Research Square

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

Cross-platform transportability of ligand–receptor communication scores for prognosis in lung adenocarcinoma: a systematic benchmark of three scoring constructs across four cohorts

Ning Zhou

AlistairZhou_Zinc@outlook.com

East China Normal University https://orcid.org/0009-0008-6178-0895

Article

Keywords:

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11587680/v1

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

Additional Declarations: There is **NO** Competing Interest.

Cross-platform transportability of ligand–receptor communication scores for prognosis in lung adenocarcinoma: a systematic benchmark of three scoring constructs across four cohorts

Authors: Ning Zhou[1]*

[1] School of Life Sciences, East China Normal University, Shanghai 200241, China

*Correspondence: Ning Zhou, AlistairZhou_Zinc@outlook.com

Running title: Transportability of LR communication scores in LUAD

Keywords: ligand–receptor; communication competence; bulk RNA-seq; prognostic model; cross-platform; lung adenocarcinoma

Abstract

Ligand–receptor communication scores derived from bulk tumour transcriptomes have been proposed as prognostic markers in lung adenocarcinoma, but whether they survive a platform change has never been benchmarked head-to-head. We quantified communication competence for 2,459 measurable CellChatDB v2 pairs in four cohorts (1,355 tumours, 529 deaths) using three scoring constructs — a z-based construct (Z-CP), a rank co-elevation construct invariant to monotone transformations (RCE) and a cell-type-resolved construct built on marker-anchored deconvolution (CTRC) — with nested cross-validation, formal equivalence testing (two one-sided tests, margin $\Delta C = 0.05$) and

a graded platform-shift simulation. In the discovery cohort the constructs performed comparably (cross-validated C-index 0.543–0.585); in the three external cohorts every construct landed in the same modest band (C-index 0.46–0.56). RCE was statistically equivalent to Z-CP in both adequately powered cohorts (TOST p = 0.011 and 0.015) and directionally superior in a third; CTRC lost discovery power and bought no transport benefit. Under simulated platform shift all constructs degraded smoothly and comparably, with no rescue regime. Bulk-derived ligand–receptor scores transport weakly and construct-independently; future benchmarks should report formal equivalence rather than post hoc superiority claims.

Introduction

Single-cell transcriptomics has made cell–cell communication inference routine, and ligand–receptor (LR) databases such as CellChatDB now anchor a family of methods that "project" LR interactions back onto bulk tumour transcriptomes, producing per-sample communication competence (CC) scores [1,2]. Several groups have reported prognostic LR signatures in lung adenocarcinoma (LUAD), the most common histological subtype of lung cancer [3–5].

There is a catch. Bulk-derived CC scores are sensitive to analytic choices that are essentially arbitrary: how expression is scaled (z-scores or ranks), how multi-subunit complexes are handled, and the standing assumption that a pair's expression "co-elevation" reflects genuine intercellular signalling rather than tumour purity or proliferation.

A second issue matters more and is examined less: **transportability**. Does a CC scoring rule — and the prognostic model fitted on it — keep its discrimination when the data come from a different measurement platform? Each construct on the market embodies a different hypothesis about why transport fails. Z-score constructs are cohort-calibrated by construction. Rank-based constructs are invariant to monotone transformations. Deconvolution-based constructs try to absorb platform effects into explicit cell-composition estimates. Yet no study has benchmarked them on identical cohorts with pre-specified statistical criteria. That is what we do here.

We benchmark in LUAD using the full CellChatDB v2 database (3,233 interactions), four public cohorts spanning RNA-seq and three microarray platforms (1,304 tumours), three scoring constructs, nested cross-validation, formal equivalence testing (TOST, margin $\Delta C = 0.05$) and a graded platform-shift simulation. Our aim is not to nominate another prognostic signature. It is to quantify — honestly and reproducibly — how much prognostic information LR-CC scores carry across platforms, and which scoring construct, if any, travels better.

Results

Four cohorts, 2,459 measurable LR pairs from the full CellChatDB v2

The four cohorts comprised 1,304 tumours with 508 deaths (Table 1; Fig 1). After intersecting genes across RNA-seq and three microarray platforms (10,391 shared genes), 2,459 of 3,233 CellChatDB v2 interactions were measurable, and the pre-registered evaluability rule kept 2,297 pairs (93.4%) for survival screening. Marker-anchored deconvolution returned biologically plausible mean compositions everywhere (epithelial

0.20–0.23, fibroblast 0.12–0.17, T/NK 0.12–0.13) — the bulk matrices do support compartment resolution (Fig 5A).

