Supplementary Material: Consistent PTGDR downregulation has limited clinical specificity for sepsis across transcriptomic cohorts

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

[1]Department of Health Informatics, 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]Parul Institute of Allied and Healthcare Sciences, Faculty of Medicine, Parul University, Vadodara, Gujarat, India.

Supplementary Methods

Historical-method reconciliation

All distinct historical method families were compared before this revision: the archived April/May submission branch, the July baseline and revision-v2/v3 branches, the July failure-boundary rebuild and technical appendix, the August 14 corrected-manuscript plan, and the August 21 targeted analysis. Repeated and ZIP-packaged copies were identified by content comparison. Method text was carried forward only when supported by an executed checkpoint, sample manifest, result table, or stored scoring record. The earlier sample-size-scaled uncertainty calculation, cohort-refitted external evaluation, and exhaustive fold-specific search were executed in isolation on 24 August 2026 before classification. They remained excluded because the uncertainty scaling is invalid, target-label refitting is not external validation, and the exhaustive multigene choices were fold dependent. Planned RMA wording and the cross-disease panel derivation were likewise classified as superseded rather than silently omitted. The complete reconciliation is supplied in the supplementary provenance records.

Primary-sample selection

GEO metadata were reconciled to CEL identifiers in a predefined sample manifest. The primary GSE57065 analysis selected the H00 observation for each sepsis participant and one observation for each healthy control. The primary GSE95233 analysis selected D01 sepsis observations and independent D00 healthy controls. Alignment checks confirmed 53 and 73 unique participants, respectively. Longitudinal observations were excluded before the primary design matrix was fitted. Ambiguous samples were not assigned fallback phenotypes.

Executed preprocessing and annotation

The primary analysis applied MAS5 summarization followed by log2 transformation to 231 raw CEL files. This corrects planned-RMA wording and removes fallback-path ambiguity. Probe-to-symbol mapping used `hgu133plus2.db` 3.2.3 with the first returned symbol; unresolved probes were removed and the highest-mean probe was retained per gene without phenotype information. Both gene-level matrices contained 20,962 features. The GSE57065 S100A9 annotation record was regenerated from the authoritative collapse table: the correct selected probe is 203535_at, not the erroneous 217871_s_at record (which maps to MIF). The fitted matrix and reported S100A9 result had already used 203535_at.

Table 1 Executed preprocessing and primary-sample verification

Cohort	CEL samples/features	Primary cases/controls	Executed method
GSE57065	107/20,962	28/25	MAS5 then log2
GSE95233	124/20,962	51/22	MAS5 then log2

Alternative-probe sensitivity

Every mapped probe for the four historical genes was fitted under the same primary contrast. PTGDR direction was unchanged by the representative-probe rule: GSE57065 effects were −2.103 (215894_at; selected), −1.395 (234165_at), and −1.487 (215937_at); GSE95233 effects were −1.926 (215937_at; selected), −1.840 (215894_at), and −1.664 (234165_at). VPS8 showed materially greater probe dependence. Complete key-gene results are supplied in Additional file 7.

Meta-analytic inference

For gene g in cohort k, the study effect was the moderated linear-model coefficient $\hat{\beta}_{gk}$. Its sampling standard error was obtained directly from the fitted `limma` object as

$$SE(\hat{\beta}_{gk}) = \sqrt{s^2_{gk,\text{post}}}\, u_{gk},$$

where u_{gk} is the unscaled coefficient standard deviation. Effects were combined by REML random-effects meta-analysis. Ordinary Hartung–Knapp inference was retained because only two cohorts contributed to the primary meta-analysis. Its scale was $q_{\mathrm{HK}} = (k-1)^{-1}\sum_i w_i(y_i - \hat{\mu})^2$. A final source-first audit additionally applied the modified Knapp–Hartung safeguard $q^*_{\mathrm{HK}} = \max\{1, q_{\mathrm{HK}}\}$ when the ordinary scale was below one. This was reported as a conservative sensitivity rather than substituted retrospectively into the historical adjudication rule. No sample-size-derived multiplier was applied to the standard error.

Before the focused predictive comparison, the four historical genes were adjudicated under the corrected-analysis rule: direction agreement, a pooled interval excluding zero, nominal ordinary Hartung–Knapp $P < 0.05$, and estimated $I^2 \leq 50\%$. This adjudication evaluated a previously defined historical candidate set and was not a genome-wide de novo PTGDR discovery test. A Holm family-wise sensitivity gave adjusted $P = 0.100$ for PTGDR, so no gene met adjusted $P < 0.05$. Its ordinary Hartung–Knapp interval was -2.957 to -0.987 ($P = 0.025$), whereas its modified Knapp–Hartung interval was -4.264 to 0.320 ($P = 0.058$). PTGDR was therefore treated only as an exploratory benchmark. The I^2 threshold was an operational criterion, not evidence of precisely known homogeneity; with two cohorts, I^2 and both small-sample procedures remain imprecise. Complete conventional, ordinary Hartung–Knapp, and modified Knapp–Hartung estimates are supplied in Additional file 15.

