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.

Longitudinal multicohort analysis of tissue composition and continuous malignant states in recurrent glioblastoma

Pengran Yu[1]
Shanghai Ninth People's Hospital
Kaihua Cao[1]
Shanghai Ninth People's Hospital
Zhihua Cheng[1]
Shanghai Ninth People's Hospital
Yu Zeng[1]
Shanghai Ninth People's Hospital
WenChuan Zhang[1]
zhangwench88@sjtu.edu.cn
Shanghai Ninth People's Hospital

Research Article

Keywords: glioblastoma, recurrence, deconvolution, malignant cell states, Geneformer

Posted Date: November 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11192692/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

Bulk expression changes in recurrent glioblastoma can arise from tissue composition or malignant-cell variation. We sought to distinguish these sources and evaluate the external validity of virtual perturbation.

Methods

This retrospective analysis integrated 84 paired patients from the Glioma Longitudinal Analysis Consortium (GLASS), 28 single-nucleus discovery pairs and 43 malignant-eligible pairs from an external single-nucleus cohort. Brain-reference deconvolution was benchmarked in 104 pseudobulks from 52 held-out donors. Patient-paired tests, frozen external hypotheses, baseline-adjusted continuous-state models and nine public KAT5-knockout cell models were evaluated separately.

Results

GLASS identified 1,432 differentially expressed genes. Calibrated estimates supported increased oligodendrocyte RNA weight (+ 8.25 percentage points; false discovery rate 0.00054) and decreased vascular weight (− 6.18 points; 0.0036), with concordant directions from a second method. Malignant RNA weights failed calibration. Expanded malignant programs showed no corrected average recurrence shift. Exploratory analysis identified 293 concordant gene-axis associations in the external cohort, reduced to 33 after baseline-state adjustment in both cohorts. None of 36 previously frozen gene-state hypotheses replicated. Geneformer discrimination declined from an internal area under the receiver operating characteristic curve of 0.696 to an external value of 0.486. Virtual KAT5 deletion did not outperform matched deletion controls against real knockout (P = 0.488).

Conclusions

Recurrent specimens show selected tissue-composition changes alongside heterogeneous malignant-state variation. The continuous-state associations remain exploratory, and the tested prediction and perturbation endpoints do not establish a therapeutic target.

Introduction

Radiotherapy with temozolomide established an important treatment framework for glioblastoma, with O6-methylguanine-DNA methyltransferase (MGMT) promoter methylation informing treatment-associated outcome [1, 2]. Nevertheless, recurrence remains a central biological problem. Longitudinal glioma studies have shown that recurrence reflects genetic evolution, cellular plasticity, and interactions with non-malignant tissue [3, 4]. A bulk expression difference can therefore arise from altered malignant states, different proportions of non-malignant cells, regional sampling, or several processes together. This distinction matters when a recurrence-associated gene list is used to propose a treatment: an association with the recurrent specimen does not by itself identify a dependency of recurrent malignant cells.

Bulk subtype studies and their longitudinal extensions describe variation at the tissue level [5, 6]. Single-cell profiling demonstrates heterogeneity within individual glioblastomas [7], and glioblastoma stem-cell studies connect

developmental programs with plasticity [8]. Single-cell studies provide a framework for describing astrocyte-like, oligodendrocyte-progenitor-like, neural-progenitor-like, and mesenchymal-like malignant states [9]. However, recovering these states in a dataset does not establish that all patients move in the same direction under therapy. Likewise, a large number of nuclei cannot compensate statistically for a small number of independent patients. A recurrence analysis must preserve longitudinal pairing and distinguish a change within malignant cells from a change in the composition of the sampled tissue.

Geneformer represents transcriptomes using a pretrained transformer and can support context-dependent virtual gene perturbation [10, 11, 12]. Its use for state reversal introduces an additional assumption: the reference representation must meaningfully distinguish the observed origin and desired goal. If that assumption fails in held-out patients, a precise perturbation statistic may characterize the geometry of an unsuitable reference rather than a therapeutically useful transition. Network-based virtual knockout provides complementary information about regulatory disruption [13], but agreement between computational outputs is not equivalent to an independent biological intervention.

We examined these assumptions in a public-data study of recurrent glioblastoma. A longitudinal bulk program was frozen in Glioma Longitudinal Analysis Consortium (GLASS) before external analyses. We then tested its malignant-cell localization, state-reference separability, spatial context, prognostic association, and relationship with experimentally measured cell-line dependencies and drug response. This design was intended to retain both corroborating and discordant evidence, with an explicit boundary between descriptive computational ranking and therapeutic prioritization.

Recent integrative work expands malignant programs beyond the conventional four states to include glial-progenitor, neuronal, and cilia-associated variation [14]. These signatures allow direct testing of whether a negative four-state comparison reflects an incomplete reference. Brain-reference deconvolution additionally requires empirical calibration because reference choice and platform differences can bias estimated composition [15, 16, 17, 18, 19]. The expanded analysis incorporates the external CARE cohort (GSE274546) [20] and a published KAT5 intervention dataset [21]. It evaluates sample recovery, statistical-model sensitivity, external prediction, and real-intervention alignment as distinct questions. Post-validation exploratory analyses are reported separately from hypotheses frozen before external gene testing.

Methods

Study design and cohort selection

This retrospective observational study reanalyzed de-identified public records. The exposure contrast was the within-patient change from primary diagnosis to first recurrence; there was no assigned intervention. GLASS was the longitudinal bulk discovery source; GSE174554 and its related GSE138794 matrices formed the reconciled single-nucleus discovery series. CARE GSE274546 served as the separately deposited external human cohort. Original authors' metadata, clinical timepoints, sample names, gene identifiers, and file checksums were reconciled before comparisons. CARE counts were checked against all 121 author sample files and annotations. Recovery of a missing discovery sample was not counted as an independent cohort. Only primary-to-first-recurrence comparisons entered the main human longitudinal analysis. Patient pairs, not cells, regions, spots, or technical libraries, supplied biological replication.

Bulk and single-nucleus analysis

The strict GLASS analysis retained 84 adult IDH-wildtype grade IV patients with documented intervening radiotherapy or temozolomide. Log2(transcripts per million [TPM] + 1) expression for 17,923 genes was modeled with patient-blocked limma-trend, and the 50 qualifying genes in each direction were frozen. Original inclusion, sensitivity, and external bulk analyses were retained [3, 22].

Raw unique molecular identifier (UMI) counts, author tumor labels, and documented quality control (QC) rules were used for discovery and CARE. Recovery scripts preserved annotation provenance and reused original QC cells when only labels were missing. Discovery quality control used sample-specific lower gene and UMI thresholds with floors of 200 genes and 300 UMIs; the mitochondrial threshold was bounded between 5% and 20%. Scrublet was attempted when at least 200 cells passed initial filters, with expected doublet rate 0.06 and seed 20260913; estimated doublets were excluded. Per-sample thresholds and failures are recorded in the quality-control tables. CARE retained author-QC nuclei. A 21,810-symbol assay intersection standardized log1p(CP10k) module scoring; discovery-derived control genes were fixed before CARE scoring. Four malignant labels were assigned by the maximal harmonized Neftel-like score [9]. Author-defined malignancy and a stricter copy-number variation (CNV) sensitivity remained distinct. Pseudobulk counts were summed within patient, specimen, and compartment. Main thresholds were 50 malignant cells per sample, or 20 cells for a state/normal compartment. edgeR expression filtering, trimmed mean of M-values (TMM) normalization, robust dispersion estimation, patient-blocked quasi-likelihood (QL) testing, and limma-voom sensitivity were applied [22, 23]. Patient-level aggregation avoids treating nuclei from the same patient as independent replicates, a problem emphasized in single-cell differential-expression benchmarks [24, 25, 26, 27]. QC and normalization choices were documented following general single-cell analysis principles [28]. Full tested universes and paired denominators accompany the results.

Mean cell scores and composition fractions were reduced to specimen values before patient-paired tests. Two-sided t tests, 95% t intervals, and Wilcoxon sensitivities were retained. Composition denominators were all retained QC nuclei; missing communication cell groups were instead kept unobserved. The CARE 43-pair and 52-pair composition analyses address different selection rules and are not independent replications. Frozen external gene tests used 36 hypotheses together for BH correction, preserving discovery directions; missing or unestimable genes could not count as successful replication.

Brain-reference construction and deconvolution

Author-labeled CARE pseudobulks from 52 clinical pairs defined malignant, oligodendrocyte, neuronal, astrocyte, myeloid, OPC, vascular (endothelial and pericyte), and lymphocyte compartments. Each reference contributor required at least 20 nuclei of that type, and each type required at least five contributing donors. Reference profiles averaged sample-type counts per million (CPM) with equal donor weight. Within the common measured universe, 60 specific genes per compartment were selected with mean CPM ≥ 1, excluding mitochondrial and ribosomal genes, the frozen GLASS program, and every four-state signature/control gene. The full reference contained 480 markers. Every donor-held-out fold rebuilt both marker selection and reference profiles from the remaining donors.

The CIBERSORT v1.03 CoreAlg implemented nu-support-vector regression with nu = 0.25, 0.50, or 0.75, coefficient clipping, sum normalization, and minimum-RMSE selection [15]. Its core was used unchanged, while the wrapper explicitly standardized reference and mixture without quantile normalization or automatic antilog transformation. No mixture-permutation fit P values were generated or used for filtering. Weighted non-negative least squares (NNLS) used the same reference with inverse-square-root mean-reference weights. This is an independent estimator sensitivity, not an independent reference cohort. Observed held-out pseudobulks were assessed against captured-UMI RNA fractions; synthetic held-out-donor mixtures provided additional coverage. Compartment-level gates

required observed Spearman rho ≥ 0.50 and mean absolute error (MAE) ≤ 0.15. They were fixed in the extension protocol before deconvolution outputs were inspected.

The full CARE reference was applied to 168 GLASS bulk specimens. Two-sided paired t tests, 95% t intervals, Wilcoxon sensitivities, and BH correction across eight compartments within each method were retained. All 84 patient pairs remained included. Correlations between paired weight changes and the frozen program or ten original target-expression changes used Spearman statistics with BH correction across all 176 tests. These same-sample correlations were interpreted as exploratory context.

