Published in final edited form as:

Crit Care Med. 2017 January ; 45(1): 1–10. doi:10.1097/CCM.0000000000002021.

Benchmarking sepsis gene expression diagnostics using public data

Timothy E. Sweeney, MD, PhD^{1,2,*} and Purvesh Khatri, PhD^{1,2,*}

¹Stanford Institute for Immunity, Transplantation and Infection, Stanford University School of Medicine

²Division of Biomedical Informatics Research, Department of Medicine, Stanford University School of Medicine

Abstract

Objective—In response to a need for better sepsis diagnostics, several new gene expression classifiers have been recently published, including an 11-gene ‘Sepsis MetaScore’, the *FAIM3:PLAC8* ratio, and the Septicyte Lab. We performed a systematic search for publicly available gene expression data in sepsis, and tested each gene expression classifier in all included datasets. We also created a public repository of sepsis gene expression data to encourage their future re-use.

Data Sources—We searched NIH GEO and EBI ArrayExpress for human gene expression microarray datasets. We also included the Glue Grant trauma gene expression cohorts.

Study Selection—We selected clinical, time-matched, whole blood studies of sepsis and acute infections as compared to healthy and/or non-infectious inflammation patients. We identified 39 datasets composed of 3,241 samples from 2,604 patients.

Data Extraction—All data were renormalized from raw data, when available, using consistent methods.

Data Synthesis—Mean validation areas under the ROC curve (AUCs) for discriminating septic patients from patients with non-infectious inflammation for the Sepsis MetaScore, the *FAIM3:PLAC8* ratio, and the Septicyte Lab were 0.82 (range 0.73–0.89), 0.78 (range 0.49–0.96) and 0.73 (range 0.44–0.90), respectively. Paired t-tests of validation datasets showed no significant differences in AUCs. Mean validation AUCs for discriminating infected patients from healthy controls for the Sepsis MetaScore, *FAIM3:PLAC8* ratio, and Septicyte Lab were 0.97 (range 0.85–1.0), 0.94 (range 0.65–1.0) and 0.71 (range 0.24–1.0), respectively. There were few significant differences in any diagnostics due to pathogen type.

Conclusions—The three diagnostics do not show significant differences in overall ability to distinguish non-infectious SIRS from sepsis, though the performance in some datasets was low (AUC<0.7) for the *FAIM3:PLAC8* ratio and Septicyte lab. The Septicyte Lab also demonstrated significantly worse performance in discriminating infections as compared to healthy controls.

*Corresponding authors: tes17@stanford.edu, pkhatri@stanford.edu, **Address for Reprints:** 1265 Welch Rd., MSOB X223, Stanford, CA 94305.

Overall, public gene expression data is a useful tool for benchmarking gene expression diagnostics.

Keywords

Sepsis; diagnosis; gene expression; microarray

Introduction

There is no rapidly available gold-standard test that can determine whether a patient with systemic inflammation has an underlying infection. Missed diagnoses of sepsis lead to delayed treatment and increased mortality, while inappropriate antibiotics increase antibiotic resistance and can lead to complications (1–3). There is thus an urgent and unmet need for new diagnostics that can separate patients with non-infectious inflammation from patients with sepsis (4).

Diagnostics that can distinguish sepsis from non-infectious inflammation are difficult to derive, as many of the cellular pathways that are activated in response to infections are also activated in response to tissue trauma and non-infectious inflammation. High-throughput ‘omics’ technologies, such as gene expression microarrays, are a good way to study sepsis, as they allow for the simultaneous examination of tens of thousands of genes. However, such datasets always have more variables than samples, and so are prone to non-reproducible, overfit results (5, 6). Moreover, in an effort to increase statistical power, biomarker discovery is usually performed in a clinically homogeneous cohort using a single type of microarray. Although this homogeneous design does result in a greater statistical power, the results are less likely to remain true in different clinical cohorts using different laboratory techniques. As a result, multiple independent validations are necessary for any new classifier derived from high-throughput studies.

To the best of our knowledge, there are three gene expression diagnostics that have been specifically developed to separate patients with sepsis from those with non-infectious inflammation. Each hypothesizes that a conserved set of host genes in whole blood are transcriptionally regulated in response to infection. These are an 11-gene set hereafter referred to as the ‘Sepsis MetaScore’ (SMS) (7), the *FAIM3:PLAC8* ratio (8), and the Septicyte Lab (9). In addition, there are now dozens of publicly available datasets examining patients with sepsis or acute infections. They span a broad range of clinical conditions, including different age groups, infection types, comorbid conditions, and control (non-infectious) conditions. This public resource can thus be used to estimate the relative strengths and weaknesses of different diagnostics across an enormous number of patient samples. Here we used all available public gene expression data to study and directly compare the diagnostic power of the three sepsis gene expression diagnostics.

Methods

We completed a systematic search on Dec. 10, 2015 of two public gene expression repositories (NIH Gene Expression Omnibus and EBI ArrayExpress) using the following terms: sepsis, SIRS, pneumonia, trauma, ICU, infection, acute, shock, and surgery. We

Author Manuscript

automatically excluded non-microarray, non-human data. Next, using the abstracts of the corresponding manuscripts for screening, we eliminated non-clinical and non-time-matched datasets. Finally, we selected only datasets performed in whole blood, total blood leukocytes, or neutrophils (since neutrophils make up ~75% of white blood cells during acute inflammation in adults(7)). The remaining datasets were then sorted according to whether the reference group (compared to sepsis) was healthy controls or non-infected SIRS patients. A schematic is shown in Figure 1. In addition to the above cohorts, we included the two longitudinal trauma cohorts from the Inflammation and Host Response to Injury (Glue Grant) cohorts, as described previously (7). Cohorts from the same study run on different types of microarrays were treated as independent. Use of the Glue Grant was approved by both the Glue Grant Consortium and the Stanford University IRB (protocol 29798). All other data are publicly available and so exempt from review.