Cross-cohort development benchmark and information flow

Each reciprocal fold estimated gene centring/scaling parameters and a binomial model in its training dataset and applied them unchanged to the other dataset. Standard generalized-linear-model fitting was attempted first; an L2 penalty of 10^{-8} was the numerical fallback. Primary performance was evaluated separately within each held-out cohort because separately fitted fold models can yield different probability scales.

The modelling mechanics were outcome-isolated within each fold, but feature priority was not: PTGDR had been prioritized by the corrected meta-analysis of both GSE57065 and GSE95233. The two model forms were historical/predefined, yet the designation of PTGDR as the exploratory benchmark was informed by both datasets. We therefore describe these estimates as cross-cohort development performance, not independent validation.

Within each held-out cohort, fixed predictions were resampled 5,000 times within outcome strata. PTGDR-minus-panel AUC differences were 0.356 (95% CI 0.256–0.441) in GSE57065 and 0.453 (0.407–0.488) in GSE95233. These intervals condition on the fixed prediction rows and do not propagate cohort selection, candidate prioritization, or model redevelopment uncertainty. Concatenated prediction metrics remain available for provenance but are scale-sensitive secondary descriptions.

External prediction-record verification

The external PTGDR object was fitted to the pooled primary samples from both development cohorts and fixed before external scoring. Discovery-derived means and

sample standard deviations, the model intercept and coefficient, score orientation, and the 0.5 probability threshold were fixed; external refitting, recalibration, and threshold optimization were prohibited. For each external contrast, labels and PTGDR probabilities were read from stored prediction-level rows. ROC AUC, average precision, and Brier score were recomputed using scikit-learn. External ROC-AUC 95% intervals were regenerated using 5,000 outcome-stratified participant bootstrap replicates per contrast, percentile limits, and seed 20260824. All included external prediction rows represented unique participants; the participant-level implementation nevertheless preserves the intended resampling unit. These intervals are conditional on the stored prediction rows and not on raw-data reconstruction or model redevelopment. Nine of nine AUC and Brier values agreed with the stored evidence table to floating-point precision. GSE65682 average precision differed because the earlier value used a trapezoidal precision–recall area, whereas verification used the tie-aware step-wise definition; the latter value (0.959) is reported. The present verification does not claim raw-data reconstruction of all external cohorts.

Calibration was assessed by regressing outcome on the logit of the stored probability; no recalibration was applied. Scaled Brier score was $1 - Brier/[\pi(1-\pi)]$, where π is case prevalence. Complete threshold results showed specificity 0 in GSE65682, GSE134347, and GSE137340 and 0.032 in GSE74224. The full machine-readable table, including all GSE154918 contrasts and historical-panel comparisons, is Additional file 8; prediction rows are Additional file 9.

PTGDR expression extension specified before outcome inspection

The baseline analysis was finalized as version `v1.0.8-public-release`. On 31 August 2026, before inspecting PTGDR in the newly selected cohorts, a separate protocol defined the lower-PTGDR hypothesis, eligibility rules, participant-level resampling, and prohibition of candidate reselection, multigene modelling, threshold optimization, and result-dependent cohort exclusion. During final metadata reconciliation, GSE222393 was recognized as an eligible ICU comparator and GSE185263 was found to contain 44 healthy controls rather than no comparator. Each correction was documented in a dated protocol addendum before its relevant contrast was computed; the timing and the linked-cohort overlap were retained in the provenance record.

PTGDR was quantified as $\log_2(\mathrm{CPM} + 0.5)$ from processed gene-level RNA-seq counts or as $\log_2(\mathrm{RPM} + 0.5)$ for GSE63042. Welch mean-difference intervals, Hedges' g, two-sided Welch and Mann–Whitney tests, and AUC with 5,000 group-stratified bootstrap replicates were reported. Integer-count cohorts additionally used an NB2 negative-binomial regression of PTGDR count on case status with log library size as offset. For the linked GSE222393/GSE185263 analysis, participant identifiers and day suffixes were reconciled and the earliest available observation was retained, yielding 22 non-COVID sepsis and 20 severe COVID-19 ICU participants.

