

Cross-cohort transcriptomic analysis identifies a reproducible immune-active program associated with neoadjuvant treatment response in rectal cancer

Xichang FEI

Department of Hematologic Oncology, Zhabei Central Hospital, Jing'an District, Shanghai

Haimin CHEN

Department of Hematologic Oncology, Zhabei Central Hospital, Jing'an District, Shanghai

Ying ZHANG

Department of Hematologic Oncology, Zhabei Central Hospital, Jing'an District, Shanghai

Fan ZHOU

zhoufanzzx@163.com

Department of Hematologic Oncology, Zhabei Central Hospital, Jing'an District, Shanghai

Research Article

Keywords: rectal cancer, neoadjuvant chemoradiotherapy, transcriptomics, gene-set enrichment analysis, tumor microenvironment, external validation

Posted Date: September 2nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10645533/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

Response to neoadjuvant treatment in locally advanced rectal cancer varies, and transcriptomic biomarkers often fail to reproduce across cohorts. We sought pathway-level signals stable across platforms and response definitions.

Methods

Pretreatment transcriptomes were analyzed independently in an RNA-sequencing discovery cohort (GSE190826; primary n = 104), two primary microarray validation cohorts (GSE150082, n = 39; GSE45404 GPL570, n = 42), and a supplementary platform-specific cohort (GSE45404 GPL4133, n = 38). Differential expression and preranked Hallmark gene-set enrichment were followed by leading-edge and overlap-free immune-score analyses.

Results

Eight pathways stable across three overlapping discovery sensitivity analyses were prespecified. All eight matched the discovery direction in GSE150082, whereas only two matched in GPL570. Allograft rejection alone was directionally concordant and FDR-significant in both primary validations (NES/FDR: GSE190826, 1.74/6.21 × 10⁻⁴; GSE150082, 2.27/1.10 × 10⁻¹¹; GPL570, 2.37/8.50 × 10⁻¹³); GPL4133 supported the direction without significance (NES = 1.13; FDR = 0.231). Thirty-two leading-edge genes were shared by all three primary cohorts. After removing all MCP-counter marker genes from score construction, correlations with T-cell, cytotoxic-lymphocyte, and B-lineage scores remained FDR-significant across the primary cohorts. The sample-level Allograft score did not differ by response after FDR correction.

Conclusions

A coordinated immune-active program was reproducibly associated with favorable response across heterogeneous rectal cancer cohorts. Six other pathways were heterogeneous, and the absence of validated sample-level discrimination precludes using the current signature as a clinical predictor.

Key points

- Only the Hallmark Allograft rejection program achieved directionally concordant, FDR-significant replication in both primary external validation cohorts.

- Six of eight discovery pathways reversed direction in GSE45404 GPL570, providing direct evidence of substantial cohort-dependent heterogeneity.
- Overlap-free correlations supported a coordinated immune-active state, but the sample-level pathway score was not a validated response classifier.

1. Introduction

Neoadjuvant treatment is central to risk-adapted management of locally advanced rectal cancer, including long-course chemoradiotherapy, short-course radiotherapy, and total neoadjuvant therapy in selected clinical settings [1]. Tumor response varies widely: some patients achieve pathological or durable clinical complete response, whereas others retain substantial viable tumor. A pathological complete response (pCR) is associated with favorable long-term outcomes, but it is known only after treatment and surgery [2]. Reliable pretreatment biomarkers could improve risk stratification, trial design, and the selection of organ-preserving strategies.

Gene-expression studies have proposed response-associated genes and pathways, yet few findings have remained reproducible across cohorts [3]. Major obstacles include small samples, different treatment regimens and response criteria, variable tumor purity and microenvironmental composition, and platform-specific measurement. A gene-level classifier discovered in one dataset may therefore capture a cohort-specific technical or biological state. Pathway-level analysis can reduce dependence on individual probes or transcripts by asking whether coordinated gene programs occupy the extremes of a ranked differential-expression profile.

Three public datasets provide complementary opportunities to examine this problem. GSE190826 profiled pretreatment rectal cancer biopsies by RNA sequencing and linked inflammatory cancer-associated fibroblasts to poor chemoradiotherapy response [4]. GSE150082 used Agilent microarrays and reported that pretreatment B-cell infiltration and immune-related transcriptional features accompanied successful tumor regression [5]. GSE45404 used two microarray platforms in the same source study and identified a response-associated molecular network centered on XRCC3 [6]. These studies differ in endpoint, platform, treatment context, and biological emphasis, making concordance across them a stringent test of reproducibility.

We performed a staged reanalysis that deliberately avoided direct merging of cross-platform expression matrices. First, Hallmark pathways robust to three overlapping discovery sensitivity analyses were fixed as external hypotheses. Second, those pathways were evaluated independently in two primary validation cohorts, with the second GSE45404 platform retained as supplementary evidence rather than counted as an additional independent study. Third, the pathway showing reproducible cross-cohort enrichment was interpreted through leading-edge genes and immune-microenvironment scores, including an overlap-removal sensitivity analysis. Our aim was to distinguish reproducible biology from cohort-dependent findings, not to construct an individual-level clinical prediction model.

2. Materials and methods

2.1 Study design and public datasets

This was a secondary analysis of deidentified, publicly available pretreatment transcriptomic data from the National Center for Biotechnology Information Gene Expression Omnibus (GEO) [8]. The analytical hierarchy was fixed before the cross-cohort synthesis: GSE190826 was the discovery cohort; GSE150082 and GSE45404 GPL570 were primary external validation cohorts; and GSE45404 GPL4133 was a supplementary platform-specific cohort. The two GSE45404 matrices were processed separately and had no overlapping patient-name fields in the downloaded metadata, but they originated from one GEO study and were not treated as two independent studies. Expression values from different platforms were never concatenated.

GSE190826 contains 105 pretreatment, formalin-fixed paraffin-embedded biopsies from patients receiving neoadjuvant chemoradiotherapy followed by surgery, including 26 pCR and 79 non-pCR samples in the source record [4]. One prominent quality-control outlier (GSM5732642) was excluded from the primary analysis, leaving 104 samples (26 pCR and 78 non-pCR). The complete 105-sample set and a stricter 102-sample set were retained as sensitivity analyses.

GSE150082 includes 39 pretreatment biopsies from consecutive patients with locally advanced rectal cancer analyzed on the Agilent GPL13497 two-color platform [5]. Sixteen were classified as favorable and 23 as unfavorable. Favorable response comprised pathological tumor-regression grade 0–1 or clinical complete response; unfavorable response comprised grade 2–3 or unresectability. Because this endpoint is broader than pCR, GSE150082 was interpreted as validation of response-associated biology rather than strict pCR prediction.