Discovery: a laminin–integrin axis under both expression constructs; deconvolution costs power

In TCGA-LUAD, Z-CP nominated 503 pairs at nominal $p < 0.05$, 111 at BH $q < 0.05$ and 35 at maxT $q < 0.05$ (minimum $p = 1.2\times10^{-11}$; permutation null 95th percentile $|z| = 4.67$). RCE ran almost identically (505/88/27). CTRC lost an order of magnitude of power (130/0/6; minimum $p = 6.5\times10^{-4}$) — the price of the variance that deconvolution adds (Table 2; Fig 2).

The models themselves tell a coherent story. The four-pair Z-CP model selected a laminin–integrin axis (LAMA3–ITGA2/ITGB1, ANGPTL4–ITGA5/ITGB1, LAMA3–ITGA6/ITGB1, LAMB3–ITGA6/ITGB1; apparent C = 0.623); the RCE model independently picked laminin–integrin pairs (LAMC2/LAMA3–ITGA6, apparent C = 0.613); the CTRC model went elsewhere, to desmosomal and neuroendocrine pairs (DSC1–DSG3, CRH–CRHR2; apparent C = 0.609). After nested cross-validation, apparent optimism was 0.039 (Z-CP; CV C = 0.585 ± 0.014), 0.035 (RCE; 0.579 ± 0.013) and 0.066 (CTRC; 0.543 ± 0.030). Two mechanistically different constructs converging on the same biology is reassuring. The CTRC penalty is not.

Transport: modest external discrimination for every construct

Transported to the three external cohorts, the Z-CP model achieved C = 0.483 (95% CI 0.387–0.589) in GSE31210, 0.553 (0.511–0.594) in GSE68465 and 0.593 (0.514–0.674) in GSE50081; RCE achieved 0.528, 0.558 and 0.569, and CTRC 0.462, 0.525 and 0.521

(Table 3; Fig 3, Fig 6). Raw coefficients versus cohort-wise z-standardisation changed C by ≤ 0.008 in every cohort–construct combination; we report the cohort-wise z-standardised scaling throughout (Table 3 lists these values). These external C-indices (0.46–0.59) are what you get when cross-validated discovery performance (0.585) is discounted by platform change and case mix. No construct reached conventional "strong" discrimination in any external cohort.

Equivalence testing: RCE is statistically equivalent to Z-CP where powered; CTRC is not separable from Z-CP

Paired-bootstrap TOST at margin ΔC = 0.05 declared RCE equivalent to Z-CP in TCGA (ΔC = −0.010, TOST p = 0.011, 90% CI −0.039 to +0.018) and GSE68465 (ΔC = +0.005, p = 0.015, 90% CI −0.029 to +0.039). GSE31210 (226 tumours, 35 events) was underpowered for a verdict (ΔC = +0.045, p = 0.462, 90% CI −0.042 to +0.132); the direction favours RCE but equivalence is unprovable here. In GSE50081 the direction mildly favoured Z-CP (ΔC = −0.025, p = 0.207) without reaching either conclusion. CTRC was inconclusive in all four cohorts (ΔC from −0.072 to −0.020; p = 0.138–0.657) (Table 4).

So the defensible claims are narrow. Rank-based scoring preserves prognostic content — equivalent in two of two adequately powered cohorts, directionally positive in a third. Cell-type-resolved scoring adds no detectable transport benefit. Beyond that, the data do not let us go.

Platform-shift simulation: smooth, construct-independent degradation with no rescue regime

Unperturbed baselines reproduced the observed C-indices exactly (Z-CP 0.627, RCE 0.617, CTRC 0.607), confirming the corrected implementation (Fig 4). Under graded distortion, discrimination eroded smoothly with intensity ε: by ε = 1.0–1.2 all three constructs had lost ≈ 0.03–0.06 C-index units, with maximal losses of 0.095 (Z-CP), 0.080 (RCE) and 0.067 (CTRC). The 2-D phase diagram (ε × σ_bias) contains no rescue regime — no corner where any construct keeps baseline performance under high distortion — and degradation is dominated by ε rather than by the bias-to-offset ratio. CTRC degraded least in absolute terms but started lowest, so its relative position never improved. Mean ΔC across the phase plane: −0.012 (Z-CP), −0.011 (RCE), −0.010 (CTRC). Platform shift imposes a common ceiling on LR-score transportability, and none of the tested constructs circumvents it.

Discussion

This benchmark yields three findings with direct methodological consequences. **First**, bulk-derived LR communication scores carry real but modest prognostic information in LUAD (cross-validated C ≈ 0.58, external C ≈ 0.46–0.59), and the signal is construct-independent: two mechanistically different constructions — geometric means of z-scores and of within-cohort ranks — converged on the same laminin–integrin biology and on statistically equivalent discrimination. **Second**, formal equivalence testing changes the interpretive frame. The literature is full of "our construct outperformed" claims built on point estimates alone; TOST shows that the honest claim for RCE is *equivalence* where powered, with a positive direction where it is not, and for CTRC *indistinguishability* at available sample sizes. **Third**, platform distortion degrades all constructs smoothly and

comparably, with no perturbation regime in which cell-type resolution or rank invariance comes to the rescue.

