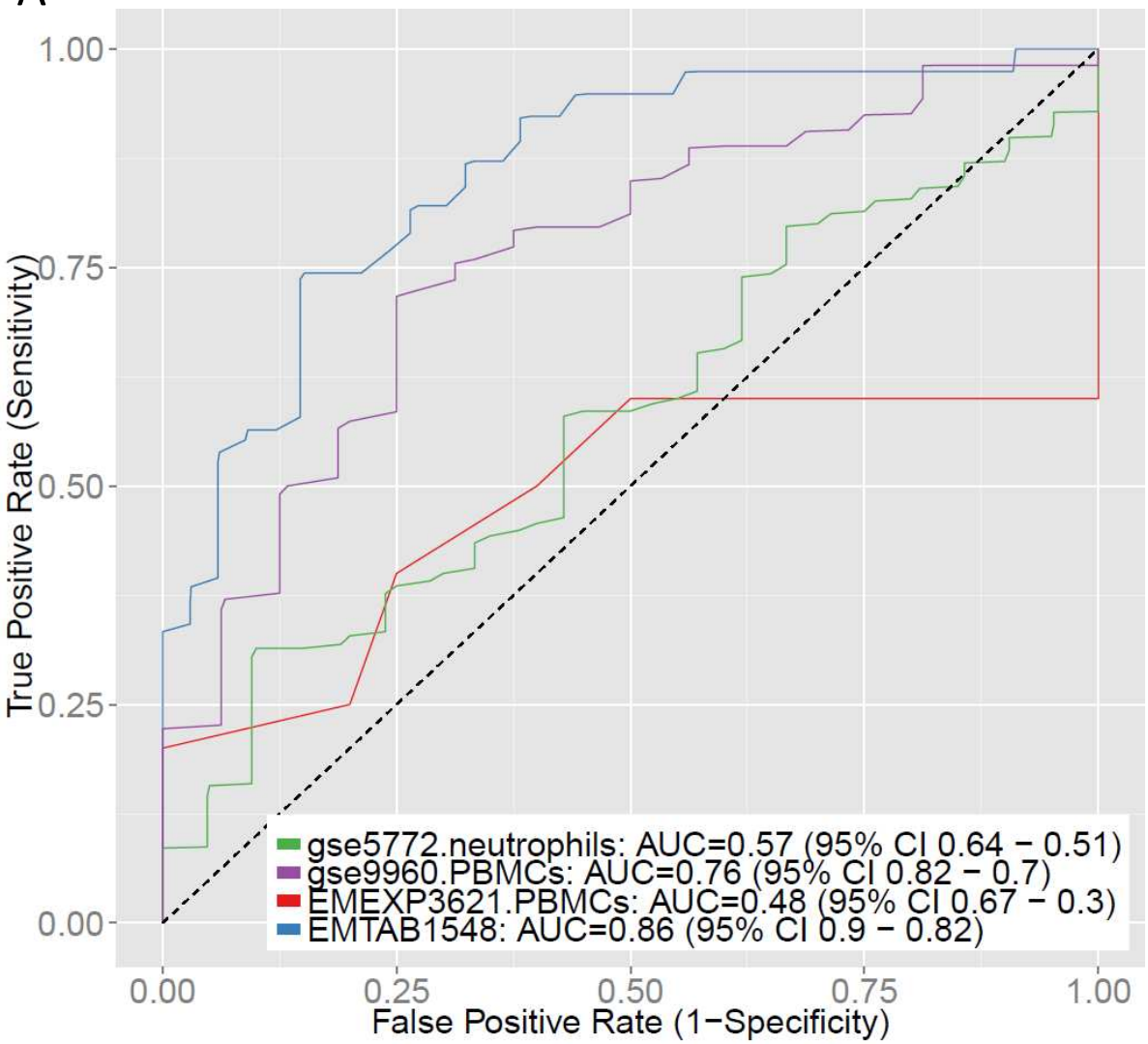
Binary SIRS Criteria Only					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-1.77613	0.47528	-3.737	0.000186	***
SIRS.Temp	1.0799	0.34552	3.125	0.001776	**
SIRS.Resp	0.02356	0.32952	0.072	0.942998	
SIRS.WBC	0.55216	0.3269	1.689	0.0912	
SIRS.HR	0.05884	0.44389	0.133	0.894549	