GSE45404 comprises 80 samples profiled on GPL570 (42 samples; 19 responders and 23 non-responders) or GPL4133 (38 samples; 16 responders and 22 non-responders) [6]. The GPL570 subset was assigned as primary validation 2 and was harmonized as Mandard tumor-regression grade 1–2 versus 3–5 [7]. GPL4133 was retained as supplementary evidence. In all datasets, positive effect directions were harmonized to indicate higher expression or pathway enrichment in the more favorable response group.

2.2 Discovery-cohort quality control and differential expression

Processed gene-count data and sample metadata were imported using GEO-compatible workflows [8, 9]. Low-abundance genes were removed using the prespecified edgeR expression filter. Library sizes were normalized by the trimmed mean of M-values, and differential expression was estimated with robust quasi-likelihood negative-binomial generalized linear models in edgeR [10, 11]. Positive log₂ fold changes represented higher expression in pCR. Principal-component analysis of the 1,000 most variable

genes was used for quality review, together with targeted review of gastric epithelial markers that were prominent among high-effect genes.

Three overlapping discovery analyses were defined. The all105 analysis used every pretreatment sample. The primary104 analysis excluded GSM5732642, an isolated multivariate outlier, and served as the main discovery dataset. The strict102 analysis additionally excluded GSM5732587 and GSM5732624, two pCR samples flagged during gastric epithelial-marker review. The latter two analyses were sensitivity analyses of overlapping individuals, not independent cohorts.

2.3 Microarray processing and differential expression

Normalized GEO Series Matrix files were read separately for each platform. Platform annotation was used to map probes to HUGO gene symbols. Probes without a valid symbol were removed, and multiple probes mapping to the same symbol were collapsed by retaining the probe with the largest across-sample variance. This yielded 21,753 genes for GSE150082, 21,750 for GSE45404 GPL570, and 19,260 for GSE45404 GPL4133. Differential expression was estimated with empirical-Bayes linear models in limma [12]. Contrasts were oriented as favorable minus unfavorable response.

For quality assurance of the GSE45404 direction, the original GPL570 probes reported in the source study were examined. The exact XRCC3 probe (216299_s_at) was higher in responders, and the gene-level XRCC3 coefficient from the analysis was positive. This concordance argued against a simple reversal of response labels as the explanation for pathway-level direction changes.

2.4 Hallmark enrichment and prespecified pathway validation

Genes were ranked by signed statistics from the corresponding differential-expression model. Preranked gene-set enrichment analysis (GSEA) was performed with the 50 human Hallmark gene sets [13, 14] using the multilevel algorithm implemented in fgsea [15]. Gene sets with 15–500 represented genes were tested. Normalized enrichment scores (NES) were oriented so that positive values denoted enrichment in favorable responders. P values were adjusted by the Benjamini–Hochberg procedure [16].

Eight pathways were prespecified for external evaluation because they had $FDR < 0.05$ and the same direction in all105, primary104, and strict102. External results were reported both across the full Hallmark collection and after correction across the eight prespecified hypotheses; the latter was used for the main validation classification. Robust replication required direction concordance with discovery and $FDR < 0.05$ in both primary validations. Partial replication required direction concordance in both primary validations and significance in one. A reversal in GPL570 was classified as substantial cohort-dependent heterogeneity. GPL4133 could support but did not define these primary classes.

2.5 Leading-edge and cross-cohort core genes

Leading-edge genes—the subset encountered before the maximum deviation of the GSEA running-sum statistic—were extracted from each dataset. For the reproducible Allograft rejection pathway, a primary

cross-cohort core was defined as genes present in at least two of the three primary leading edges. Genes present in all three primary cohorts were reported as the most stringent shared set. Support from GPL4133 and from the all105/primary104/strict102 internal consensus was tabulated separately.

2.6 Sample-level pathway and microenvironment scoring

For the Allograft rejection sample score, expression values were converted to within-sample fractional ranks and averaged across available pathway genes. Scores were standardized within cohort only for visualization. Immune and stromal scores were calculated as the mean expression of available genes from the fixed HUGO marker sets published for MCP-counter [17]; no cross-cohort scaling or pooling was performed. Five immune populations—T cells, CD8 T cells, cytotoxic lymphocytes, natural killer (NK) cells, and B lineage—were prioritized a priori for interpretation, while all ten MCP-counter populations were tested in response and correlation analyses.

Response-group differences were tested with two-sided Wilcoxon rank-sum tests and summarized using Cliff's delta [18], with positive values indicating higher scores in favorable responders. Spearman rank correlations quantified associations between Allograft rejection and cell-population scores. For MCP-counter response comparisons and correlations, P values were adjusted within each dataset across all ten populations. P values for the Allograft sample-score response comparisons were adjusted across the four matrices. To exclude mechanical correlation caused by shared scoring genes, all 111 unique MCP-counter markers were removed from the 200-gene Hallmark Allograft rejection set. Nine genes overlapped directly, leaving 191 genes for the overlap-free score; correlations and response comparisons were then repeated.

2.7 Statistical interpretation and reproducibility

All tests were two-sided. Unless stated otherwise, $FDR < 0.05$ defined statistical significance. Effect direction and multiplicity-adjusted significance were considered jointly. No formal pooled meta-analysis was performed because response endpoints, platforms, and source-study designs were not sufficiently homogeneous. No LASSO or other multivariable clinical classifier was fitted. Analyses were performed in R 4.6.1 with code organized as numbered, reproducible scripts. Key packages were edgeR 4.10.1, limma 3.68.4, fgsea 1.38.0, GEOquery 2.80.0, and msigdb 7.5.1; complete session information was archived with the analysis materials.

3. Results

3.1 Cohort composition and discovery quality control

The main analysis comprised one RNA-sequencing discovery cohort, two primary external microarray validations, and one supplementary microarray subset (Table 1; Fig. 1). The primary sample sizes were 104, 39, and 42, with favorable-response proportions of 25.0%, 41.0%, and 45.2%, respectively. The supplementary GPL4133 subset contained 38 samples. Cross-platform expression matrices were processed independently.

Principal-component analysis identified one marked non-pCR outlier and two pCR samples requiring targeted review (Supplementary Fig. S1 (ESM 1)). Several gastric epithelial genes, including PGC, ATP4A, ATP4B, CLDN18, PSCA, and TFF2, contributed strongly to the initial differential-expression pattern (Fig. 2a; Supplementary Fig. S2 (ESM 1)). Excluding the marked outlier defined primary104; excluding the additional two flagged pCR samples defined strict102. The pathway findings described below were required to remain significant and directionally stable across all three discovery analyses.