These results temper expectations for bulk-derived LR signatures. The laminin–integrin axis selected here is biologically plausible in LUAD — basement-membrane remodelling, invasive phenotype — and it has surfaced in earlier prognostic studies. Its discrimination ceiling, though, is low. Claims that bulk CC scores can approach single-cell-informed models should be read with that ceiling in mind.

Limitations. (i) The deconvolution anchors are a fixed seven-compartment marker set; finer-grained CTRC variants (e.g., single-cell-signature projection [14]) may behave differently, although the power loss observed here is structural — variance inflation from solving P and H — rather than marker-specific. (ii) External cohorts are microarray-based with heterogeneous event counts; GSE31210 (35 events) limits equivalence testing, which we report rather than hide. (iii) The simulation emulates multiplicative, additive, compressive and quantisation distortions but not batch-specific sample handling. (iv) We benchmarked OS only; disease-free survival endpoints may interact differently with tumour-purity-driven scores.

Conclusions. Across 1,304 LUAD tumours and 2,459 CellChatDB v2 interactions, ligand–receptor communication scores transport across platforms weakly and construct-independently: rank-based scoring neither rescues nor harms transportability, and deconvolution-based scoring costs discovery power without a transport benefit. Studies proposing bulk LR prognostic signatures should pre-specify equivalence margins and report transport performance, not just discovery performance.

Methods

Study design and cohorts

We analysed four independent LUAD cohorts with overall survival (OS) follow-up. Discovery: TCGA-LUAD (RNA-seq, HiSeqV2 log2(norm_count+1); n=506 tumours, 183 deaths). External validation: three GEO series-matrix cohorts — GSE31210 (Affymetrix GPL570, n=226, 35 events), GSE68465 (GPL96/HG-U133A, n=442, 236 events; follow-up in months, converted to days by ×30.44; of 443 samples with vital status, one with invalid survival time was excluded) and GSE50081 (GPL570, n=130, 54 events; follow-up in years, converted to days by ×365.25) [6]. GSE50081 is a mixed-histology NSCLC series; to keep all four cohorts adenocarcinoma-only, we retained the 130 samples annotated as adenocarcinoma (including adenosquamous and NSCLC-favour-adenocarcinoma) out of 181. For GSE31210, no sample in the original metadata was hit by the pre-registered exclusion flag (incomplete resection or adjuvant therapy); all 226 tumours had complete survival data and were analysed. Cohort characteristics: Table 1.

Cohort substitution. The originally planned validation cohort GSE72094 [7] was replaced by GSE31210. GSE72094 runs on a custom Rosetta/Merck HuRSTA_2a520709.CDF array whose probe identifiers (merck-\<EST accession\>_at) had no public gene annotation usable for consistent mapping at the time of analysis: bridging through GenBank accessions recovered only ~13% of probes, and batch queries to mygene.info [16] returned no usable hits. GSE31210 preserves the intended design (East-Asian resected stage I–II LUAD) and keeps external validation at 798 tumours.

Probe annotation. Probe-to-gene maps for GPL96 and GPL570 were built by bridging Bioconductor platform packages (probe → GenBank accession via the chip `accessions` table) with org.Hs.eg.db (accession → Entrez ID → HGNC symbol), recovering 15,283 (GPL96) and 24,281 (GPL570) annotated probes. Multiple probes per symbol were collapsed to the probe with highest mean expression. The four cohorts were intersected to 10,391 shared genes; expression was log2-transformed where raw intensities indicated a linear scale (max > 50).

Ligand–receptor interaction database

The full CellChatDB v2 human database (3,233 interactions: 1,920 carried over from v1 and 1,313 v2-specific; Secreted Signaling 1,280, ECM-Receptor 424, Cell–Cell Contact 535, Non-protein Signaling 994 in the source file) was obtained from the CellChat repository [1,2]. An interaction counted as *measurable* in a cohort when every ligand and receptor subunit mapped to a gene in the 10,391-gene intersection; this yielded 2,459 measurable pairs (Secreted 896, Non-protein 903, ECM-Receptor 360, Cell–Cell Contact 300). Multi-subunit weights were set to 1/(number of subunits). A pre-registered evaluability rule (SD > 0 and detection rate ≥ 10% in the discovery cohort) retained 2,297 pairs for survival screening.

Communication-competence (CC) scoring constructs

For each sample *i* and LR pair *m*:

1. **Z-CP** (z-based construct): sample-wise z-scores of gene expression, positive part $z^+ = \max(z, 0)$; score = $\sqrt{(z^+_L \cdot z^+_R)}$, multi-subunit contributions averaged by weights.