Expanded malignant programs and continuous axes

Published metaprogram and biological-process genes were read from the published Nomura supplementary tables [14], with source hashes preserved. Cell-level log1p(CP10k) expression was averaged within each malignant specimen; each signature score subtracted frozen expression-matched controls. Controls used 24 discovery expression bins, 30 controls per signature gene, exclusion of all signature genes, and seed 20260920. A minimum 80% coverage rule was applied before paired tests. Biological and technical signature families were corrected separately, with an additional heteroscedasticity-consistent covariance estimator type 3 (HC3) regression sensitivity for paired UMI change. Expanded CARE state fractions used author annotations and malignant-cell denominators; these labels were not inferred from an argmax over the new signatures.

For continuous axes, discovery means and standard deviations of patient-level four-state changes and principal component analysis (PCA) loadings were fixed before projecting CARE. After excluding all directly defining score and control genes, expression filtering and TMM normalization preceded logCPM calculation with prior count two. For each gene and axis, a limma model related recurrent-minus-primary expression to axis score and mean-log-UMI change, with robust, trend-aware empirical Bayes inference. Discovery BH correction jointly covered both axes; CARE correction covered all 627 selected hypotheses, assigning untestable results P = 1. Concordance required the same slope direction and external false discovery rate (FDR) < 0.05. Sensitivities added the initial state projected with fixed loadings, performed leave-one-patient-out slope refits, and tested partial-rank associations after residualizing ranks on UMI change with n − 3 degrees of freedom. These analyses were specified as a post-validation exploratory extension, not as blinded or preregistered validation.

Representation comparison and attribution

Verified local Geneformer-V1-10M weights used raw counts mapped to model Ensembl identifiers, collapsed duplicates, division by model gene medians, and descending rank truncation to 2,048 tokens. Penultimate gene embeddings were mean-pooled. The AC comparison sampled at most 64 cells per specimen with fixed seed 20260915. Geneformer, four state scores, fixed 100-gene RNA, and training-selected 2,000-gene RNA models used identical cells and 20 repeated patient-grouped five-fold splits. Standardization and variance selection were fitted only within training folds. Logistic regression used fixed C = 1 with liblinear; full-discovery fits were then frozen for CARE. No CARE tuning or orientation reversal was allowed. AUROC intervals used 2,000 patient-pair bootstraps; external label-swap tests used 2,000 within-patient swaps with BH correction across four models.

Linear interventional Shapley additive explanations (SHAP) values [29] were computed in log-odds units from standardized features and fitted coefficients using each training fold's mean background. Only held-out patients contributed to attribution summaries. A fixed first patient in identifier order supplied the single-fold waterfall example; it was not selected for a favorable prediction. Correlated-feature attribution remains associative.

Real knockout calibration

GSE198522 control and KAT5-knockout matrices were downloaded with source receipts [21]. QC required at least 200 detected genes, 500 UMIs, and at most 20% mitochondrial UMIs. A fixed random 128-cell sample per library established real-KO and control embedding centroids. The initial benchmark required at least ten target-token-positive control cells. After documenting limited coverage, an explicitly labeled sensitivity searched all QC control cells and selected at most 64 eligible cells per model. Three other tokens within 64 rank positions provided matched deletion controls. Contextual effects compared post-deletion mean embeddings with original embeddings of the same remaining genes. Model-mean virtual vectors were aligned with the fixed observed KO-minus-control direction. Model-level exact sign-flip tests and 10,000 model bootstraps assessed the control-adjusted alignment. The sorted 827 model was excluded in a predefined-setting sensitivity; no new target was selected by these results.

Communication pathways and networks

CellChat 2.2.0.9001 [30, 31] used protein interactions whose ligand/receptor complex subunits and cofactors were all in the common assay universe. Non-protein interactions were excluded. Each paired-patient object used a minimum of 20 and a maximum of 300 cells per available cell group, fixed seed 20260915, truncated mean with trim 0.1, raw normalized signaling data, no population-size weighting, and 100 cell-label permutations. Only within-stage scores were extracted. All 12,201 eligible pathway-source-target comparisons were corrected together; cell-label permutation P values were not biological recurrence tests.

Ranked gene-set analysis and Hallmark summaries followed established enrichment frameworks [32, 33]. KEGG 2021 Human was downloaded from Enrichr with a checksum, and ranked fgsea analyses used gene sets with 15–500 measured genes. Joint correction covered five bulk or malignant analyses. CAMERA sensitivity [34] used patient-blocked voom expression, comparing estimated non-negative inter-gene correlation with fixed correlation 0.01. Weighted gene co-expression network analysis (WGCNA) reference parameters were retained [35], with additional powers 8 and 16 and five fixed random 67-of-84 patient subsamples. Adjusted Rand indices and best-matching module Jaccard indices quantify partition sensitivity. DEG/module overlap remains descriptive because both use the same discovery data.

The archived scTenifold analysis used 22 patients, 32 malignant nuclei per specimen, 1,508 genes, ten resampled networks per stage, and ten fixed virtual targets [13]. It was retained as a limited original computational analysis, with its original patient denominator retained. Per-target figures exclude the deleted gene itself and distinguish unsigned network FC from expression fold change.

Spatial clinical and drug context

Spatial scores retained the original deposited coordinates, coverage rules, UMI-adjusted within-section correlations, and patient aggregation. Program genes were excluded from comparison marker sets. Core and edge regions were not inferred without deposited annotations. Survival models used continuous standardized program scores and clinical covariates, with complete-case counts and proportional-hazards checks retained. DepMap 24Q4, PRISM 24Q2, and GDSC2 release 8.5 used archived verified files and biological-model aggregation; these were not labeled the latest releases. Complete source tables and analysis scripts accompany the manuscript [36, 37, 38, 39, 40, 41].

Multiple testing and evidential status

Benjamini–Hochberg FDR control [42] was applied within explicitly stated biological test families. Pooled cell counts were not used as patient replication. Frozen tests, exploratory extensions, method sensitivities, and descriptive figures were distinguished throughout. The 36 frozen gene-state tests, 627 later axis-association tests, and 176 deconvolution-context tests answer different questions and were not merged or selectively substituted for one another. Full tables include failed calibration gates and non-significant tests.

No prospective sample-size calculation was performed for this secondary analysis; all eligible pairs meeting the stated criteria were retained.

Results

A longitudinal tissue program provides the discovery context

The GLASS snapshot contained 533 RNA profiles and 214 RNA silver-set pairs. Clinical and biospecimen reconciliation identified 84 eligible adult patients with IDH-wildtype, baseline grade IV disease, a primary-to-first-recurrence comparison, and documented intervening radiotherapy or temozolomide exposure. Eighty patients had evidence of both treatments. The strict cohort represented 11 collection sites. Median age was 55 years (interquartile range, 48–64), with 53 male and 31 female patients as recorded in the source clinical metadata. Median time between sampled operations was 11 months (interquartile range, 7–16). Age, sex and sampling interval were available for all 84 patients; MGMT status was available for 53 (20 methylated and 33 unmethylated). An extended IDH-wildtype cohort comprised 103 pairs (cohort manifests; Supplementary Figs. S1–S3).

Paired limma-trend analysis of 17,923 expressed genes identified 1,432 genes at false discovery rate (FDR) below 0.05, including 792 increased and 640 decreased genes. The strongest bulk results did not support a uniform increase in a mesenchymal program. E2F targets and G2/M checkpoint were negatively enriched, with normalized enrichment scores (NES) of approximately − 3.10; angiogenesis was also negatively enriched (NES − 2.39). Neural and synaptic genes were represented among recurrence-increased genes, whereas several vascular or perivascular genes decreased. Patient-paired state scores and sensitivity analyses were retained separately from the gene-set enrichment results (Fig. 1; Supplementary Figs. S4–S5).

We froze the 50 strongest qualifying genes in each direction, requiring FDR below 0.05 and direction agreement in at least 60% of discovery patients. In the external GSE339484 grade IV, IDH1-wildtype subset, ten primary–recurrent pairs showed a mean program increase of 0.287 (95% confidence interval [CI] 0.158–0.416; paired P = 0.00072; Wilcoxon P = 0.0081). Nine of ten pairs increased. The paired gene-statistic rank correlation with GLASS was 0.662, and all 96 program genes entering the cross-cohort differential-expression comparison were directionally concordant. Including one additional legacy grade III pair produced a similar mean score change of 0.286 (Supplementary Fig. S6).

The external abundance file contained fractional values and was analyzed accordingly, rather than treated as integer sequencing counts. An expression-identity audit found no near-identical profile among the deposited external samples and 533 GLASS profiles; the highest cross-cohort correlation was 0.777. This audit reduces concern about an exact expression duplicate but cannot establish the independence of anonymized patients. Two additional NanoString panels, GSE254873 and GSE254874, represented the same 17 pairs and measured only nine and eight frozen program genes, respectively. They were not counted as two independent cohorts or used to claim validation of the complete score.

Sample recovery improves coverage without establishing a universal malignant recurrence signal

Reconciliation of GSE174554 with the related GSE138794 deposit recovered six missing paired patients, yielding 28 pairs. These recovered specimens are part of the original discovery series, not a new validation cohort. Two primary regions from one patient were retained as one biological timepoint. Additional author annotations recovered previously unannotated malignant cells; an annotation-origin sensitivity excluded the affected pair. Twenty pairs met the strict copy-number-variation (CNV) subset requirements. A common universe of 21,810 gene symbols standardized cell-state scoring across discovery, CARE, and the supplementary patient-derived xenograft (PDX) matrices.

The reconciled all-malignant edgeR quasi-likelihood analysis found no gene at FDR below 0.05, as did the strict-CNV sensitivity. Four-state mean scores and proportions showed substantial individual variation without a consistent cohort-wide shift. The frozen bulk program changed by 0.00326 equal-cell rank-score units (95% CI -0.00288 to 0.00940; P = 0.286; nine-test family FDR = 0.866). Continuous axes were fitted to patient-level state changes and saved for external application; these retrospective axes are not baseline predictors and were not converted into forced trajectory classes (Supplementary Fig. S7).