Table 1 Cohort and analysis summary

Dataset / platform	Analysis role	Response endpoint	n	Favorable n	Unfavorable n	Genes
GSE190826 RNA sequencing	Discovery cohort	pCR versus non-pCR	104	26	78	19,481
GSE150082 Agilent microarray (GPL13497)	Primary external validation 1	Good versus Poor response	39	16	23	21,753
GSE45404 Affymetrix HG-U133 Plus 2.0 (GPL570)	Primary external validation 2	Mandard TRG 1–2 versus TRG 3–5	42	19	23	21,750
GSE45404 Agilent whole-genome microarray (GPL4133)	Supplementary platform-specific cohort	Responder versus Non-responder	38	16	22	19,260

Favorable and unfavorable are harmonized analytical labels; the original clinical endpoints were retained within each cohort. GSE45404 GPL4133 is supplementary and is not counted as an independent primary validation.

3.2 Eight Hallmark pathways were internally robust

Eight Hallmark pathways met the internal robustness rule (Fig. 2b; Fig. 3; Supplementary Table S1 (ESM 2)). Allograft rejection (primary104 NES 1.74) and TNFA signaling via NF-κB (NES 1.58) were enriched in pCR. MYC targets V1 (NES -2.26), mTORC1 signaling (-2.06), E2F targets (-1.94), unfolded protein response (-1.87), G2M checkpoint (-1.81), and glycolysis (-1.50) were enriched in non-pCR. Each pathway retained FDR < 0.05 and the same direction in all105, primary104, and strict102, demonstrating robustness to the prespecified sample exclusions but not constituting independent replication.

3.3 External validation separated reproducible from heterogeneous pathways

In GSE150082, all eight prespecified pathways matched the discovery direction and seven achieved FDR < 0.05 across the eight tests. Glycolysis remained directionally concordant but was not significant (NES -1.21; FDR 0.085). Thus, this cohort provided broad cross-platform support despite its wider Good-versus-Poor endpoint.

The GSE45404 GPL570 results differed sharply. Allograft rejection (NES 2.37; FDR 8.50×10^{-13}) and glycolysis (NES -1.63; FDR 3.28×10^{-4}) matched the discovery direction. The six other pathways reversed, including strongly significant positive NES values for MYC targets V1, mTORC1 signaling, E2F targets, unfolded protein response, and G2M checkpoint; TNFA signaling via NF- κ B also reversed but was not significant. The XRCC3 probe-direction check supported the response-label orientation, making a simple labeling error an unlikely explanation.

Across the two primary validations, Allograft rejection was therefore the sole robustly replicated pathway (Table 2; Fig. 4). Its NES/FDR values were 1.74/ 6.21×10^{-4} in GSE190826 primary104, 2.27/ 1.10×10^{-11} in GSE150082, and 2.37/ 8.50×10^{-13} in GPL570. GPL4133 showed the same direction (NES 1.13) without significance (FDR 0.231). Glycolysis was classified as partial replication because both primary validations matched discovery but only GPL570 was significant. The remaining six pathways were classified as substantially heterogeneous. Only Allograft rejection had the same direction across all four data matrices.

Table 2 Cross-cohort evidence for the eight prespecified Hallmark pathways. Values are NES (FDR)

Pathway	GSE190826 primary104	GSE150082	GSE45404 GPL570	GSE45404 GPL4133	Primary evidence class
Allograft rejection	1.74 (6.21 × 10 ⁻⁴)	2.27 (1.10 × 10 ⁻¹¹)	2.37 (8.50 × 10 ⁻¹³)	1.13 (0.23)	Robust cross-cohort replication
Glycolysis	-1.50 (0.02)	-1.21 (0.09)	-1.63 (3.28 × 10 ⁻⁴)	1.21 (0.13)	Partial replication with consistent direction
TNFA signaling via NF-κB	1.58 (0.01)	1.68 (9.22 × 10 ⁻⁵)	-1.00 (0.47)	-1.68 (1.05 × 10 ⁻⁴)	Substantial cohort-dependent heterogeneity
MYC targets V1	-2.26 (4.76 × 10 ⁻¹⁰)	-2.74 (6.00 × 10 ⁻²⁰)	2.29 (8.75 × 10 ⁻¹¹)	2.38 (8.49 × 10 ⁻¹⁴)	Substantial cohort-dependent heterogeneity
mTORC1 signaling	-2.06 (1.34 × 10 ⁻⁷)	-1.87 (2.60 × 10 ⁻⁶)	1.85 (4.21 × 10 ⁻⁶)	1.73 (5.34 × 10 ⁻⁵)	Substantial cohort-dependent heterogeneity
E2F targets	-1.94 (1.47 × 10 ⁻⁶)	-3.31 (3.52 × 10 ⁻³⁶)	2.93 (6.89 × 10 ⁻²⁴)	2.49 (3.65 × 10 ⁻¹⁵)	Substantial cohort-dependent heterogeneity
Unfolded protein response	-1.87 (4.30 × 10 ⁻⁴)	-1.77 (4.39 × 10 ⁻⁴)	1.46 (0.03)	0.88 (0.71)	Substantial cohort-dependent heterogeneity
G2M checkpoint	-1.81 (4.68 × 10 ⁻⁵)	-2.65 (2.92 × 10 ⁻¹⁸)	2.67 (7.99 × 10 ⁻¹⁸)	2.43 (4.99 × 10 ⁻¹⁴)	Substantial cohort-dependent heterogeneity

NES, normalized enrichment score. Positive NES indicates enrichment in favorable responders. For GSE190826, FDR was adjusted across all 50 Hallmark pathways; for the external validation datasets, FDR was adjusted across the eight prespecified hypotheses. GPL4133 was supplementary and did not define the primary evidence class.

3.4 A recurrent immune leading edge drove Allograft rejection enrichment

The three primary Allograft rejection leading edges contained 138 unique genes. Thirty-two genes were present in all three primary cohorts and 84 in at least two (Table 3; Supplementary Table S2 (ESM 2)). Thirty-three of the 84-gene core were also supported by GPL4133, 44 were present in the internal all105/primary104/strict102 consensus, and 14 were shared by all four matrices.