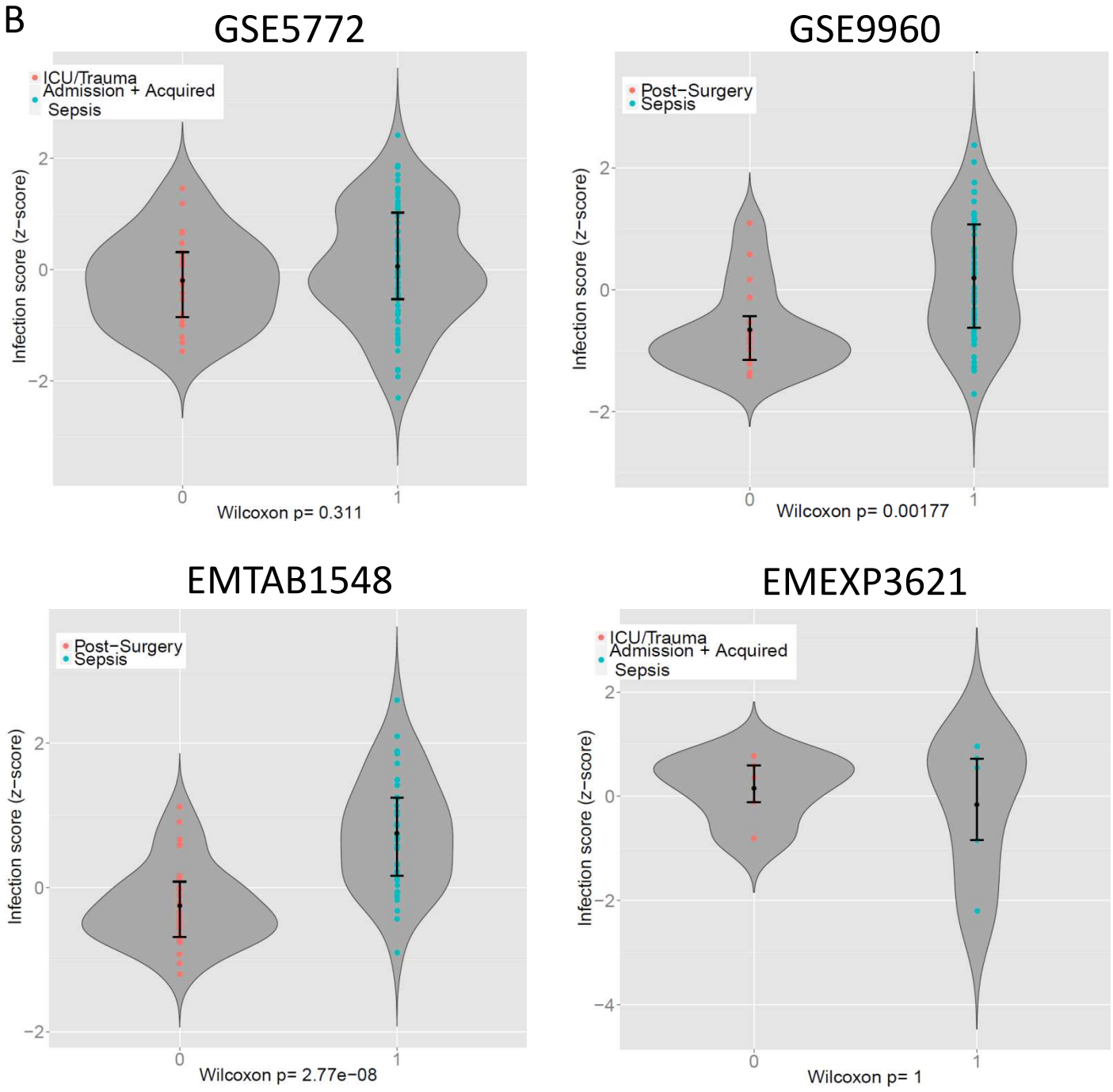
Binary SIRS criteria PLUS Infection *z* score					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-1.5709	0.5361	-2.93	0.00338	**
Infection *z* score	1.2989	0.2446	5.309	1.10E-07	***
SIRS.Temp	1.2583	0.3925	3.206	0.00134	**
SIRS.Resp	-0.1986	0.37	-0.537	0.59152	
SIRS.WBC	0.3003	0.3664	0.82	0.41237	
SIRS.HR	-0.4683	0.5141	-0.911	0.36229	

B.

