

Cross-Platform Evaluation of a Locked Two-Pathway Transcriptomic Score for Early Mortality in Diffuse Large B-Cell Lymphoma

Xichang Fei

Zhabei Central Hospital, Jing'an District

Haimin Chen

Zhabei Central Hospital, Jing'an District

Ying Zhang

Zhabei Central Hospital, Jing'an District

Fan Zhou

zhoufanzzx@163.com

Zhabei Central Hospital, Jing'an District

Research Article

Keywords: diffuse large B-cell lymphoma, early mortality, transcriptomics, pathway score, immune depletion, external validation

Posted Date: October 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11038697/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Early mortality remains clinically important in diffuse large B-cell lymphoma (DLBCL), but transcriptomic associations are vulnerable to platform effects and overfitting. We evaluated a locked 2-pathway score across prespecified cohorts.

Patients and Methods

GSE181063 was the development cohort, GSE10846 the prior cross-platform validation cohort, and GSE31312 the confirmatory cohort. Within-sample rank scores represented MYC activity and depletion of a 195-gene immune program. Genes, directions, weights, cutoff, endpoint, and covariates were frozen before confirmatory testing. Cox models estimated 0–2-year all-cause mortality with adjustment for age, sex, and cell of origin.

Results

The primary sets included 1314 patients and 300 deaths within 2 years. In GSE31312, the equal-weight composite was associated with 0–2-year mortality (hazard ratio [HR] per standard deviation, 1.29; 95% confidence interval [CI], 1.07–1.56; $P = .008$). The immune-depletion axis showed a stronger association (HR, 1.38; 95% CI, 1.15–1.66; $P < .001$), whereas the MYC axis was not associated. Composite results were stable in quality-control sensitivity analyses but attenuated after International Prognostic Index adjustment (HR, 1.15; 95% CI, 0.93–1.42; $P = .193$). The development cutoff transported directionally but was borderline in GSE31312, and the fixed model was poorly calibrated. Descriptive pooling supported immune depletion (HR, 1.39; 95% CI, 1.08–1.79; $I^2 = 0\%$).

Conclusion

A locked 2-pathway score retained an association with 0–2-year mortality in an additional cohort, primarily through a reproducible immune-depletion signal. It is not yet a calibrated clinical predictor or validated treatment threshold.

Introduction

Diffuse large B-cell lymphoma (DLBCL) is curable for many patients treated with rituximab-based immunochemotherapy, yet early treatment failure and death remain clinically consequential. Event-free survival status at 24 months (EFS24) is a validated landmark for disease-related outcome in DLBCL [1]. The International Prognostic Index (IPI) remains the dominant clinical risk framework [2]. The endpoint in the present study was different: all-cause mortality within 2 years, defined from recorded overall survival rather than EFS24. Clinical risk indices incompletely capture the molecular diversity that shapes treatment response.

Gene expression profiling established cell-of-origin (COO) classes and revealed prognostically relevant tumor and stromal programs [3–5]. Subsequent genomic taxonomies further resolved pathogenic diversity [6, 7], and microenvironmental studies have linked immune and stromal states to biologically and clinically distinct lymphoma phenotypes [8]. Translating these observations into a reproducible prognostic signal is difficult because high-dimensional analyses are sensitive to cohort composition, assay platform, preprocessing, endpoint choice, and post hoc score optimization.

We therefore used a staged, locked analysis in which discovery, prior validation, and additional confirmation had distinct roles. A 2-pathway adverse score was defined in GSE181063, assessed without redefinition in GSE10846, and

then tested under a frozen protocol in GSE31312. The primary question was whether the locked equal-weight composite was associated with 0–2-year all-cause mortality after adjustment for age, sex, and COO. Secondary analyses examined component consistency, unchanged cutoff transport, incremental model behavior, calibration, and an explicitly exploratory decomposition of the locked immune program. The design and reporting boundaries emphasized reproducibility and transparent limitations consistent with biomarker-reporting principles [9].

Patients and Methods

Study design and cohort hierarchy

We performed a secondary analysis of publicly available gene-expression and clinical data from the NCBI Gene Expression Omnibus (GEO) [10]. The hierarchy was fixed before confirmatory testing: GSE181063 was the discovery and score-development cohort; GSE10846 was the prior cross-platform outcome-validation cohort; and GSE31312 was an additional locked-score confirmatory cohort. GSE173263 contrasted primary refractory disease or relapse within 12 months with remission in 39 samples and was used only as related-phenotype support; its endpoint was not equivalent to 0–2-year all-cause mortality, and it was not pooled with the 3 overall-survival cohorts.

The GSE181063 material derives from the United Kingdom Haematological Malignancy Research Network setting [11, 12]. GSE10846 is the Lenz et al R-CHOP-era cohort [5], and GSE31312 is the International DLBCL Rituximab-CHOP Consortium cohort reported by Visco et al [13]. GSE173263 was reported by Dlouhy et al [14]. GEO accession identifiers were audited across datasets. No accession overlap was detected between GSE31312 and the earlier outcome cohorts, but patient-level independence could not be proven from public metadata alone. Analyses were performed separately by cohort; expression matrices were never merged.

Endpoints, Eligibility, and Locked Analysis Sets

The primary endpoint was all-cause mortality within 2 years from the recorded overall-survival origin. For time-to-event models, follow-up was administratively truncated at 2 years: deaths by 2 years were events, and all other observations were censored at the earlier of last follow-up and 2 years. Patients censored alive before 2 years remained in the Cox risk sets. For the binary discovery contrast, early death was death by 2 years, the comparator group required observation through 2 years, and patients censored before 2 years were unclassified.

Cohort-specific eligibility rules retained adults with DLBCL, rituximab-based curative intent therapy when documented, pretreatment tumor expression, and usable survival time and status. Outcome blinded quality control (QC) defined primary, all-sample, and strict-no-flag sets before pathway scoring. Primary early-window models included 615 patients in GSE181063, 232 in GSE10846, and 467 in GSE31312. The primary adjustment set was age, sex, and COO. IPI plus COO complete-case models were prespecified sensitivity analyses because IPI was incompletely observed.