The all-three set included antigen-presentation genes (CD74, HLA-DMA, HLA-DMB, HLA-DRA, HLA-DOA, and HLA-DQA1), T- and natural-killer-cell signaling or cytotoxic genes (KLRD1, TRAT1, CD247, CD8A,

GZMA, CCL5, IL2RB, ITK, CD2, CD7, IFNG, and CD96), and B-lineage or pan-leukocyte components (CD79A, PRKCB, FCGR2B, PTPRC, and HCLS1). All 32 occupied the favorable-response side of the ranked statistic in each primary cohort (Fig. 5a). These genes define a recurrent pathway-driving module; they were not evaluated as independent single-gene biomarkers.

Table 3 Allograft rejection leading-edge genes shared by all three primary cohorts

Gene 1	Gene 2	Gene 3	Gene 4
CD74	KLRD1	HLA-DMB	HLA-DMA
HLA-DRA	HLA-DOA	TRAT1	ITGB2
CXCL9	GZMA	CD247	CCL5
CD8A	IL16	EGFR	IL2RB
PRKCB	ITGAL	HLA-DQA1	ITK
LY86	CD2	CD79A	CD7
HCLS1	IFNG	PTPRC	IL15
CD86	FCGR2B	LTB	CD96

All 32 genes were members of the Allograft rejection leading edge and had favorable-direction ranks in GSE190826 primary104, GSE150082, and GSE45404 GPL570. These are pathway-driving candidates, not independently validated single-gene biomarkers.

3.5 Immune associations persisted after removal of shared marker genes

The original sample-level Allograft score correlated positively with multiple microenvironment scores, particularly T cells, cytotoxic lymphocytes, B lineage, and monocytic lineage. Direct gene overlap with the fixed MCP-counter markers was modest: nine of 200 Allograft genes overlapped the 111-marker union. Removing every MCP-counter marker from the Allograft score left 191 genes and changed the prioritized correlations only slightly.

After overlap removal, T-cell correlations were $\rho = 0.62, 0.75,$ and 0.68 in GSE190826, GSE150082, and GPL570; cytotoxic-lymphocyte correlations were $0.51, 0.58,$ and 0.73 ; and B-lineage correlations were $0.25, 0.63,$ and 0.53 (Table 4; Fig. 5b). All nine primary-cohort tests remained positive and FDR-significant. These results exclude direct sharing of scoring genes as the sole explanation and support a broader coordinated immune-active state.

Response-group comparisons of microenvironment scores were directionally positive for the five prioritized immune populations in all three primary cohorts, but statistical support was limited: only

cytotoxic lymphocytes and B lineage in GSE150082 reached within-dataset FDR < 0.05 (Supplementary Fig. S7 (ESM 1)). Similarly, the overlap-free Allograft score was higher on average in favorable responders in every cohort but did not achieve FDR < 0.05 in any of the four matrices (Supplementary Fig. S10 (ESM 1)). The pathway-level GSEA signal therefore did not translate into a validated sample-level response discriminator.

Table 4 Correlations between overlap-free Allograft rejection scores and prioritized immune-cell scores in the three primary cohorts

Population	Directly shared genes	GSE190826 ρ (FDR)	GSE150082 ρ (FDR)	GPL570 ρ (FDR)	Evidence
T cells	4	0.62 (7.96×10^{-12})	0.75 (3.11×10^{-7})	0.68 (2.17×10^{-6})	Robust after removal of shared marker genes
Cytotoxic lymphocytes	2	0.51 (8.24×10^{-8})	0.58 (2.46×10^{-4})	0.73 (1.64×10^{-7})	Robust after removal of shared marker genes
B lineage	1	0.25 (0.02)	0.63 (5.31×10^{-5})	0.53 (4.50×10^{-4})	Robust after removal of shared marker genes

All 111 unique MCP-counter marker genes were excluded from the 200-gene Hallmark Allograft rejection set, leaving 191 genes. FDR was calculated within each dataset across all ten tested cell populations. ρ , Spearman rank correlation.

4. Discussion

This cross-cohort analysis produced a deliberately narrower conclusion than the discovery analysis alone. Eight Hallmark pathways were internally robust in GSE190826, and GSE150082 reproduced the direction of all eight. However, six pathways reversed in GSE45404 GPL570. Only Allograft rejection was concordant and FDR-significant in both primary external validations, while glycolysis received partial directional support. Reporting the reversals is central to the result: it shows that within-cohort stability and even validation in one external dataset do not guarantee generalizability.

The Hallmark Allograft rejection label should be interpreted as a transcriptional program rather than literal transplantation biology. Its recurrent leading edge combined class II antigen presentation, leukocyte signaling, interferon-responsive genes, cytotoxic effectors, and B-lineage components. The same broad immune architecture was suggested by the original GSE150082 study, which associated pretreatment B-cell infiltration with successful tumor regression [5]. Our leading-edge and overlap-free

correlation analyses extend that observation by showing a coordinated T-cell, cytotoxic, and B-lineage state across the three primary data matrices.

The findings also need to be reconciled with the source GSE190826 study, in which inflammatory cancer-associated fibroblasts were linked to poor response [4]. Bulk tumors can contain simultaneous favorable and unfavorable microenvironmental programs. A leukocyte-rich immune-active state may accompany response, while specific fibroblast states, spatial exclusion, or therapy-induced stromal remodeling contribute resistance. The positive correlations with several immune and stromal estimates underscore that the Allograft signal is composite. Neither bulk expression nor marker-score correlation identifies the causal cell type.

The persistence of correlations after removing shared marker genes is important because gene-set scores can correlate mechanically when they contain the same genes. Here, only nine genes overlapped directly, and removal of the full MCP-counter marker union preserved significant positive correlations for T cells, cytotoxic lymphocytes, and B lineage in every primary cohort. This supports coordinated transcription beyond direct score construction. It does not prove immune-cell abundance, spatial organization, or functional activity; those require orthogonal histology, flow cytometry, spatial profiling, or single-cell data.

Why did pathway-level GSEA replicate while a simple sample score did not? GSEA tests whether many pathway genes collectively occupy one extreme of a cohort-specific differential-expression ranking. A per-sample mean-rank score compresses that information, gives equal weight to available genes, and may be affected by tumor content and platform coverage. A reproducible group-level enrichment can therefore coexist with overlapping individual-level score distributions. The absence of FDR-significant group separation in all four matrices precludes claims of clinical classification, calibration, or decision utility.

The six reversed pathways illustrate several plausible sources of heterogeneity. GSE190826 used a strict pCR endpoint, GSE150082 used a broader Good-versus-Poor definition that included clinical complete response and unresectability, and GSE45404 used Mandard regression grades. Treatment regimens, sample preservation, tumor cellularity, patient mix, geographic setting, and platform annotation also differed. Proliferation and metabolic programs may further reflect differences in viable tumor fraction. Because these factors could not be harmonized reliably from public metadata, we did not pool raw expression values or force a summary meta-effect.