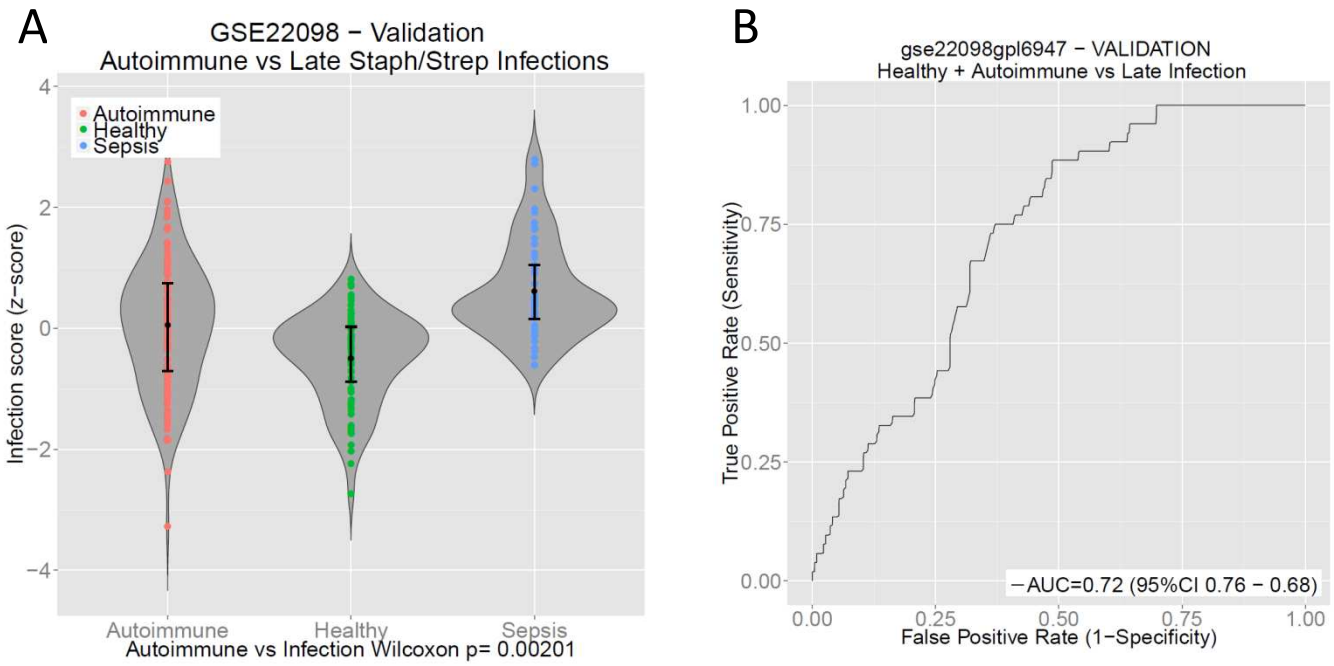
Supplemental Figure 6. Linear models of SIRS criteria and the infection *z* score. (A) Logistic regression models for Glue Grant patients with both SIRS data and microarray data available. SIRS criteria are represented as binary variables. First model shows SIRS criteria in combination; the second model adds the infection *z* score. Significance codes: $p < 0.001$ '***' ; 0.01 '**' ; 0.05 '*'. (B) Boxplots are shown of predicted log odds of infection for patients as output by the logistic regression models in (A).