Expression Processing, Annotation, and Outcome-Blinded Quality Control

Depositor-supplied processed expression scales were retained: Lumi-normalized values for GPL14951 in GSE181063, MAS 5.0 globally scaled values for GPL570 in GSE10846, depositor-supplied DQN3 signals for GPL570 in GSE31312, and robust multiarray-average values for GPL13667 in GSE173263. No additional log transformation or between-array normalization was applied. Platform annotation packages mapped probes to gene symbols; unmapped or

ambiguous probes were excluded, and multiple probes mapping to 1 symbol were averaged with limma [15]. Final gene matrices contained 19,418 genes for GSE181063, 20,823 for each GPL570 cohort, and 18,488 for GSE173263.

QC used sample median and interquartile range, missingness, median intersample correlation, and robust principal component distance. Robust z scores greater than 5 flagged individual metrics; multimetric or extreme findings received high priority review. Overall survival variables were not used in QC decisions. COO could color principal component plots but did not determine exclusion. The two high priority GSE181063 samples and three high priority GSE31312 samples were excluded from the respective primary sets, while all-samples were restored in mandatory sensitivity analyses.

Pathway Discovery and Locked-Score Construction

In GSE181063, gene-level moderated models adjusted the binary early-death contrast for age, sex, and COO. Hallmark gene sets were obtained from the Molecular Signatures Database [16]. Competitive gene-set testing used CAMERA with an intergene correlation of 0.01 [17]. The Benjamini-Hochberg false-discovery rate (FDR) controlled multiplicity across 50 Hallmark sets [18]. Discovery-significant pathways were evaluated in GSE10846 with multiplicity correction restricted to the discovery set and a required matching effect direction. GSE173263 was examined as related-phenotype context only.

Two pathways survived the discovery and prior validation hierarchy: HALLMARK_MYC_TARGETS_V2, higher in early death, and HALLMARK_ALLOGRAFT_REJECTION, lower in early death. The latter name is an MSigDB annotation label and was interpreted as an immune expression program, not as transplant rejection. Scores were calculated from within-sample percentile ranks over the locked 16,502 gene universe shared by GSE181063 and GSE10846. The MYC adverse axis was the mean rank of 51 MYC target genes; the immune depletion axis was the negative mean rank of the 195 gene immune program. The raw equal-weight composite was one half of the sum of these two adverse-oriented axes. Continuous effects were expressed per within-cohort standard deviation. No GSE31312 outcome was used to select genes, directions, weights, cutoff, time point, endpoint, or covariates.

Early-Window Models, Validation, and Model Assessment

Cox proportional hazards models estimated hazard ratios (HRs) and 95% confidence intervals (CIs) [19]. Primary models included 1 standardized score and age, sex, and COO. Schoenfeld residual tests assessed proportional hazards [20]. A prespecified piecewise model in GSE181063 compared effects from 0 to 2 years with effects after a 2-year landmark among survivors, using patient-clustered robust standard errors.

The single confirmatory test was the equal-weight composite in the GSE31312 primary QC set, with a 2-sided significance threshold of .05 and an adverse HR greater than 1. Prespecified sensitivities reintroduced all samples, excluded every QC-flagged sample, and substituted IPI plus COO complete-case adjustment. Model fit was compared with likelihood-ratio tests. Harrell concordance was reported for clinical and expanded models, with paired differences treated as descriptive. The development cutoff was the median raw composite in the GSE181063 primary set (0.0122577831956875) and was transferred unchanged. Fixed GSE181063 coefficients were also applied to GSE31312 without coefficient refitting or recalibration; external discrimination, calibration slope, and agreement between mean predicted and Kaplan-Meier estimated 2-year survival were summarized following external-validation principles [21].

Descriptive Evidence Synthesis and Exploratory Immune Mechanism

For each locked score, the 3 cohort-specific log HRs were summarized descriptively with restricted maximum-likelihood random-effects models and modified Knapp-Hartung inference [22, 23]. Because the synthesis contained only 3 cohorts with different inferential roles, pooled estimates did not replace the confirmatory GSE31312 test. Heterogeneity was described with I^2 .

Mechanistic follow-up was limited to five prespecified, overlapping Hallmark annotations within the unchanged 195 gene immune program: interferon gamma response, inflammatory response, TNFA/NFkB signaling, IL2/STAT5 signaling, and complement. Each depletion-oriented module score was tested in the three primary early-window models and synthesized descriptively. Benjamini-Hochberg adjustment covered the five module summaries. Gene level moderated effects were also triangulated across cohorts. An exploratory core required lower expression in early death in all three cohorts and nominal $P < .05$ in at least two; it was interpretive only and did not rebuild the score.

Statistical Analysis and Integrity Controls

All tests were 2-sided. Except for the single locked confirmatory test, analyses were explicitly labeled discovery, validation, sensitivity, secondary, descriptive, or exploratory. Missing covariates were handled by complete-case analysis for the relevant model; no outcome or covariate imputation was performed. Each analytical stage wrote analysis-set counts and technical gates before the next stage. At manuscript finalization, 259 reviewed source files were fingerprinted by path, byte size, and MD5, and the repeated freeze matched the prior manifest. Publication assets were generated only from the frozen summary results; no model was fitted or updated.

Results

Cohort Accounting and Quality Control

The 3 primary outcome cohorts comprised 1,314 patients and 300 deaths within 2 years (Table 1). GSE181063 contributed 615 patients and 132 early deaths; GSE10846 contributed 232 and 54; and GSE31312 contributed 467 and 114.

Table 1
Cohort design and locked early overall survival analysis sets

Cohort	Prespecified role	Platform	Primary n	Deaths ≤ 2 y	Deaths > 2 y	Observed > 2 y	Censored ≤ 2 y
GSE181063	Discovery/development; pathway definition originated here	GPL14951	615	132	131	481	2
GSE10846	Prior cross-platform validation; assessed before the confirmatory cohort	GPL570	232	54	5	121	57
GSE31312	Additional locked-score confirmation; no detected accession overlap; patient-level independence not proven	GPL570	467	114	55	304	49
<i>Endpoint: all-cause mortality within 2 years. Primary models adjusted for age, sex, and COO. COO, cell of origin</i>							

GSE31312 began with 498 expression samples, of which 470 matched audited clinical records with usable overall survival. Three multimetric QC samples were excluded from its primary set, leaving 467; the all-sample and strict-no-flag sets included 470 and 460, respectively. Outcome blinded QC findings and analysis set audits are provided in Supplementary Tables S1–S3.

