Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Consistent PTGDR downregulation has limited clinical specificity for sepsis: a retrospective multicohort transcriptomic analysis

Hemantkumar Patadia
Parul University
Ben Tweve
Parul University
Yellapu Sampreeth Naidu
Parul University
Nikhilesh Dangi
nikhilesh.dangi38493@paruluniversity.ac.in
Parul University

Research Article

Keywords: sepsis, blood transcriptomics, PTGDR, biomarker evaluation, transportability, external prediction verification

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11576550/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Consistent PTGDR downregulation has limited clinical specificity for sepsis: a retrospective multicohort transcriptomic analysis

Hemantkumar Patadia[1], Ben Tweve[2], Yellapu Sampreeth Naidu[3], Nikhilesh Dangi[4*]

[1]Parul Institute of Allied and Healthcare Sciences, Faculty of Medicine, Parul University, Vadodara, Gujarat, India.
[2]Department of Medical Technology & Health Sciences, Parul University, Vadodara, Gujarat, India.
[3]Department of Surgical Technology and Emergency Sciences, Parul Institute of Allied and Healthcare Sciences, Faculty of Medicine, Parul University, Vadodara, Gujarat, India.
[4]Department of Health Informatics, Parul Institute of Allied and Healthcare Sciences, Faculty of Medicine, Parul University, Vadodara, Gujarat, India.

*Corresponding author. E-mail: nikhilesh.dangi38493@paruluniversity.ac.in;

Abstract

Background: Blood transcriptomic signatures can separate sepsis from health, but cross-cohort consistency does not establish specificity against clinically realistic inflammatory alternatives. We corrected the discovery meta-analysis of a historical sepsis panel and evaluated its surviving signal across cohorts.
Methods: GSE57065 and GSE95233 were reprocessed from 231 raw microarray CEL files. Moderated standard errors were combined by restricted maximum-likelihood random-effects meta-analysis with ordinary and modified Hartung–Knapp sensitivities. Four historical genes were evaluated, with a family-wise multiplicity sensitivity. Because both cohorts informed candidate prioritization,

reciprocal modelling was classified as development rather than independent validation and reported by held-out cohort. Metrics were recomputed from stored prediction records for nine contrasts. A PTGDR-only expression extension specified before outcome inspection evaluated bulk RNA-seq, donor-level single-cell data, covariates and composition.

Results: PTGDR was downregulated (pooled effect $\mathbf{-1.972}$, conventional 95% CI $\mathbf{-2.326}$ to $\mathbf{-1.619}$; ordinary Hartung–Knapp $\boldsymbol{P = 0.025}$), but the modified Hartung–Knapp interval crossed zero ($\mathbf{-4.264}$ to 0.320; $\boldsymbol{P = 0.058}$) and the four-hypothesis Holm-adjusted $\boldsymbol{P}$ was 0.100. In reciprocal development folds, PTGDR AUC was 0.896 and 0.972 versus 0.540 and 0.520 for the historical panel. External prediction-record AUC was 0.873 for healthy versus infection but 0.554 for infection versus sepsis; against post-surgical inflammation AUC was 0.790, yet fixed-threshold specificity was 0.032. In newly analysed cohorts, sepsis-versus-healthy AUCs were 0.901–0.970, whereas AUCs were 0.582 versus adult ICU controls and 0.589 versus severe COVID-19 ICU patients. A three-cohort synthesis had $\boldsymbol{I^2 = 67.4\%}$ and a prediction interval crossing zero. In 133-donor single-cell data, adult classical-monocyte AUC was 0.831 versus health but 0.630 versus ICU controls.

Conclusions: Correcting the historical uncertainty calculation rejected the three-gene panel. PTGDR remained the strongest exploratory signal across cohorts, probes, platforms, and selected immune-cell compartments, but did not show sepsis-specific diagnosis, calibration, or fixed-threshold transportability. Biomarker consistency must therefore be distinguished from clinical utility.

Keywords: sepsis, blood transcriptomics, PTGDR, biomarker evaluation, transportability, external prediction verification

1 Introduction

Sepsis is a heterogeneous syndrome in which infection-triggered host responses overlap with sterile inflammation, surgery, shock, and other forms of critical illness. Blood transcriptomics can capture these responses and has supported molecular stratification, yet transfer across cohorts and clinical contexts remains difficult [1–4]. Strong retrospective discrimination, particularly in comparisons with healthy controls, does not by itself demonstrate that a signature can distinguish sepsis from clinically plausible alternatives. Prior multicohort and critical-care studies likewise show that apparent gene-expression diagnostic performance depends on cohort composition and the comparator being tested [5–7].

Compact transcriptomic panels are appealing because they could be easier to interpret and assay than large signatures. Their apparent simplicity can conceal several validation problems. Feature estimates can vary with cohort composition and platform annotation; small panels are sensitive to a single unstable gene; and a model that ranks observations may still be poorly calibrated or unusable at a fixed decision threshold. Cross-platform integration therefore requires explicit annotation, uncertainty, and validation controls [8]. More generally, translation requires a defined clinical use case and evidence beyond discovery cohorts [9, 10].

The present work arose from re-evaluation of a historical three-gene hypothesis comprising *PTGDR*, *S100A9*, and *VPS8*. That hypothesis had been assembled using an uncertainty calculation that increased apparent precision beyond the direct moderated standard errors. We therefore repeated the primary sepsis analysis from raw inputs, reconstructed participant-level contrasts, used direct uncertainty estimates, and established candidate status before predictive evaluation. PTGDR has plausible immunomodulatory context through prostaglandin D2 receptor signalling [11], but biological plausibility was not treated as evidence of diagnostic specificity.

We asked three linked questions: which historical gene provides the strongest exploratory signal after corrected two-cohort meta-analysis and multiplicity sensitivity; whether that benchmark outperforms the historical panel within each reciprocal development fold; and whether stored external predictions and newly analysed expression cohorts retain discrimination across disease-versus-health and clinically harder inflammatory comparisons. Our aim was not to claim a deployment-ready diagnostic, but to identify the boundary between consistent host-response biology and clinically specific transportability. We use *benchmark* to mean a one-gene host-response signal showing cross-cohort consistency and useful for studying transportability, not a clinical diagnostic reference standard.

Four evidence labels are used consistently. *Development* denotes GSE57065/GSE95233, because both cohorts informed candidate selection. *External prediction-record verification* denotes metric recomputation from stored participant-level predictions without raw-cohort reconstruction. The *expression extension* denotes newly analysed public bulk RNA-seq cohorts, while *localization evidence* denotes sorted-cell and donor-level single-cell analyses. Independent prospective or assay-level clinical validation was not performed.

2 Methods

2.1 Study design and evidence hierarchy

This retrospective secondary analysis used public, de-identified transcriptomic data from the NCBI Gene Expression Omnibus (GEO) [12, 13]. The corrected discovery analysis was repeated from raw CEL files in a clean computational environment. Claims were restricted to regenerated primary outputs, metrics recomputed from stored prediction-level records, or newly analysed expression cohorts. The workflow comprised participant reconstruction, raw preprocessing, differential expression, random-effects meta-analysis, candidate adjudication, reciprocal cross-cohort development modelling, external prediction-record verification, an expression extension, and cell-localization analyses (Fig. 1). No prospective sample-size calculation was performed; all eligible participants in the defined contrasts were included. Reporting was informed by relevant TRIPOD transparency principles, while recognizing that this development and transportability study is not an independent model-validation study [14].