2. **RCE** (rank co-elevation): within-cohort percentile ranks u = 2·rank_pct − 1 clipped at 0; score = √(u_L · u_R). RCE is invariant to any monotone transformation of expression — a direct answer to the cross-platform calibration failure mode of Z-CP.

3. **CTRC** (cell-type-resolved communication): seven compartments (Epithelial, Fibroblast, Endothelial, Myeloid, T/NK, B/Plasma, Mast) deconvolved per sample by non-negative least squares (NNLS) [13] against a marker-gene anchor matrix (14/12/10/11/11/9/6 markers; compartment omitted if < 4 markers present); compartment-level expression H solved per gene by NNLS on the estimated proportions P; score = Σ_ab P_ia P_ib √(H_aL · H_bR) over all compartment pairs.

Statistical analysis

Discovery (TCGA-LUAD). Each evaluable pair's association with OS was screened by the Breslow score test; p-values were corrected by Westfall–Young maxT permutation (1,000 permutations) [10] and Benjamini–Hochberg FDR [11]. The top k=4 pairs by |z| went into a multivariable Cox model (Efron ties [9]); performance was quantified by Harrell's C-index [8] with optimism corrected by nested cross-validation (5-fold × 20 repeats, pairs re-selected inside each training fold).

Transport and equivalence testing. Each discovery model was applied to the three external cohorts under two scalings (raw coefficients; cohort-wise z-standardised scores). Discrimination: C-index with 500-bootstrap 95% CIs. For each external cohort and each alternative construct, ΔC (alternative − Z-CP) was estimated with a paired bootstrap (2,000 resamples, identical resampling indices for both constructs), and equivalence at

margin $\Delta C = 0.05$ was assessed by TOST [12] — equivalence declared when the 90% CI of ΔC sits entirely within $(-0.05, +0.05)$.

Platform-shift simulation. On the discovery cohort we applied a graded distortion operator emulating cross-platform differences: (i) gene-specific multiplicative bias (σ_bias), (ii) gene-specific additive offset, (iii) dynamic-range compression (exponent $1 - 0.6\varepsilon$), (iv) quantisation (step $0.05 + 0.35\varepsilon$); overall intensity $\varepsilon \in [0, 1.2]$ scaled all components linearly. A 2-D phase diagram ($\varepsilon \times \sigma_bias \in [0.1, 0.8]$) maps each construct's ΔC against its unperturbed baseline. Deconvolution in every simulated condition used the full 10,391-gene matrix, so marker genes were never lost to subsetting.

Software. Python 3.14 (NumPy/SciPy/pandas). Cox models (Efron ties), score tests, maxT permutations and TOST were implemented directly in NumPy and validated against published benchmark results (MKI67 HR = 1.24, $P = 1.8\times10^{-4}$ in TCGA-LUAD, concordant with the literature).

Ethics approval

All data are from public repositories (TCGA, GEO); no new human-subject data were generated.

Declarations

Ethics approval and consent to participate: Not applicable (public data).

Consent for publication: Not applicable.

Availability of data and materials: TCGA-LUAD: GDC/Xena [15] (HiSeqV2, LUAD_clinicalMatrix). GEO: GSE31210, GSE68465, GSE50081 (series-matrix files). CellChatDB v2: CellChat repository (CellChatDB.human.rda, v2 flag). All datasets are publicly available. Code: Python 3.14; scripts s1–s7 (cohort construction, CC scoring, discovery, transport/TOST, simulation, figures, tables) accompany this manuscript. All analysis code and derived result tables are deposited in Zenodo (doi:10.5281/zenodo.23836993).

Competing interests: The author declares no competing interests.

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Authors' contributions: Ning Zhou: conceptualisation, methodology, software, formal analysis, data curation, validation, visualisation, writing — original draft, writing — review and editing. The author read and approved the final manuscript.

Acknowledgements: The author thanks the TCGA Research Network, the Gene Expression Omnibus contributing studies (GSE31210, GSE68465, GSE50081) and the CellChat developers for making data and annotation resources publicly available.

References (Vancouver)

1. Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021;12:1088. doi:10.1038/s41467-021-21246-9.

2. Jin S, Plikus MV, Nie Q. CellChat for systematic analysis of cell-cell communication from single-cell transcriptomics. Nat Protoc. 2025;20:180–219. doi:10.1038/s41596-024-01045-4.

3. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511:543–50. doi:10.1038/nature13385.

4. Okayama H, Kohno T, Ishii Y, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. Cancer Res. 2012;72(1):100–11. doi:10.1158/0008-5472.CAN-11-1403.

5. Shedden K, Taylor JM, Misek DE, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. Nat Med. 2008;14(8):822–7. doi:10.1038/nm.1790.

6. Der SD, Sykes J, Pintilie M, et al. Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. J Thorac Oncol. 2014;9(1):59–64. doi:10.1097/JTO.0000000000000042.