Strengths include prespecification of external hypotheses after internal sensitivity analysis, independent platform-specific processing, explicit distinction between primary and supplementary evidence, response-label quality checks, correction across prespecified tests, and transparent reporting of contradictory results. The leading-edge intersection provided a biologically interpretable core, and the overlap-removal analysis addressed an important score-correlation artifact. Together, these choices reduce—but do not eliminate—the risk of post hoc selection.

4.1 Limitations

Limitations are substantial. This was a retrospective analysis of public data, with modest cohort sizes and incomplete harmonization of clinicopathological covariates. The response endpoints were not identical; GPL570 and GPL4133 came from the same source study; and bulk expression cannot resolve cell states or spatial interactions. Probe-to-gene collapse may discard isoform- or probe-specific information. The discovery sensitivity sets shared most samples and must not be counted as independent replications. No formal prospective assay, multivariable clinical model, calibration analysis, or survival validation was performed. Accordingly, the findings establish cross-cohort biological associations rather than clinical utility.

4.2 Future directions

A prospective validation study should use a uniform response endpoint, prespecified tissue-quality criteria, and a locked assay for the recurrent immune module. Spatial or single-cell profiling could determine whether the favorable program reflects organized lymphoid aggregates, antigen-presenting cells, intratumoral cytotoxic lymphocytes, or a broader inflamed microenvironment. Any clinical model should be developed only after assay simplification and should be compared with established clinical and imaging predictors in a genuinely independent cohort.

5. Conclusions

Across one RNA-sequencing discovery cohort and two primary microarray validations, Hallmark Allograft rejection was the only one of eight prespecified pathways to show robust directionally concordant, FDR-significant replication. A recurrent antigen-presentation and lymphocyte-related leading edge and overlap-free correlations with T-cell, cytotoxic-lymphocyte, and B-lineage scores support a coordinated immune-active state associated with favorable response. The pathway is a reproducible biological correlate, not yet an individual-level predictor. The six direction reversals in GPL570 emphasize that cross-cohort heterogeneity is itself a major result and should guide cautious biomarker development.

Declarations

Ethics approval and consent to participate

This study used only deidentified data from publicly accessible repositories and involved no new participant recruitment, intervention, or contact. Ethics approval and informed consent were obtained in the original studies. No additional ethics approval or participant consent was required for this secondary analysis.

Consent for publication

Not applicable to this secondary analysis of deidentified public data.

Availability of data and materials

All analyzed data are publicly available from the NCBI Gene Expression Omnibus under accession numbers GSE190826 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE190826>), GSE150082 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150082>), and GSE45404 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE45404>). Analysis-derived supplementary figures and tables are provided as ESM 1 and ESM 2.

Code availability

The curated R scripts used in this study, together with execution instructions and a static audit report, are provided as ESM 3.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Funding

This research received no external funding.

Author contributions

Xichang Fei: writing - original draft. Haimin Chen and Ying Zhang: methodology. Fan Zhou: writing - review and editing. All authors read and approved the final manuscript.

Acknowledgments

The authors acknowledge the investigators and participants who generated and shared GSE190826, GSE150082, and GSE45404.

References

1. ESMO Guidelines Committee, Hofheinz RD, Fokas E, Benhaim L, Price TJ, Arnold D, et al. Localised rectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann Oncol*. 2025;36(9):1007–24. <https://doi.org/10.1016/j.annonc.2025.05.528>.

2. Maas M, Nelemans PJ, Valentini V, Das P, Rödel C, Kuo LJ, et al. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *Lancet Oncol.* 2010;11(9):835–44. [https://doi.org/10.1016/S1470-2045\(10\)70172-8](https://doi.org/10.1016/S1470-2045(10)70172-8).
3. Dayde D, Tanaka I, Jain R, Tai MC, Taguchi A. Predictive and prognostic molecular biomarkers for response to neoadjuvant chemoradiation in rectal cancer. *Int J Mol Sci.* 2017;18(3):573. <https://doi.org/10.3390/ijms18030573>.
4. Nicolas AM, Pesic M, Engel E, Ziegler PK, Diefenhardt M, Kennel KB, et al. Inflammatory fibroblasts mediate resistance to neoadjuvant therapy in rectal cancer. *Cancer Cell.* 2022;40(2):168–e18413. <https://doi.org/10.1016/j.ccell.2022.01.004>.
5. Sendoya JM, Iseas S, Coraglio M, Golubicki M, Robbio J, Salanova R, et al. Pre-existing tumoral B cell infiltration and impaired genome maintenance correlate with response to chemoradiotherapy in locally advanced rectal cancer. *Cancers (Basel).* 2020;12(8):2227. <https://doi.org/10.3390/cancers12082227>.
6. Agostini M, Zangrando A, Pastrello C, D'Angelo E, Romano G, Giovannoni R, et al. A functional biological network centered on XRCC3: a new possible marker of chemoradiotherapy resistance in rectal cancer patients. *Cancer Biol Ther.* 2015;16(8):1160–71. <https://doi.org/10.1080/15384047.2015.1046652>.
7. Mandard AM, Dalibard F, Mandard JC, Marnay J, Henry-Amar M, Petiot JF, et al. Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma: clinicopathologic correlations. *Cancer.* 1994;73(11):2680–6. [https://doi.org/10.1002/1097-0142\(19940601\)73:11%3C2680::AID-CNCR2820731105%3E3.0.CO;2-C](https://doi.org/10.1002/1097-0142(19940601)73:11%3C2680::AID-CNCR2820731105%3E3.0.CO;2-C).
8. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002;30(1):207–10. <https://doi.org/10.1093/nar/30.1.207>.
9. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and Bioconductor. *Bioinformatics.* 2007;23(14):1846–7. <https://doi.org/10.1093/bioinformatics/btm254>.
10. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139–40. <https://doi.org/10.1093/bioinformatics/btp616>.
11. Chen Y, Lun ATL, Smyth GK. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res.* 2016;5:1438. <https://doi.org/10.12688/f1000research.8987.2>.
12. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47. <https://doi.org/10.1093/nar/gkv007>.

13. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50. <https://doi.org/10.1073/pnas.0506580102>.
14. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark gene set collection. *Cell Syst*. 2015;1(6):417–25. <https://doi.org/10.1016/j.cels.2015.12.004>.
15. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. *bioRxiv*. 2021. <https://doi.org/10.1101/060012>.
16. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Stat Methodol*. 1995;57(1):289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
17. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol*. 2016;17:218. <https://doi.org/10.1186/s13059-016-1070-5>.
18. Cliff N. Dominance statistics: ordinal analyses to answer ordinal questions. *Psychol Bull*. 1993;114(3):494–509. <https://doi.org/10.1037/0033-2909.114.3.494>.