A

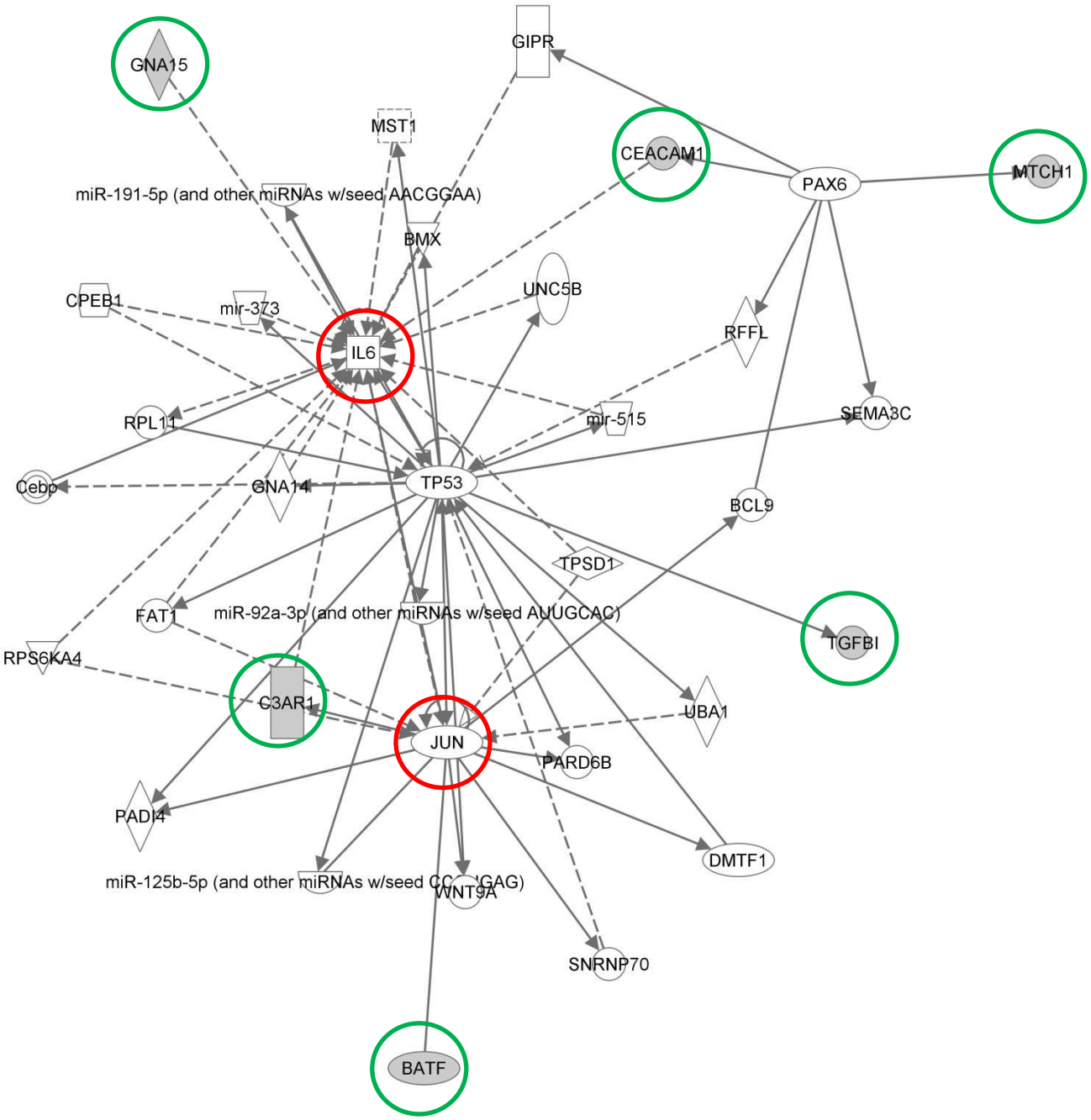

1.00
0.75
0.50
0.25
0.00
True Positive Rate (Sensitivity)
0.00
0.25
0.50
0.75
1.00
False Positive Rate (1−Specificity)
gse5772.neutrophils: AUC=0.57 (95% CI 0.64 − 0.51)
gse9960.PBMCs: AUC=0.76 (95% CI 0.82 − 0.7)
EMEXP3621.PBMCs: AUC=0.48 (95% CI 0.67 − 0.3)
EMTAB1548: AUC=0.86 (95% CI 0.9 − 0.82)