7. Schabath MB, Welsh EA, Fulp WJ, et al. Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. Oncogene. 2016;35(24):3209–16. doi:10.1038/onc.2015.375.

8. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15:361–87. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

9. Efron B. Logistic regression, survival analysis, and the Kaplan–Meier curve. J Am Stat Assoc. 1988;83:414–25. doi:10.1080/01621459.1988.10478612.

10. Westfall PH, Young SS. Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment. New York: Wiley; 1993.

11. 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. doi:10.1111/j.2517-6161.1995.tb02031.x.

12. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J Pharmacokinet Biopharm. 1987;15:657–80. doi:10.1007/BF01068419.

13. Lawson CL, Hanson RJ. Solving Least Squares Problems. Englewood Cliffs: Prentice-Hall; 1974.

14. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12:453–7. doi:10.1038/nmeth.3337.

15. Grossman RL, Heath AP, Ferretti V, et al. Toward a shared vision for cancer genomic data. N Engl J Med. 2016;375:1109–12. doi:10.1056/NEJMp1607591.

16. Xin J, Mark A, Afrasiabi C, et al. High-performance web services for querying gene and variant annotation. Genome Biol. 2016;17(1):91. doi:10.1186/s13059-016-0953-9.

Figure legends

Figure 1. Study design. Four LUAD cohorts (TCGA-LUAD discovery; GSE31210/GSE68465/GSE50081 external validation), full CellChatDB v2 projection, three CC scoring constructs, discovery → transport → equivalence → simulation workflow.

Figure 2. Discovery landscape in TCGA-LUAD. (A) Number of LR pairs passing each screening threshold per construct (nominal $P < 0.05$, BH-FDR $q < 0.05$, maxT $P < 0.05$); (B) effect-size landscape: log2 hazard ratio versus nominal Cox score-test P for the evaluable pairs of each construct.

Figure 3. Transport performance and equivalence. External-cohort C-indices (500-bootstrap CIs) for the three constructs; paired-bootstrap ΔC distributions with TOST 90% CIs against the ±0.05 margin.

Figure 4. Platform-shift phase diagram. (A) Degradation curves of C-index vs. perturbation intensity ε on shifted discovery data; (B–C) phase maps of ΔC over ε × σ_bias for Z-CP and RCE (CTRC analogous), showing smooth, construct-independent degradation with no rescue regime.

Figure 5. Deconvolution composition and top pairs. (A) Mean estimated compartment proportions per cohort; (B) top-ranked CTRC pairs in the discovery cohort.

Figure 6. Transport summary. C-indices across all cohorts and constructs, illustrating the 0.46–0.59 external range.

Tables

- **Table 1** (results/tables/Table1_cohorts.csv): cohort, platform, n, deaths, median OS.

- **Table 2** (results/tables/Table2_discovery.csv): screening counts per construct (nominal / BH / maxT), apparent and CV C.
- **Table 3** (results/tables/Table3_transport.csv): per-cohort C-indices with 95% CIs.
- **Table 4** (results/tables/Table4_TOST.csv): ΔC, bootstrap SE, TOST p, 90% CI, equivalence verdict.

Table 1. Cohorts

Cohort	Platform	n	deaths	median_OS_m	OS_	months_25	stage_I	stage_II	stage_III	stage_IV
TCGA-LUAD	GA (HiSeqV2, lo	506	183	653.000		13.7–37.0	271.000	120.000	81.000	26.000
GSE31210	70 (HG-U133 P	226	35	1744.500		40.9–67.3				
GSE68465	PL96 (HG-U13	442	236	1430.680		24.0–73.7				
GSE50081	70 (HG-U133 P	130	54	1590.664		22.9–68.5				

Table 2. Discovery screening in TCGA-LUAD

Construct	pairs_evaluate	nominal	BH_q05	maxT_p05	apparent_C	nested_CV_C	nested_CV_sd
ZCP	2297	503	111	35	0.623	0.585	0.014
RCE	2297	505	88	27	0.613	0.579	0.013
CTRC	2297	130	0	6	0.609	0.543	0.030

Table 3. External transport C-indices

cohort	construct	n	deaths	C	CI_lo	CI_hi
GSE31210	CTRC	226	35	0.462	0.368	0.546
GSE31210	RCE	226	35	0.528	0.427	0.624
GSE31210	ZCP	226	35	0.483	0.387	0.589
GSE50081	CTRC	130	54	0.521	0.441	0.609
GSE50081	RCE	130	54	0.569	0.480	0.648
GSE50081	ZCP	130	54	0.593	0.514	0.674
GSE68465	CTRC	442	236	0.525	0.487	0.568
GSE68465	RCE	442	236	0.558	0.519	0.598
GSE68465	ZCP	442	236	0.553	0.511	0.594
TCGA-LUAD	CTRC	506	183	0.607	0.559	0.648
TCGA-LUAD	RCE	506	183	0.617	0.566	0.663
TCGA-LUAD	ZCP	506	183	0.627	0.580	0.677