Recovered tissue composition changes consistently across cohorts

In 28 discovery pairs, the mean malignant fraction among retained quality-control (QC) nuclei declined by 25.64 percentage points (95% CI -40.32 to -10.96; FDR = 0.0158), while the oligodendrocyte fraction increased by 16.08 points (4.36–27.80; FDR = 0.0384). Neuronal fractions increased by 6.94 points (1.83–12.06; FDR = 0.0384). These are patient-weighted estimates, not pooled-cell comparisons (Fig. 2).

CARE contained 121 specimens from 59 patients; 52 patients had a clinical primary-to-first-recurrence pair, and 43 pairs had at least 50 author-annotated malignant nuclei at both timepoints. Among the 43 malignant-eligible pairs, the malignant fraction decreased by 14.02 percentage points (95% CI -23.29 to -4.76; FDR = 0.0195) and the oligodendrocyte fraction increased by 8.90 points (3.57–14.23; FDR = 0.0161). In all 52 clinical pairs, the corresponding changes were − 18.40 and + 11.05 points, each FDR = 0.00158. The neuronal fraction did not pass correction in CARE. Thus, only part of the discovery composition pattern was consistent externally.

Normal-cell differential expression was estimable in 14 astrocyte, 19 myeloid, and 16 oligodendrocyte pairs. One gene-compartment association remained significant across the combined test family; other normal compartments lacked eight eligible pairs. Normal subtypes in discovery were marker-inferred, whereas CARE used author labels; donor-equal marker expression is displayed for annotation transparency (Supplementary Fig. S8). Composition is sensitive to specimen region, dissociation, cell recovery and annotation (Supplementary Fig. S9), and cannot be equated with unbiased histological purity or causal replacement of tumor by normal cells. The CARE composition analysis followed inspection of other CARE outcomes and is exploratory cross-cohort consistency.

Brain-specific deconvolution supports selected tissue changes

We completed brain-reference deconvolution using the CIBERSORT support-vector regression core and a separate weighted non-negative least-squares (NNLS) estimator. The reference covered eight broad compartments and was constructed from author-labeled CARE nuclei. Marker selection and reference construction were repeated without the test donor in each of 52 folds; both timepoints were excluded together. Evaluation used 104 observed donor

pseudobulks with known captured-RNA contributions, supplemented by synthetic mixtures. This estimates RNA contributions rather than nuclear fractions.

In observed held-out pseudobulks, CIBERSORT oligodendrocyte weights had Spearman rho = 0.840 and mean absolute error (MAE) = 0.028. Vascular weights had rho = 0.682 and MAE = 0.025. These compartments, neurons, and myeloid cells passed both the rho ≥ 0.50 and MAE ≤ 0.15 criteria in both methods. Malignant weights correlated with truth (rho = 0.747) but had MAE = 0.276 and failed calibration. Astrocyte, OPC, and lymphocyte estimates also failed at least one CIBERSORT criterion. Accordingly, a significant estimated malignant change cannot be interpreted as validated tumor-purity loss (Fig. 3).

Applied to all 84 GLASS pairs, CIBERSORT estimated an oligodendrocyte increase of 8.25 percentage points (95% CI 4.34–12.17; eight-compartment FDR = 0.00054) and a vascular decrease of 6.18 points (95% CI − 10.00 to − 2.37; FDR = 0.0036). Weighted NNLS gave concordant changes of + 7.57 points (FDR = 0.0127) and − 2.11 points (FDR = 0.0257), respectively. The numerical magnitudes are method-dependent. Neuronal and myeloid changes did not pass correction in both methods. No patient was excluded on a deconvolution fit P value.

Across patients, change in the frozen program correlated positively with oligodendrocyte weight change (rho = 0.643 with CIBERSORT and 0.729 with NNLS), and negatively with vascular change (− 0.502 and − 0.419). All four associations passed correction across the full 176 feature-compartment-method family (Supplementary Fig. S10). Because both quantities were calculated from the same bulk samples, these associations provide tissue context rather than independent validation or causal mediation. Single-nucleus reference CPM and bulk TPM also differ in measurement and gene-length effects; held-out reference accuracy does not eliminate cross-platform bias.

Expanded malignant programs do not show a uniform recurrence shift

To address programs omitted by the four-state representation, we scored published malignant metaprograms and biological processes, including GPC, cilia-associated, and neuronal programs [14]. Seventeen of eighteen gene sets met the frozen 80% common-gene coverage rule; a low-quality metaprogram failed coverage. Fifteen biological signatures were tested together within each cohort, while ribosomal and microenvironment-associated technical signatures were reported separately. Gene controls were selected from discovery expression and reused unchanged in CARE.

No biological program passed FDR < 0.05 or 0.10 in 28 discovery pairs, the strict-CNV subset, or 43 CARE pairs. Minimum biological-program FDRs were 0.858, 0.914, and 0.730, respectively. Adjustment for change in mean log UMI did not uncover a corrected cohort-wide shift. CARE author-defined expanded states, including GPC, Cilia, Neuron, and Stress, likewise showed no corrected paired composition difference across 43 malignant-eligible patients (Fig. 4). The published signatures were developed using data that include CARE, so CARE scoring is application of the reference rather than independent validation of its discovery. These results do not exclude patient-specific program changes.

Continuous malignant variation reveals reproducible associations with substantial baseline sensitivity

We next evaluated whether heterogeneous patient changes carry expression associations despite the absence of a uniform average shift. The first two axes were fitted to standardized four-state changes in the 28 discovery patients and applied to CARE with the discovery scaling and loadings fixed. PC1 contrasted OPC/NPC with MES-associated

change, whereas PC2 was dominated by AC-associated change. To avoid direct arithmetic self-association, all 3,855 genes used in the defining scores or their expression controls were excluded before gene-level testing.

Models related each gene's paired log-expression change to one state axis, adjusting for change in mean log UMI. Joint discovery correction across both axes yielded 627 gene-axis hypotheses, comprising 576 PC1 and 51 PC2 associations. External tests across this fixed family identified 293 concordant associations: 274 with PC1 and 19 with PC2. These counts refer to gene-axis tests and need not represent distinct genes (Fig. 5).

Adding each patient's initial projected state reduced the associations passing correction in both cohorts to 33 (32 PC1 and one PC2). All 293 original associations retained their slope direction in leave-one-patient-out refits in both cohorts; 275 also passed the external partial-rank sensitivity. Direction stability therefore did not imply robustness to baseline adjustment. The baseline-sensitive reduction is central to interpretation, because change scores can couple to initial state and regression to the mean. These analyses were added after earlier CARE outcomes had been inspected and are explicitly exploratory cross-cohort associations. They do not repair the failed frozen 36-hypothesis test, predict future recurrence from baseline, or identify causal dependencies.

State-specific candidates fail the frozen external gene test

Discovery state-specific differential expression was sensitive to the count model. Astrocyte-like malignant cells yielded 49 within-state FDR-significant genes with edgeR QL and 152 with limma-voom across 21 pairs. Joint correction across all genes in four states removed the AC associations at the 0.05 threshold; the strongest OPC signal was model-sensitive. In the stricter CNV-defined AC analysis, no gene passed genome-wide FDR below 0.05. Additional sample recovery alone therefore does not explain every increase in detected genes.

Before CARE gene testing, 36 exploratory gene-state hypotheses representing 34 genes were frozen: 34 hypotheses passed the discovery joint-state $q < 0.10$ eligibility rule, and two passed the all-malignant $q < 0.10$ rule. This was an exploratory eligibility threshold, not a claim of $q < 0.05$ discovery. CARE supplied 28–43 paired patients depending on compartment. None of the 36 hypotheses passed external FDR below 0.05. The smallest external FDR was 0.099 for MOG in AC-like malignant cells; RGCC changed in the opposite direction. No candidate was promoted to a replicated therapeutic target (Supplementary Fig. S11; Supplementary Figs. S12–S13).

Internal representation performance does not transfer to CARE

The initial whole-malignant centroid area under the receiver operating characteristic curve (AUROC) was 0.302 in 22 pairs. A patient-label-swap audit did not establish meaningful inverse biological discrimination, and direct inspection confirmed the implemented token ordering against the official rank rule. State-conditioned analysis then evaluated AC-like malignant cells using identical sampled cells, patients, and held-out folds across four fixed classifiers. In 21 discovery pairs, AUROCs were 0.696 for Geneformer embeddings, 0.753 for frozen GLASS RNA genes, 0.705 for training-selected RNA features, and 0.594 for four state scores.

All four classifiers were frozen and applied without refitting to 33 CARE AC-eligible pairs. Geneformer AUROC was 0.486 (95% patient-bootstrap CI 0.379–0.597), the frozen RNA model achieved 0.579 (0.462–0.702), training-selected RNA achieved 0.552, and state scores achieved 0.389. None passed correction across four patient-label-swap tests. No label reversal or outcome-selected threshold was applied (Fig. 6). Linear SHAP plots [29] explain held-out predictions of the fixed RNA model (Supplementary Fig. S14), but do not validate its biological mechanism or justify selecting new therapeutic targets. Original perturbation endpoints are shown in Supplementary Fig. S15; token-order checks accompany the audit tables.

Public KAT5 knockout data do not support virtual perturbation alignment

GSE198522 provided nine in-vitro cell models with matched CD8-control and KAT5-knockout libraries [21]. A random 128-cell reference per sample yielded at least ten KAT5-token-positive control cells in only three models. This initial benchmark did not show control-adjusted superiority. A documented coverage sensitivity then searched all QC-passing control cells and sampled up to 64 KAT5-token-positive cells, making all nine models evaluable while retaining the original reference centroids and results.

Contextual virtual-deletion vectors were compared with each model's observed real-knockout minus control embedding direction. The mean improvement over three local-rank deletion controls was 0.000126 cosine units (95% model-bootstrap CI -0.00693 to 0.00678; exact model sign-flip P = 0.488). Excluding the sorted GSC827 setting gave P = 0.629. Raw UMI detection substantially exceeded token retention, showing that transcript detection and perturbation eligibility are different quantities (Fig. 7). This benchmark concerns one known intervention with one library per condition per model; it does not establish general failure of Geneformer or validate any new patient target.