2.2 Cohort reconstruction and participant selection

The development cohorts were GSE57065 and GSE95233, both Affymetrix Human Genome U133 Plus 2.0 whole-blood studies. GEO series metadata were reconciled

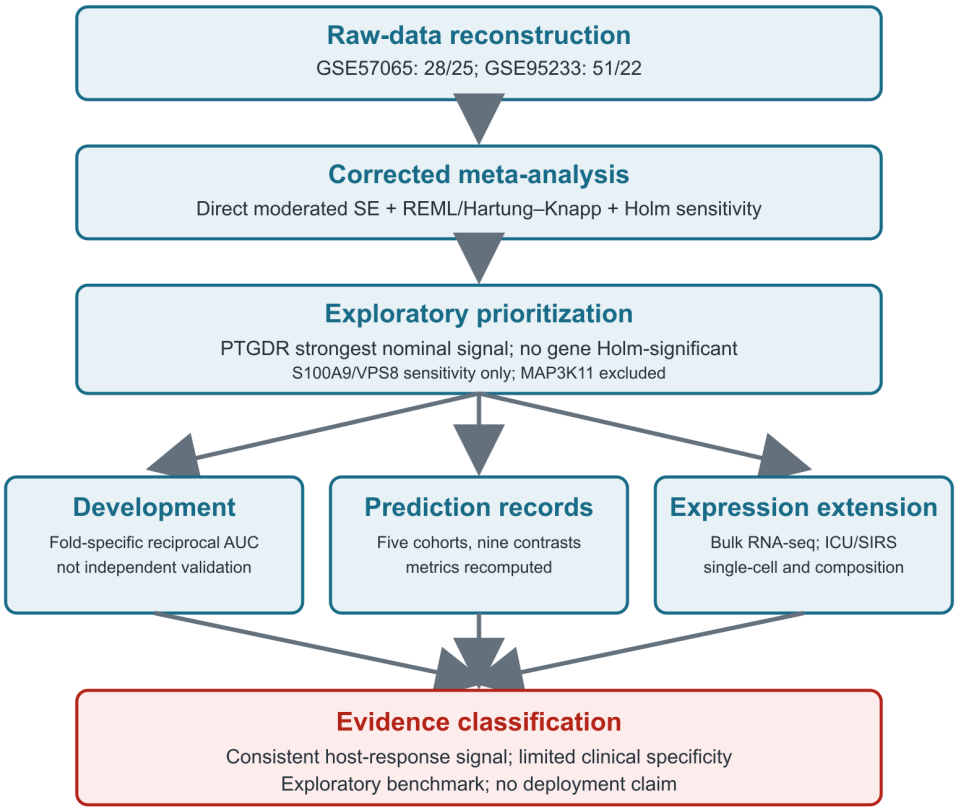


Fig. 1 Study design and evidence hierarchy

with CEL-file identifiers; sample identifiers, participant identifiers, phenotype, and time point were recorded in a predefined manifest. Expression columns and manifest rows were required to match one-to-one. Ambiguous or non-target observations were not assigned synthetic labels and were excluded from the relevant contrast.

The full resources contained repeated sepsis observations, but the primary comparison retained one early sample per participant before differential expression or modelling. GSE57065 contributed 28 H00 sepsis samples and 25 healthy controls;

Table 1 Primary development cohorts and analysis roles

Cohort	Assay and material	Primary case definition	Cases/controls	Role
GSE57065	GPL570 microarray, whole blood	H00 sepsis	28/25	Differential expression and development fold
GSE95233	GPL570 microarray, whole blood	D01 sepsis	51/22	Differential expression and development fold

GSE95233 contributed 51 D01 sepsis samples and 22 independent D00 healthy controls. The primary analysis therefore included 126 unique participants, with no repeated participant within a contrast (Table 1).

2.3 Executed preprocessing, quality control, and annotation

Raw CEL files were read with `affy` and processed by MAS5 summarization followed by log2 transformation. This corrects earlier planned-RMA wording and states the method that was executed. The complete matrices contained 107 GSE57065 and 124 GSE95233 samples. Quality control included distributional summaries, density and box plots, principal-components analysis, and sample-distance plots. No expression-outlier threshold was defined in advance and no sample was removed by an ad hoc QC rule.

Affymetrix probes were mapped to official symbols with `hgu133plus2.db` 3.2.3 (`mapIds`, `multiVals=‘‘first’’`). Unresolved mappings were removed. For multiple probes per symbol, the probe with the highest mean across all cohort samples was retained, yielding 20,962 shared genes. This unsupervised representative-probe rule used expression magnitude only and did not use outcome labels. The selected probes were reconciled against the gene-collapse table. As a sensitivity analysis, every mapped

probe for PTGDR, S100A9, VPS8, and MAP3K11 was fitted under the primary contrast; PTGDR direction was concordant for all three probes in both cohorts. Missing expression values were not imputed.

2.4 Differential expression

For each cohort, a sepsis-versus-control linear model was fitted with `limma`. The gene-level effect was the moderated model coefficient on the processed log2-expression scale. Its standard error was calculated directly as $\sqrt{s^2_{\text{post}}} \times \text{stdev.unscaled}$; the unsupported sample-size precision multiplier present in an earlier workflow was not applied. Two-sided nominal P values were adjusted across genes by the Benjamini–Hochberg method [15].

2.5 Meta-analysis and candidate adjudication

Coefficients and direct standard errors from both cohorts were combined for each of the 20,962 shared genes using restricted maximum-likelihood random-effects meta-analysis. We report the pooled effect, the normal-approximation 95% confidence interval returned by `metafor`, I^2, and ordinary Hartung–Knapp small-sample inference. Because the ordinary Hartung–Knapp residual scale can fall below one with very few studies and produce a standard error smaller than the conventional random-effects standard error, a modified Knapp–Hartung sensitivity replaced that scale by $\max\{1, q/(k-1)\}$. This conservative sensitivity was added during the final source-first audit; it did not replace the historical adjudication rule. Prediction intervals were inspected cautiously but were not treated as stable with only two cohorts. The four historical key genes were adjudicated under a rule established before the focused predictive comparison; this was evaluation of a previously defined candidate set, not a genome-wide de novo PTGDR discovery test. The historical operational rule required direction agreement, a pooled interval excluding zero, nominal ordinary Hartung–Knapp $P < 0.05$, and estimated $I^2 \leq 50\%$. Because four genes were evaluated, a Holm

family-wise adjustment was added as a multiplicity sensitivity. No gene met adjusted $P < 0.05$, and PTGDR also failed modified Knapp–Hartung $P < 0.05$; PTGDR was therefore carried forward only as the strongest exploratory benchmark rather than a statistically confirmed biomarker. With only two cohorts, the I^2 threshold served as an adjudication rule rather than evidence that true between-cohort heterogeneity was known precisely.

This was a development-stage decision, not an independent feature-selection exercise: the corrected meta-analysis used both GSE57065 and GSE95233 and prioritized PTGDR for exploratory benchmarking. Consequently, the reciprocal results below estimate cross-cohort transport within the same two-cohort development system and are not unbiased independent validation estimates.

2.6 Reciprocal cross-cohort development comparison

One fold fitted a model in GSE95233 and evaluated GSE57065; the reciprocal fold fitted GSE57065 and evaluated GSE95233. Within each fold, gene means and sample standard deviations were estimated from the training cohort and applied unchanged to the other cohort. Binomial generalized linear models were fitted with `statsmodels`; an effectively unpenalized ridge fit (L2 penalty 10^{-8}) was the predefined numerical fallback if standard fitting failed. Two historically defined model forms were compared: PTGDR alone and PTGDR/S100A9/VPS8. Thus, test-cohort outcomes did not enter fold-specific centering, scaling, or coefficient estimation, but both cohorts had informed the preceding candidate-priority decision. Concatenating non-overlapping fold predictions yielded 126 participant-level development predictions.