Table 4. TOST equivalence tests

cohort	contrast	C_ref	C_alt	delta	se	p_tost	ci90_lo	ci90_hi	equivalent	n
TCGA-LUAD	RCE vs ZCP	0.627	0.617	-0.010	0.017	0.011	-0.039	0.018	True	506
TCGA-LUAD	CTRC vs ZCI	0.627	0.607	-0.020	0.027	0.138	-0.065	0.025	False	506
GSE31210	RCE vs ZCP	0.483	0.528	0.045	0.053	0.462	-0.042	0.132	False	226
GSE31210	CTRC vs ZCI	0.483	0.462	-0.021	0.061	0.315	-0.121	0.079	False	226
GSE68465	RCE vs ZCP	0.553	0.558	0.005	0.021	0.015	-0.029	0.039	True	442
GSE68465	CTRC vs ZCI	0.553	0.525	-0.027	0.030	0.222	-0.076	0.021	False	442
GSE50081	RCE vs ZCP	0.593	0.569	-0.025	0.031	0.207	-0.076	0.026	False	130
GSE50081	CTRC vs ZCI	0.593	0.521	-0.072	0.055	0.657	-0.162	0.018	False	130

Figure 1. Study design

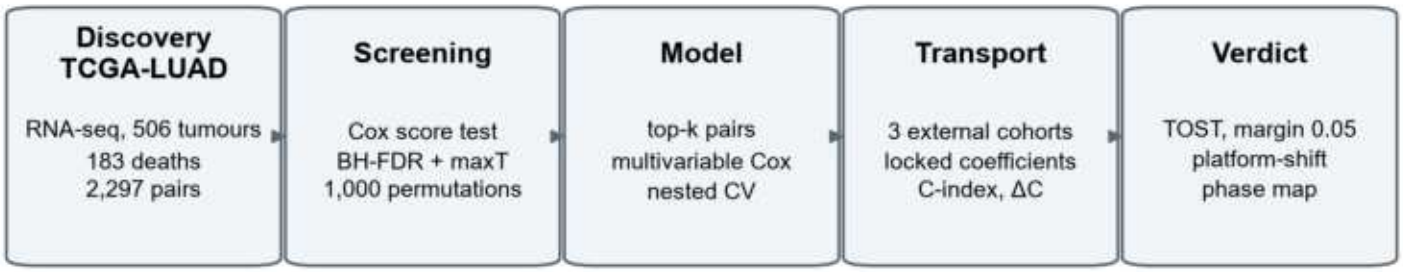