Discovery and Locking of the 2-Pathway Score

Sixteen Hallmark pathways met $FDR < .05$ in the GSE181063 discovery analysis. Only MYC_TARGETS_V2 and the 195 gene HALLMARK_ALLOGRAFT_REJECTION program met the prespecified GSE10846 direction and within discovery validation criteria. For MYC_TARGETS_V2, development FDR was 8.43×10^{-11} and validation adjusted P was 7.33×10^{-4} . For the immune program, development FDR was 1.19×10^{-5} and validation adjusted P was 5.86×10^{-7} . Neither received nominal related phenotype support in GSE173263. Global gene-level concordance was modest, emphasizing the rationale for pathway level locking (Supplementary Tables S4–S7).

The Prespecified 2-Year Window Concentrated the Development Signal

In GSE181063, each score was more strongly associated with mortality in the zero to 2-year window than after the 2-year landmark. The early versus late HRs were 1.53 versus 0.96 for MYC (interaction $P = 1.19 \times 10^{-4}$), 1.36 versus 0.91 for immune depletion ($P = .0010$), and 1.56 versus 0.94 for the composite ($P = 1.50 \times 10^{-5}$). These estimates supported the pre-existing 2-year boundary rather than a data-selected time point (Supplementary Fig. S1 and Supplementary Table S17).

Primary Associations Across Cohorts

The immune-depletion axis showed the most consistent association with 2-year mortality: HR 1.36 (95% CI 1.14–1.62; $P < .001$) in GSE181063, 1.50 (1.14–1.97; $P = .004$) in GSE10846, and 1.38 (1.15–1.66; $P < .001$) in GSE31312 (Fig. 1; Table 2). Its descriptive pooled HR was 1.39 (1.08–1.79; $P = .030$; $I^2 = 0\%$). The GSE31312 immune model had a nominal Schoenfeld residual signal ($P = .046$), so a constant HR interpretation for that component warrants caution.

The MYC adverse axis was associated with early mortality in GSE181063 (HR 1.61, 95% CI 1.31–1.97; $P < .001$) but not in GSE10846 or GSE31312. The descriptive pooled estimate was imprecise and heterogeneous (HR 1.25, 95% CI 0.69–2.27; $I^2=77.2\%$). The equal-weight composite was associated in GSE181063 (HR 1.61, 95% CI 1.34–1.94; $P < .001$), directionally similar but borderline in GSE10846 (HR 1.32, 95% CI 0.99–1.78; $P = .061$), and significant in the prespecified GSE31312 confirmatory test (HR 1.29, 95% CI 1.07–1.56; $P = .008$). The composite proportional hazards test in GSE31312 was not significant ($P = .223$). Its descriptive pooled HR was 1.42 (95% CI 1.01–2.00; $P = .048$; $I^2=33.2\%$).

Table 2
Primary adjusted 0–2-year pathway associations

Score	Cohort or summary	Role	n / early deaths	HR (95% CI)	P Value	Diagnostic / Heterogeneity
MYC adverse axis	GSE181063	Discovery/development	615 / 132	1.61 (1.31–1.97)	< .001	PH $P = .807$
MYC adverse axis	GSE10846	Prior validation	232 / 54	1.11 (0.81–1.54)	.513	PH $P = .906$
MYC adverse axis	GSE31312	Prespecified confirmatory validation	467 / 114	1.07 (0.88–1.30)	.514	PH $P = .910$
MYC adverse axis	Descriptive pooled estimate	Descriptive synthesis; not a new confirmatory test	—	1.25 (0.69–2.27)	.244	$I^2 = 77.2\%$
Immune-depletion axis	GSE181063	Discovery/development	615 / 132	1.36 (1.14–1.62)	< .001	PH $P = .556$
Immune-depletion axis	GSE10846	Prior validation	232 / 54	1.50 (1.14–1.97)	.004	PH $P = .484$
Immune-depletion axis	GSE31312	Prespecified confirmatory validation	467 / 114	1.38 (1.15–1.66)	< .001	PH $P = .046$
Immune-depletion axis	Descriptive pooled estimate	Descriptive synthesis; not a new confirmatory test	—	1.39 (1.08–1.79)	.030	$I^2 = 0.0\%$
Equal-weight composite	GSE181063	Discovery/development	615 / 132	1.61 (1.34–1.94)	< .001	PH $P = .936$
Equal-weight composite	GSE10846	Prior validation	232 / 54	1.32 (0.99–1.78)	.061	PH $P = .645$
Equal-weight composite	GSE31312	Prespecified confirmatory validation	467 / 114	1.29 (1.07–1.56)	.008	PH $P = .223$
Equal-weight composite	Descriptive pooled estimate	Descriptive synthesis; not a new confirmatory test	—	1.42 (1.01–2.00)	.048	$I^2 = 33.2\%$

Cohort estimates are adjusted for age, sex, and COO and expressed per 1-SD higher adverse score. Pooled rows are descriptive REML/mKH summaries. PH, proportional hazards; mKH, modified Knapp-Hartung.

Confirmatory Robustness, Incremental Value, and Cutoff Transport

The GSE31312 composite result was stable when all 470 samples were included (HR 1.31, 95% CI 1.09–1.58; $P = .004$) and when all 10 QC-flagged samples were excluded ($n = 460$; HR 1.31, 95% CI 1.08–1.58; $P = .006$) (Fig. 2; Table 3). In 421 IPI complete-cases, however, the composite attenuated after IPI and COO adjustment (HR 1.15, 95% CI 0.93–1.42; $P = .193$), whereas the immune depletion axis retained borderline nominal evidence (HR 1.22, 95% CI 1.00–1.49; $P = .046$). The data therefore do not establish composite independence beyond IPI.