Performance measures were receiver operating characteristic area under the curve (ROC AUC), average precision, and Brier score. ROC AUC quantified ranking discrimination [16]; average precision was calculated with the tie-aware step-wise `scikit-learn` definition because prevalence differed across cohorts; and Brier score

summarized probability error. Primary results were reported separately for each held-out cohort because probabilities from separately fitted models need not share a common scale. Concatenated prediction metrics were retained only as secondary descriptive records and were not used as the primary ranking comparison. No probability threshold was selected for this development comparison.

Within each held-out cohort, 95% intervals and paired AUC differences were quantified by 5,000 outcome-stratified participant bootstrap replicates (seed 20261026). These intervals condition on the fitted development predictions; they do not represent uncertainty from cohort sampling, candidate selection, or model redevelopment.

2.7 External prediction-record verification

The historical external scoring object was distinct from the reciprocal models. Before external scoring, a PTGDR-only binomial model was fitted to the pooled GSE57065 and GSE95233 primary samples. Its coefficient, intercept, discovery-sample centring and scaling parameters, score direction, and probability threshold of 0.5 were fixed in the model registry. External refitting, recalibration, and threshold optimization were prohibited by the scoring specification.

Not all external cohorts were reconstructed from raw data in the present study. Instead, stored sample labels and probabilities were re-scored for nine defined contrasts from GSE74224, GSE154918, GSE65682, GSE134347, and GSE137340. ROC AUC, average precision, Brier score, scaled Brier score, calibration intercept and slope, and sensitivity and specificity at 0.5 were recomputed. Ninety-five percent confidence intervals for external ROC AUC were estimated separately within each contrast using 5,000 outcome-stratified participant bootstrap replicates and percentile limits (seed 20260824); these intervals condition on the stored prediction rows and do not propagate raw-data reconstruction or model-development uncertainty. Historical-panel metrics were also recomputed wherever all three features had been scored. All nine

PTGDR ROC AUC and Brier estimates agreed with stored evidence. GSE65682 average precision was corrected from 0.916 to 0.959 under the step-wise definition; other conclusions were unchanged.

Contrasts were interpreted by clinical question rather than pooled. Healthy-versus-disease comparisons assessed broad host-response ranking. Infection versus sepsis, sepsis versus shock, and sepsis versus post-surgical inflammation were treated as more clinically informative failure boundaries. Fixed-threshold sensitivity and specificity were reported only where supported by the stored scoring records.

2.8 PTGDR expression extension specified before outcome inspection

The baseline analysis was finalized before a separate PTGDR-only extension was specified on 31 August 2026, before its PTGDR outcomes were inspected. The extension tested lower PTGDR in early sepsis; candidate reselection, multigene modelling, threshold optimization, and exclusion based on PTGDR outcome were prohibited. Additional public RNA-seq evidence comprised whole blood in GSE63042 (106 sepsis/23 noninfectious SIRS) and GSE185263 (348 sepsis/44 healthy), PBMCs in GSE243217 (22 sepsis/35 COVID-19/15 healthy) and GSE279448 (116 adult sepsis/23 adult ICU control/43 adult healthy; 16 pediatric sepsis/5 pediatric ICU control/4 pediatric healthy), and a linked earliest-participant whole-blood contrast across GSE222393/GSE185263 (22 non-COVID sepsis/20 severe COVID-19 ICU). The linked contrast overlaps GSE185263 and was analysed separately rather than treated as an independent cohort. Two sorted-cell RNA-seq cohorts, GSE133822 and GSE136200, and one PBMC single-cell cohort, GSE167363, addressed localization.

RNA-seq counts were reduced to $\log_2(\mathrm{CPM} + 0.5)$ for the specified PTGDR gene identifier. Each contrast reported the mean difference with a Welch 95% confidence interval, Hedges' g, two-sided Welch and Mann–Whitney tests, and AUC oriented

so lower PTGDR predicted sepsis with 5,000 group-stratified participant bootstrap replicates. Integer-count cohorts additionally used a one-gene negative-binomial NB2 model with log library-size offset. Adult and pediatric strata and healthy and clinically ill comparators were kept separate.

For GSE63042, seven cell-composition scores were defined before PTGDR inspection: neutrophil, monocyte, T cell, B cell, NK cell, platelet, and erythroid. Each was the mean standardized log-expression of its available predefined markers; all entered an HC3 linear model together. In GSE167363, only baseline/T0 samples from five sepsis and two healthy donors were used. Cells with 200–6,000 detected genes and less than 20% mitochondrial UMIs were retained, assigned to the highest predefined marker score without PTGDR, and summarized by participant within cell type. Cells were never treated as independent replicates. Benjamini–Hochberg adjustment was applied across tested cell types within each localization dataset.

On 1 September 2026, before computing pooled, adjusted, GSE279452, or matched-composition outcomes, a second protocol specified four analyses. First, the three adult sepsis-versus-healthy RNA-seq estimates (GSE185263, GSE243217, and adult GSE279448) were combined as Hedges' g using REML random effects, conventional and Hartung–Knapp intervals, heterogeneity statistics, a prediction interval, and leave-one-cohort-out sensitivity. Clinically different comparator families were not pooled. Second, available age and sex were entered without outcome-based selection in HC3 OLS and NB2 count models; predefined common-location/site restrictions were also evaluated where both groups occurred.

Third, GSE279452 expression libraries comprised 93 adult sepsis, 8 adult ICU-control, 7 adult healthy, 17 pediatric sepsis, 4 pediatric ICU-control, and 4 pediatric healthy PBMC donors. V(D)J records were excluded. The same cell QC was applied; outcome-blind predefined markers assigned eight compartments, with zero-score cells unclassified and PTGDR excluded from annotation. Inference used donor-by-cell-type

pseudobulk $\log_2(\mathrm{CPM}+0.5)$, with adult/pediatric and healthy/ICU families kept separate and Benjamini–Hochberg adjustment within each family. Raw expression triplets were resolved by stable GSM accession; an observed two-record title/file donor-suffix exchange and six capitalization differences were preserved in an identity map. Finally, GSE279448 bulk and GSE279452 single-cell donors were matched by archive donor suffix. Models compared case status alone, case status plus age/sex, and case status plus age/sex and six isometric-log-ratio coordinates representing seven measured broad compartments after a 0.5-cell replacement. The eligibility gate required at least 50 matched donors, five controls, 90% compartment representation, and full-rank HC3 models.

2.9 Longitudinal sensitivity

Paired H24-versus-H00, H48-versus-H00, and H48-versus-H24 contrasts in GSE57065 were retained as descriptive temporal sensitivity analyses. The analogous GSE95233 D02-versus-D01 and D03-versus-D01 comparisons and outcome-by-time comparisons were rebuilt from explicit within-patient expression differences after the historical `duplicateCorrelation` implementation segfaulted under the legacy runtime. For the interaction analyses, patient-level changes were compared between survivors and non-survivors. These longitudinal results were not pooled with the independent-participant primary contrast and did not influence candidate selection.

Superseded cross-disease, target-cohort-refitted, and exploratory feature-selection branches were not used for inference. Their execution status and exclusion rationale are documented in the supplementary methods and provenance records.