Figure 2. Discovery landscape in TCGA-LUAD

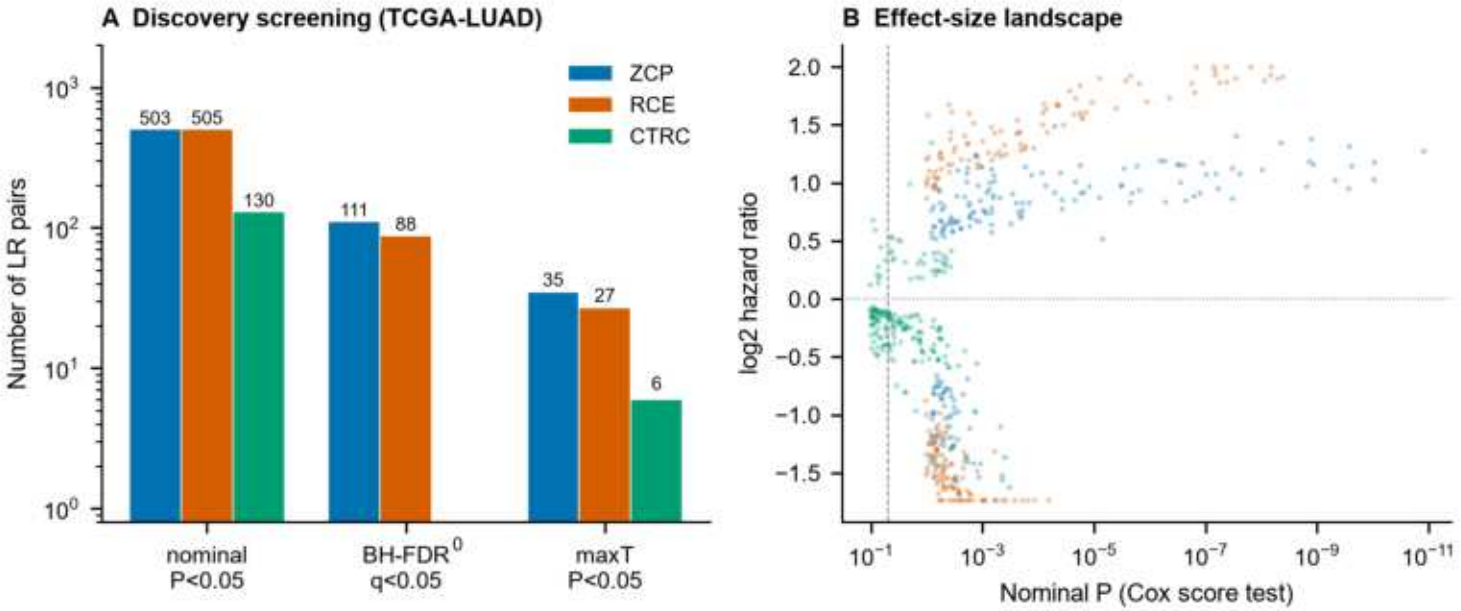


Figure 3. Transport performance and equivalence

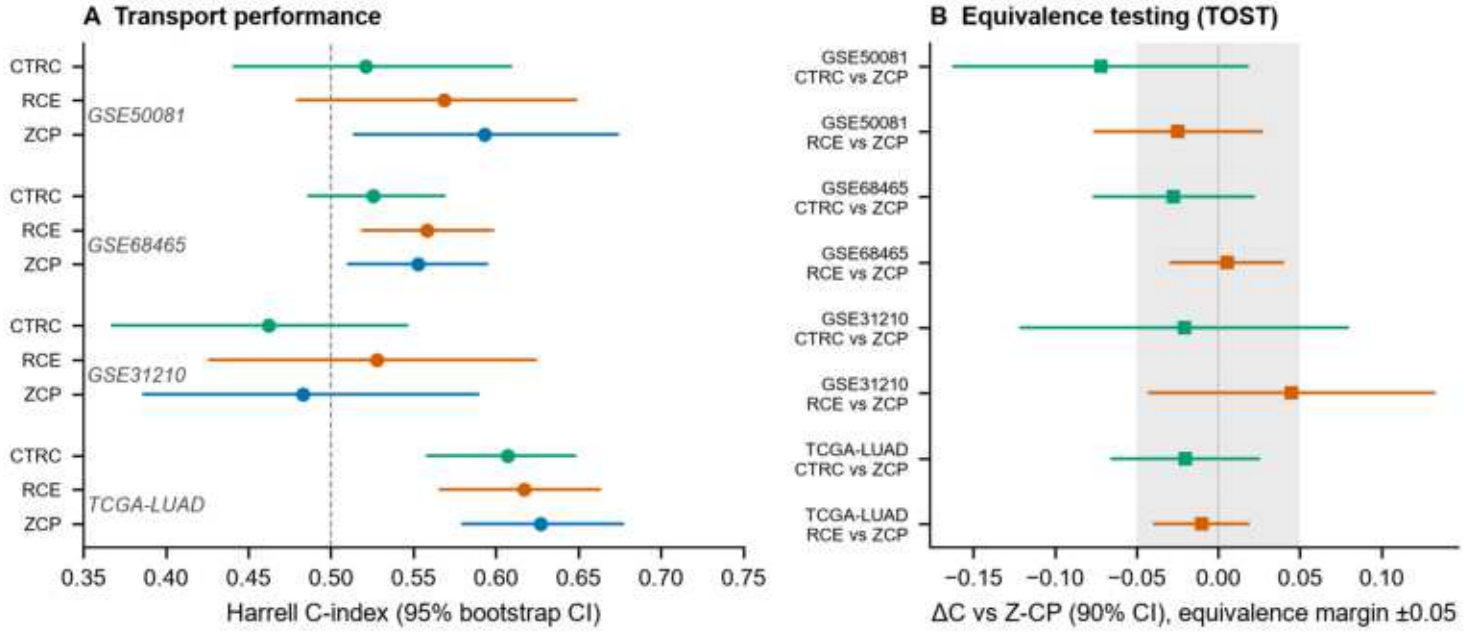


Figure 4. Platform-shift phase diagram

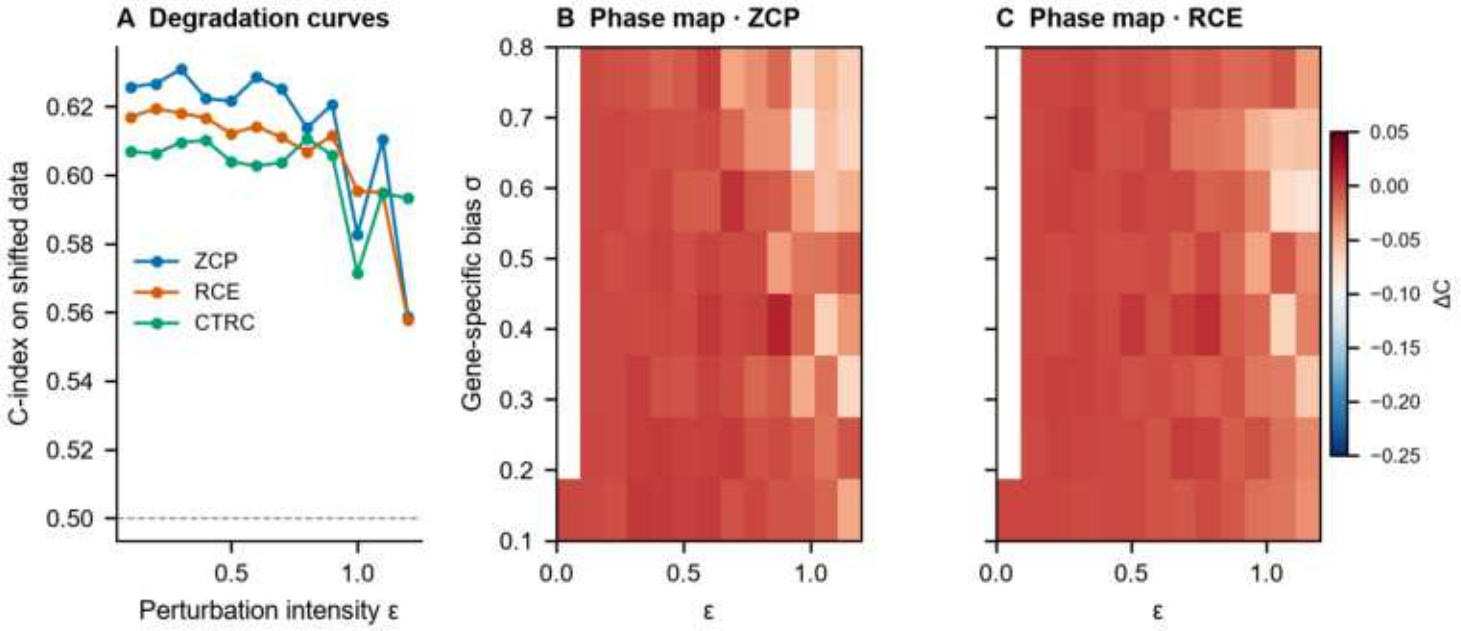


Figure 5. Deconvolution composition and top pairs

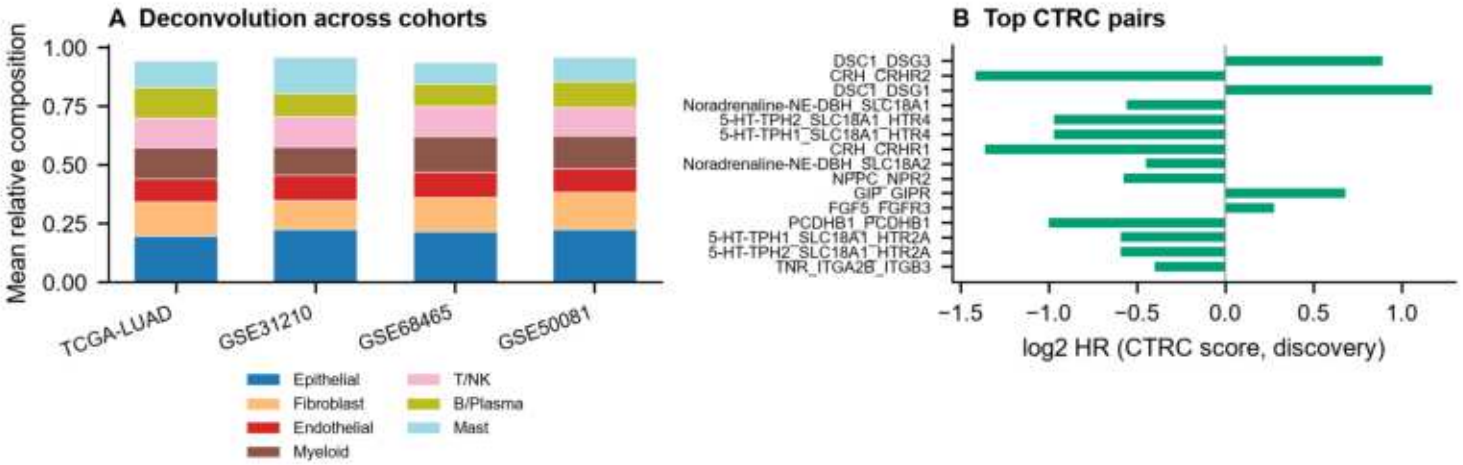


Figure 6. Transport summary