Author Manuscript

Given the limited phenotype data available in the public domain, it was impossible to for us to review determinations of infection. Thus, in all cases, we accepted the determination of infection supplied by the original study authors, including cases of ‘clinical sepsis’, where no pathogen was confirmed but independent adjudicators retrospectively classified a patient as likely having an infection.

Author Manuscript

We have included some patients at multiple timepoints both to serve as controls, and to ensure a robust diagnostic effect. The duplicated patients are: (1) Glue Grant non-infected controls: all patients are duplicated over time; in other words, in the non-infected class, the same group of patients serve as controls for the infected patients at later time-points. As we previously described (7), this is necessary to prevent bias due to change in gene expression over time of recovery. (2) GSE68310: Patients were followed over time for a full year; we have included controls from both initial baseline and following seasonal baseline. (3) Four cohorts included multiple timepoints within 48 hours of admission: GSE20346, GSE40012, GSE57065, GSE68310.

Author Manuscript

Microarray data exists in both ‘raw’ form (fluorescence intensities) and ‘normalized’ form (corrected for background and chip effects). Different normalization techniques between datasets can lead to extra technical differences (10). Here, we renormalized all datasets for which raw data was available using standardized techniques to minimize technical variation. Affymetrix arrays were renormalized using gcRMA (on arrays with perfect-match probes) or RMA (11). Illumina, Agilent, GE, and other commercial arrays were renormalized via normal-exponential background correction followed by quantile normalization. Custom arrays were not renormalized. All data were log2 transformed. Probes were summarized to genes within datasets using a fixed-effects model (10).

Author Manuscript

A literature search was conducted for gene expression signatures specifically optimized for diagnosis of sepsis as compared to non-infected hospitalized patients. The 11-gene Sepsis MetaScore is calculated according to the following formula (7):

$$\sqrt[6]{(CEACAM1 * ZDHHC19 * C9orf95 * GNA15 * BATF * C3AR1)} - \frac{5}{6} \sqrt[5]{(KIAA1370 * TGFB1 * MTCH1 * RPGRIP1 * HLADPB1)}$$

The *FAIM3:PLAC8* ratio is calculated as: *FAIM3/PLAC8* (we also added a negative coefficient to the score so that it would be increasing in the presence of infection and decreasing in controls) (8). The Septicyte Lab is calculated as: $(PLAC8 + LAMP1) - (PLA2G7 + CEACAM4)$ (9). The derivation of each score can be found in its corresponding original paper. In all cases, the calculations are performed on log₂-transformed data. The resulting scores were tested for diagnostic power as measured by the area under the receiver operating characteristic curves (AUC). Datasets for which any of the sepsis scores could not be calculated (either all up-regulated or all down-regulated genes were missing) were excluded from final results. For a given comparison (e.g. non-infectious SIRS vs. sepsis at admission), mean AUCs were calculated both for all datasets of that type, as well as for only non-discovery datasets, since discovery datasets overestimate diagnostic power as compared to independent validation. Finally, we compared the overlapping validation sets for each diagnostic score with paired t-tests (e.g., the Sepsis MetaScore and the *FAIM3:PLAC8* ratio were compared in their ability to diagnose sepsis in GSE74227, E-MEXP-3589, and the Glue Grant neutrophils, as these were the only cohorts that were validation for both datasets). To compare sensitivity and specificity at equal points, we used the R package OptimalCutpoints to select the closest attainable sensitivity to 95% for each dataset. As this sensitivity, the maximum specificity was recorded. The same paired t-test procedure was used to compare specificity levels between scores.

The patient samples in GSE28750 (12) (N=21) were re-assayed in the later dataset GSE74224 (9) (N=105), though the two datasets were run using different microarray types (Affymetrix HG 2.0 vs. Affymetrix Exon 1.0 ST). As a result, GSE28750 is not included in the validation calculation for the Septicyte Lab (discovered in GSE74224), while in computing the validation mean for the Sepsis MetaScore, the AUC in GSE74224 was penalized to account for the fact that 20% of the GSE74224 patients were present in discovery ((penalized AUC * 0.8 + actual AUC in GSE28750 * 0.2) = actual AUC).

To test confounding by infection type, each dataset was screened for presence of both (1) Gram positive and Gram negative infections, or (2) bacterial and viral infections. Cases of co-infection were not included in the confounding comparisons. In each cohort that included two classes of interest, the diagnostic scores between classes were compared via the Wilcoxon rank-sum test. Then, raw scores across cohorts were compared using paired t-tests.

Meta-analysis was performed as previously described (7). Briefly, differential gene expression between patients with non-infectious inflammation and sepsis was summarized within datasets using Hedge's *g*, and then compared between datasets using a DerSimonian-Laird random effects model.

All analyses were performed using the R statistical computing language. Significance tests were always two-tailed. Code and data to recreate the admission non-infectious SIRS vs sepsis comparisons for the examined gene sets are available at <http://khatrilab.stanford.edu/sepsis>. The uploaded data are in the renormalized form used here. Glue Grant data is available to researchers who have been approved by the Glue Grant consortium; instructions are on the website.