2.10 Software and computational verification

The primary pipeline was tested on Linux using R 3.6.1 with `affy`, `hgu133plus2.db`, `limma`, and `metafor`. To bound native-memory retention in the legacy R/Bioconductor stack, each discovery cohort was processed from CEL files in a separate single-threaded

R process before statistical analysis of the saved checkpoints. The predictive, correction, and expression-extension analyses were retested in an installable Python 3.12 environment with pandas, statsmodels, scikit-learn, scipy, and openpyxl; software version information is documented in the computational records available from the corresponding author. The analysis used versioned R and Python environments, checksum-validated raw inputs, deterministic verification tests, and archived scripts and reference outputs. A clean-host rerun completed all 231 CEL files and matched 13 of 13 whole-primary output files within a prespecified 10^{-10} tolerance (maximum absolute difference 1.39×10^{-11}). Re-execution also matched all 167 primary/secondary numerical checks (maximum absolute difference 7.28×10^{-12}), all 196 longitudinal checks, and all 146 extension checks; four core metric tests and three synthetic extension tests passed. This establishes numerical reproducibility in the tested environment, not platform independence or clinical validity.

The accompanying supporting files include corrected key-gene meta-analysis and inference sensitivities, held-out-cohort model performance, external prediction-level verification and calibration, participant-level external predictions, discovery sample provenance, probe sensitivity, and extended supplementary methods. These are supplied as individually named CSV and PDF files listed under Supplementary files in the Declarations. The files are a selected reporting subset, not a complete executable pipeline or a replacement for historical verification archives. Additional analysis code, software environments, and validation-extension outputs should be obtained through the code-availability arrangement described below.

3 Results

3.1 Corrected meta-analysis prioritized PTGDR as an exploratory benchmark

The two participant-level discovery contrasts agreed that PTGDR expression was lower in sepsis. Cohort effects were −2.103 (SE 0.353) in GSE57065 and −1.926 (SE 0.210) in GSE95233. The pooled effect was −1.972 (conventional 95% CI −2.326 to −1.619); the ordinary Hartung–Knapp interval was −2.957 to −0.987 with $P = 0.025$ (Fig. 2; Table 2). Because its ordinary Hartung–Knapp scale was 0.185, the modified safeguard widened the interval to −4.264 to 0.320 and gave $P = 0.058$. No between-cohort heterogeneity was detected ($I^2 = 0\%$), although heterogeneity is poorly determined with only two cohorts. PTGDR met the historical operational rule, but its Holm-adjusted P across the four evaluated genes was 0.100 and the modified Knapp–Hartung sensitivity crossed zero. It was therefore prioritized only as the strongest exploratory benchmark, not classified as a confirmatory biomarker.

S100A9 was consistently positive but highly heterogeneous: the cohort effects were 1.280 and 2.418, and the pooled estimate was 1.849 (95% CI 0.733 to 2.964; $I^2 = 98.31\%$; Hartung–Knapp $P = 0.190$). VPS8 had effects of 0.099 and 0.473 and a pooled estimate of 0.280 (95% CI −0.087 to 0.647; $I^2 = 76.43\%$; Hartung–Knapp $P = 0.375$). MAP3K11 changed direction between cohorts and had a pooled effect near zero (0.001; 95% CI −0.207 to 0.210; $I^2 = 55.73\%$; Hartung–Knapp $P = 0.991$). Thus, S100A9 and VPS8 were demoted to sensitivity-only observations, and MAP3K11 was excluded.

3.2 PTGDR exceeded the historical panel within both development folds

When GSE57065 was held out, PTGDR achieved ROC AUC 0.896 (95% conditional bootstrap CI 0.799–0.969), compared with 0.540 (0.500–0.600) for the historical panel

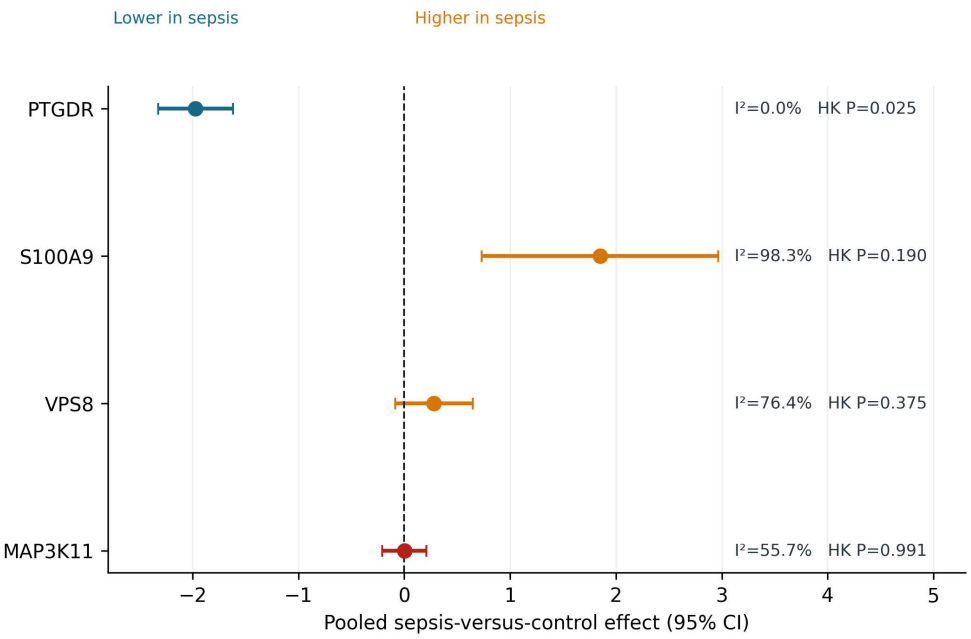


Fig. 2 Corrected two-cohort meta-analysis of historical key genes. Points and bars show the random-effects pooled estimates and normal-approximation 95% confidence intervals; ordinary and modified Knapp–Hartung sensitivities are reported in the text, table, and `Meta_Analysis_Inference_Sensitivity.csv`.

(Fig. 3; Table 3). When GSE95233 was held out, the corresponding AUCs were 0.972 (0.936–0.998) and 0.520 (0.500–0.549). The paired PTGDR-minus-panel AUC differences were 0.356 (0.256–0.441) and 0.453 (0.407–0.488), respectively. These intervals quantify resampling of fixed predictions within each held-out cohort and do not remove the development-stage information-flow limitation.

Average precision and Brier score showed the same within-fold ordering. PTGDR average precision was 0.912 and 0.989, versus 0.549 and 0.710 for the panel; PTGDR Brier scores were 0.261 and 0.183, versus 0.472 and 0.685. Complete fold-specific estimates and intervals are supplied in `Fold_Specific_Model_Performance.csv`.

Concatenating the two prediction sets produced scale-sensitive descriptive AUCs of 0.765 for PTGDR and 0.425 for the panel because the separately trained models

Table 2 Corrected meta-analysis and candidate classification

Gene	Pooled effect	Conv. 95% CI	I^2 (%)	HK P	mKH P
PTGDR	−1.972	−2.326, −1.619	0.00	0.025	0.058
S100A9	1.849	0.733, 2.964	98.31	0.190	0.190
VPS8	0.280	−0.087, 0.647	76.43	0.375	0.375
MAP3K11	0.001	−0.207, 0.210	55.73	0.991	0.991

Note: HK is ordinary Hartung–Knapp; mKH applies the scale safeguard $\max\{1, q/(k-1)\}$. For PTGDR, the ordinary HK 95% CI was −2.957 to −0.987 and the mKH CI was −4.264 to 0.320. PTGDR was retained only as an exploratory benchmark (Holm $P = 0.100$); S100A9 and VPS8 were sensitivity-only, and MAP3K11 was excluded. Adjudication was restricted to the previously defined four-gene historical candidate set and was not a genome-wide de novo discovery test. With two cohorts, neither $I^2 = 0\%$ nor small-sample inference establishes precisely known homogeneity.