Adding the composite to age, sex, and COO improved likelihood-based fit in GSE31312 (likelihood ratio $P = .008$), but the C-index increase was modest, from 0.580 to 0.610 ($\Delta = 0.030$; paired $P = .147$). The unchanged development cutoff separated higher and lower risk in GSE181063 (HR 1.96; $P = .002$) and GSE10846 (HR 2.07; $P = .009$); in GSE31312 the direction persisted but the result was borderline (HR 1.47, 95% CI 0.99–2.19; $P = .055$). Thus, the continuous association transported more convincingly than the dichotomous threshold.

When all GSE181063 coefficients were transported without refitting, GSE31312 concordance increased descriptively from 0.572 for the clinical model to 0.610 with the score ($\Delta = 0.038$; $P = .081$). Calibration was inadequate: the slope was 0.383 (95% CI 0.168–0.598), mean predicted 2-year survival was 0.836, and Kaplan-Meier estimated survival was 0.745 (95% CI 0.702–0.783). No clinical nomogram or calibrated risk calculator was therefore developed. Full sensitivity, model comparison, cutoff, and calibration outputs are in Supplementary Tables S12–S19.

Table 3
GSE31312 confirmation robustness and transportability

Section	Analysis	n / early deaths	Effect Estimate or Statistic (95% CI)	P Value	Additional value	Interpretation
Adjusted 0–2- year association	Composite: primary QC, age + sex + COO	467 / 114	1.29 (1.07– 1.56)	.008	—	Prespecified confirmatory test
Adjusted 0–2- year association	Composite: all samples, age + sex + COO	470 / 115	1.31 (1.09– 1.58)	.004	—	Prespecified QC sensitivity
Adjusted 0–2- year association	Composite: strict no flags, age + sex + COO	460 / 111	1.31 (1.08– 1.58)	.006	—	Prespecified QC sensitivity
Adjusted 0–2- year association	Composite: IPI complete cases, IPI + COO	421 / 102	1.15 (0.93– 1.42)	.193	—	Association attenuated; independence beyond IPI not confirmed
Adjusted 0–2- year association	Immune axis: primary QC, age + sex + COO	467 / 114	1.38 (1.15– 1.66)	< .001	—	Most consistent component
Adjusted 0–2- year association	Immune axis: IPI complete cases, IPI + COO	421 / 102	1.22 (1.00– 1.49)	.046	—	Nominal association retained
Incremental fit	Composite added to age + sex + COO	467 / 114	Likelihood- ratio χ^2 = 6.95	.008	ΔC = 0.030; paired P = .147	Model fit improved; C-index increase was modest and not statistically established
Fixed cutoff	Unchanged GSE181063 median cutoff: high versus low	467 / 114	1.47 (0.99– 2.19)	.055	342 low / 125 high; cutoff = 0.012257783	Adverse direction reproduced; primary result borderline
Fixed-model transportability	Development coefficients applied without refitting	467 / 114	Calibration slope = 0.38 (0.17–0.60)	—	Clinical C = 0.572; clinical + score C = 0.610; ΔC = 0.038; slope CI excludes 1	External calibration inadequate; nomogram remains on hold

The cutoff was 0.0122577831956875 and was not optimized in GSE31312. IPI complete-case models are sensitivities. External coefficients were applied without refitting or recalibration.

Exploratory Immune Mechanism

All 5 depletion-oriented immune subprograms had HRs greater than 1 in each outcome cohort. Descriptive pooled evidence was strongest for the interferon gamma response overlap (37 genes; HR 1.36, 95% CI 1.05–1.76; P = .037;

BH adjusted $P = .097$; $I^2=7.9\%$) and complement overlap (15 genes; HR 1.34, 95% CI 1.04–1.72; $P = .039$; BH adjusted $P = .097$; $I^2 = 0\%$) (Fig. 3; Table 4). Inflammatory response, IL2/STAT5 signaling, and TNFA/NFKB signaling were directionally concordant but did not meet the exploratory BH $< .10$ criterion.

Gene-level triangulation identified 26 of the 195 locked immune genes with lower expression in early death in all three cohorts and nominal $P < .05$ in at least two (Fig. 4; Supplementary Tables S23–S24). The set included multiple T-cell receptor and immune signaling genes, including CD2, CD3D, CD3G, CD247, LCK, CD8A, TRAT1, CD28, and ITGAL. These patterns support an immune-depleted transcriptional state associated with early mortality, but they do not quantify immune cell fractions, establish causal mechanisms, or justify a replacement score.

Table 4
Exploratory immune mechanism synthesis

Overlapping module	Genes	Pooled HR (95% CI)	P Value	BH-Adjusted P Value	I^2	Evidence
IFN-gamma response overlap	37	1.36 (1.05–1.76)	.037	.097	7.9%	Exploratory BH-adjusted $P < .10$ support
Complement overlap	15	1.34 (1.04–1.72)	.039	.097	0.0%	Exploratory BH-adjusted $P < .10$ support
Inflammatory response overlap	37	1.24 (0.95–1.62)	.074	.108	0.0%	Directional only
IL2/STAT5 signaling overlap	17	1.22 (0.93–1.59)	.086	.108	0.0%	Directional only
TNFA/NFKB signaling overlap	17	1.14 (0.88–1.47)	.171	.171	0.0%	Directional only

Hazard ratios are per 1-SD higher depletion score. Modules overlap and remain part of the unchanged 195-gene program. BH adjustment covers five module summaries.

Discussion

This staged analysis confirmed the prespecified association between a locked equal-weight 2-pathway score and mortality within 2 years in an additional public DLBCL cohort. The confirmatory GSE31312 HR of 1.29 per standard deviation was directionally aligned with both earlier cohorts and remained stable under opposing QC sensitivities. The central biological observation was more specific: depletion of the locked 195 gene immune program was consistently adverse across all three outcome cohorts, whereas the MYC component was strongly associated only in development and showed substantial between-cohort heterogeneity.

The immune finding is compatible with decades of work showing that nonmalignant stromal and immune programs contribute prognostic information beyond tumor-intrinsic cell of origin [5, 8]. The exploratory overlap results implicate interferon gamma response, complement, inflammatory signaling, and lymphocyte-associated genes, but their interpretation must remain at the level of bulk tissue transcription. Lower expression could reflect cellular composition, functional suppression, sampling, tissue handling, or correlated tumor biology. The Hallmark label used to define the parent program should not be read literally as transplant rejection.