Results

We performed a systematic search of public gene expression databases (Figure 1), and also we used the two independent Glue Grant trauma cohorts, broken up into time-matched bins of never-infected patients and patients within $\pm$ 24 hours of diagnosis of sepsis, as previously described (7). This yielded a total of 39 cohorts that matched criteria, composed of 3,241 samples from 2,604 patients (8, 9, 12–41).

The robustness and reproducibility of each of the three sepsis scores depends on robust and reproducible change in expression for each of their constituent genes. Therefore, we explored how consistently individual genes in each of the three tests changed across 12 whole-blood cohorts comparing non-infected SIRS/trauma patients to sepsis patients. Our meta-analysis of these datasets revealed that each of the 16 genes included in any of the 3 gene scores (except *CEACAM4*) changed in the hypothesized direction (FDR < 5%; Figure 2, Table S1). Notably, *CEACAM4*, one of the genes in the Septicyte Lab, was significantly down-regulated only in its corresponding discovery cohort (Figure 2).

Next, we divided the datasets into two broad types of comparison: patients with non-infectious SIRS or trauma vs. sepsis or acute infections (Table 1 & Table S2); and healthy controls vs. patients with sepsis or acute infection (Table 2 & Table S3). For both of these types of comparison, we calculated both the overall mean AUC, as well as the AUC when including only independent validation datasets; for each of the three signatures, we excluded their corresponding discovery datasets.

In the non-infectious SIRS/trauma vs. sepsis datasets (16 cohorts, 1148 samples from 835 patients, Table 1 & Table S2), there were no significant differences in paired t-tests between the AUCs of the three gene expression diagnostic scores comparing overlapping validation datasets (all $p > 0.1$; Figures S1–S2). When comparing the AUCs from all 16 cohorts (i.e. including the discovery cohorts), the Sepsis MetaScore AUCs were significantly higher than those of the other two gene scores (both $p < 0.05$), with no significant difference between the *FAIM3:PLAC8* ratio and the Septicyte Lab. However, these results do not necessarily point to better overall performance of the Sepsis MetaScore, as the Sepsis MetaScore used 9 of these cohorts in discovery. The *FAIM3:PLAC8* ratio showed decreased performance in GSE32707 and GSE40012, but as discussed previously, it is specifically designed for testing the presence of CAP and may not be generalizable to other forms of non-infectious inflammation (42, 43). Finally, the Septicyte Lab had significantly reduced performance (AUC < 0.5) in separating both pediatric SIRS/sepsis patients, as well as hospitalized COPD patients with and without infections. It possible that this reduction in AUC for the Septicyte Lab is due to the differences in clinical circumstances or microarray types compared to the initial discovery cohort for the Septicyte Lab. Since the ROC curve measures a large potential space, we also obtained the sensitivity closest to 95% for each test, and recorded the maximum specificity at that level (Table S3). At sensitivities near 95%, the mean validation specificities were 53%, 45%, and 38%, respectively, for the Sepsis MetaScore, *FAIM3:PLAC8* ratio, and Septicyte Lab. There were no significant differences in paired t-tests in validation specificities among the three scores.

We next examined datasets that compared healthy controls to patients with sepsis or acute infections (26 datasets, 2417 samples from 2075 patients, Table 2 & Table S4). Most, but not all, of these patients had sepsis; however, it is reasonable to expect that a sepsis diagnostic should be able to distinguish most infections from healthy controls. Here, both the Sepsis MetaScore and the *FAIM3:PLAC8* ratio performed well, with mean validation AUCs of 0.96 $\pm$ 0.05 and 0.94 $\pm$ 0.09 (Table 2). However, the Septicyte Lab had an AUC $<$ 0.7 in 12 cohorts (43% of all cohorts) composed of 1562 samples (64% of total samples), resulting in a mean validation AUC = 0.71 $\pm$ 0.20 (Table 2), significantly lower than both other scores (both $P < 1e-5$). Although there is no clinical need for a diagnostic to separate healthy controls from patients with sepsis, poor performance in this area may be indicative of deeper biases and may increase the risk of non-generalizability.

An ideal sepsis diagnostic would not show varying performance depending on the type of infection present. In order to study whether any of the diagnostics is biased by pathogen, we searched through all of the included datasets to find those comparing patients with bacterial and viral infections, and those comparing Gram positive and Gram negative infections. We then compared both the diagnostic power (as compared to healthy controls) and raw score distributions by infection type (via paired t-test across cohorts). Score distributions that are consistently lower in one infection type may indicate decreased diagnostic performance for that type, even if a change in diagnostic performance is not detected as compared to healthy controls.

There were 8 datasets that provided information about whether a patient had bacterial or viral infection. In general, there were few differences between the AUCs for bacterial and viral infections for any of the three scores (Table S5); however, this may be due to small numbers, and the relatively high AUCs in comparing these infections to healthy controls. However, despite these caveats, both the Sepsis MetaScore and the *FAIM3:PLAC8* ratio showed higher raw mean scores in patients with bacterial infections as compared to viral infections (both $P < 0.05$) (Table S6). The Septicyte Lab, in contrast, did not show a strong trend in comparing bacterial and viral infections.

