Benchmarking Published Gene Expression Signatures for Infection Classification

Nicholas Bodkin[1], Melissa Ross[2], Ricardo Henao[2], Ephraim L. Tsalik[2,3]

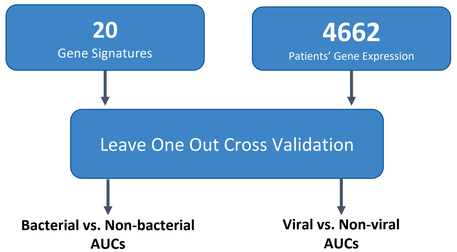
[1]Trinity College, Duke University, Durham, NC, USA
[2]Duke University School of Medicine, Durham, NC, USA
[3]Durham Veterans Affairs Health Care System, Durham, NC, USA

Motivation & Objective

Host gene expression has emerged as a promising diagnostic strategy to discriminate bacterial and viral infection. Multiple gene signatures of varying size and complexity have been developed in various clinical populations. However, there has been no systematic comparison of these signatures. It is also unclear how these signatures apply to different clinical populations.

This meta-analysis examined 20 published signatures, validated in 49 publicly available datasets for a total of 4662 patients. The objectives were to understand how the signatures compared to each other with respect to composition and performance, and to evaluate their performance in different patient subgroups.

Experimental Design



- We identified 20 published host response signatures that included both bacterial and viral patient samples in either their discovery or validation cohorts.
- We then curated a standardized validation cohort that consisted of 49 publicly available datasets with 4662 samples that met the following criteria:
  - Dataset included at least one infection phenotype (bacterial or viral) and at least one other phenotype (bacterial, viral, healthy, or non-infectious illness), as defined by the data contributor
  - Minimum of 10 total samples and 10% infected cases
  - Data was generated using either whole blood or PBMCs
  - Gene expression was measured using a commercial microarray or RNA-seq platform

Methods

Signatures were characterized with respect to size, platform, and discovery population. Genes in each signature were mapped to Ensemble Gene IDs, where duplicates and those without an ID were removed. Subjects with repeated measurements were excluded.

Samples in each validation dataset were annotated with the patient's adjudicated infection phenotype, identified pathogen, age group, and infection severity (as assessed by ICU admission).

AUCs were generated for each signature's ability to classify bacterial vs. non-bacterial and viral vs. non-viral classification in each validation dataset. Classification was determined by nested leave-one-out cross validation in Scikit Learn with LASSO.

We applied dataset-specific thresholds (Youden index) to generate signature accuracies. This allowed us to pool patients across datasets and assess performance in patient subpopulations.

We created a composite signature ("All") from the union of all other signatures.

Overall Signature Performance

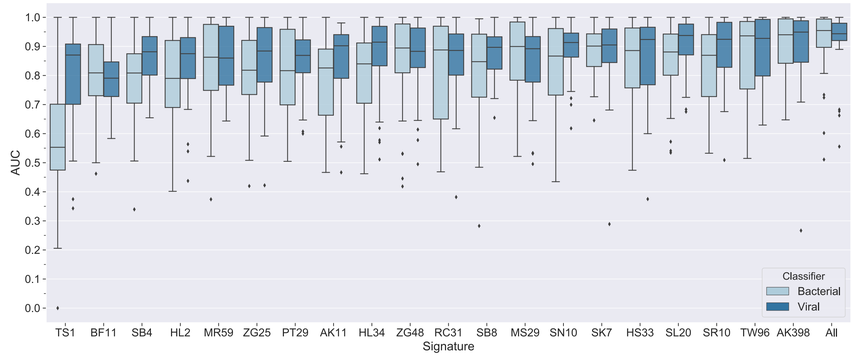


Figure 1: Distribution of AUCs for each signature's classification of bacterial and viral patients across 49 datasets.

Median AUCs ranged from:
- Bacterial classification: 0.55-0.94
- Viral classification: 0.79-0.96

Signature size varied (1-398 genes) and smaller signatures generally performed more poorly.
- Bacterial: R = 0.66, P < .001
- Viral: R = 0.45, P = 0.02

Viral infection was easier to classify than bacterial infection
- 85% vs. 80% overall accuracy
- P < .001

Differential Performance in Certain Populations

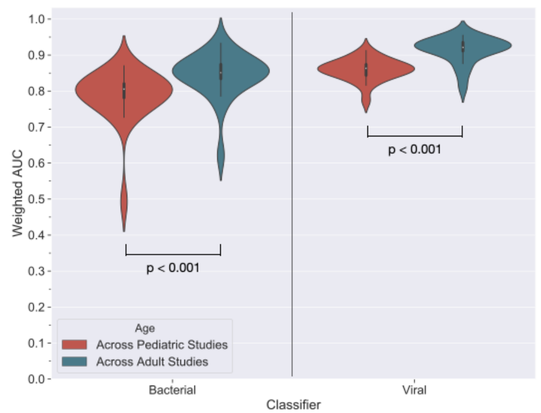


Figure 2: Distribution of weighted average AUCs for each signature as measured in datasets with either pediatric (red) or adult (blue) subjects, as measured for bacterial classification (left) or viral classification (right).

While analyzing the results at the dataset-level, we noticed that across all signatures, datasets comprised of only pediatric patients had lower AUCs than datasets with only adult patients. This significant finding was present in both bacterial and viral classification.