The contrast between continuous and fixed threshold results is informative. A standardized continuous score transported to GSE31312, while the unchanged raw development cutoff produced a borderline result and different high-versus-low group proportions. Rank based construction reduced dependence on absolute platform scale, but it

could not eliminate differences in specimen processing, platform annotation, case mix, follow-up, or baseline hazards. The poor external calibration slope and overoptimistic mean survival prediction further show that an associated biomarker is not automatically a deployable clinical model.

Adjustment choices also constrain the claim. The primary model used age, sex, and COO because these variables were available across cohorts and fixed before confirmation. In the GSE31312 IPI complete-case subset, the composite estimate attenuated and was not significant. Although the immune axis retained a borderline association, this was a sensitivity result rather than a separate confirmatory test. Accordingly, the present data do not demonstrate prognostic value beyond IPI and do not support a nomogram.

Several design features strengthen interpretability. The cohorts had separate discovery, prior validation, and confirmatory roles; the endpoint, weights, directions, cutoff, and adjustment were frozen; QC was outcome-blinded; analyses remained cohort-specific; and a source manifest fingerprinted the reviewed outputs. The additional cohort was processed independently and no accession overlap was detected. These controls reduce common sources of analytical flexibility, although they cannot substitute for prospective validation.

Limitations

This study was a retrospective secondary analysis of public datasets with heterogeneous tissue processing and microarray platforms and without centralized pathology review. The endpoint was all-cause mortality within 2 years of the recorded overall-survival origin; it was neither lymphoma-specific mortality nor a direct measure of primary refractory disease. Clinical covariate availability differed among cohorts, and IPI adjustment required complete-case analysis and attenuated the composite association in GSE31312. Although no accession overlap was detected, public metadata could not establish patient-level independence across all source cohorts.

Only three outcome cohorts were available, so the random-effects summaries and I^2 estimates are descriptive and imprecise, particularly for the heterogeneous MYC axis. The transferred cutoff was not fully confirmed, and the fixed development model was poorly calibrated; no recalibration, clinical decision-curve analysis, or prospective utility assessment was performed. Finally, bulk expression cannot identify immune-cell fractions or establish causal pathways. The immune-module and 26-gene analyses were exploratory and were not used to modify the locked score.

Conclusion

A locked rank-based composite of MYC activation and immune-program depletion was associated with 0–2-year all-cause mortality in a prespecified additional DLBCL cohort. The reproducible signal was dominated by immune depletion, whereas the MYC effect was cohort-dependent. The score should currently be viewed as a cross-platform biological association rather than a calibrated clinical predictor. Prospective validation with complete IPI data, standardized tissue processing, and prespecified calibration assessment is required before clinical use.

Abbreviations

ABC
activated B-cell-like
BH
Benjamini-Hochberg
CAMERA

correlation adjusted mean rank
CI
confidence interval
COO
cell of origin
DLBCL
diffuse large B-cell lymphoma
FDR
false discovery rate
GEO
Gene Expression Omnibus
GCB
germinal center B-cell-like
HR
hazard ratio
IPI
International Prognostic Index
mKH
modified Knapp-Hartung
OS
overall survival
PH
proportional hazards
QC
quality control
REML
restricted maximum likelihood
R-CHOP
rituximab,cyclophosphamide,doxorubicin,vincristine,and prednisone.

Declarations

Funding

This research received no external funding.

Ethics approval

Not applicable. This study was a secondary analysis conducted entirely using publicly available, de-identified datasets from the NCBI Gene Expression Omnibus. The authors did not recruit human participants, collect human tissues, conduct interventions, or perform animal experiments; therefore, formal ethics committee approval was not required.

Consent to participate

Not applicable. This study used publicly available, de-identified datasets and did not involve direct participation of identifiable individuals; therefore, informed consent to participate was not required.

Consent for publication

Not applicable.

Data Availability Statement

The gene-expression and associated clinical or phenotypic data analyzed in this study were obtained from the NCBI Gene Expression Omnibus (GEO) under the following accession numbers: GSE181063 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE181063>), GSE10846 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE10846>), GSE31312 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31312>), and GSE173263 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE173263>). These publicly available datasets constituted the sources of the gene-expression and clinical or phenotypic data used in this study. No new participant-level data were generated.

Author Contributions

Methodology, H.C. and Y.Z.; writing—original draft preparation, X.F.; writing—review and editing, F.Z. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors have no competing interests.

References

1. Maurer MJ, Ghesquières H, Jais JP, Witzig TE, Haioun C, Thompson CA, et al. Event-free survival at 24 months is a robust end point for disease-related outcome in diffuse large B-cell lymphoma treated with immunochemotherapy. *J Clin Oncol*. 2014;32:1066–73. <https://doi.org/10.1200/JCO.2013.51.5866>.
2. International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med*. 1993;329:987–94. <https://doi.org/10.1056/NEJM199309303291402>.
3. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000;403:503–11. <https://doi.org/10.1038/35000501>.
4. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med*. 2002;346:1937–47. <https://doi.org/10.1056/NEJMoa012914>.
5. Lenz G, Wright G, Dave SS, Xiao W, Powell J, Zhao H, et al. Stromal gene signatures in large-B-cell lymphomas. *N Engl J Med*. 2008;359:2313–23. <https://doi.org/10.1056/NEJMoa0802885>.
6. Schmitz R, Wright GW, Huang DW, Johnson CA, Phelan JD, Wang JQ, et al. Genetics and pathogenesis of diffuse large B-cell lymphoma. *N Engl J Med*. 2018;378:1396–407. <https://doi.org/10.1056/NEJMoa1801445>.
7. Chapuy B, Stewart C, Dunford AJ, Kim J, Kamburov A, Redd RA, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med*. 2018;24:679–90. <https://doi.org/10.1038/s41591-018-0016-8>.
8. Kotlov N, Bagaev A, Revuelta MV, Phillip JM, Cacciapuoti MT, Antysheva Z, et al. Clinical and biological subtypes of B-cell lymphoma revealed by microenvironmental signatures. *Cancer Discov*. 2021;11:1468–89. <https://doi.org/10.1158/2159-8290.CD-20-0839>.

9. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies. *J Natl Cancer Inst.* 2005;97:1180–4. <https://doi.org/10.1093/jnci/dji237>.
10. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002;30:207–10. <https://doi.org/10.1093/nar/30.1.207>.
11. Painter D, Barrans S, Lacy S, Smith A, Crouch S, Westhead D, et al. Cell-of-origin in diffuse large B-cell lymphoma: findings from the UK's Haematological Malignancy Research Network. *Br J Haematol.* 2019;185:781–4. <https://doi.org/10.1111/bjh.15519>.
12. Lacy SE, Barrans SL, Beer PA, Painter D, Smith AG, Roman E, et al. Targeted sequencing in DLBCL, molecular subtypes, and outcomes: a Haematological Malignancy Research Network report. *Blood.* 2020;135:1759–71. <https://doi.org/10.1182/blood.2019003535>.
13. Visco C, Li Y, Xu-Monette ZY, Miranda RN, Green TM, Li Y, et al. Comprehensive gene expression profiling and immunohistochemical studies support application of immunophenotypic algorithm for molecular subtype classification in diffuse large B-cell lymphoma: a report from the International DLBCL Rituximab-CHOP Consortium Program Study. *Leukemia.* 2012;26:2103–13. <https://doi.org/10.1038/leu.2012.83>.
14. Dlouhy I, Karube K, Enjuanes A, et al. Revised International Prognostic Index and genetic alterations are associated with early failure to R-CHOP in patients with diffuse large B-cell lymphoma. *Br J Haematol.* 2022;196:589–98. <https://doi.org/10.1111/bjh.17858>.
15. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47. <https://doi.org/10.1093/nar/gkv007>.
16. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark gene set collection. *Cell Syst.* 2015;1:417–25. <https://doi.org/10.1016/j.cels.2015.12.004>.
17. Wu D, Smyth GK. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res.* 2012;40:e133. <https://doi.org/10.1093/nar/gks461>.
18. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B.* 1995;57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
19. Cox DR, Stat Soc B. 1972;34:187–220. <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>.
20. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika.* 1994;81:515–26. <https://doi.org/10.1093/biomet/81.3.515>.
21. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol.* 2013;13:33. <https://doi.org/10.1186/1471-2288-13-33>.
22. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw.* 2010;36:1–48. <https://doi.org/10.18637/jss.v036.i03>.
23. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Stat Med.* 2003;22:2693–710. <https://doi.org/10.1002/sim.1482>.

Figures

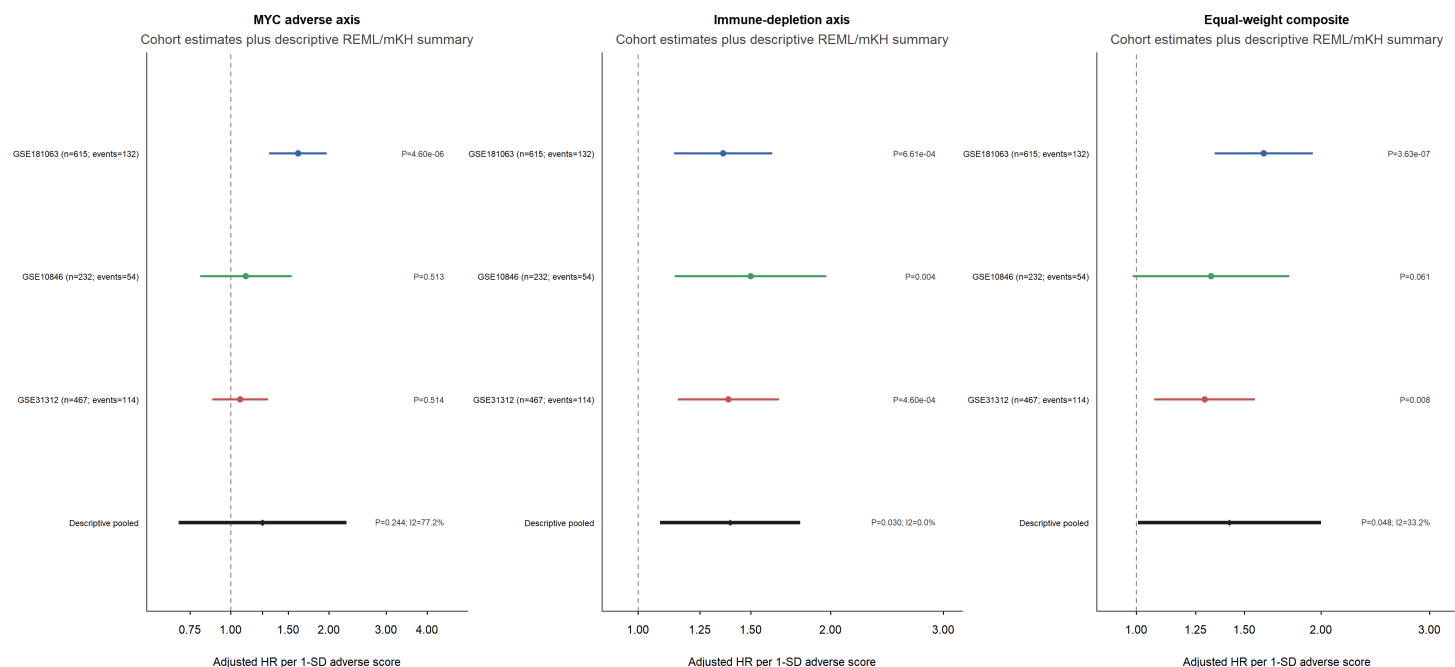


Figure 1

Adjusted associations of the MYC adverse axis, immune-depletion axis, and locked equal-weight composite with all-cause mortality within 2 years. Cohort-specific estimates are hazard ratios per 1-SD higher adverse score from Cox models adjusted for age, sex, and COO. Black diamonds show descriptive REML random-effects summaries with modified Knapp-Hartung confidence intervals; these pooled estimates do not replace the prespecified GSE31312 confirmatory test