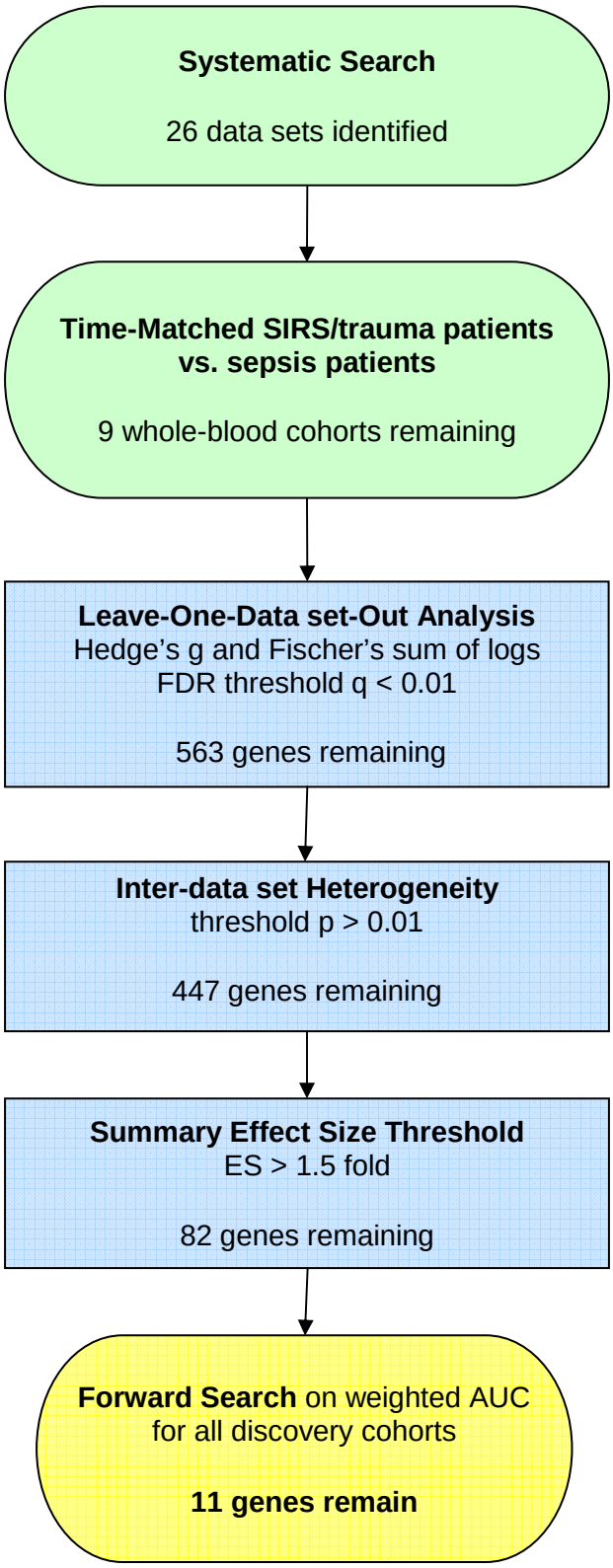
Supplemental Figure 7. The infection *z* score in non–time-matched data sets. (A) Four data sets compared SIRS/ICU/trauma patients to sepsis patients at non-matched time points. These data sets tested neutrophils (GSE5772, N=93), whole blood (EMTAB1548, N=73), and PBMCs (GSE9960, N=30; EMEXP3621, N=10). See Table 7 for further data set details. (B) Violin plots for the non-matched time-point data sets. Error bars show middle quartiles. Tested with Wilcoxon rank-sum test.



Supplemental Figure 8. Comparison of the infection *z* scores in patients with acute infections to healthy controls and patients with autoimmune diseases. GSE22098 compares healthy controls to patients with acute autoimmune inflammation or acute infections. The infection *z* score shows good discrimination of infection from both healthy patients and those with autoimmune inflammation. (A) Violin plots; error bars show middle quartiles. Patients with autoimmune inflammation vs. those with sepsis tested with Wilcoxon rank-sum test. (B) ROC plot of autoimmune patients or healthy controls vs. septic patients.