Communication and network analyses constrain mechanistic claims

CellChat was run separately within 28 paired-patient objects, with a fixed common-gene-compatible universe of 2,186 protein ligand-receptor interactions. Primary and recurrent samples within a patient shared expression scaling. Within-stage pathway scores were compared only where both cell groups were observed at both timepoints. Across 12,201 pathway-source-target combinations with at least eight eligible pairs, none passed FDR below 0.05 or 0.10. The smallest nominal P was 0.00208, but the adjusted FDR was 1.00. The patient-level comparison is presented in Supplementary Fig. S16. Consequently, visually prominent pooled networks cannot establish a general recurrence-associated communication change.

The original GLASS WGCNA network used 84 patient-change vectors and 3,000 genes. Its power of 12 was a fallback because the scale-free fit criterion was not reached. Module adjusted Rand indices were 0.813 and 0.740 at powers 8 and 16, respectively, and 0.693-0.800 in five 67-patient subsamples. This supports partial partition stability, not independent mechanism validation (Supplementary Figs. S17–S18). The archived 20 scTenifold knockouts across ten targets and two stages were retained with per-target visualizations. Their unsigned network statistics are not expression fold changes or experimental perturbation effects (Supplementary Figs. S19–S28).

KEGG ranked enrichment identified pathway signals in bulk and malignant compartments, including selected neural/calcium-related terms. These findings were method-dependent: no pathway passed FDR below 0.05 in CAMERA analyses that estimated inter-gene correlation in either single-cell cohort or malignant compartment. Full enrichment contrasts and the sensitivity caveat are shown in Supplementary Fig. S29. Thus, enrichment was not treated as a substitute for successful target replication.

Supplementary models and spatial observations provide context

The GSE271619 PDX series contains 12 samples but only two patient-derived models. Their state-composition changes differed (Supplementary Fig. S30); technical or model replicates were not counted as independent patients. The associated GSE271620 matrix is transformed expression and was not passed to a raw-count model. GeoMx GSE254145 yielded five GFAP-region and three CD64-region paired patients. The deposited normalized panel covered

only 14 of 50 recurrence-up and 24 of 50 recurrence-down program genes, failing the predefined program coverage rule. Eligible state-module tests did not pass correction; no full-program validation was claimed.

Spatial analyses used deposited count matrices and physical coordinates from three cohorts. GSE237183 contributed 13 IDH-wildtype sections from five patients and 24,089 retained spots; GSE318562 contributed 22 sections from nine patient identifiers and 46,207 spots; GSE286413 contributed two sections from two patients and 7,735 spots. Marker sets were made non-overlapping with the frozen program before correlation analysis, and UMI-adjusted section correlations were summarized within patient (Fig. 8).

The recurrence program was inversely associated with the vascular marker score in both inferential cohorts: mean patient partial correlation − 0.375 (95% CI − 0.519 to − 0.231; FDR 0.015) in GSE237183 and − 0.204 (95% CI − 0.284 to − 0.123; FDR 0.0031) in GSE318562. A positive neuronal association was supported in GSE318562 (mean correlation 0.102; FDR 0.038), whereas MES association did not survive correction in either cohort (Supplementary Fig. S31). The two-patient series was reported descriptively without inferential P values or confidence intervals.

These cross-sectional associations are consistent with a spatially varying tissue program, but they do not establish that vascular abundance caused the longitudinal bulk change. They also do not provide paired primary–recurrence spatial evidence. Unannotated spots were not assigned core, edge, or niche labels. This interpretation is compatible with the heterogeneous spatial organization described in prior glioblastoma studies [36, 37].

Clinical and pharmacological data do not establish a validated intervention

External survival analysis included 134 TCGA, 73 CGGA325, and 104 CGGA693 adults with primary IDH-wildtype histological GBM and valid outcome data. TCGA cases represented in the GLASS case list were removed. Adjusted complete-case models included 108, 68, and 88 patients, respectively. Hazard ratios per standard deviation of the frozen program were 1.18 (95% CI 0.88–1.58), 0.74 (0.56–0.97), and 0.86 (0.69–1.06). None survived FDR correction across the three adjusted models, and directions were inconsistent (Supplementary Fig. S32). TCGA also showed evidence against a constant program effect over time; a time-varying sensitivity model estimated a one-year hazard ratio of 0.97 (95% CI 0.68–1.38).

DepMap 24Q4 provided 51 GBM biological units for dependency description and 47 with joint program-expression and dependency measurements. The ten computational follow-up genes did not display strong mean GBM dependency in this release. No genome-wide program–dependency correlation survived FDR below 0.05. PRISM 24Q2 yielded 1,597 compound-dose association tests across up to 38 GBM biological units; GDSC2 yielded 280 compound tests across up to 23 units. Neither resource contained a program–drug association surviving FDR below 0.05 (Supplementary Fig. S33). These cell-line resources do not uniformly resolve IDH status or clinical recurrence and were interpreted as functional context rather than matched recurrent-GBM experiments.

Discussion

The analyses support a distinction between a reproducible tissue-level recurrence program and heterogeneous changes within malignant cells. Oligodendrocyte-related composition changes agree across the recovered single-nucleus cohorts and two brain-reference deconvolution methods. Decreased vascular RNA weight in GLASS is also directionally compatible with spatial program associations. These results provide a coherent tissue context, but do not establish whether treatment, sampling region, nuclear recovery, or other processes produced the observed

changes. Histological localization remains relevant because glioblastoma transcriptomes vary across anatomical structures [43].

The malignant-cell result is more nuanced than either a universal transition or an absence of plasticity. Expanded published programs did not reveal a common shift, while continuous axes captured cross-cohort expression co-variation. The reduction from 293 concordant associations to 33 surviving baseline adjustment argues against presenting all associations as robust recurrence mechanisms. Initial state, correlated programs, and regression to the mean remain alternative explanations. The strong contrast with earlier reports of mesenchymal recurrence [4, 6] should be examined in terms of patient selection, available nuclei, regional sampling, annotation, and the precise estimand; this reanalysis does not invalidate those studies.

Developmental hierarchies also appear in oligodendroglioma [44], but that disease is biologically distinct and its findings cannot directly validate an IDH-wildtype glioblastoma trajectory. Macrophage-associated induction of mesenchymal-like glioblastoma states offers a plausible mechanistic precedent [45]; the present patient-paired CellChat results did not establish a general communication change supporting that mechanism in these samples. Null corrected tests may reflect modest patient counts and heterogeneous treatment effects. They should not be recast as proof that a pathway is absent.

The benchmark exposes an important limitation of deconvolution interpretation. Correlation with known RNA contributions was insufficient to ensure acceptable absolute error for malignant weights. Relative RNA contributions are further constrained to sum to one, so an increase in one compartment necessarily affects other estimates. Formal compositional models provide a complementary framework [46]; here, paired weight changes are descriptive effect estimates with explicit calibration gates, not absolute cell counts. Alternative reference-based methods such as MuSiC and Bisque illustrate additional normalization and modeling approaches [16, 17], but were not implemented in this comparison. Brain-specific and general benchmarking studies motivate testing reference transfer rather than assuming it [18, 19].

Internal Geneformer discrimination did not transfer to CARE, and one known real knockout failed to establish virtual-deletion advantage over matched controls. These findings constrain this particular representation, endpoint, and sample construction. They do not demonstrate that all adaptations of Geneformer fail. Similarly, scTenifold network statistics, SHAP contributions, or aggregation of several gene rankings cannot substitute for successful external or experimental validation. Rank aggregation is designed to combine ordered evidence [47]; it cannot create causal evidence from dependent computational outputs. No newly validated therapeutic target is claimed.

Limitations

Patient identity is anonymized across some deposits, reference signatures partly overlap their application datasets, and post-validation extensions are exploratory. The deconvolution benchmark is internal to the nuclear-reference platform; true tissue proportions and cross-platform calibration remain unknown. Legacy clinical labels do not fully harmonize to current molecular classification. Patient-bootstrap AUROC intervals condition on fitted predictions, and real-knockout comparisons have one library per condition per model. Spatial cohorts are cross-sectional and lack usable core-edge labels in the analyzed metadata. These limitations define the scope of the completed analyses and motivate independent validation.

These results indicate that recurrent specimens combine changes in sampled tissue composition with diverse malignant-state trajectories. The computational results show that external prediction, calibrated deconvolution, and intervention benchmarking must remain separate tests. Positive tissue associations and exploratory state

associations can be reported together with unsuccessful target and model validation without converting either into a treatment claim.

Conclusions

Selected tissue-composition changes accompany heterogeneous malignant-state variation in recurrent glioblastoma. Calibrated RNA-weight estimates and baseline-adjusted exploratory associations refine interpretation, but the tested models and perturbations do not establish a therapeutic target.

Abbreviations

AC
astrocyte-like
AUROC
area under the receiver operating characteristic curve
BH
Benjamini–Hochberg
CARE
external single-nucleus cohort GSE274546
CI
confidence interval
CNV
copy-number variation
CPM
counts per million
FDR
false discovery rate
GBM
glioblastoma
GLASS
Glioma Longitudinal Analysis Consortium
GPC
glial progenitor cell
HC3
heteroscedasticity-consistent covariance estimator type 3
MAE
mean absolute error
MES
mesenchymal-like
MGMT
O6-methylguanine-DNA methyltransferase
NNLS
non-negative least squares
NPC
neural-progenitor-like

OPC
oligodendrocyte-progenitor-like
PCA
principal component analysis
PDX
patient-derived xenograft
QC
quality control
QL
quasi-likelihood
SHAP
Shapley additive explanations
TMM
trimmed mean of M-values
TPM
transcripts per million
UMI
unique molecular identifier
WGCNA
weighted gene co-expression network analysis.

Declarations

Ethical Approval and Accordance

Not applicable. Ethical approval was not required because this study consisted entirely of secondary computational analyses of publicly available, de-identified datasets. The present study did not recruit or involve human participants, collect or use human tissues, or conduct animal experiments. Analyses concerning established in vitro cell line models used previously published datasets only; no new cell culture or other biological experiments were performed.

Consent to participate

Not applicable. This study used publicly available, de-identified datasets and involved no direct participation of identifiable human participants, collection or use of human tissues, or animal subjects. Informed consent to participate was therefore not required for the present secondary analysis.

Consent to publish

Not applicable. This article contains no identifiable individual-level human information, images, or other materials requiring consent for publication. The analyses used publicly available, de-identified datasets.