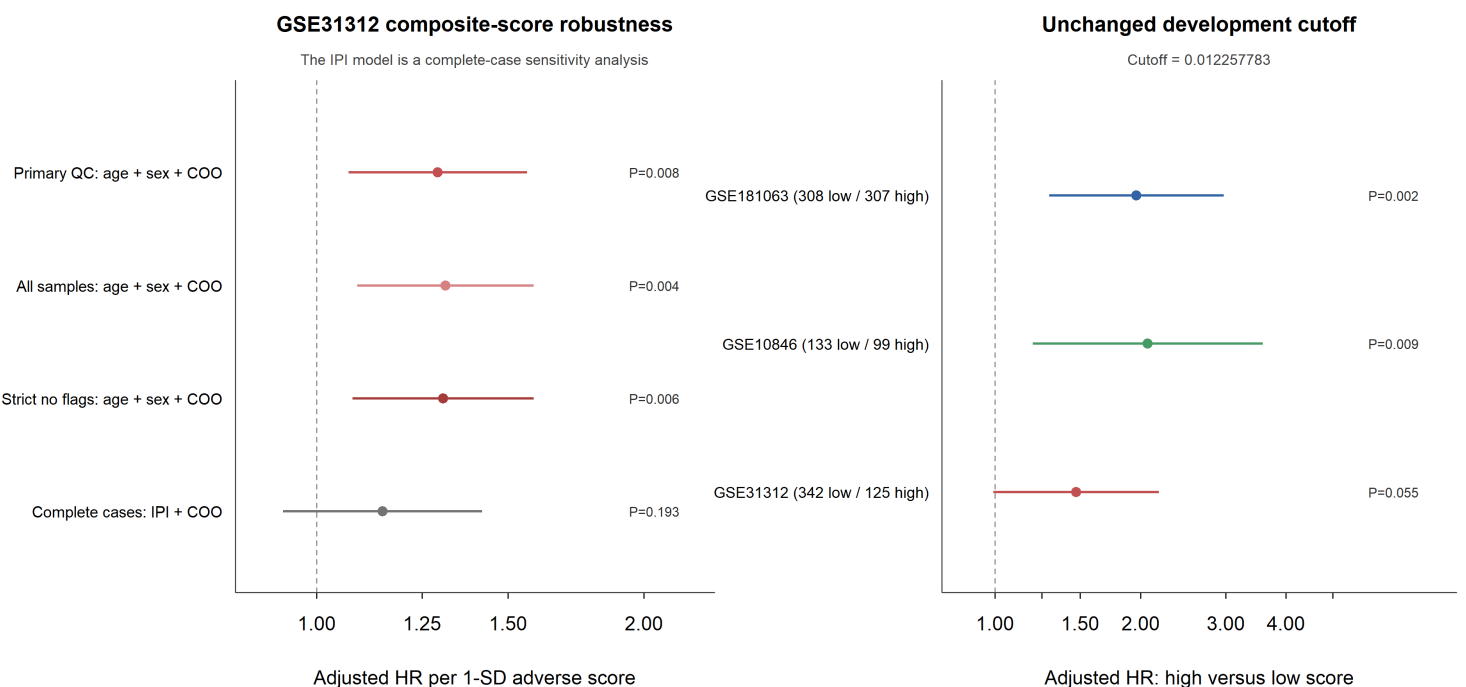


Figure 2

Robustness and transportability analyses. Left: the locked equal-weight composite in GSE31312 under the primary QC set, all-sample sensitivity, strict no-flag sensitivity, and IPI-plus-COO complete-case sensitivity. Right: adjusted high-versus-low hazard ratios using the unchanged GSE181063 median cutoff in all three cohorts. The cutoff was not optimized in either validation cohort