Supplemental Figure 9. Ingenuity Pathway Analysis results for the 11-gene set. Genes that are part of the 11-gene set are circled in green; *IL-6* and *JUN* are circled in red.



Supplemental Figure 10: Schematic of the entire integrated multicohort analysis.

Supplemental Table 1. Summary spreadsheet of all data sets referenced in the manuscript. Attached as separate Excel file.

Supplemental Table 2. Design matrix of individual phenotypes for multicohort analysis. Includes accession numbers, clinical time-points, and infection status (attached as separate CSV file).

Supplemental Table 3. Summary statistics for the 82 genes that passed significance, heterogeneity, and effect-size filtering after multicohort analysis. Attached as separate CSV file.

Supplemental Table 4. Probe-level data for all 11 genes in the diagnostic set for all patients in the multicohort analysis. The first page is all probe-to-gene mappings. The following pages show normalized probe (or probeset) expression data for each of the 12 cohorts in the multicohort analysis (attached as separate Excel file).

A. Repeated-measures ANOVA						
Buffy Coat – DISCOVERY SET						
	DF	**Sum Sq**	**Mean Sq**	**F value**	**P value**	
Time since injury	1	38.84	38.84	65.182	2.85E-14	***
Infection Status	1	59.83	59.83	100.413	< 2e-16	***
Time:Infection status	1	0.97	0.97	1.635	0.202	
Residuals	251	149.56	0.6			
Neutrophils – VALIDATION SET						
	DF	**Sum Sq**	**Mean Sq**	**F value**	**P value**	
Time since injury	1	3.33	3.33	4.822	2.92E-02	*
Infection Status	1	32.32	32.32	46.743	8.32E-11	***
Time:Infection status	1	1.41	1.41	2.044	0.1543	
Residuals	214	147.99	0.69			

B. Linear Regression					
Buffy Coat – DISCOVERY SET					
	Estimate	**Std Error**	**T stat**	**P value**	
(Intercept)	-0.281325	0.188184	-1.495	0.13639	
Injury Severity Score (ISS)	0.020229	0.006067	3.33	0.00101	**
Eventual Infection	0.913283	0.275058	3.32	0.00106	**
ISS:Eventual Infection	-0.019907	0.008214	-2.423	0.0162	*
Residual standard error: 0.7764 on 215 degrees of freedom					
F-statistic: 7.484 on 3 and 215 DF, p-value: 8.673e-05					
Neutrophils – VALIDATION SET					
	Estimate	**Std Error**	**T stat**	**P value**	
(Intercept)	-0.740711	0.241641	-3.07	0.00253	**
Injury Severity Score (ISS)	0.029675	0.007585	3.91	0.000132	***
Eventual Infection	1.1357	0.372217	3.051	0.002645	**
ISS:Eventual Infection	-0.030582	0.011067	-2.763	0.006353	**
Residual standard error: 0.8582 on 170 degrees of freedom					
F-statistic: 6.19 on 3 and 170 DF, p-value: 0.0005129					

Supplemental Table 5. Linear models of infection score in the Glue Grant data. (A) Repeated-measures ANOVA of Glue Grant cohorts examining the effects of time since injury and infection status on infection *z* score. (B) Linear regression of admission time point (Day 0-to-1 since injury) infection score data versus injury severity score and infection status. Significance levels: P less than: 0.001 '***'; 0.01 '**'; 0.05 '*'.