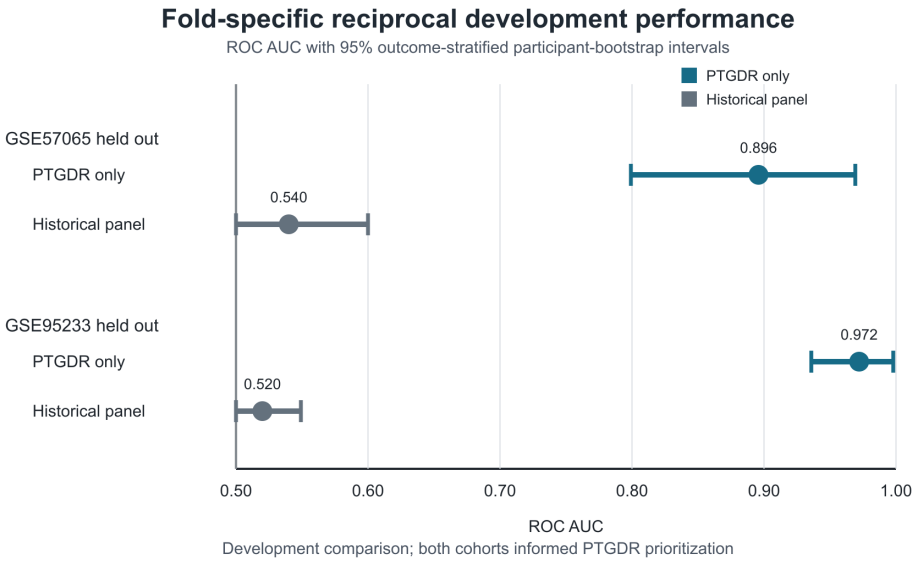


Fig. 3 Fold-specific reciprocal development AUC; not independent validation

occupied different probability ranges. Those pooled rank statistics are retained in the machine-readable provenance but are not used as the primary model comparison.

Table 3 Fold-specific reciprocal development comparison; not independent validation

Held-out cohort	Model	n	ROC AUC	Average precision	Brier score
GSE57065	PTGDR only	53	0.896	0.912	0.261
GSE57065	Historical panel	53	0.540	0.549	0.472
GSE95233	PTGDR only	73	0.972	0.989	0.183
GSE95233	Historical panel	73	0.520	0.710	0.685

3.3 External prediction-record performance depended on the clinical question

All nine external contrasts matched their stored ROC AUC and Brier values when recomputed from prediction-level rows (Fig. 4; Table 4). In GSE154918, PTGDR separated healthy participants from uncomplicated infection (AUC 0.873) and from sepsis (AUC 0.860). However, infection-versus-sepsis discrimination was close to chance (AUC 0.554), arguing against sepsis specificity. Severity-oriented comparisons were intermediate: sepsis versus shock yielded AUC 0.747 and infection versus shock yielded AUC 0.816.

Against post-surgical infection-negative inflammation in GSE74224, PTGDR achieved AUC 0.790, average precision 0.914, and Brier score 0.275. Despite that ranking, the fixed threshold had sensitivity 1.000 and specificity 0.032, indicating near-universal positive classification. The large GSE65682 healthy-control comparison had AUC 0.616 and Brier score 0.052, but its fixed-threshold specificity was 0.0. Its corrected average precision of 0.959 reflects the 760:42 case-control imbalance and should not be read as evidence of strong specificity.

GSE134347 and GSE137340 produced AUCs of 0.703 and 0.688, respectively. Taken together, the external results were compatible with a broad disease-associated host-response signal, but they did not establish reliable discrimination between sepsis and related infectious or inflammatory states.

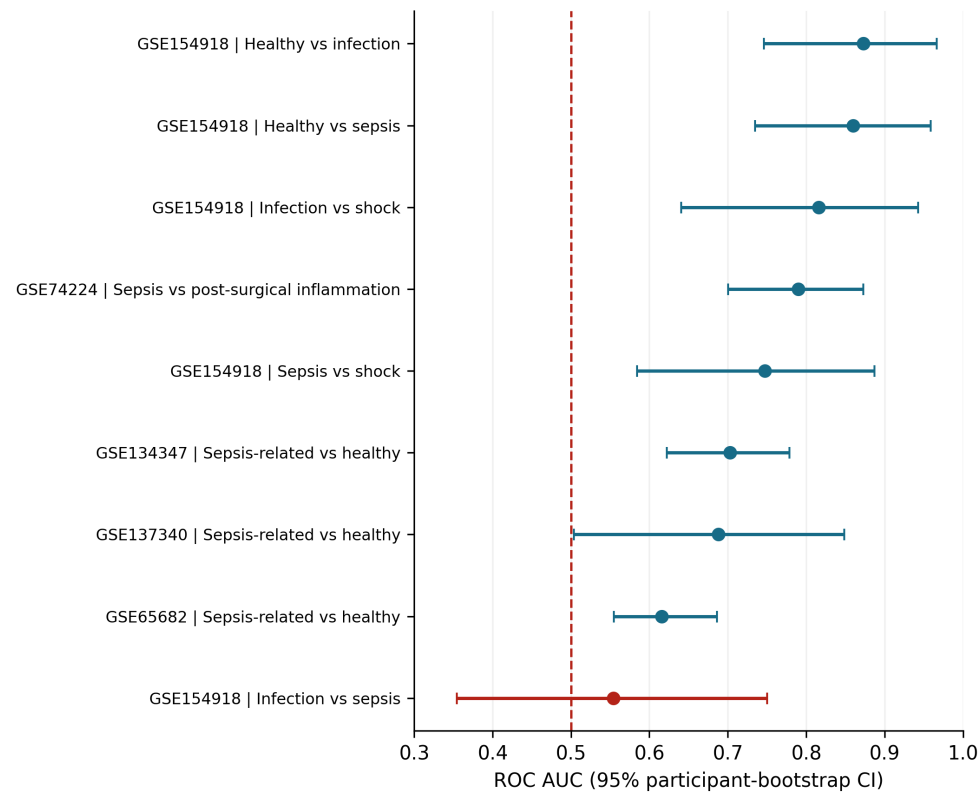


Fig. 4 External PTGDR prediction-record verification across nine clinical contrasts

Threshold and calibration checks strengthened this boundary. Specificity at the fixed 0.5 threshold was zero in GSE65682, GSE134347, and GSE137340 and 0.032 in GSE74224. Calibration intercept/slope pairs were −41.016/3.190, −5.471/0.792, −3.113/0.433, and −2.219/0.674, respectively. Scaled Brier scores were negative in all four cohorts, indicating worse probability error than the prevalence-only reference despite some ranking ability.