The GSE63042 composition model used all seven marker scores defined before PTGDR inspection. For each component, available predefined genes were log transformed, standardized across samples, and averaged; all scores entered one HC3 linear model. GSE167363 single-cell processing retained 200–6,000 detected genes per cell

and less than 20% mitochondrial UMIs, used baseline/T0 samples only, normalized each cell to 10,000 UMIs with log1p transformation for outcome-blind marker scoring, and summarized PTGDR counts by participant and assigned cell type. Exact participant-level tests were used; cells were not treated as independent replicates.

A second protocol was finalized on 1 September 2026 before computing its outcomes. The three-cohort expression synthesis used the standard large-sample Hedges' g sampling variance, REML random effects, conventional and Hartung–Knapp intervals, Cochran's Q, I^2, tau-squared, a t_{k-2} prediction interval, and leave-one-cohort-out sensitivity. It included only the adult sepsis-versus-healthy contrasts from GSE185263, GSE243217, and GSE279448. HC3 OLS and NB2 models then added available age and sex and predefined common-location/site restrictions without outcome-based covariate selection.

GSE279452 expression records excluded matched V(D)J libraries. Per donor, the defined single-cell QC and outcome-blind marker-score assignment were applied; PTGDR was absent from the annotation markers. PTGDR was pseudobulk summarized within donor and cell type, and all inference was donor-level. Adult/pediatric and healthy/ICU families remained separate, with Benjamini–Hochberg adjustment across eight tested compartments within family. Expression triplets were keyed by GSM accession. The identity file retains six capitalization discrepancies and a two-accession 'Abd-S8'/'Abd-S80' exchange between GEO titles and archive donor suffixes. Bulk GSE279448 was matched to the archive suffix. Seven broad measured compartments used a 0.5-cell replacement and six orthonormal isometric-log-ratio coordinates; crude, age/sex, and age/sex-plus-composition HC3 models were required to be full rank and to pass the predefined donor/control/representation gate.

Supplementary Results

The corresponding negative-binomial estimates agreed with the contrast pattern. The log2 fold changes were -1.483 (95% CI -1.870 to -1.097; $P = 5.65 \times 10^{-14}$) in GSE185263, -0.403 (-1.206 to 0.400; $P = 0.325$) in the linked ICU contrast, -1.567 (-1.876 to -1.259; $P = 2.31 \times 10^{-23}$) for GSE279448 adult sepsis versus healthy, and -0.102 (-0.529 to 0.324; $P = 0.638$) for adult sepsis versus ICU control. The pediatric ICU and healthy estimates were 0.342 ($P = 0.427$) and -0.978 ($P = 0.0400$), respectively.

GSE167363 erythroid cells were classified as non-informative because only four sepsis and one healthy participant contributed assigned cells. The GSE63042 unadjusted sepsis coefficient was -0.307 (HC3 95% CI -0.806 to 0.192; $P = 0.228$); after all seven predefined composition scores it was -0.229 (-0.661 to 0.203; $P = 0.299$). Full sample-level values, intervals, marker availability, QC counts, and negative results are retained in the validation-extension machine-readable files.

The three-cohort random-effects synthesis had material heterogeneity ($I^2 = 67.4\%$) and a wide prediction interval crossing zero (-7.315 to 3.570), so a consistent effect in a future comparable cohort cannot be assumed. Its average pooled Hedges' g was -1.872 (conventional 95% CI -2.363 to -1.382; Hartung–Knapp CI -2.958 to -0.787; $P = 0.0177$), with tau-squared 0.121. Because standardized effects were combined

Table 2 Complete PTGDR bulk RNA-seq expression-extension contrasts

Dataset	Contrast	Cases/controls	Difference	AUC (95% CI)
GSE63042	Sepsis vs noninfectious SIRS	106/23	−0.307	0.582 (0.447–0.712)
GSE243217	Sepsis vs healthy	22/15	−2.248	0.970 (0.900–1.000)
GSE243217	Sepsis vs COVID-19	22/35	−1.044	0.745 (0.614–0.862)
GSE279448	Adult sepsis vs adult ICU control	116/23	−0.202	0.582 (0.458–0.704)
GSE279448	Adult sepsis vs adult healthy	116/43	−1.739	0.933 (0.894–0.965)
GSE279448	Pediatric sepsis vs pediatric ICU control	16/5	0.121	0.413 (0.150–0.663)
GSE279448	Pediatric sepsis vs pediatric healthy	16/4	−1.081	0.844 (0.594–1.000)
GSE185263	Sepsis vs healthy	348/44	−1.795	0.901 (0.863–0.935)
GSE222393/185263	Non-COVID sepsis vs severe COVID-19 ICU	22/20	−0.513	0.589 (0.405–0.755)

across whole-blood and PBMC datasets and platforms, this is a cross-study summary of disease-versus-health contrast magnitude rather than evidence of a common assay-scale effect. Age/sex and common-location/site models retained the disease-versus-health pattern, but adult sepsis-versus-ICU and the linked severe-COVID ICU contrast remained uncertain or model-sensitive.