Gram Positive vs Gram Negative

Study	N, Gram Negative	N, Gram Positive	Gram Negative Mean Score	Gram Positive Mean Score	T Statistic	DF	P value	Outcome
GSE9960	18	17	0.32	0.12	0.58	33.0	0.5672	Not different
GSE13015-gpl6106	32	13	0.64	-0.13	3.10	31.0	0.0041	Higher Gram Pos
GSE33341	19	32	0.83	0.77	0.34	42.4	0.7336	Not different
GPSSSI Unique	56	87	0.26	0.61	-2.44	128.1	0.0162	Higher Gram Neg

Bacterial vs Viral

Study	N, Bacterial Infection	N, Viral Infection	Bacterial Infection Mean Score	Viral Infection Mean Score	T Statistic	DF	P value	Outcome
GSE20346	26	19	0.64	0.56	0.29	39.3	0.7770	Not different
GSE40012	74	25	0.48	0.52	-0.22	76.4	0.8230	Not different
GSE40396	8	35	1.07	0.25	2.67	12.7	0.0194	Bacterial Higher
GPSSSI Unique	143	16	0.47	0.04	1.74	17.9	0.0994	Not different

Supplemental Table 6. Comparison of infection *z* score across infection types. Shown are the infection classes present in the studied data sets for which n > 20 within 1 day of infection diagnosis. Student's t-tests were used for comparisons, $p < 0.05$ was considered significant.

A

ENCODEQT - POSITIVE GENES **+/- 5000 bp**

Factor	Total Genes with Factor	POSITIVE Observed Genes	Q-value	POSITIVE Factor Rank
Max	14735	6	0.00E+00	1

ENCODEQT - NEGATIVE GENES **+/- 5000 bp**

Factor	Total Genes with Factor	NEGATIVE Observed Genes	Q-value	NEGATIVE Factor Rank
No Significant Transcription Factor Interactions Found (q < 0.05)				

B

PASTAA - Positive Genes **- 200 bp from TSS, conserved human/mouse**

Rank	Matrix	Transcription Factor	Association Score	P-Value
1	ZBRK1_01	N/A	3.353	1.28E-03
2	PAX_Q6	Pax-1 , Pax-2	2.967	3.70E-03
3	IRF_Q6_01	Irf-1 , Irf-10	2.72	6.74E-03
4	CREL_01	C-rel	2.647	7.42E-03
5	GATA4_Q3	Gata-4	2.522	1.02E-02
6	PAX4_03	Pax-4a	2.522	1.02E-02
7	PPAR_DR1_Q2	Ppar-alpha , Ppar-beta	2.521	1.02E-02
8	STAT5A_04	Stat5a	2.503	1.02E-02
9	PTF1BETA_Q6	N/A	2.372	1.43E-02
10	MYB_Q3	C-myb	2.371	1.43E-02

PASTAA - Negative Genes **- 200 bp from TSS, conserved human/mouse**

Rank	Matrix	Transcription Factor	Association Score	P-Value
1	KAISO_01	N/A	3.264	1.28E-03
2	PAX5_01	Pax-5	3.236	1.28E-03
3	TCF11_01	Lcr-f1	3.066	1.78E-03
4	STRA13_01	Stra13	2.823	4.06E-03
5	HNF4ALPHA_Q6	Hnf-4 , Hnf-4alpha	2.473	9.47E-03
6	ARNT_02	Arnt	2.346	1.31E-02
7	USF_Q6	Usf1 , Usf2a	2.346	1.31E-02
8	PAX4_01	Pax-4a	2.221	1.70E-02
9	TFIII_Q6	Tfii-i	2.22	1.70E-02
10	AP1_Q6_01	Fosb , Fra-1	2.204	1.70E-02

Supplemental Table 7. In silico transcription factor binding analyses for the 11-gene set. The positive and negative genes in the 11-gene set were analyzed separately using (A) EncodeQT or (B) PASTAA. For EncodeQT, default settings were used. For PASTAA, input was -200 base pairs from transcription start site, searching over conserved human/mouse sequences. P-values are from hypergeometric tests; Q-values were Benjamini-Hochberg corrected.

Supplemental Table 8: Design matrix for cell type enrichment analyses. Shown are all GEO data set IDs, GSM IDs, and cell type names for each sample (attached as separate CSV file).

Supplemental Video. Rotation of a time–course-labeled PCA of trauma patients. The rotation of a 3D representation of the first three components from a time-course labelled PCA (same as in Figure 2), showing the ‘corkscrew’ nature of the progression of gene expression over time since injury in multiple data sets (attached as separate file).