We then applied thresholds, assigned accuracies, and pooled patients to explore differential performance in more heterogenous datasets. Through this analysis, we found differences in overall signature performance between identified pathogens, more granular age categories, and infection severity.

Parameter	Bacterial vs. non-Bacterial			Viral vs. non-Viral		
	Accuracy	p-value	N	Accuracy	p-value	N
All subjects	80 [79-81]	-	2952	85 [84-86]	-	3584
Pathogen class						
All Bacterial*	80 [78-82]	-	1030	-	-	-
S. aureus	74 [70-79]	**< 0.001**	184	-	-	-
E. coli	85 [80-90]	0.210	64	-	-	-
All Viral*	-	-	-	83 [81-84]	-	1679
Rhinovirus	-	-	-	74 [70-78]	**< 0.001**	179
RSV	-	-	-	82 [80-84]	0.119	406
Influenza	-	-	-	90 [88-92]	**< 0.001**	431
Age						
Adult*	83 [81-85]	-	1183	89 [87-90]	-	1268
12-18 years	82 [78-85]	0.087	144	90 [86-94]	0.6	95
2-11 years	70 [67-73]	**< 0.001**	412	80 [77-83]	**< 0.001**	352
3 mo - 1 year	73 [69-77]	**< 0.001**	195	80 [78-83]	**< 0.001**	576
<3 months	85 [82-88]	0.056	322	82 [80-85]	**< 0.001**	547
Severity						
non-ICU*	63 [58-68]	-	99	85 [81-89]	-	117
ICU	71 [67-75]	**0.002**	191	87 [83-91]	0.493	107

Table 1: Overall accuracies (with 95% CI and p-value) of bacterial and viral classification, separated by different patient populations. N is equal to the number of patients validated in the classifier that were annotated with the specific parameter.
* = Reference category used for p-value calculations

Additional Findings

Gene Importance:
- The five most common genes in the 20 signatures and the frequency with which they were included are: *IFI27 (63%)*, *OASL (44%)*, *RSAD2 (38%)*, *ISG15 (38%)*, and *LY6E (38%)*
- The most discriminating genes across all signatures were:
  - Bacterial classification: *CEPT* (+), *ANKRD20A11P* (-), *IFI27* (-), *RPGRIP1* (-), *PDE9A* (-)
  - Viral classification: *IFI27* (+), *FCER1A* (-), *XAF1*(+), *OTOF* (+), *LARP1* (+)

Sample Type:
- Signatures performed better on Whole Blood samples compared to PBMCs
  - Bacterial classification: 0.84 vs. 0.70 median weighted average AUC, *P* < 0.001
  - Viral classification: 0.89 vs. 0.82 median weighted average AUC, *P* = 0.016

Binary Classification:
- Classifying 2 groups (e.g., bacterial or viral) was easier than classifying >2 groups (e.g., bacterial, viral, or SIRS)
  - Bacterial classification: 0.87 vs. 0.82 median weighted average AUC, *P* = 0.002
  - Viral classification: 0.91 vs. 0.86 median weighted average AUC, *P* = 0.016

COVID-19:
- Strong signature performance generalized to datasets that include COVID-19 patients
  - Median weighted average AUC for viral classification: 0.86
- Signatures can accurately differentiate between COVID-19 and other viral infections
  - Median weighted average AUC for COVID-19 classification: 0.83

Conclusions

We observed significant differences among gene expression signatures for bacterial and viral discrimination, though these were not due to variations in the discovery methods or populations. The number of unique genes in a signature directly correlated with test performance.

Across all signatures, performance varied significantly among different age groups, pathogen class, and infection severity. Future clinical validation of host gene expression classifiers must take such population differences into account.

Our analysis was limited by the breadth of sample annotations in GEO and ArrayExpress. There may be additional characteristics imparting differences in signature performance that we could not explore due to limitations in the availability of such data.

References

Signatures were named using the initials of the first and last author followed by number of unique genes in the signature.

Abbrv	First Author	Last Author	PubMed ID
TS1	Tang	Schughart	28619954
HL2	Herberg	Levin	27552617
SB4	Sampson	Brandon	28588308
SK7	Sweeney	Khatri	27384347
SB8	Sampson	Brandon	28588308
SR10	Suarez	Ramilo	25637350
SN10	Sampson	Noursadeghi	32690014
BF11	Bhattacharya	Falsey	28747714
AK11	Andres-Terre	Khatri	26682989
SL20	Song	Lei	28771541

Abbrv	First Author	Last Author	PubMed ID
ZG25	Zaas	Ginsburg	19664979
PT29	Parnell	Tang	22898401
MS29	Mayhew	Sweeney	32132525
RC31	Ramilo	Chaussabel	17105821
HS33	Hu	Storch	23858444
HL34	Herberg	Levin	27552617
ZG48	Zaas	Ginsburg	24048524
MR59	Mahajan	Ramilo	27552618
TW96	Tsalik	Woods	26791949
AK398	Andres-Terre	Khatri	26682989
All	-	-	-

Acknowledgments:
I'd first like to thank my research mentor, Dr. Ephraim Tsalik, for his support and guidance over the past year. I'd also like to thank Dr. Ricardo Henao who helped us immensely with the machine learning and statistical analyses.