There were 9 datasets that provided information about whether a patient had Gram positive and Gram negative infection. The comparison of Gram positive and Gram negative infections revealed small differences in AUC for both the Sepsis MetaScore or the *FAIM3:PLAC8* ratio; the Septicyte Lab showed greater variability, but this may be due to a high variability in diagnostic performance vs. healthy controls rather than differences between Gram positive and Gram negative infections (Table S7). None of the three tests had raw scores that were significantly different overall according to Gram status, though the Sepsis MetaScore showed a trend towards higher scores in Gram negative infections ($P = 0.055$; Table S8).

Discussion

Here we compared three sepsis gene expression diagnostics (the Sepsis MetaScore, the *FAIM3:PLAC8* ratio, and the Septicyte Lab) in all available time-matched, whole blood clinical sepsis datasets. There were no significant differences among the distribution of

AUCs comparing all validation non-infectious SIRS/trauma and sepsis datasets. However, in four of these cohorts the *FAIM3:PLAC8* ratio and the Septicyte Lab showed AUCs < 0.7, showing a need for further prospective testing of all three scores. Notably, the Septicyte Lab also had significantly reduced performance in validation comparing healthy controls to patients with sepsis or acute infections, showing AUCs < 0.7 in 43% of these cohorts. However, the Septicyte Lab was initially validated in a large, independent cohort of patients from the MARS consortium using targeted qPCR, and showed an overall validation AUC of 0.88 (9); thus, the reduced performance in our analysis may be indicative of either differences in clinical conditions, difference in technology, or both. Meta-analysis shows that *CEACAM4* (one of four genes in the Septicyte Lab) was down-regulated only in its discovery cohort, which may be contributing to the relatively worse generalizability.

There is evidence of higher scores in bacterial infection as opposed to viral infection for the Sepsis MetaScore and the *FAIM3:PLAC8* ratio, while the Septicyte Lab showed no significant differences. There were no significant differences in Gram negative as opposed to Gram positive infections for any of the three scores. However, the differing pathogen types were not matched for illness severity, age, gender, or other clinical confounders in their individual datasets; hence, these trends must be interpreted with caution. For instance, if bacterial infections were generally more severe than viral infections, and Gram negative generally more severe than Gram positive, then these scores might point to a confounding by severity. Still, when used in clinical practice, patients will not be part of a matched cohort, and the differences in bacterial and viral illness found here are worthy of further investigation. Further testing against all confounders will be necessary.

Previously the Sepsis MetaScore was validated in the Glue Grant neutrophils cohort and in a subset of the healthy vs. infections cohorts (7). In the additional cohorts tested here, the Sepsis MetaScore continues to show results similar to prior validation. The *FAIM3:PLAC8* ratio was validated in some of these data in a follow-up publication (43), though the authors pointed out that their gene set was initially designed for a very narrow question of determining the presence of CAP in patients admitted to the ICU suspected of having CAP (42). The Septicyte Lab was tested in cohort E-TABM-1548 (9), but we have previously shown that because of expected changes in the baseline gene expression profile due to recovery from surgery, it is not appropriate to use for testing sepsis diagnostics (7).

The fact that the public data are not used more often for validation of new diagnostics may reflect the difficulty and knowledge curve that some researchers face in accessing and using these data. Given this difficulty, we have provided a hand-curated, unified repository of these data, along with an R script to easily apply a classifier of interest to the datasets (<http://khatrilab.stanford.edu/sepsis>). We recommend a practice that any new gene expression classifiers for sepsis should be tested in these data to allow for easy benchmarking and comparison between classifiers. We recognize that the simple measure of the AUC does not account for all potential measures of clinical utility, and that a score that repeatedly performs well in a single clinical area can still have great clinical utility even if it fails in a different clinical area. Nevertheless, it is important to elucidate the strengths and weaknesses of any new sepsis diagnostic in order to help focus resources on applications that show the most promise.

In all datasets used here, we accepted the determination of infections assigned by the original study authors. A true positive (such as Gram-negative bacteremia) is often easy to assign from clinical data. However, the lack of gold standard in determining presence of infection can make the culture-negative patient difficult to correctly diagnose. For instance, if a culture-negative patient still clinically classified as 'septic' were to show very low probabilities of infection in all three gene scores (or vice-versa), it is not immediately clear which designation would be correct. When such tests are ready for clinical use, treating clinicians will need to integrate clinical judgement with test results to ensure patient safety.

Another issue in translating these tests to practice is measurement platform. Accurate and rapid quantitation of gene expression is traditionally done with quantitative PCR, which can be optimized to return in a couple of hours for a limited target set. Commercial optimization could take many potential forms; the goal will be a rapid sample-to-answer system that can give a clear, interpretable outcome, perhaps as a range of likelihood of infection. We conclude from the data presented here that these engineering problems are worth solving, given the relatively good performance of these tests at detecting the presence of infection.

Each of the diagnostic gene sets tested here has both strengths and weaknesses. In general, for any sepsis diagnostic to become useful clinically, it must retain good diagnostic power in a broad range of patient settings in its final form. Public microarray data allow for head-to-head comparisons of different gene expression diagnostics, but may underestimate the diagnostic performance of any test compared to using a targeted assay. Thus, further prospective validation of any gene set will be needed prior to their application in clinical practice; such clinical trials will be most effective by calculating not just the AUC of the new diagnostics, but rather the net reclassification of those data given the circumstances at the time. The ultimate goal would be a randomized interventional trial to assess the effect of the host diagnostic on patient outcome; this will likely have to wait until an appropriate platform is available. Given the increasing accuracy of molecular profiling of the host response to infections, these tests will likely become a valuable part of the clinical toolset in diagnosing, treating, and potentially preventing sepsis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the authors of the public datasets used here. Without their continuing contributions to open science, none of this would be possible. We thank the Glue Grant authors for sharing their data; they are supported in this by NIGMS Glue Grant Legacy Award R24GM102656.