Data availability

All analyses used previously deposited public data. The following accession-specific records and release-specific download pages identify the data and reference resources used in the main and supplementary analyses. Derived results, sample eligibility records, and analysis scripts accompany this article. Supplementary Data 7 provides an expanded resource index, archived provenance records, and input-file checksums. Access to third-party resources remains subject to their original access and reuse conditions.

GLASS: Glioma Longitudinal Analysis Consortium, Synapse project syn17038081 (https://www.synapse.org/Synapse:syn17038081). The expression input was gene_tpm_matrix_all_samples.tsv; its repository record is syn69961520 (https://www.synapse.org/Synapse:syn69961520). Pairing and clinical inputs were analysis_rna_silver_set.csv, analysis_pairs.csv, analysis_tumor_pairs.csv, analysis_platinum_set.csv, biospecimen_aliquots.csv, biospecimen_samples.csv, clinical_cases.csv, and clinical_surgeries.csv. The analysis used an archived local snapshot whose historical remote release label could not be established; the exact local input checksums and cohort-selection audit are supplied in Supplementary Data 7. A later repository revision should not be assumed identical to these archived inputs.

Single-nucleus discovery: GSE174554 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE174554) and related GSE138794 matrices (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE138794). CARE external cohort: GSE274546 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE274546). External bulk assessment: GSE339484 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE339484). Published KAT5-knockout calibration data: GSE198522 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE198522).

Spatial datasets: GSE237183 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE237183), GSE318562 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE318562), and GSE286413 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE286413). GSE200984 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE200984) was considered in the supplementary cohort-eligibility and context assessments; it was not an additional successful malignant-state validation cohort.

Additional targeted-panel and model resources: GSE254873 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE254873) and GSE254874 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE254874), under parent series GSE254875 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE254875); GSE254145 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE254145); GSE271619 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE271619); and GSE271620 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE271620). Their eligibility and coverage limitations are retained in the supplementary results.

TCGA survival context used the CGGA-hosted TCGA RNAseq collection, identifier TCGA_mRNAseq_702, release dated 14 June 2020 (https://www.cgga.org.cn/download_other.jsp): TCGA_mRNAseq_702.txt and TCGA_mRNAseq_702_clinical.txt. The exact expression download is https://www.cgga.org.cn/download?file=download/20200614/TCGA_mRNAseq_702.txt.zip&type=TCGA_mRNAseq_702&time=20200614 and the clinical download is https://www.cgga.org.cn/download?file=download/20200614/TCGA_mRNAseq_702_clinical.txt.zip&type=TCGA_mRNAseq_702_clinical&time=20200614. Only the eligible adult IDH-wildtype GBM subset entered the survival analysis, with GLASS-overlapping cases excluded.

CGGA survival cohorts were mRNAseq_325 (https://www.cgga.org.cn/download.jsp#mRNAseq_325) and mRNAseq_693 (https://www.cgga.org.cn/download.jsp#mRNAseq_693), using the 6 May 2020 RSEM expression and clinical files: CGGA.mRNAseq_325.RSEM-genes.20200506.txt, CGGA.mRNAseq_325_clinical.20200506.txt,

CGGA.mRNAseq_693.RSEM-genes.20200506.txt, and CGGA.mRNAseq_693_clinical.20200506.txt. The dataset-specific sections provide these downloads; the individual file URLs are also listed in Supplementary Data 7.

DepMap 24Q4 Public, Figshare record 27993248 version 1 (https://doi.org/10.25452/figshare.plus.27993248.v1): Model.csv, CRISPRGeneEffect.csv, and OmicsExpressionProteinCodingGenesTPMLogp1.csv. PRISM Repurposing Public 24Q2, Figshare record 25917643 version 1 (https://doi.org/10.6084/m9.figshare.25917643.v1): Repurposing_Public_24Q2_Cell_Line_Meta_Data.csv, Repurposing_Public_24Q2_LFC_COLLAPSED.csv, and Repurposing_Public_24Q2_Treatment_Meta_Data.csv. GDSC2 release 8.5: GDSC2_fitted_dose_response_27Oct23.xlsx (https://cog.sanger.ac.uk/cancerrxgene/GDSC_release8.5/GDSC2_fitted_dose_response_27Oct23.xlsx). These were archived releases, not claims about the latest available data.

Published signature and annotation resources: Neftel et al., Supplementary Table S2 (PMC6703186; https://pmc.ncbi.nlm.nih.gov/articles/PMC6703186/#SD11); Wang et al., Supplementary Tables 1-12, file 43018_2022_475_MOESM2_ESM.xlsx (https://pmc.ncbi.nlm.nih.gov/articles/PMC9767870/#MOESM2); and CARE author data (https://github.com/dravishays/GBM-CARE-WT/tree/main/data). Expanded malignant metaprograms and biological-process gene lists came from Nomura supplementary Tables 2 and 3 in nomura_supptables.xlsx (https://github.com/dravishays/GBM-CARE-WT/blob/main/data/nomura_supptables.xlsx). The implemented signatures and frozen controls are provided with the analysis outputs.

Enrichment libraries were the Enrichr collections MSigDB_Hallmark_2020 (https://maayanlab.cloud/Enrichr/geneSetLibrary?mode=text&libraryName=MSigDB_Hallmark_2020), Reactome_2022 (https://maayanlab.cloud/Enrichr/geneSetLibrary?mode=text&libraryName=Reactome_2022), GO_Biological_Process_2023 (https://maayanlab.cloud/Enrichr/geneSetLibrary?mode=text&libraryName=GO_Biological_Process_2023), and KEGG_2021_Human (https://maayanlab.cloud/Enrichr/geneSetLibrary?mode=text&libraryName=KEGG_2021_Human). The analyzed library snapshots and checksums are included in Supplementary Data 7.

Pretrained models: Geneformer-V1-10M (https://huggingface.co/ctheodoris/Geneformer/tree/main/Geneformer-V1-10M) and the archived Geneformer-V2-104M_CLcancer sensitivity model (https://huggingface.co/ctheodoris/Geneformer/tree/main/Geneformer-V2-104M_CLcancer); archived model-integrity records identify the analyzed weights. Gene annotation used GENCODE human release 47, GRCh38 (https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_47/gencode.v47.annotation.gtf.gz). Communication analysis used the human CellChatDB distributed with CellChat 2.2.0.9001 (https://github.com/jinworks/CellChat/blob/main/data/CellChatDB.human.rda); the tested interaction universe is provided in the supplementary results.

Code availability

Custom analysis scripts and the figure provenance index are provided in the supplementary code archive. The archive documents software dependencies and data-path configuration. Original public matrices, pretrained Geneformer weights and third-party CIBERSORT software must be obtained separately under their respective terms; they are not redistributed in the code archive.

References

1. Stupp R, van den Mason WP, et al. Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. N Engl J Med. 2005;352:987–96. https://doi.org/10.1056/NEJMoa043330.
2. Hegi ME, Diserens AC, Gorlia T, et al. MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma. N Engl J Med. 2005;352:997–1003. https://doi.org/10.1056/NEJMoa043331.
3. Varn FS, Johnson KC, Martinek J, et al. Glioma progression is shaped by genetic evolution and microenvironment interactions. Cell. 2022;185:2184–e219916. https://doi.org/10.1016/j.cell.2022.04.038.
4. Wang L, Jung J, Babikir H, et al. A single-cell atlas of glioblastoma evolution under therapy reveals cell-intrinsic and cell-extrinsic therapeutic targets. Nat Cancer. 2022;3:1534–52. https://doi.org/10.1038/s43018-022-00475-x.
5. Verhaak RG, Hoadley KA, Purdom E, et al. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell. 2010;17:98–110. https://doi.org/10.1016/j.ccr.2009.12.020.
6. Wang Q, Hu B, Hu X, et al. Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. Cancer Cell. 2017;32:42–e566. https://doi.org/10.1016/j.ccell.2017.06.003.
7. Patel AP, Tirosh I, Trombetta JJ, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 2014;344:1396–401. https://doi.org/10.1126/science.1254257.
8. Richards LM, Whitley OKN, MacLeod G, et al. Gradient of Developmental and Injury Response transcriptional states defines functional vulnerabilities underpinning glioblastoma heterogeneity. Nat Cancer. 2021;2:157–73. https://doi.org/10.1038/s43018-020-00154-9.
9. Neftel C, Laffy J, Filbin MG, et al. An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. Cell. 2019;178:835–e84921. https://doi.org/10.1016/j.cell.2019.06.024.
10. Theodoris CV, Xiao L, Chopra A, et al. Transfer learning enables predictions in network biology. Nature. 2023;618:616–24. https://doi.org/10.1038/s41586-023-06139-9.
11. Zhang Y, Venkatesh MS, Theodoris CV. Discovery of candidate therapeutic targets with Geneformer. Nat Protoc. 2026. https://doi.org/10.1038/s41596-026-01364-8.
12. Chen H, Venkatesh MS, Gómez Ortega J, et al. Scaling and quantization of large-scale foundation model enables resource-efficient predictions in network biology. Nat Comput Sci. 2026;6:450–63. https://doi.org/10.1038/s43588-026-00972-4.
13. Osorio D, Zhong Y, Li G, et al. scTenifoldKnk: An efficient virtual knockout tool for gene function predictions via single-cell gene regulatory network perturbation. Patterns. 2022;3:100434. https://doi.org/10.1016/j.patter.2022.100434.
14. Nomura M, Spitzer A, Johnson KC, et al. The multilayered transcriptional architecture of glioblastoma ecosystems. Nat Genet. 2025;57:1155–67. https://doi.org/10.1038/s41588-025-02167-5.
15. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12:453–7. https://doi.org/10.1038/nmeth.3337.
16. Wang X, Park J, Susztak K, et al. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat Commun. 2019;10:380. https://doi.org/10.1038/s41467-018-08023-x.
17. Jew B, Alvarez M, Rahmani E, et al. Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. Nat Commun. 2020;11:1971. https://doi.org/10.1038/s41467-020-15816-6.
18. Sutton GJ, Poppe D, Simmons RK, et al. Comprehensive evaluation of deconvolution methods for human brain gene expression. Nat Commun. 2022;13:1358. https://doi.org/10.1038/s41467-022-28655-4.