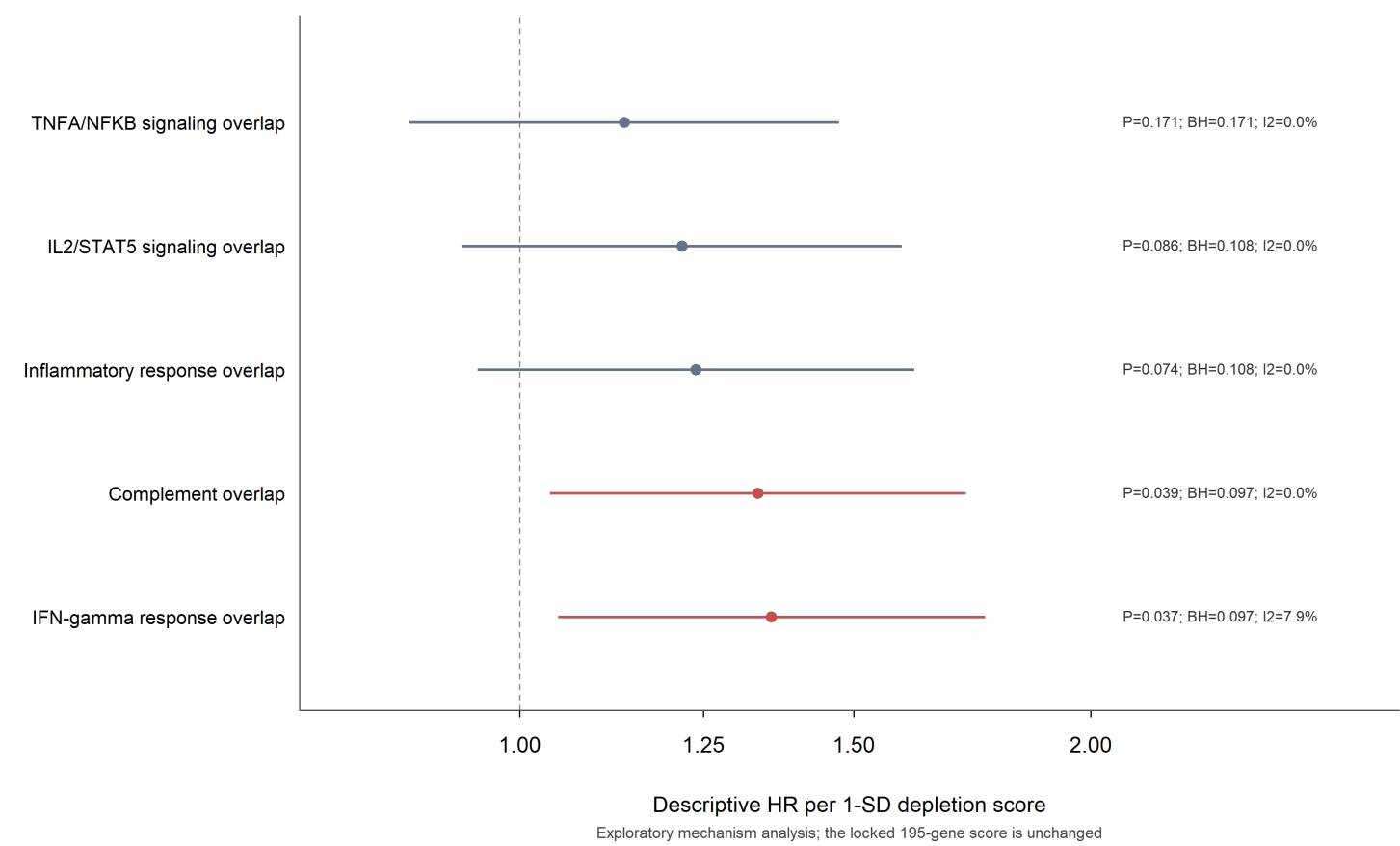


Figure 3

Descriptive three-cohort meta-analysis of five prespecified overlapping functional annotations within the locked 195-gene immune program. Hazard ratios are per 1-SD higher depletion-oriented score. Red indicates exploratory BH-adjusted $P < .10$ across five modules. Modules are overlapping annotations and are not immune-cell fractions or replacement predictors

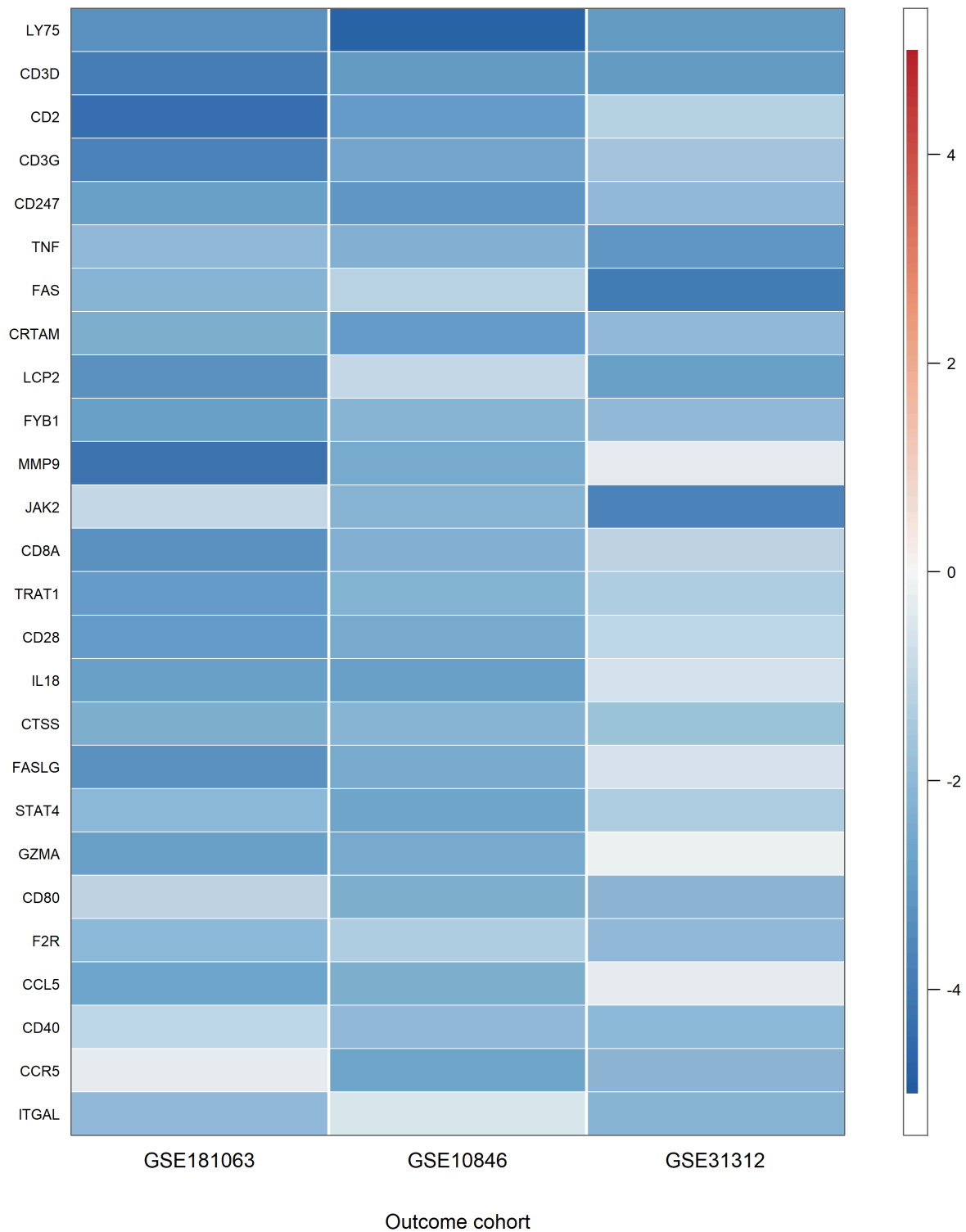


Figure 4

Moderated gene-level t statistics for the 26 genes meeting the prespecified exploratory core criterion: negative log fold change in all three cohorts and nominal $P < .05$ in at least two. Blue denotes lower expression in early death. Genes remain part of the unchanged 195-gene immune score and were not used to refit or rebuild the predictor

Supplementary Files

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

- [Supplementaryinformationlegends.docx](#)
- [Supplementary.zip](#)