Figures

Study design and analytical workflow

One RNA-seq discovery cohort, two primary external validations, and one supplementary platform-specific cohort

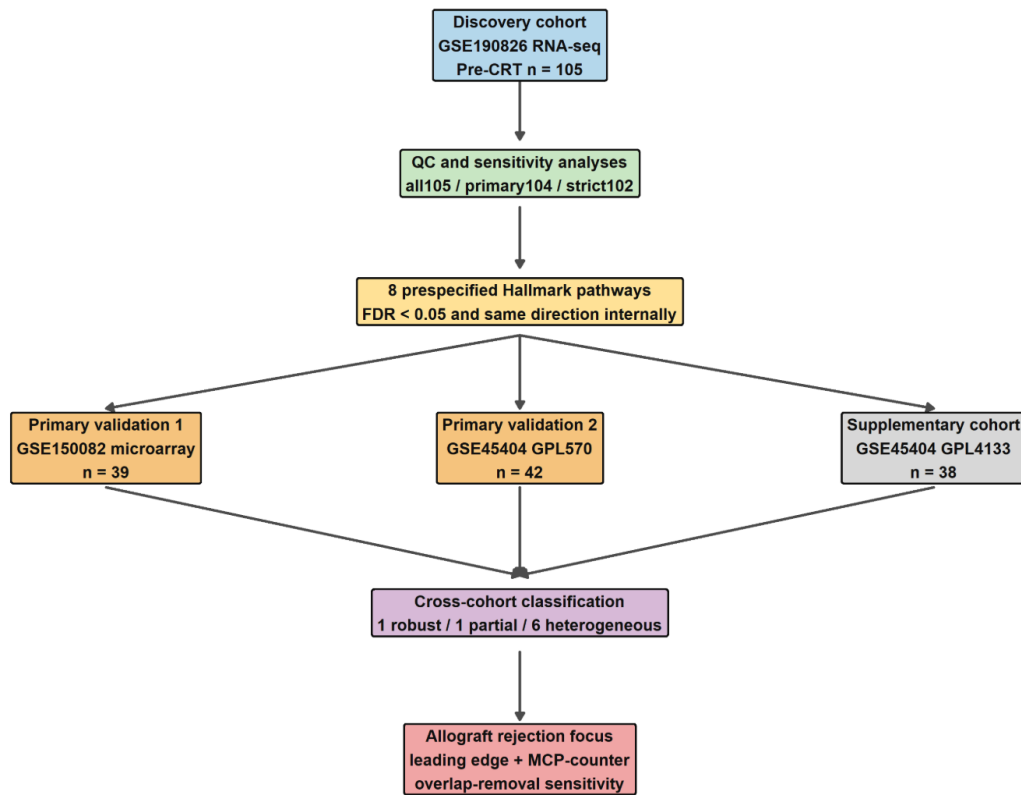


Figure 1

Study design and analytical workflow. The RNA-sequencing discovery cohort underwent quality control, differential expression, Hallmark GSEA, and internal sensitivity analyses. Eight prespecified pathways were evaluated in two primary external validation cohorts, while GPL4133 was retained as a supplementary platform-specific cohort. Cross-cohort synthesis identified Allograft rejection as the principal reproducible pathway, followed by leading-edge and immune-microenvironment analyses

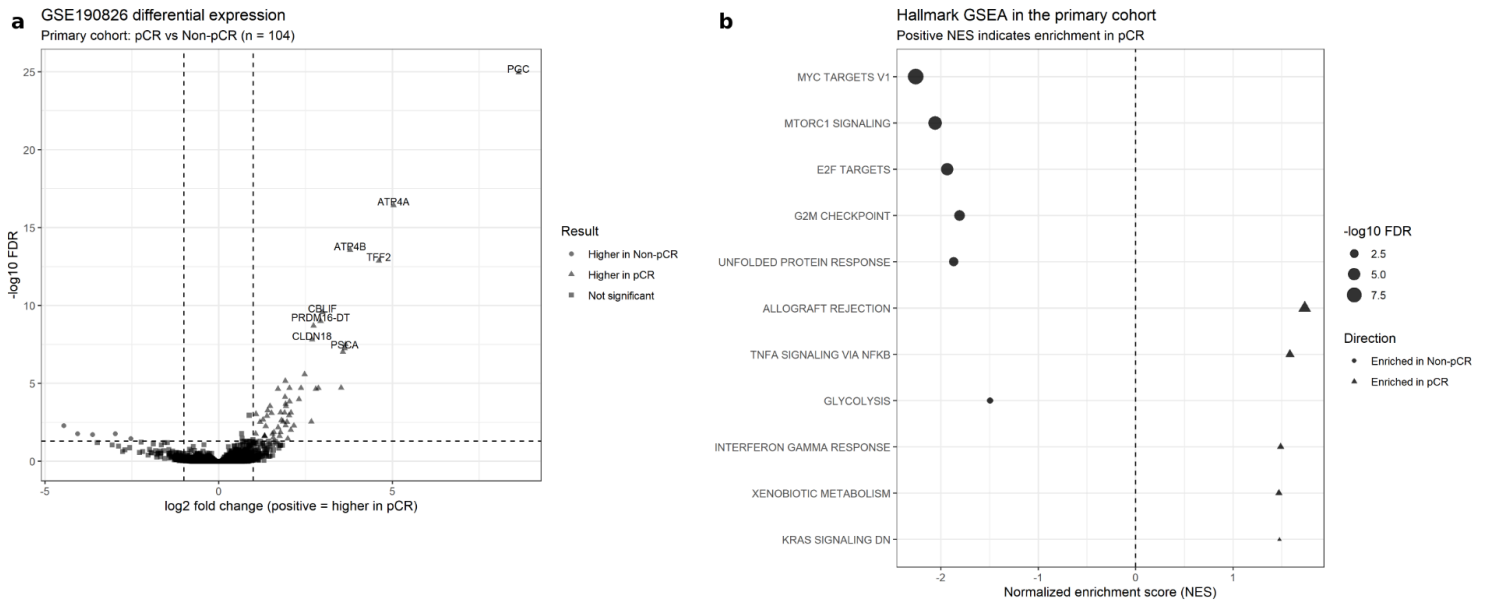


Figure 2

Discovery-cohort differential expression and Hallmark enrichment. (a) Volcano plot of pCR versus non-pCR differential expression in primary104; positive log2 fold changes indicate higher expression in pCR. (b) Hallmark pathways reaching FDR < 0.05 in primary104; positive NES indicates enrichment in pCR