19. Avila Cobos F, Alquicira-Hernandez J, Powell JE, et al. Benchmarking of cell type deconvolution pipelines for transcriptomics data. Nat Commun. 2020;11:5650. https://doi.org/10.1038/s41467-020-19015-1.
20. Spitzer A, Johnson KC, Nomura M, et al. Deciphering the longitudinal trajectories of glioblastoma ecosystems by integrative single-cell genomics. Nat Genet. 2025;57:1168–78. https://doi.org/10.1038/s41588-025-02168-4.
21. Mihalas AB, Arora S, O'Connor SA, et al. KAT5 regulates neurodevelopmental states associated with G0-like populations in glioblastoma. Nat Commun. 2025;16:4327. https://doi.org/10.1038/s41467-025-59503-w.
22. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47. https://doi.org/10.1093/nar/gkv007.
23. Chen Y, Chen L, Lun ATL, et al. edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. Nucleic Acids Res. 2025;53:gkaf018. https://doi.org/10.1093/nar/gkaf018.
24. Squair JW, Gautier M, Kathe C, et al. Confronting false discoveries in single-cell differential expression. Nat Commun. 2021;12:5692. https://doi.org/10.1038/s41467-021-25960-2.
25. Zimmerman KD, Espeland MA, Langefeld CD. A practical solution to pseudoreplication bias in single-cell studies. Nat Commun. 2021;12:738. https://doi.org/10.1038/s41467-021-21038-1.
26. Crowell HL, Soneson C, Germain PL, et al. muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. Nat Commun. 2020;11:6077. https://doi.org/10.1038/s41467-020-19894-4.
27. Soneson C, Robinson MD. Bias, robustness and scalability in single-cell differential expression analysis. Nat Methods. 2018;15:255–61. https://doi.org/10.1038/nmeth.4612.
28. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol. 2019;15:e8746. https://doi.org/10.15252/msb.20188746.
29. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems. 2017;30:4765–4774. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
30. Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021;12:1088. https://doi.org/10.1038/s41467-021-21246-9.
31. Jin S, Plikus MV, Nie Q. CellChat for systematic analysis of cell–cell communication from single-cell transcriptomics. Nat Protoc. 2025;20:180–219. https://doi.org/10.1038/s41596-024-01045-4.
32. Subramanian A, Tamayo P, Mootha VK et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005;102:15545–15550. https://doi.org/10.1073/pnas.0506580102
33. Liberzon A, Birger C, Thorvaldsdóttir H, et al. The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst. 2015;1:417–25. https://doi.org/10.1016/j.cels.2015.12.004.
34. 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.
35. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559. https://doi.org/10.1186/1471-2105-9-559.
36. Al-Dalahmah O, Argenziano MG, Kannan A, et al. Re-convolving the compositional landscape of primary and recurrent glioblastoma reveals prognostic and targetable tissue states. Nat Commun. 2023;14:2586. https://doi.org/10.1038/s41467-023-38186-1.

37. Greenwald AC, Darnell NG, Hoefflin R, et al. Integrative spatial analysis reveals a multi-layered organization of glioblastoma. Cell. 2024;187:2485–e250126. https://doi.org/10.1016/j.cell.2024.03.029.
38. Corsello SM, Nagari RT, Spangler RD, et al. Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. Nat Cancer. 2020;1:235–48. https://doi.org/10.1038/s43018-019-0018-6.
39. Yang W, Soares J, Greninger P, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2013;41:D955–61. https://doi.org/10.1093/nar/gks1111.
40. Zhao Z, Zhang KN, Wang Q, Genomics, et al. Proteom Bioinf. 2021;19:1–12. https://doi.org/10.1016/j.gpb.2020.10.005.
41. Brennan CW, Verhaak RG, McKenna A, et al. The Somatic Genomic Landscape of Glioblastoma. Cell. 2013;155:462–77. https://doi.org/10.1016/j.cell.2013.09.034.
42. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J Royal Stat Soc Ser B: Stat Methodol. 1995;57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
43. Puchalski RB, Shah N, Miller J, et al. An anatomic transcriptional atlas of human glioblastoma. Science. 2018;360:660–3. https://doi.org/10.1126/science.aaf2666.
44. Tirosh I, Venteicher AS, Hebert C, et al. Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature. 2016;539:309–13. https://doi.org/10.1038/nature20123.
45. Hara T, Chanoch-Myers R, Mathewson ND, et al. Interactions between cancer cells and immune cells drive transitions to mesenchymal-like states in glioblastoma. Cancer Cell. 2021;39:779–e79211. https://doi.org/10.1016/j.ccell.2021.05.002.
46. Büttner M, Ostner J, Müller CL, et al. scCODA is a Bayesian model for compositional single-cell data analysis. Nat Commun. 2021;12:6876. https://doi.org/10.1038/s41467-021-27150-6.
47. Kolde R, Laur S, Adler P, et al. Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics. 2012;28:573–80. https://doi.org/10.1093/bioinformatics/btr709.

Figures

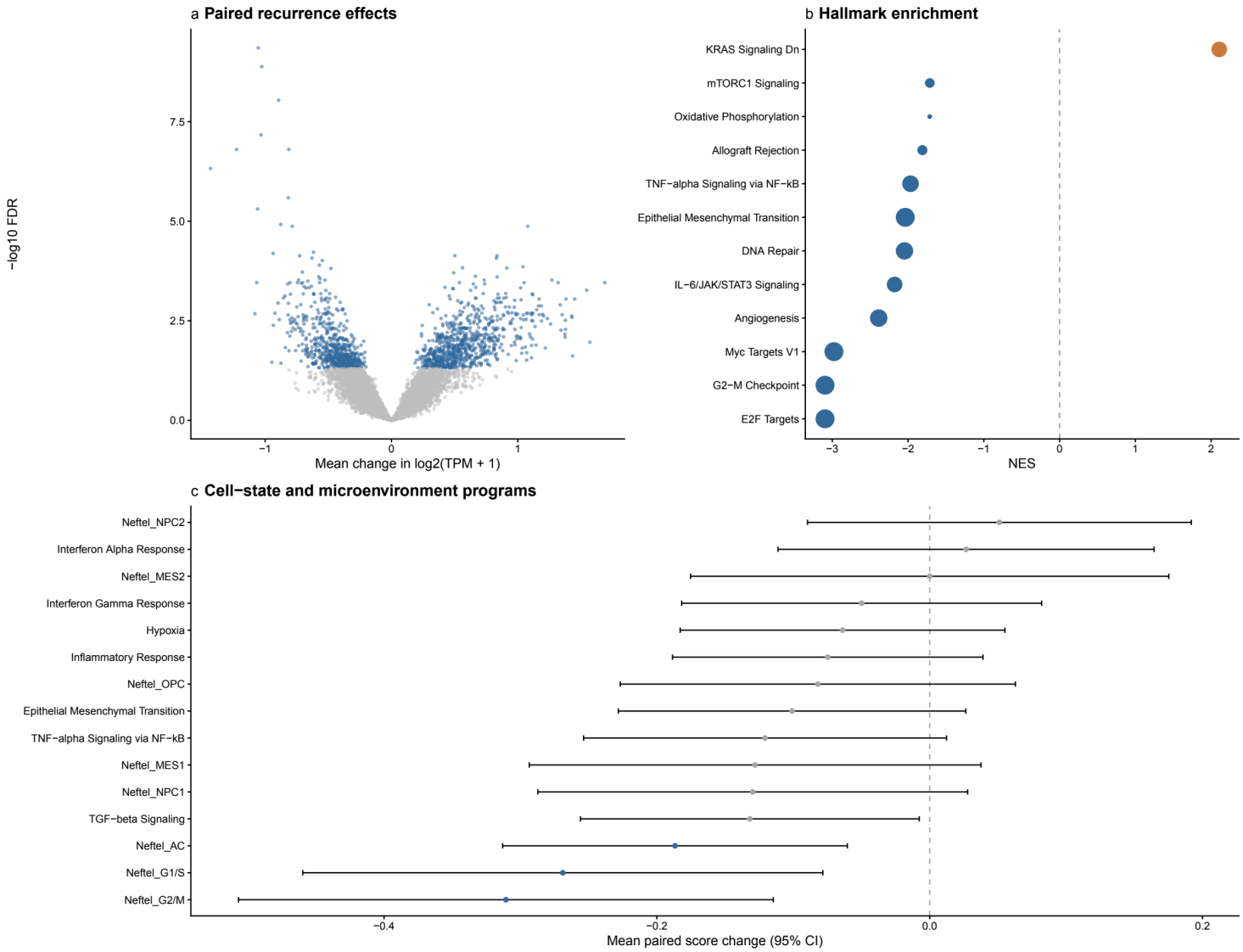
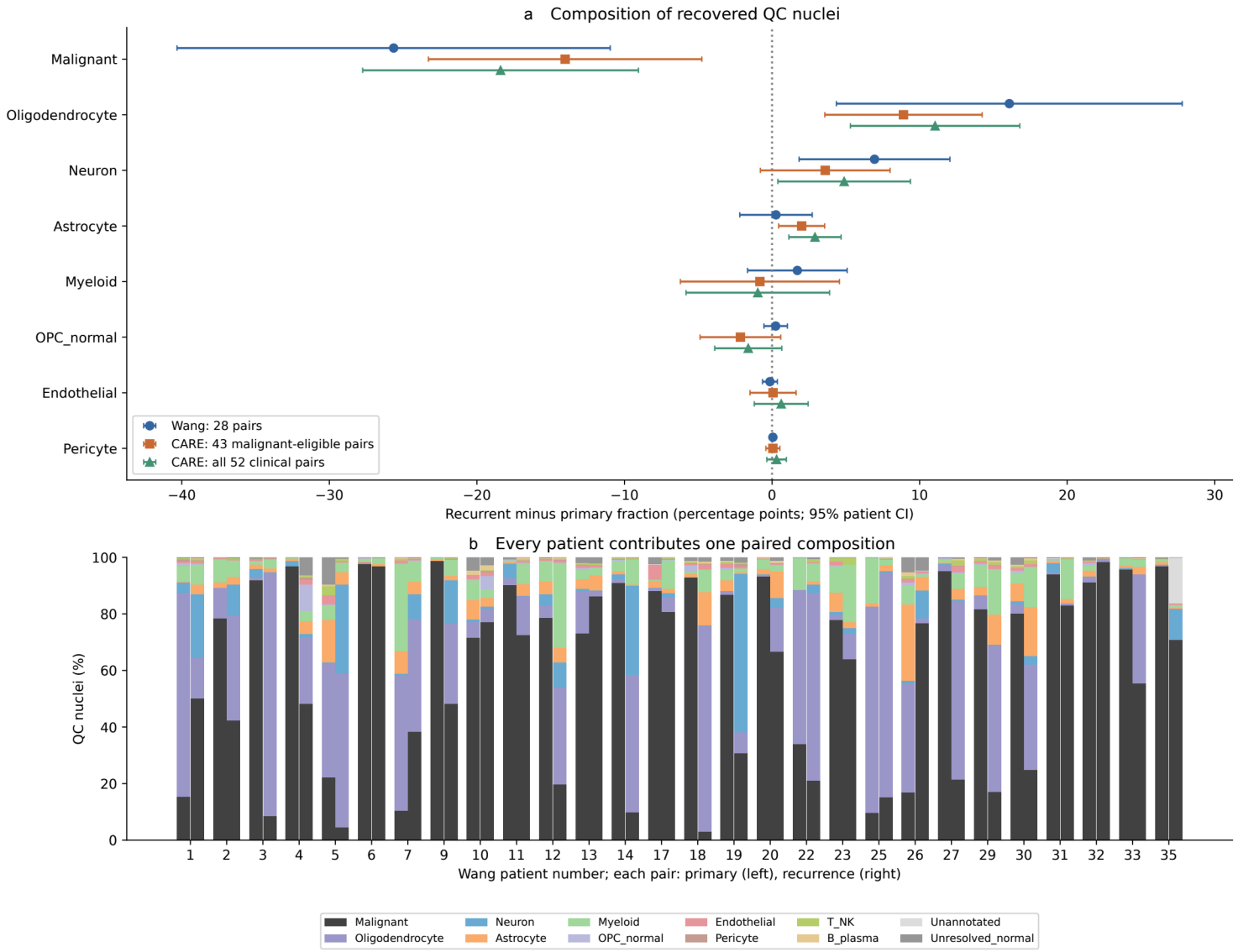