Where the complete historical panel had stored predictions, it underperformed PTGDR in GSE74224 (AUC 0.606 vs 0.790) and GSE137340 (0.374 vs 0.688). The panel ranked disease-versus-health contrasts strongly in GSE154918 (AUC 1.000 and

Table 4 External PTGDR-only prediction-record performance

Dataset	Contrast	Cases/controls	AUC	Brier	Interpretation
GSE74224	Sepsis vs post-surgical inflammation	74/31	0.790	0.275	Ranking present; threshold specificity 0.032
GSE154918	Healthy vs infection	12/40	0.873	0.128	Broad disease-versus-health signal
GSE154918	Healthy vs sepsis	20/40	0.860	0.136	Broad disease-versus-health signal
GSE154918	Infection vs sepsis	20/12	0.554	0.285	Weak sepsis specificity
GSE154918	Sepsis vs shock	19/20	0.747	0.271	Limited severity transport
GSE154918	Infection vs shock	19/12	0.816	0.192	Exploratory severity contrast
GSE65682	Sepsis-related vs healthy	760/42	0.616	0.052	Imbalanced; threshold specificity 0.0
GSE134347	Sepsis-related vs healthy	156/59	0.703	0.273	Moderate broad-response ranking
GSE137340	Sepsis-related vs healthy	43/12	0.688	0.218	Limited by small control group

0.986) but had Brier scores of 0.630 and 0.546 with specificity 0, demonstrating non-transportable probability scale. For infection versus sepsis, both models were weak (AUC 0.600 for the panel and 0.554 for PTGDR).

3.4 The expression extension showed consistent disease-versus-health differences but preserved the clinical boundary

PTGDR was lower in three cross-platform sepsis-versus-healthy contrasts: GSE185263 (difference −1.795, 95% CI −2.034 to −1.556; AUC 0.901, 95% CI 0.863–0.935), GSE243217 (−2.248, −2.798 to −1.697; AUC 0.970, 0.900–1.000), and adult GSE279448 (−1.739, −1.977 to −1.500; AUC 0.933, 0.894–0.965). The pediatric healthy comparison was directionally similar (16 sepsis/4 healthy; AUC 0.844, 0.594–1.000) but imprecise.

Clinically harder comparisons were weaker and not uniformly concordant. In GSE243217, sepsis remained lower than COVID-19 (difference −1.044, 95% CI −1.661 to −0.426; AUC 0.745, 0.614–0.862), but adult sepsis versus adult ICU control in GSE279448 was uncertain (difference −0.202, −0.593 to 0.189; AUC 0.582, 0.458–0.704). The linked earliest-participant non-COVID sepsis versus severe COVID-19 ICU contrast was also weak (difference −0.513, −1.252 to 0.225; AUC 0.589, 0.405–0.755). Pediatric sepsis versus pediatric ICU control reversed slightly (difference 0.121, −0.510 to 0.751; AUC 0.413, 0.150–0.663). GSE63042 sepsis versus noninfectious SIRS was directional but uncertain (difference −0.307, −0.816 to 0.202; AUC 0.582, 0.447–0.712).

In GSE63042, the sepsis coefficient changed from −0.307 (HC3 95% CI −0.806 to 0.192; $P = 0.228$) to −0.229 (−0.661 to 0.203; $P = 0.299$) after all seven predefined composition scores entered the model. Thus adjustment attenuated but did not reverse the point estimate, while neither model established an association independent of composition.

Localization was compartment dependent. In GSE133822, early-sepsis PTGDR was lower in sorted CD14 cells than in healthy controls (difference −1.319; Hedges' $g = -1.443$; within-dataset $q = 0.00160$; AUC 0.821), with a weaker CD4 pattern ($q = 0.0920$); CD8 results were null or reversed. GSE136200 was small (6 sepsis/4 healthy per population), with lower CD14, CD19, and CD4 point estimates but a slight CD8 reversal and no multiplicity-adjusted result. GSE167363 retained 34,922 cells from five sepsis and two healthy donors. Participant-level monocyte pseudobulk showed the strongest exploratory direction (difference −0.967; exact $P = 0.0952$; AUC 1.000), while B cells reversed slightly and erythroid cells were non-informative. The two healthy donors preclude confirmatory single-cell inference.

The three-cohort adult sepsis-versus-healthy synthesis, specified before pooling, showed material heterogeneity ($I^2 = 67.4\%$) and a wide prediction interval crossing

zero (−7.315 to 3.570); a consistent effect in a future comparable cohort therefore cannot be assumed. The average pooled effect was Hedges' $g = -1.872$ (Hartung–Knapp 95% CI −2.958 to −0.787; $P = 0.0177$). Because the synthesis combined standardized effects across whole-blood and PBMC datasets and platforms, it summarizes cross-study disease-versus-health contrast magnitude rather than a common assay-scale effect. Age/sex and common-location/site sensitivities retained the large disease-versus-health coefficients in GSE185263 and adult GSE279448, whereas adult ICU and linked severe-COVID ICU contrasts remained uncertain or model-sensitive.

GSE279452 processed 133 donors and retained 941,852 cells. Against seven adult healthy donors, all eight annotated compartments had lower PTGDR point estimates; classical-monocyte donor pseudobulk differed by −1.136 log2 CPM (Hedges' $g = -0.934$; within-family Welch $q = 0.00365$; AUC 0.831, 95% CI 0.731–0.919). Against eight adult ICU controls the classical-monocyte difference was smaller and uncertain (−0.533; $q = 0.275$; AUC 0.630, 0.448–0.805), and no compartment had Welch $q < 0.05$. Pediatric healthy comparisons were mostly directionally lower but had four controls; pediatric ICU directions were mixed and no adjusted result was retained as positive.

The matched measured-composition gate passed for 54 adult sepsis/healthy donors (47/7) and 53 sepsis/ICU donors (47/6). For sepsis versus healthy, the bulk PTGDR coefficient changed from −1.700 crude to −1.172 after age, sex, and all six composition coordinates (HC3 95% CI −1.754 to −0.590; $P = 7.95 \times 10^{-5}$). For sepsis versus ICU, the fully adjusted coefficient remained near null (−0.217, −0.817 to 0.384; $P = 0.479$). Thus measured composition attenuated but did not remove disease-versus-health separation and did not rescue ICU specificity.

3.5 Longitudinal sensitivity results did not alter candidate selection

In GSE57065, longitudinal changes after the H00 sepsis sample frequently opposed the primary sepsis-versus-health direction. PTGDR increased at H24 and H48 relative to H00, while S100A9 decreased; VPS8 was positive at H24 and H48 relative to H00 but nearly unchanged between H24 and H48. These within-sepsis temporal shifts show that sampling time can change apparent effects and support retaining the independent-participant early contrast as primary.

The repaired GSE95233 paired-difference analysis showed the same temporal caution. From D01 to D03, PTGDR increased by 0.442 log2 units (nominal $P = 0.0045$, FDR 0.054) and S100A9 decreased by 0.247 units (FDR 0.014); VPS8 was not significant. None of the survivor-versus-non-survivor change differences survived genome-wide adjustment. These results were descriptive and did not modify PTGDR's exploratory benchmark status.

3.6 Alternative probes preserved the PTGDR direction

The highest-mean rule selected PTGDR probes 215894_at in GSE57065 and 215937_at in GSE95233. All alternatives agreed: the three GSE57065 probe effects were −2.103, −1.395, and −1.487, and the three GSE95233 effects were −1.926, −1.840, and −1.664. Thus, PTGDR prioritization was not created by choosing different representative probes across cohorts. VPS8 probe effects were more variable, reinforcing its sensitivity-only status.