GSE279452 retained 941,852 of 980,095 cells from all 133 expression donors; all 35 predefined genes were available in every donor. Adult results are summarized below. Pediatric sepsis-versus-healthy directions were lower in seven of eight compartments but based on four controls; pediatric sepsis-versus-ICU directions were mixed and no within-family adjusted result was positive. All four pediatric ICU-control dendritic pseudobulks were at the PTGDR zero-count floor, so the associated Welch calculation generated a near-identical-values precision warning and was not interpreted as positive evidence.

The matched-composition gate passed with 54 adult sepsis/healthy donors and 53 sepsis/ICU donors. All designs were full rank. For sepsis versus healthy, crude, age/sex, and age/sex-plus-composition coefficients were −1.700, −1.707, and −1.172 (fully adjusted 95% CI −1.754 to −0.590; $P = 7.95 \times 10^{-5}$). For sepsis versus ICU, they were −0.242, −0.309, and −0.217 (fully adjusted CI −0.817 to 0.384; $P = 0.479$). These annotation-dependent proportions are neither clinical differential counts nor causal mediation estimates.

Table 3 Sorted-cell and single-cell PTGDR localization summary

Dataset	Cell type	Contrast	Cases/controls	Difference	AUC	Adjusted q
GSE133822	CD4	Early sepsis vs critically ill	13/21	−0.659	0.645	0.198
GSE133822	CD4	Early sepsis vs healthy	13/19	−0.864	0.721	0.0920
GSE133822	CD8	Early sepsis vs critically ill	14/21	0.038	0.507	0.875
GSE133822	CD8	Early sepsis vs healthy	14/19	−0.124	0.575	0.672
GSE133822	CD14	Early sepsis vs critically ill	14/21	−0.471	0.609	0.247
GSE133822	CD14	Early sepsis vs healthy	14/20	−1.319	0.821	0.00160
GSE136200	CD14	Early sepsis vs healthy	6/4	−0.685	0.667	0.749
GSE136200	CD19	Early sepsis vs healthy	6/4	−1.060	0.771	0.749
GSE136200	CD4	Early sepsis vs healthy	6/4	−0.405	0.667	0.749
GSE136200	CD8	Early sepsis vs healthy	6/4	0.035	0.417	0.921
GSE167363	Monocyte	Sepsis vs healthy	5/2	−0.967	1.000	0.476
GSE167363	T cell	Sepsis vs healthy	5/2	−0.662	0.800	0.635
GSE167363	NK cell	Sepsis vs healthy	5/2	−0.729	0.800	0.635
GSE167363	B cell	Sepsis vs healthy	5/2	0.162	0.600	0.857
GSE167363	Platelet	Sepsis vs healthy	5/2	−1.860	0.600	0.857

Runtime qualification

The primary pipeline began in an empty output directory, read all 231 CEL files, executed MAS5/log2 processing, regenerated 20,962-gene matrices and primary contrasts, reran the two-cohort meta-analysis, and regenerated the focused model benchmark. A 27 October clean-host audit isolated the two large cohorts in separate single-threaded R processes to prevent retained native memory in the legacy R 3.6.1/Bioconductor

Table 4 GSE279452 adult donor-level PTGDR localization

	Sepsis vs healthy (93/7)			Sepsis vs ICU control (93/8)		
Cell type	Difference	AUC	Welch q	Difference	AUC	Welch q
Classical monocyte	−1.136	0.831	0.00365	−0.533	0.630	0.275
Nonclassical monocyte	−1.169	0.702	0.0247	−0.382	0.586	0.561
CD4 T cell	−1.074	0.848	0.00365	−0.712	0.718	0.132
CD8 T cell	−0.703	0.717	0.0477	−0.435	0.603	0.132
B cell	−0.924	0.741	0.130	−0.387	0.599	0.489
NK cell	−0.427	0.724	0.0171	−0.299	0.652	0.177
Dendritic	−1.461	0.805	0.0477	−1.058	0.648	0.274
Platelet	−0.875	0.720	0.0171	−0.611	0.673	0.0951