Conflicts of Interest and Sources of Funding:

TES and PK have filed a preliminary patent on the 11-gene Sepsis MetaScore through the Stanford Office of Technology Licensing. TES is a scientific advisor to Multerra Biosciences, and is funded by a Stanford Child Health Research Institute Young Investigator Award (through the Institute for Immunity, Transplantation and Infection), and the Society for University Surgeons. PK is funded by NIAID grants 1U19AI109662, U19AI057229, U54I117925, U01AI089859, and the Bill and Melinda Gates Foundation.

Copyright form disclosures: Dr. Sweeney is a scientific advisor to Multerra Bio. He received support for article research from Stanford Child Health Research Institute Young Investigator Award (through the Institute for

Immunity, Transplantation and Infection), and the Society for University Surgeons. Dr. Khatri received support for article research from the National Institutes of Health and Bill & Melinda Gates Foundation. He is funded by National Institute for Allergy and Infectious Disease grants 1U19AI09662, U19AI057229, U54I117925, and U01AI089859. Drs. Sweeney and Khatri have filed a provisional patent on the 11-gene Sepsis MetaScore through the Stanford Office of Technology Licensing; they have since founded Inflammatrix, Inc. to commercialize the technology.

References

1. Ferrer R, Martin-Loeches I, Phillips G, et al. Empiric antibiotic treatment reduces mortality in severe sepsis and septic shock from the first hour: results from a guideline-based performance improvement program*. *Crit Care Med.* 2014; 42(8):1749–1755. [PubMed: 24717459]
2. McFarland LV. Antibiotic-associated diarrhea: epidemiology, trends and treatment. *Future Microbiol.* 2008; 3(5):563–578. [PubMed: 18811240]
3. Gaieski DF, Mikkelsen ME, Band RA, et al. Impact of time to antibiotics on survival in patients with severe sepsis or septic shock in whom early goal-directed therapy was initiated in the emergency department. *Crit Care Med.* 2010; 38(4):1045–1053. [PubMed: 20048677]
4. Cohen J, Vincent JL, Adhikari NK, et al. Sepsis: a roadmap for future research. *Lancet Infect Dis.* 2015; 15(5):581–614. [PubMed: 25932591]
5. Shi L, Jones WD, Jensen RV, et al. The balance of reproducibility, sensitivity, and specificity of lists of differentially expressed genes in microarray studies. *BMC Bioinformatics.* 2008; 9(Suppl 9):S10.
6. Ioannidis JP, Ntzani EE, Trikalinos TA, et al. Replication validity of genetic association studies. *Nat Genet.* 2001; 29(3):306–309. [PubMed: 11600885]
7. Sweeney TE, Shidham A, Wong HR, et al. A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. *Sci Transl Med.* 2015; 7(287):287r–ra271.
8. Scicluna BP, Klein Klouwenberg PM, van Vught LA, et al. A molecular biomarker to diagnose community-acquired pneumonia on intensive care unit admission. *Am J Respir Crit Care Med.* 2015; 192(7):826–835. [PubMed: 26121490]
9. McHugh L, Seldon TA, Brandon RA, et al. A Molecular Host Response Assay to Discriminate Between Sepsis and Infection-Negative Systemic Inflammation in Critically Ill Patients: Discovery and Validation in Independent Cohorts. *PLoS Med.* 2015; 12(12):e1001916. [PubMed: 26645559]
10. Ramasamy A, Mondry A, Holmes CC, et al. Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med.* 2008; 5(9):e184. [PubMed: 18767902]
11. Wu Z, Irizarry R, Gentleman R, et al. A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. *Journal of the American Statistical Association.* 2004; 99(468):909–917.
12. Sutherland A, Thomas M, Brandon RA, et al. Development and validation of a novel molecular biomarker diagnostic test for the early detection of sepsis. *Crit Care.* 2011; 15(3):R149. [PubMed: 21682927]
13. Dolinay T, Kim YS, Howrylak J, et al. Inflammasome-regulated cytokines are critical mediators of acute lung injury. *Am J Respir Crit Care Med.* 2012; 185(11):1225–1234. [PubMed: 22461369]
14. Parnell GP, McLean AS, Booth DR, et al. A distinct influenza infection signature in the blood transcriptome of patients with severe community-acquired pneumonia. *Crit Care.* 2012; 16(4):R157. [PubMed: 22898401]
15. Wong HR, Shanley TP, Sakthivel B, et al. Genome-level expression profiles in pediatric septic shock indicate a role for altered zinc homeostasis in poor outcome. *Physiol Genomics.* 2007; 30(2):146–155. [PubMed: 17374846]
16. Wynn JL, Cvijanovich NZ, Allen GL, et al. The influence of developmental age on the early transcriptomic response of children with septic shock. *Mol Med.* 2011; 17(11–12):1146–1156. [PubMed: 21738952]
17. Wong HR, Cvijanovich N, Allen GL, et al. Genomic expression profiling across the pediatric systemic inflammatory response syndrome, sepsis, and septic shock spectrum. *Crit Care Med.* 2009; 37(5):1558–1566. [PubMed: 19325468]