4 Discussion

4.1 Principal findings

Correcting the uncertainty calculation materially changed the scientific conclusion. The historical PTGDR/S100A9/VPS8 panel did not survive discovery repair and performed worse than PTGDR alone in a reciprocal cross-cohort development comparison. The expression extension, specified before inspection of its outcomes, showed lower PTGDR across additional public RNA-seq disease-versus-health cohorts and selected CD14/monocyte and T-cell compartments. Its three-cohort synthesis nevertheless showed material between-cohort heterogeneity and a prediction interval crossing zero. In matched donors, measured composition attenuated but did not remove healthy-control separation, while ICU separation remained near null. Both the stored external predictions and newly analysed expression data therefore depended strongly on comparator choice: disease-versus-health ranking was often strong, whereas adult ICU, severe COVID-19 ICU, noninfectious SIRS, and pediatric ICU boundaries were weak, uncertain, or reversed. The defensible output is a PTGDR-centred host-response benchmark and an explicit failure boundary, not a validated sepsis diagnostic.

4.2 Why corrected uncertainty changed the panel

Meta-analysis depends on both study effects and their uncertainty. The repaired workflow used the direct moderated standard error from each `limma` fit and removed an unsupported extra precision scaling. This exposed substantial between-cohort heterogeneity for S100A9 and VPS8. With only two cohorts, apparently precise conventional pooled intervals can be misleading, and neither $I^2 = 0\%$ nor the ordinary Hartung–Knapp result establishes precisely known homogeneity. For PTGDR, ordinary Hartung–Knapp inference excluded zero, but its residual scale was below one; the modified safeguard widened the interval across zero and gave $P = 0.058$. PTGDR also failed the four-hypothesis Holm sensitivity. Its status is therefore exploratory, and its

heterogeneity estimate remains weakly determined. This illustrates why transparent derivation of standard errors, alternative small-sample safeguards, and multiplicity are as important as the stability of point estimates.

4.3 Ranking is not diagnostic specificity

The external contrast pattern is clinically instructive. AUCs of 0.873 and 0.860 for disease versus health could look encouraging in isolation, but AUC fell to 0.554 for infection versus sepsis. Similarly, PTGDR ranked sepsis above post-surgical inflammation with AUC 0.790 while the fixed threshold classified almost all comparators as positive. Ranking, calibration, threshold utility, and clinical specificity are different properties. A biomarker cannot be considered deployable when only the first is partly supported.

The GSE65682 result also shows why prevalence-sensitive metrics require context. Average precision was high after correction, but cases outnumbered controls by approximately 18 to one, AUC was only 0.616, and specificity at the fixed threshold was zero. In such settings, average precision can largely reflect the high positive-class prevalence rather than clinically useful rule-in ability. Negative scaled Brier values in several cohorts likewise indicate greater probabilistic prediction error than the prevalence-only reference model.

4.4 Biological interpretation of PTGDR

PTGDR encodes the DP1 receptor for prostaglandin D2 and participates in immunomodulatory signalling; experimental work supports context-dependent anti-inflammatory and vascular effects [11, 17]. PTGDR itself is not a newly implicated sepsis gene: a recent multi-omics study of sepsis-associated liver injury included it among four T-cell-linked hub genes and combined public transcriptomics with murine and single-cell analyses [18]. The contribution here is a controlled evaluation of the evidence boundary after correction of a historical panel.

The cross-platform and cell-localization results make a purely platform-specific artifact less plausible and identify CD14/monocyte and, less consistently, T-cell compartments as candidate contributors. Reduced whole-blood PTGDR could reflect altered leukocyte abundance, within-cell transcription, immune-cell trafficking, treatment, or sampling time. The larger single-cell cohort and matched-composition analysis make a purely bulk-mixture explanation less likely in the analysed adult subset, but outcome-blind marker scores are not orthogonally validated labels and compositional adjustment is not causal mediation. The present data cannot distinguish these mechanisms or establish causality. Attenuation after adjustment, clinically ill comparator uncertainty, pediatric reversals, and longitudinal rebound instead support state- and context-dependent interpretation. PTGDR is best regarded as a cross-cohort-consistent one-gene test case for transportability, not as a diagnostic reference standard.

4.5 Implications for transcriptomic biomarker development

The study supports a sequence of evidence gates. Candidate effects and uncertainty should first be transparently verifiable. Eligibility should then be defined before model evaluation. Cohort-level rather than randomly mixed validation is needed to expose dataset shifts. External cohorts should be stratified by clinical question, with realistic alternatives carrying more weight than healthy controls. Finally, discrimination, probability error, and fixed-threshold behaviour should be reported separately. These practices can prevent a biologically consistent signal from being overstated as a clinically validated test.

4.6 Strengths and limitations

Strengths include reprocessing 231 raw CEL files, one early observation per participant, direct moderated standard errors, small-sample meta-analytic inference, corrected annotation traceability, alternative-probe sensitivity, candidate criteria

defined before predictive comparison, conditional paired uncertainty estimates, external prediction-record verification, complete fixed-threshold reporting, and preservation of negative contrasts. The expression extension added checksum-verified RNA-seq counts, a PTGDR-only protocol, explicit addenda when eligibility errors were found, earliest-participant reconstruction across linked deposits, count-scale sensitivity, predefined composition scores, and participant-level rather than cell-level single-cell inference.

Several limitations are substantial. Only two discovery cohorts were available, making heterogeneity and Hartung–Knapp inference imprecise even when used conservatively. PTGDR priority was established from a meta-analysis containing both cohorts; although each reciprocal fold used training-only normalization and coefficient fitting, its performance is an internal development estimate rather than independent validation. All analyses were retrospective and derived from public datasets with heterogeneous enrolment, treatment, timing, platforms, and phenotype definitions. The expression extension used processed public counts rather than uniform raw-read realignment, and GSE185263 overlaps the linked GSE222393 contrast, so those estimates are not independent. The three-cohort expression synthesis included heterogeneous cohorts; its wide prediction interval precludes assuming a common future-study effect. The GSE63042 composition analysis used marker scores, while GSE279452 provided measured cell proportions but still depended on outcome-blind marker-score annotations rather than investigator-validated labels or clinical differential counts. Its adult healthy and ICU groups contained only seven and eight donors, pediatric comparator groups only four, and two accession-keyed archive filenames exchanged donor suffixes relative to GEO display titles; this was preserved in an identity map and did not change group, age, or sex assignment. The earlier GSE167363 single-cell analysis had only two healthy donors. The external prediction-record evaluation recomputed metrics from stored prediction-level records rather than

reconstructing every external cohort from raw data. Threshold results depended on the historical fixed scoring scale and were not prospective decision-curve evaluations. No independent assay, prospective cohort, prospectively defined clinical workflow, or impact analysis was available.

4.7 Future work

PTGDR should next be tested prospectively with a fixed assay and a prospectively defined clinical question, ideally infection versus sepsis or infectious versus sterile critical illness rather than disease versus health. Validation should include representative prevalence, time since presentation, measured differential cell counts, calibration, decision-curve analysis, and comparison with routine clinical variables. Confirmatory single-cell or sorted-cell work requires substantially more control donors and orthogonal cell annotation. A multigene extension should be nominated anew in independent development data and should not inherit S100A9 or VPS8 merely from the rejected historical panel.

5 Conclusions

Corrected two-cohort analysis from raw CEL files rejects the historical PTGDR/S100A9/VPS8 sepsis panel. PTGDR was directionally consistent across cohorts, alternative probes, RNA-seq platforms, and selected immune-cell compartments and exceeded the panel in reciprocal development modelling. Nevertheless, ICU and inflammatory comparators, composition sensitivity, calibration, and fixed-threshold analyses revealed limited specificity and utility. These findings provide evidence against interpreting disease-versus-health discrimination as sepsis-specific diagnostic validation. PTGDR is supported as an exploratory host-response benchmark, while its clinically relevant failure boundaries demonstrate why biological consistency alone is insufficient for deployment.