Figure 1

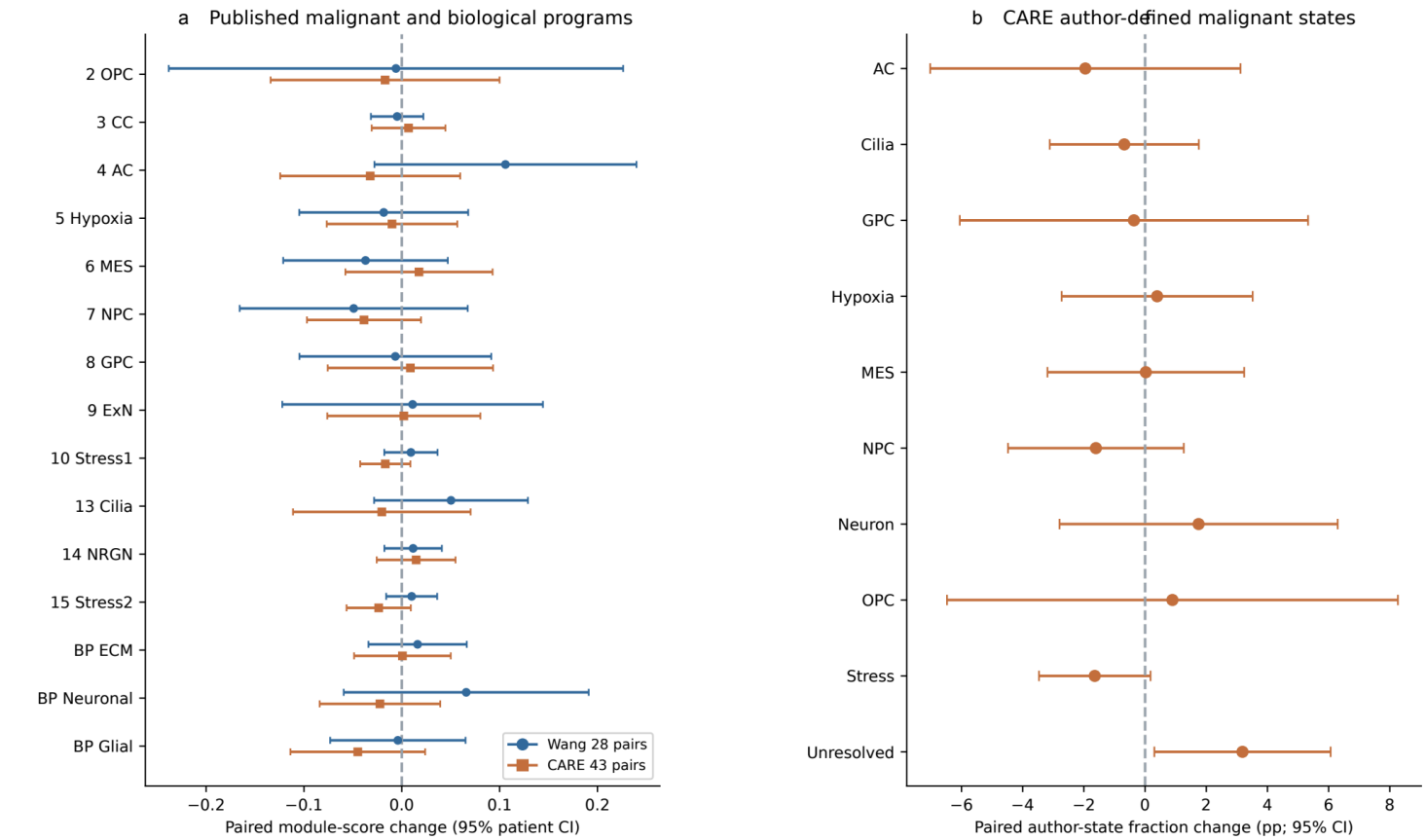
Longitudinal GLASS discovery. a, Patient-blocked limma-trend results for 17,923 expressed genes across 84 GLASS primary-to-first-recurrence pairs. Blue marks gene-level BH FDR < 0.05; 1,432 genes pass. b, The twelve lowest-FDR Hallmark terms, with normalized enrichment scores and bubble area reflecting significance. c, Patient-paired state and microenvironment score changes with 95% confidence intervals; blue marks corrected significance in the tested program family. Gene tests, ranked gene-set enrichment, and paired score differences address distinct estimands



Sampling and cell recovery affect these proportions; they are not histological tumor-purity estimates.
Wang normal types are marker-inferred; CARE uses author annotations. CARE composition is exploratory cross-cohort consistency.

Figure 2

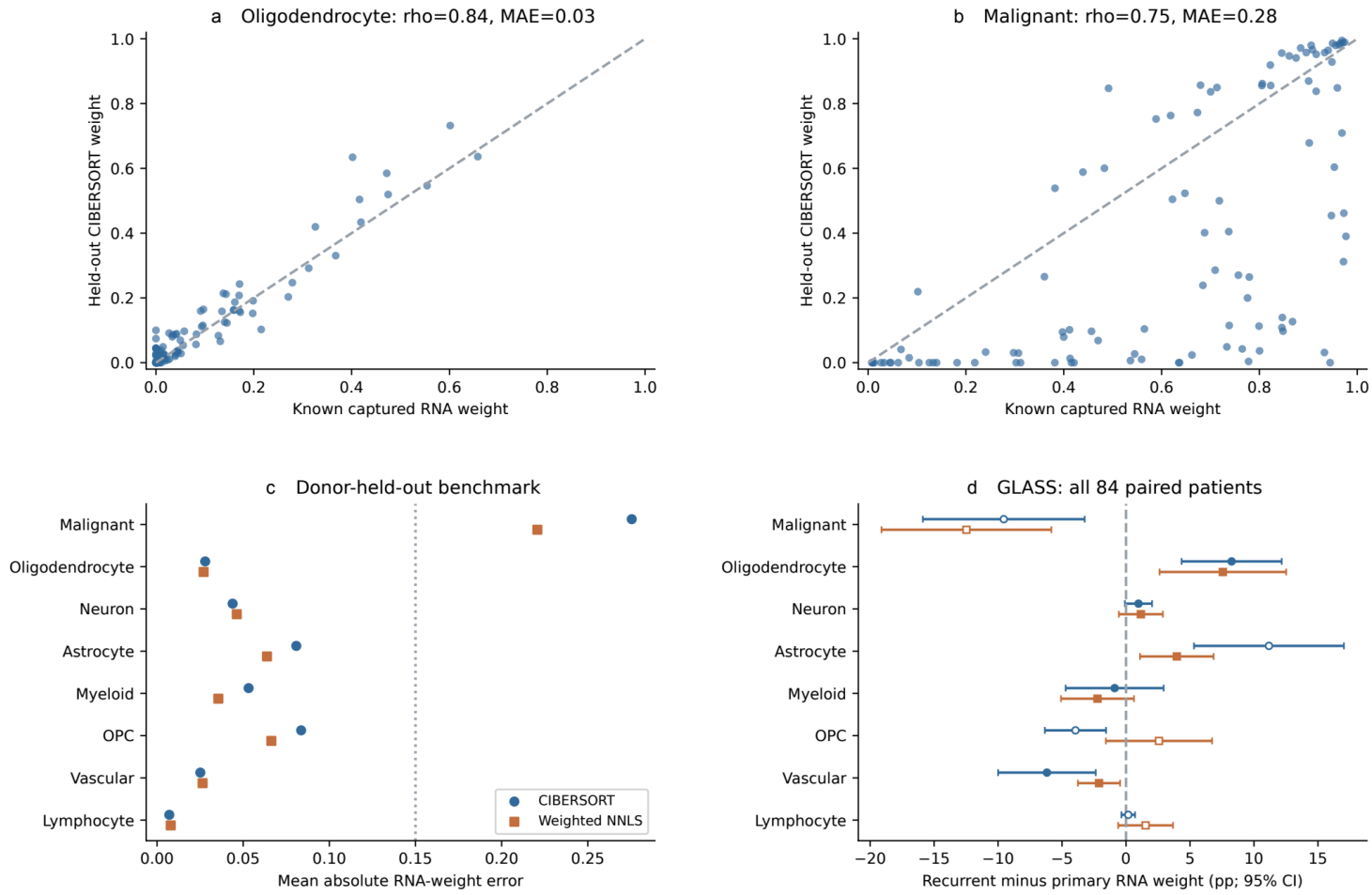
Patient level cellular composition. a, Patient-equal mean changes in fractions among retained QC nuclei, with 95% paired t intervals, in 28 discovery pairs, 43 malignant-eligible CARE pairs, and all 52 clinical CARE pairs. The two CARE subsets are overlapping sensitivities. b, Each discovery patient contributes adjacent primary and recurrent composition bars. Normal labels in discovery are marker-inferred, while CARE uses author labels. These are sampled-nucleus proportions, not histological purity. CARE composition was added after other CARE outcomes were inspected and is exploratory consistency



All 15 biological program tests per cohort have FDR >= 0.73; no author-state fraction passes FDR < 0.05.
Mean single-cell logCP10k minus discovery-frozen controls; GPC, Cilia and neuronal programs are explicitly included.
CARE contributed to publication of these signatures; this is exploratory application, not independent signature derivation.