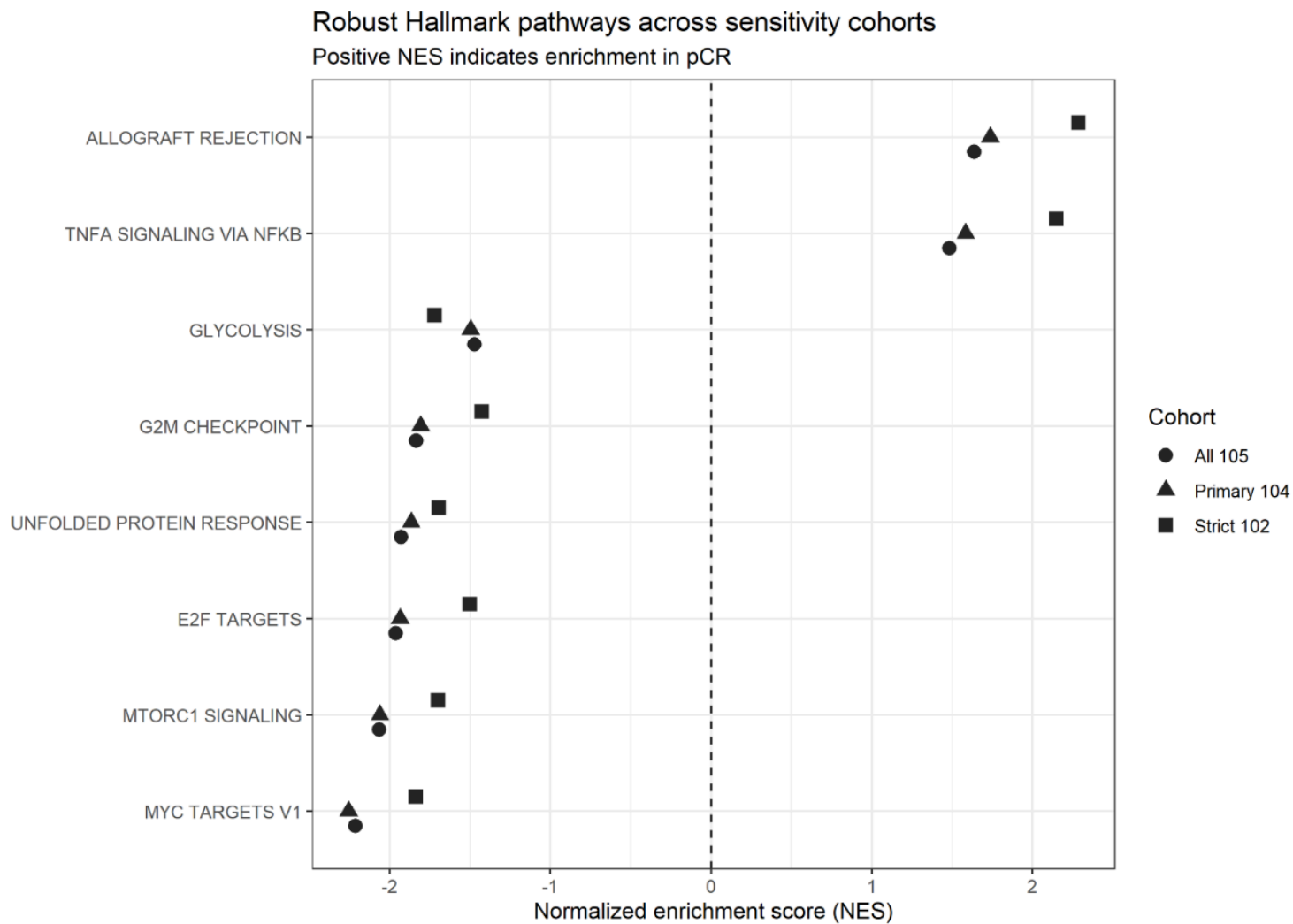


Figure 3

Internal robustness of the eight prespecified Hallmark pathways. NES values across all105, primary104, and strict102. All eight pathways remained FDR-significant with consistent directions. The three analyses contain overlapping samples and are sensitivity analyses, not independent cohorts