18. Shanley TP, Cvijanovich N, Lin R, et al. Genome-level longitudinal expression of signaling pathways and gene networks in pediatric septic shock. *Mol Med*. 2007; 13(9–10):495–508. [PubMed: 17932561]
19. Cvijanovich N, Shanley TP, Lin R, et al. Validating the genomic signature of pediatric septic shock. *Physiol Genomics*. 2008; 34(1):127–134. [PubMed: 18460642]
20. Almansa R, Socias L, Sanchez-Garcia M, et al. Critical COPD respiratory illness is linked to increased transcriptomic activity of neutrophil proteases genes. *BMC Res Notes*. 2012; 5:401. [PubMed: 22852767]
21. Irwin AD, Marriage F, Mankhambo LA, et al. Novel biomarker combination improves the diagnosis of serious bacterial infections in Malawian children. *BMC Med Genomics*. 2012; 5:13. [PubMed: 22559298]
22. van de Weg CA, van den Ham HJ, Bijl MA, et al. Time since onset of disease and individual clinical markers associate with transcriptional changes in uncomplicated dengue. *PLoS Negl Trop Dis*. 2015; 9(3):e0003522. [PubMed: 25768297]
23. Emonts, M. Ph.D. thesis. Erasmus University Rotterdam; 2008. Polymorphisms in Immune Response Genes in Infectious Diseases and Autoimmune Diseases.
24. Pankla R, Buddhisa S, Berry M, et al. Genomic transcriptional profiling identifies a candidate blood biomarker signature for the diagnosis of septicemic melioidosis. *Genome Biol*. 2009; 10(11):R127. [PubMed: 19903332]
25. Zaas AK, Chen M, Varkey J, et al. Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. *Cell Host Microbe*. 2009; 6(3):207–217. [PubMed: 19664979]
26. Parnell G, McLean A, Booth D, et al. Aberrant cell cycle and apoptotic changes characterise severe influenza A infection—a meta-analysis of genomic signatures in circulating leukocytes. *PLoS One*. 2011; 6(3):e17186. [PubMed: 21408152]
27. Bermejo-Martin JF, Martin-Loeches I, Rello J, et al. Host adaptive immunity deficiency in severe pandemic influenza. *Crit Care*. 2010; 14(5):R167. [PubMed: 20840779]
28. Berry MP, Graham CM, McNab FW, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature*. 2010; 466(7309):973–977. [PubMed: 20725040]
29. Smith CL, Dickinson P, Forster T, et al. Identification of a human neonatal immune-metabolic network associated with bacterial infection. *Nat Commun*. 2014; 5:4649. [PubMed: 25120092]
30. Berdal JE, Mollnes TE, Wæhre T, et al. Excessive innate immune response and mutant D222G/N in severe A (H1N1) pandemic influenza. *J Infect*. 2011; 63(4):308–316. [PubMed: 21781987]
31. Ahn SH, Tsalik EL, Cyr DD, et al. Gene expression-based classifiers identify *Staphylococcus aureus* infection in mice and humans. *PLoS One*. 2013; 8(1):e48979. [PubMed: 23326304]
32. Mejias A, Dima B, Suarez NM, et al. Whole blood gene expression profiles to assess pathogenesis and disease severity in infants with respiratory syncytial virus infection. *PLoS Med*. 2013; 10(11):e1001549. [PubMed: 24265599]
33. Hu X, Yu J, Crosby SD, et al. Gene expression profiles in febrile children with defined viral and bacterial infection. *Proc Natl Acad Sci U S A*. 2013; 110(31):12792–12797. [PubMed: 23858444]
34. Herberg JA, Kaforou M, Gormley S, et al. Transcriptomic profiling in childhood H1N1/09 influenza reveals reduced expression of protein synthesis genes. *J Infect Dis*. 2013; 208(10):1664–1668. [PubMed: 23901082]
35. Kwissa M, Nakaya HI, Onlamoon N, et al. Dengue virus infection induces expansion of a CD14(+)CD16(+) monocyte population that stimulates plasmablast differentiation. *Cell Host Microbe*. 2014; 16(1):115–127. [PubMed: 24981333]
36. Cazalis MA, Lepape A, Venet F, et al. Early and dynamic changes in gene expression in septic shock patients: a genome-wide approach. *Intensive Care Med Exp*. 2014; 2(1):20. [PubMed: 26215705]
37. Suarez NM, Bunsow E, Falsey AR, et al. Superiority of transcriptional profiling over procalcitonin for distinguishing bacterial from viral lower respiratory tract infections in hospitalized adults. *J Infect Dis*. 2015; 212(2):213–222. [PubMed: 25637350]