Abbreviations

AP, average precision; AUC, area under the receiver operating characteristic curve; CI, confidence interval; GEO, Gene Expression Omnibus; HK, Hartung–Knapp; REML, restricted maximum likelihood; ROC, receiver operating characteristic.

Declarations

Ethics approval and consent to participate

This secondary analysis used only publicly available, de-identified data and involved no new participant recruitment, contact, or collection of identifiable information. No additional ethics approval or informed consent was required for the present secondary analysis.

Consent for publication

Not applicable.

Availability of data and materials

The transcriptomic datasets analysed in this study are publicly available through the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/). The datasets include GSE57065, GSE95233, GSE74224, GSE154918, GSE65682, GSE134347, GSE137340, GSE63042, GSE133822, GSE136200, GSE167363, GSE185263, GSE222393, GSE243217, GSE279448, and GSE279452. Selected derived analysis results, external validation metrics, participant-level prediction records, sample provenance information, and statistical sensitivity analyses are provided in the accompanying supplementary files. Additional computational records and analysis code are available from the corresponding author upon reasonable request.

Code availability

The selected supplementary files do not include the complete executable workflow. Custom R and Python scripts, environment specifications, input manifests, verification tests, and reference outputs are available from the corresponding author upon reasonable request. The workflow was tested on Linux using R and Python; platform independence is not claimed.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors received no dedicated external funding for this study.

Authors' contributions

N.D. conceived the study, designed the analytical workflow, conducted the analyses, interpreted the data, and drafted the manuscript. B.T. contributed to study framing, interpretation, and critical revision. H.P. contributed to supervision, interpretation, and critical revision. Y.S.N. contributed to clinical interpretation and critical revision. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the investigators and participants whose public GEO datasets made this secondary analysis possible.

Supplementary files

The following individually named supporting files accompany the manuscript:

- `PTGDR_Key_Gene_Meta_Analysis.csv`: corrected two-cohort key-gene estimates and candidate classifications.
- `Meta_Analysis_Inference_Sensitivity.csv`: conventional, ordinary Hartung–Knapp, and modified Knapp–Hartung inference.
- `Fold_Specific_Model_Performance.csv`: held-out-cohort metrics and conditional uncertainty estimates.
- `External_Performance_Verification.csv`: external prediction-record performance verification.
- `External_Calibration_and_Thresholds.csv`: calibration and fixed-threshold performance.
- `External_Participant_Predictions.csv`: stored participant-level external prediction records.
- `Discovery_Sample_Provenance.csv`: discovery sample and participant traceability.
- `PTGDR_Probe_Sensitivity.csv`: key-gene probe robustness analysis.
- `Supplementary_Methods.pdf`: extended methods and supporting methodological documentation.

References

[1] Davenport EE, Burnham KL, Radhakrishnan J, Humburg P, Hutton P, Mills TC, et al. Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study. The Lancet Respiratory Medicine. 2016;4(4):259–271. https://doi.org/10.1016/S2213-2600(16)00046-1.

[2] Scicluna BP, van Vught LA, Zwinderman AH, Wiewel MA, Davenport EE, Burnham KL, et al. Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study. The Lancet Respiratory Medicine. 2017;5(10):816–826. https://doi.org/10.1016/S2213-2600(17)30294-1.

[3] Balch JA, et al. Defining critical illness using immunological endotypes in patients with and without sepsis: a cohort study. Critical Care. 2023;27(1):292. https://doi.org/10.1186/s13054-023-04571-x.

[4] Scicluna BP, Cano-Gamez K, Burnham KL, Davenport EE, Moore AR, Khan S, et al. A consensus blood transcriptomic framework for sepsis. Nature Medicine. 2025;31(12):4119–4130. https://doi.org/10.1038/s41591-025-03964-5.

[5] Sweeney TE, Shidham A, Wong HR, Khatri P. A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. Science Translational Medicine. 2015;7(287):287ra71. https://doi.org/10.1126/scitranslmed.aaa5993.

[6] Sweeney TE, Khatri P. Benchmarking sepsis gene expression diagnostics using public data. Critical Care Medicine. 2017;45(1):1–10. https://doi.org/10.1097/CCM.0000000000002021.

[7] Maslove DM, Shapira T, Tyryshkin K, Veldhoen RA, Marshall JC, Muscedere J. Validation of diagnostic gene sets to identify critically ill patients with sepsis. Journal of Critical Care. 2019;49:92–98. https://doi.org/10.1016/j.jcrc.2018.10.028.

[8] Walsh CJ, Hu P, Batt J, Santos CCD. Microarray Meta-Analysis and Cross-Platform Normalization: Integrative Genomics for Robust Biomarker Discovery. Microarrays. 2016;5(4):33. https://doi.org/10.3390/microarrays5040033.

[9] Perlis RH. Translating biomarkers to clinical practice. Molecular Psychiatry. 2011;16(11):1076–1087. https://doi.org/10.1038/mp.2011.63.

[10] Thiele M, Villesen IF, Niu L, Johansen S, Sulek K, Nishijima S, et al. Opportunities and barriers in omics-based biomarker discovery for steatotic liver diseases.

Journal of Hepatology. 2024;81(2):345–359. https://doi.org/10.1016/j.jhep.2024.03.035.

[11] Vijay R, Fehr AR, Janowski AM, Athmer J, Wheeler DL, Grunewald M, et al. Virus-induced inflammasome activation is suppressed by prostaglandin D2/DP1 signaling. Proceedings of the National Academy of Sciences of the United States of America. 2017;114(36):E7447–E7456. https://doi.org/10.1073/pnas.1704099114.

[12] Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Research. 2002;30(1):207–210. https://doi.org/10.1093/nar/30.1.207.

[13] Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Research. 2013;41(D1):D991–D995. https://doi.org/10.1093/nar/gks1193.

[14] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Annals of Internal Medicine. 2015;162(1):55–63. https://doi.org/10.7326/M14-0697.

[15] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B. 1995;57(1):289–300.

[16] Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition. 1997;30(7):1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2.

[17] Zou F, Li Y, Zhang S, Zhang J. DP1 (Prostaglandin D2 Receptor 1) Activation Protects Against Vascular Remodeling and Vascular Smooth Muscle Cell Transition to Myofibroblasts in Angiotensin II-Induced Hypertension in Mice. Hypertension. 2022;79(6):1203–1215. https://doi.org/10.1161/HYPERTENSIONAHA.121.17584.

[18] Hong Y, Chen Q, Xie H, Ma M, Gan S, Zhang Y, et al. Multi-Omics Analysis Identifies Immune Regulatory Networks in Sepsis-Associated Liver Injury: Experimental Validation and Clinical Relevance. Journal of Inflammation Research. 2025;18:10711–10722. https://doi.org/10.2147/JIR.S515615.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- MetaAnalysisInferenceSensitivity.csv
- ExternalCalibrationandThresholds.csv
- PTGDRKeyGeneMetaAnalysis.csv
- FoldSpecificModelPerformance.csv
- PTGDRProbeSensitivity.csv
- ExternalPerformanceVerification.csv
- DiscoverySampleProvenance.csv
- ExternalParticipantPredictions.csv
- SupplementaryMethods.pdf