Cross-cohort evidence for 8 prespecified Hallmark pathways

Positive NES indicates enrichment in favorable responders; * FDR < 0.05

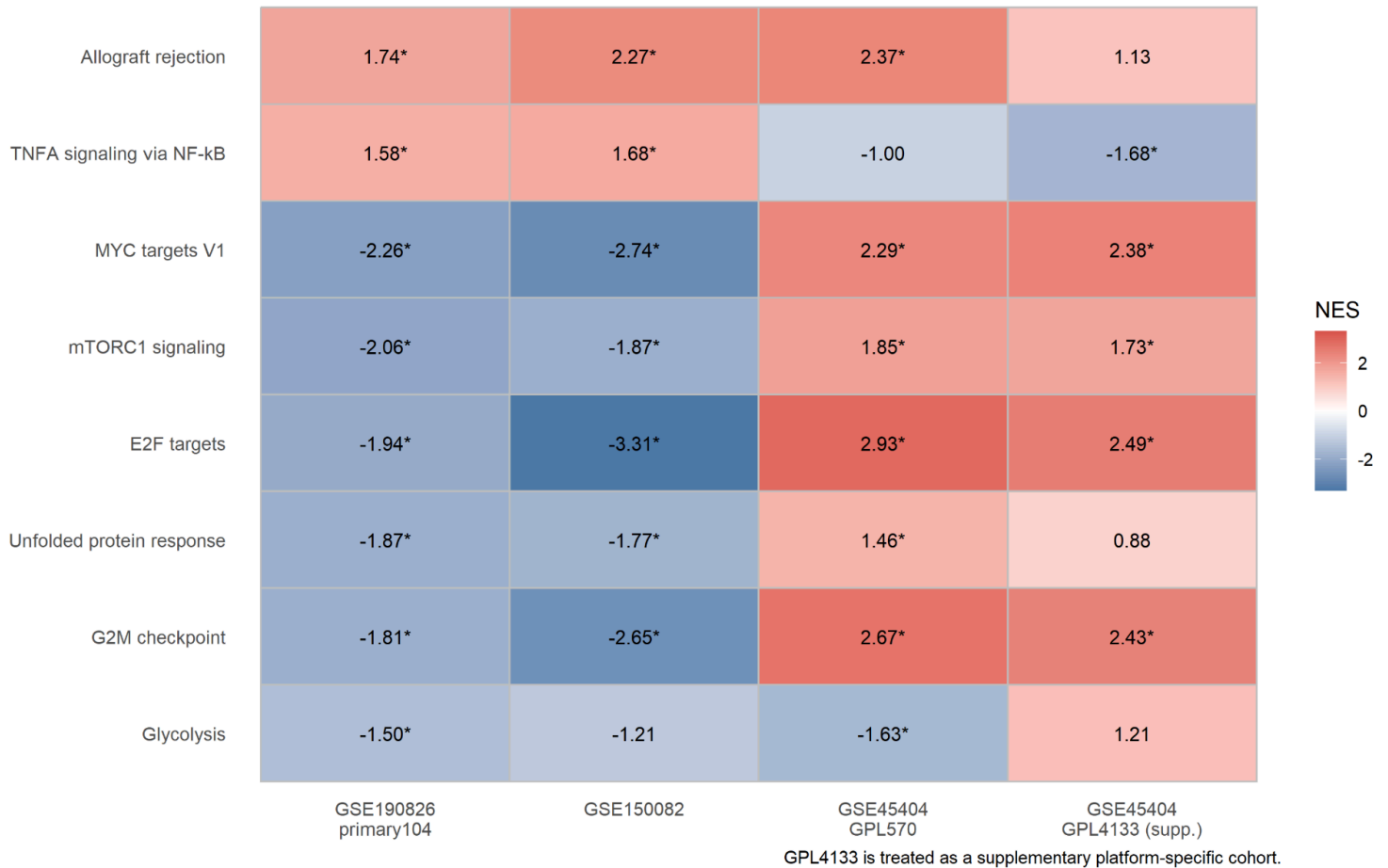
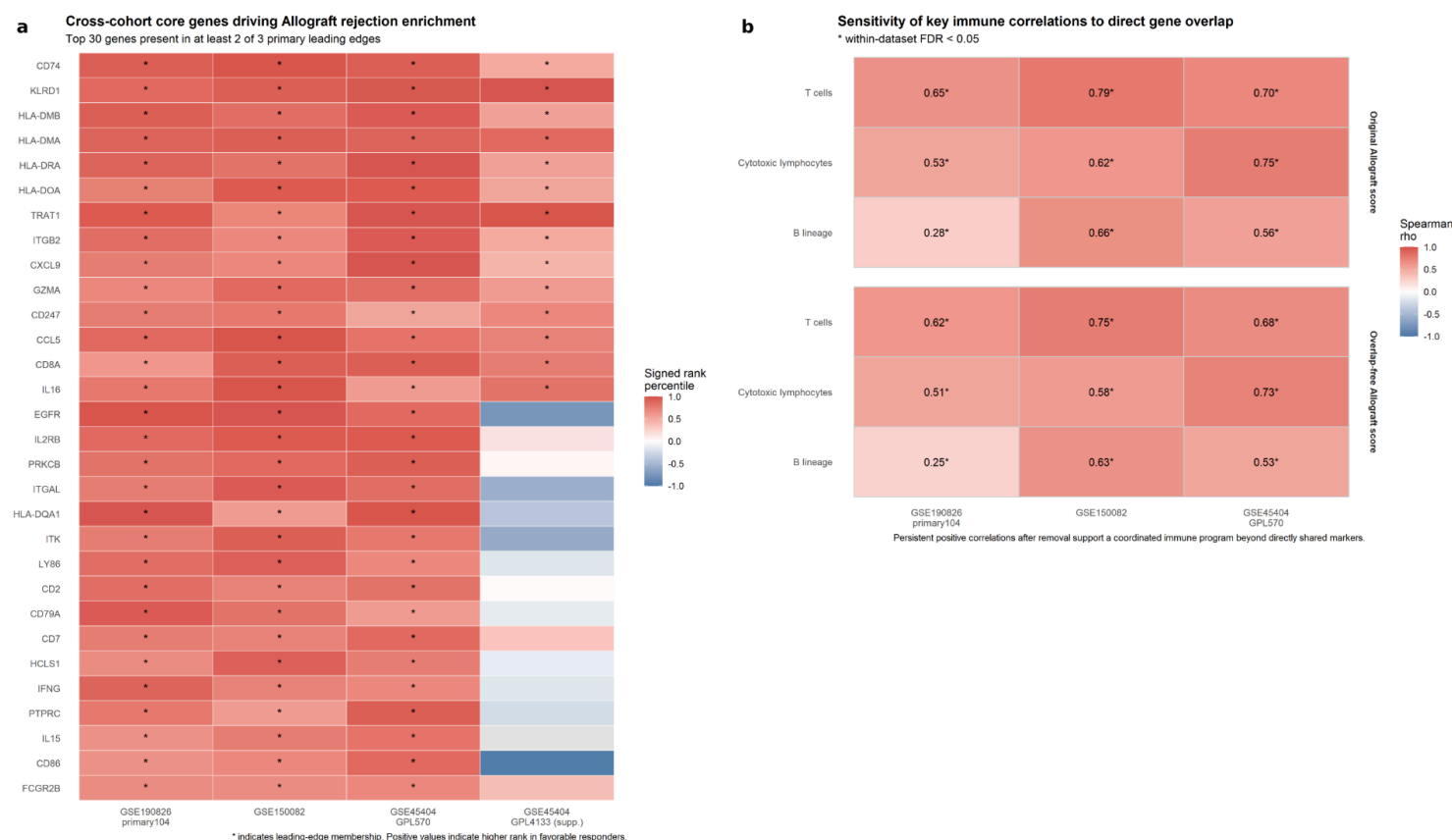


Figure 4

Cross-cohort evidence for the eight prespecified pathways. NES values across the discovery cohort, two primary external validations, and supplementary GPL4133. Positive values indicate enrichment in favorable responders. Asterisks indicate FDR < 0.05, calculated across all 50 Hallmark pathways in GSE190826 and across the eight prespecified hypotheses in each validation dataset



Supplementary Files

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

- [FigureS1discoveryPCA.png](#)
- [FigureS2gastricmarkerreview.png](#)
- [FigureS3GSE150082validation.png](#)
- [FigureS4GSE45404twoplatforms.png](#)
- [FigureS5primaryvalidationcomparison.png](#)
- [FigureS6Allograftleadingedgeoverlap.png](#)
- [FigureS7MCPresponseeffects.png](#)
- [FigureS8deoverlapallcellcorrelations.png](#)

- [FigureS9directgeneoverlapcounts.png](#)
- [FigureS10overlapfreescorerresponse.png](#)
- [SupplementaryInformation.docx](#)