38. Zhai Y, Franco LM, Atmar RL, et al. Host Transcriptional Response to Influenza and Other Acute Respiratory Viral Infections—A Prospective Cohort Study. *PLoS Pathog.* 2015; 11(6):e1004869. [PubMed: 26070066]
39. Conejero L, Potempa K, Graham CM, et al. The Blood Transcriptome of Experimental Melioidosis Reflects Disease Severity and Shows Considerable Similarity with the Human Disease. *J Immunol.* 2015; 195(7):3248–3261. [PubMed: 26311902]
40. Xiao W, Mindrinos MN, Seok J, et al. A genomic storm in critically injured humans. *J Exp Med.* 2011; 208(13):2581–2590. [PubMed: 22110166]
41. Warren HS, Elson CM, Hayden DL, et al. A genomic score prognostic of outcome in trauma patients. *Mol Med.* 2009; 15(7–8):220–227. [PubMed: 19593405]
42. Scicluna BP, Klein Klouwenberg PM, van der Poll T, et al. Reply: Comprehensive Validation of the FAIM3:PLAC8 Ratio in Time-matched Public Gene Expression Data. *Am J Respir Crit Care Med.* 2015; 192(10):1261–1262. [PubMed: 26568248]
43. Sweeney TE, Khatri P. Comprehensive Validation of the FAIM3:PLAC8 Ratio in Time-matched Public Gene Expression Data. *Am J Respir Crit Care Med.* 2015; 192(10):1260–1261. [PubMed: 26568247]

Figure 1.
Schema for systematic search and selection of clinical sepsis datasets.

Figure 2.

Forest plots for all genes used in any of the three scores, tested in all admission non-infectious SIRS vs. sepsis datasets. Genes are organized by gene set. The x axes represent standardized mean difference between non-infectious SIRS vs. sepsis samples, computed as Hedges' g , in log2 scale. The size of the blue rectangles is inversely proportional to the standard error of mean in the study. Whiskers represent the 95% confidence interval. The orange diamonds represent overall, combined mean difference for a given gene. Width of the diamonds represents the 95% confidence interval of overall combined mean difference.

Table 1
Non-infectious SIRS/trauma vs. sepsis/infections datasets, (D): Discovery dataset for the given score.

Accession	Microarray Type	Clinical comparison	N Non-infected SIRS	N Sepsis	AUC Sepsis MetaScore	AUC FAIM3: PLAC8 Ratio	AUC Septicocyte Lab
GSE28750	GPL570	post-op vs. sepsis	11	10	0.96 (0.92–1) (D)	0.87 (0.79–0.95)	0.85 (0.77–0.94) *
GSE32707	GPL10558	MICU +/-SIRS vs. sepsis	55	48	0.8 (0.75–0.84) (D)	0.66 (0.6–0.71)	0.66 (0.61–0.71)
GSE40012	GPL6947	ICU SIRS vs. CAP	24	52	0.71 (0.65–0.77) (D)	0.58 (0.51–0.65)	0.8 (0.75–0.85)
GSE65682	GPL13667	ICU non-infected vs. CAP	33	101	0.78 (0.74–0.82)	0.84 (0.8–0.87) (D)	0.74 (0.7–0.79)
GSE66099	GPL570	PICU SIRS vs. sepsis	30	199	0.79 (0.76–0.83) (D)	0.74 (0.7–0.78)	0.44 (0.38–0.5)
GSE74224	GPL5175	post-op vs. sepsis	31	74	0.90 (0.87–0.92) **	0.92 (0.9–0.95)	0.99 (0.99–1) (D)
E-MEXP-3589	GPL10332	Hosp. COPD +/- infection	14	14	0.74 (0.65–0.83)	0.49 (0.38–0.6)	0.46 (0.36–0.57)
Buffy Coat, Day [1,3]	GPL570	never-infected trauma vs. trauma with sepsis	65	9	0.91 (0.84–0.97) (D)	0.62 (0.51–0.72)	0.83 (0.75–0.92)
Buffy Coat, Day [3,6]	GPL570	never-infected trauma vs. trauma with sepsis	63	17	0.89 (0.84–0.94) (D)	0.84 (0.78–0.9)	0.73 (0.65–0.8)
Buffy Coat, Day [6,10]	GPL570	never-infected trauma vs. trauma with sepsis	50	15	0.91 (0.86–0.96) (D)	0.83 (0.77–0.9)	0.72 (0.64–0.79)
Buffy Coat, Day [10,18]	GPL570	never-infected trauma vs. trauma with sepsis	22	4	0.84 (0.72–0.97) (D)	0.78 (0.65–0.92)	0.8 (0.66–0.93)
Buffy Coat, Day [18,24]	GPL570	never-infected trauma vs. trauma with sepsis	6	4	0.96 (0.88–1) (D)	0.96 (0.88–1)	0.83 (0.69–0.97)
Neutrophils, Day [1,3]	GGH-1, GGH-2	never-infected trauma vs. trauma with sepsis	56	10	0.72 (0.63–0.82)	0.74 (0.65–0.83)	0.68 (0.58–0.77)
Neutrophils, Day [3,6]	GGH-1, GGH-2	never-infected trauma vs. trauma with sepsis	55	10	0.83 (0.75–0.91)	0.87 (0.79–0.94)	0.9 (0.84–0.97)
Neutrophils, Day [6,10]	GGH-1, GGH-2	never-infected trauma vs. trauma with sepsis	46	14	0.88 (0.82–0.94)	0.88 (0.82–0.94)	0.84 (0.77–0.91)
Neutrophils, Day [10,18]	GGH-1, GGH-2	never-infected trauma vs. trauma with sepsis	24	3	0.89 (0.77–1)	0.9 (0.79–1)	0.78 (0.62–0.94)
				Overall mean:	0.844 +/- 0.080	0.782 +/- 0.135	0.754 +/- 0.145
				mean in validation only:	0.817 +/- 0.069	0.779 +/- 0.139	0.729 +/- 0.135

* GSE28750 is a subset of GSE74224, so was counted as discovery for Septicocyte Lab.