Table 5 Fold-specific cross-cohort development performance

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

stack from exhausting the host. GSE57065 completed 107 CEL files with 6.9 GB peak resident memory and GSE95233 completed 124 with 7.9 GB; neither process swapped. The fresh run matched 13 of 13 whole-primary output files within 10^{-10} (maximum absolute difference 1.39×10^{-11}). Subsequent re-execution regenerated the bootstrap and external prediction-record tables and matched all 167 primary/secondary numerical checks (maximum absolute difference 7.28×10^{-12}), all 196 longitudinal checks, and all four deterministic metric tests. The expanded expression extension separately passed 146 numerical, sample-count, identity, negative-result, QC, and manifest checks plus three synthetic unit tests and processed all 133 GSE279452 donors. The previously crashing GSE95233 repeated-measures analysis was rebuilt using within-patient differences, generating 146,734 longitudinal rows across both cohorts. Historical imbalance and pathway analyses were also repaired and executed, but remain outside the primary evidence because their panel premise was rejected. Null, weak, and direction-reversed clinical contrasts were retained as scientific findings rather than labelled as computational failures. Input SHA-256 hashes and separate R/Python environment records accompany the analysis code. Platform independence is not claimed.

Machine-readable companion files

Additional files 1–15 contain the corrected meta-analysis, internal predictions and fold-specific uncertainty, external metrics and prediction rows, sample-level provenance,

Table 6 External calibration and fixed-threshold failures

Dataset	Contrast	AUC	Scaled Brier	Intercept	Slope	Sens.	Spec.
GSE74224	Sepsis vs post-surgical inflammation	0.790	−0.320	−2.219	0.674	1.000	0.032
GSE65682	Sepsis-related vs healthy	0.616	−0.055	−41.016	3.190	1.000	0.000
GSE134347	Sepsis-related vs healthy	0.703	−0.372	−5.471	0.792	1.000	0.000
GSE137340	Sepsis-related vs healthy	0.688	−0.279	−3.113	0.433	1.000	0.000
GSE154918	Healthy vs infection	0.873	0.276	−0.612	1.386	0.667	0.900
GSE154918	Healthy vs sepsis	0.860	0.386	−0.268	1.258	0.650	0.900
GSE154918	Infection vs sepsis	0.554	−0.214	0.416	0.140	0.650	0.333
GSE154918	Sepsis vs shock	0.747	−0.086	−0.992	0.628	0.947	0.350
GSE154918	Infection vs shock	0.816	0.191	−0.736	0.901	0.947	0.333

Table 7 Historical-panel external comparison where evaluable

Dataset	Contrast	Panel AUC	Panel Brier	PTGDR AUC	PTGDR Brier	Panel spec.
GSE74224	Sepsis vs post-surgical inflammation	0.606	0.248	0.790	0.275	0.065
GSE137340	Sepsis-related vs healthy	0.374	0.218	0.688	0.218	0.000
GSE154918	Healthy vs infection	1.000	0.630	0.873	0.128	0.000
GSE154918	Healthy vs sepsis	0.986	0.546	0.860	0.136	0.000
GSE154918	Infection vs sepsis	0.600	0.374	0.554	0.285	0.000
GSE154918	Sepsis vs shock	0.742	0.510	0.747	0.271	0.000
GSE154918	Infection vs shock	0.860	0.386	0.816	0.192	0.000

probe and longitudinal sensitivities, a file manifest, the baseline computational verification bundle, the PTGDR validation-extension protocols, code, verification and results, and the complete ordinary/modified meta-inference sensitivity. The version-controlled analysis repository is hosted at https://github.com/Nikhilesh-Dangi/CBF_PTGDRCBF. It is private during peer review; access will be provided to editors and reviewers by the corresponding author on request. The repository will be made public and archived with a persistent DOI before publication.

Table 8 Disposition of historical analysis branches

Branch	Disposition	Reason
Historical three-gene panel	Rejected	Two genes failed corrected stability criteria; cross-cohort development performance was poor
Cross-disease pathway display	Excluded from primary paper	Panel premise did not survive corrected discovery
Historical whole-blood sensitivity	Excluded	Reused defective uncertainty logic and contained a manifest inconsistency
Fold-specific exploratory feature search	Executed; excluded from primary inference	Completed after restoration of required inputs; multigene selections were fold dependent, with PTGDR as the only stable intersection
GSE57065 longitudinal contrasts	Descriptive sensitivity only	Repeated samples answer a different temporal question
GSE95233 longitudinal contrasts	Repaired descriptive sensitivity	Explicit within-patient differences replaced the crashing legacy correlation implementation