** GSE28750 is a subset of GSE74224 and is treated as described in Methods.

Table 2

Healthy vs. sepsis/acute infections. (D): Discovery dataset for the given score.

Accession	Microarray Type	Clinical cohort	N Healthy	N Infected	AUC Sepsis MetaScore	AUC FAIM3/PLAC8	AUC Septicyle Lab
E-MEXP-3567	GPL96	Children with meningococcal sepsis +/- HIV co-infection	3	12	0.97 (0.93-1)	1 (1-1)	0.94 (0.89-1)
E-MEXP-3589	GPL10332	Hospitalized COPD patients with infection	4	14	0.98 (0.95-1)	0.95 (0.9-1)	0.32 (0.16-0.48)
E-MTAB-3162	GPL570	Dengue fever (+/- severe) within 48 hours of fever	15	30	1 (1-1)	1 (1-1)	0.8 (0.73-0.86)
GSE11755	GPL570	children w/ meningococcal sepsis	3	6	1 (1-1)	1 (1-1)	0.78 (0.62-0.93)
GSE13015	GPL6106	Adults with sepsis, many from burkholderia	10	48	1 (0.99-1)	0.98 (0.97-1)	0.94 (0.9-0.97)
GSE13015	GPL6947	Adults with sepsis, many from burkholderia	10	15	1 (1-1)	1 (1-1)	0.85 (0.78-0.93)
GSE17156	GPL571	Viral challenge peak symptoms	56	27	0.91 (0.87-0.94)	0.89 (0.85-0.93)	0.51 (0.45-0.58)
GSE20346	GPL6947	Severe bacterial or influenza pneumonia	36	20	1 (1-1)	1 (1-1)	0.95 (0.92-0.99)
GSE21802	GPL6102	Severe H1N1 influenza with mechanical ventilation	4	12	0.98 (0.95-1)	1 (1-1)	0.69 (0.55-0.83)
GSE22098	GPL6947	Children with Staph and Strep infections	81	52	0.85 (0.81-0.88)	0.65 (0.6-0.7)	0.79 (0.75-0.83)
GSE25504	GPL13667	Neonatal sepsis	6	14	0.92 (0.86-0.98)	0.83 (0.74-0.92)	0.42 (0.28-0.56)
GSE25504	GPL570	Neonatal sepsis	3	2	1 (1-1)	1 (1-1)	0.83 (0.62-1)
GSE25504	GPL6947	Neonatal sepsis	35	28	0.94 (0.91-0.97)	0.88 (0.83-0.92)	0.24 (0.18-0.3)
GSE27131	GPL6244	Severe H1N1 influenza with mech. ventilation	7	7	1 (1-1)	1 (1-1)	1 (1-1)
GSE28750	GPL570	Community-acquired sepsis at admission to ICU	20	10	1 (1-1) (D)	1 (1-1)	0.74 (0.64-0.84) (D)
GSE33341	GPL571	BSIS aureus or E coli	43	51	1 (1-1)	0.99 (0.98-1)	0.69 (0.64-0.74)
GSE38900	GPL10558	Infants hospitalized with viral LRTI	8	28	0.89 (0.84-0.95)	0.7 (0.61-0.79)	0.64 (0.54-0.74)
GSE38900	GPL6884	Infants hospitalized with viral LRTI	31	153	0.91 (0.89-0.93)	0.91 (0.89-0.93)	0.41 (0.35-0.46)
GSE40012	GPL6947	Adults in ICU with CAP within 48 hr of admission	18	52	1 (1-1) (D)	1 (0.99-1)	0.89 (0.85-0.93)
GSE40396	GPL10558	Children with infection + fever	22	30	0.97 (0.94-0.99)	0.95 (0.93-0.98)	0.77 (0.71-0.83)
GSE42026	GPL6947	Children admitted with bacterial or viral infection	33	59	0.97 (0.95-0.98)	0.98 (0.97-0.99)	0.74 (0.7-0.79)
GSE51808	GPL13158	Dengue fever (+/- severe) at admission	9	28	0.98 (0.95-1)	1 (1-1)	1 (0.99-1)
GSE57065	GPL570	Septic Shock at admission, 24hr, 48hr	25	82	1 (1-1)	0.99 (0.99-1)	0.81 (0.77-0.85)
GSE60244	GPL10558	Adults hospitalized with LRTI	40	118	0.96 (0.95-0.97)	0.84 (0.81-0.87)	0.64 (0.59-0.68)
GSE65682	GPL13667	Adults in ICU with CAP	42	101	1 (0.99-1)	0.93 (0.92-0.95) (D)	0.62 (0.58-0.66)
GSE66099	GPL570	Pediatric sepsis	47	199	1 (1-1) (D)	0.99 (0.99-1)	0.54 (0.49-0.59)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Accession	Microarray Type	Clinical cohort	N Healthy	N Infected	AUC Sepsis MetaScore	AUC FAIM3/PLAC8	AUC Septicyte Lab
GSE68310	GPL10558	Outpatients with acute viral illness at Days 0 & 2	243	258	0.87 (0.85–0.88)	0.92 (0.91–0.93)	0.66 (0.64–0.69)
GSE69528	GPL10558	Adults with sepsis, many from burkholderia	55	83	0.99 (0.99–1)	0.97 (0.96–0.98)	0.72 (0.68–0.76)
			Mean Validation AUC:		0.963 +/- 0.046	0.940 +/- 0.092	0.711 +/- 0.203