Figure 3

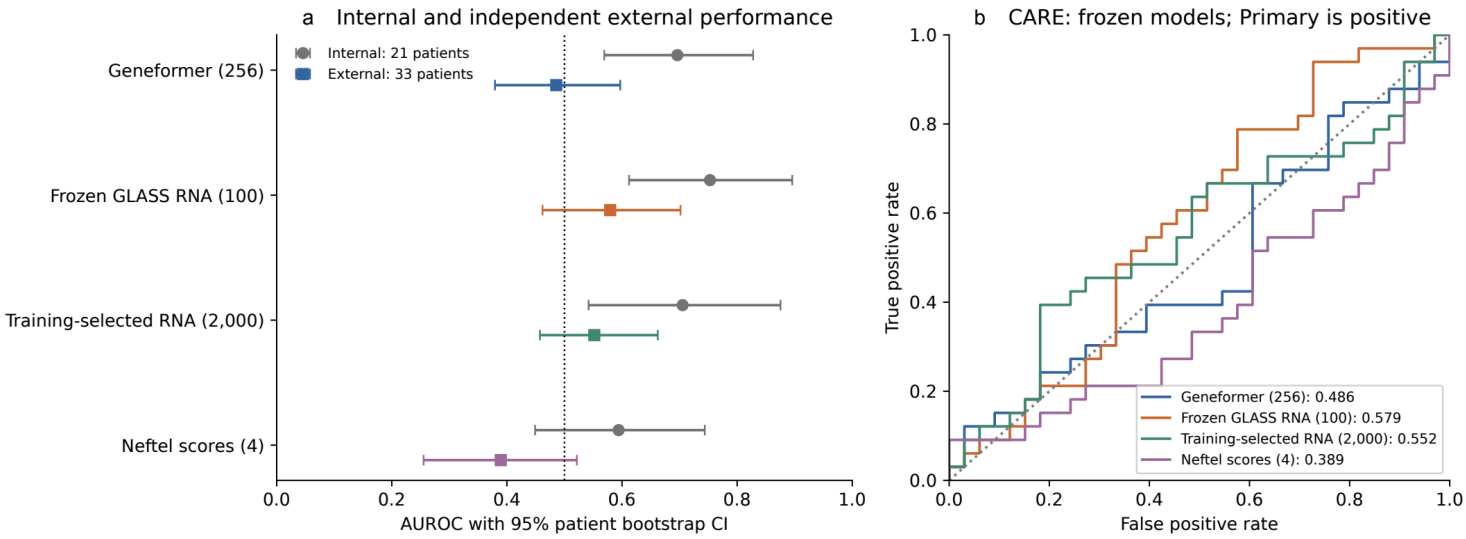
Brain-reference deconvolution with donor-held-out calibration. a–b, Estimated versus known captured-RNA weights in 104 observed CARE pseudobulks from 52 held-out donors, shown for oligodendrocytes and malignant cells. Each fold excludes both specimens of the test donor during reference and marker construction. c, Mean absolute errors for all eight compartments with CIBERSORT and weighted NNLS; the dashed threshold is 0.15. d, Mean recurrent-minus-primary weights in all 84 GLASS patients, with two-sided 95% paired t intervals. Filled markers pass observed Spearman rho ≥ 0.50 and MAE ≤ 0.15; hollow markers fail. Oligodendrocyte and vascular changes pass BH FDR < 0.05 within eight tests for each method. Failed malignant calibration precludes a tumor-purity claim. Weights are relative RNA proxies, not histological cell fractions



Held-out validation: 104 observed pseudobulks from 52 donors; signatures rebuilt without the test donor.
Filled markers in d pass both rho >= 0.50 and MAE <= 0.15; hollow markers fail. All compartments and patients remain shown.
Oligodendrocyte and vascular changes pass FDR < 0.05 in both methods. Nuclear-reference weights are not histological purity.

Figure 4

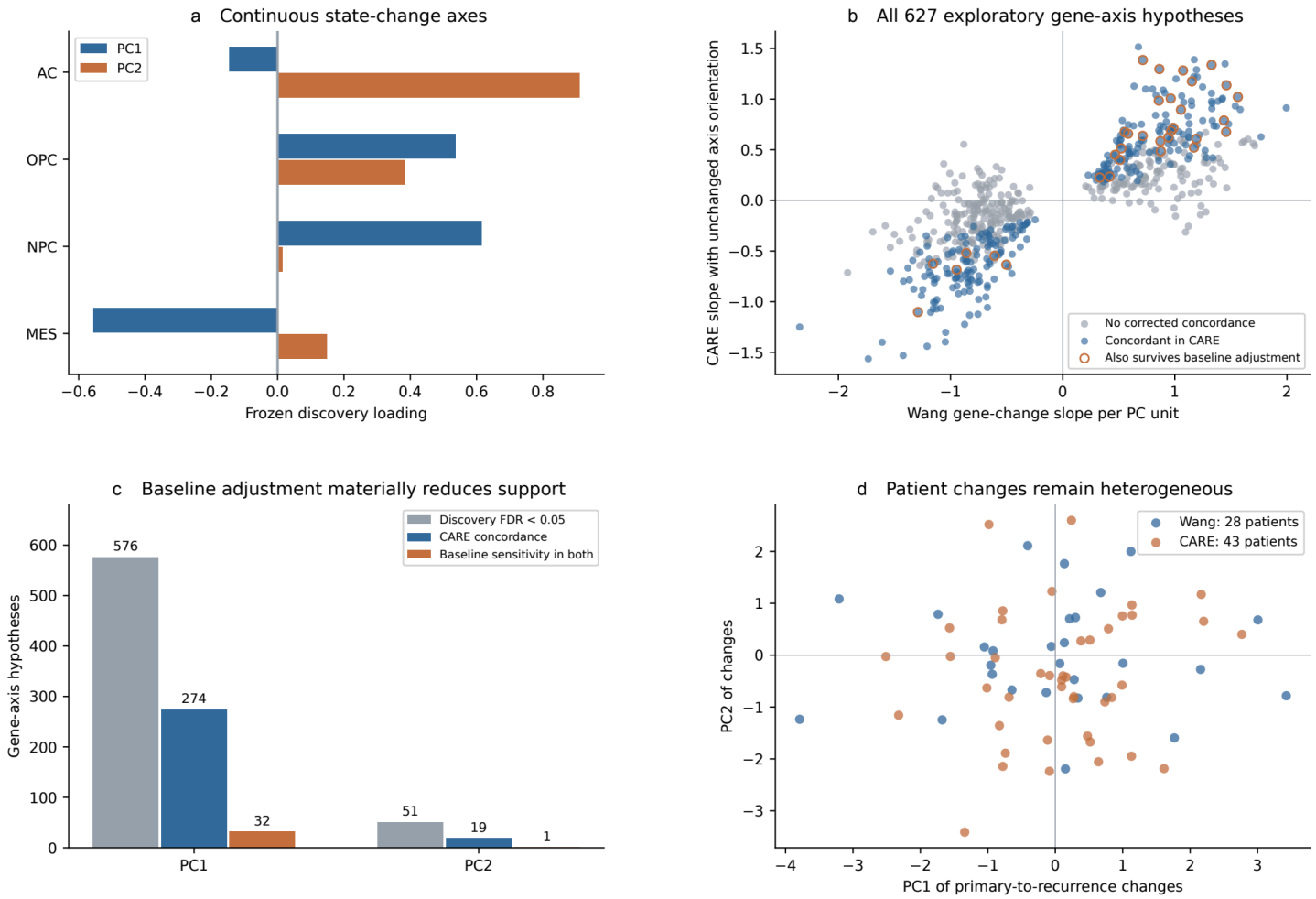
Expanded malignant programs and author-defined CARE states. a, Paired changes in published malignant metaprogram and biological-process scores in 28 discovery and 43 CARE patients. GPC, Cilia, excitatory neuronal and NRGN programs extend the four-state framework. Scores are mean per-cell log1p(CP10k) signature expression minus discovery-frozen controls. Fifteen biological programs are corrected together within each cohort; none passes FDR < 0.05. The low-coverage LQ signature is excluded by the 80% coverage rule, while technical RP and MIC signatures are tabulated separately. b, Author-defined state fractions among malignant CARE nuclei, using 43 pairs; none of ten paired state tests passes correction. Intervals are 95% paired t intervals. CARE contributed to the published signature development, so this is reference application rather than independent signature discovery



AC-state samples. No CARE refitting, label reversal or threshold search. All four external label-swap FDR values > 0.29.

Figure 5

Continuous malignant states and baseline-sensitive expression associations. a, Discovery-fitted loadings of two continuous state-change axes, projected into CARE without refitting. b, All 627 discovery-selected gene-axis slopes across 28 discovery and 43 CARE patients; blue identifies 293 same-direction CARE associations passing BH correction across 627 tests. Orange outlines indicate the 33 also surviving initial-state adjustment in both cohorts. c, The reduction after baseline adjustment, separated by axis. d, Individual paired state changes; points are patients, not cells. Gene-change models adjust for mean-log-UMI change and exclude all 3,855 defining signature/control genes. These post-validation analyses describe retrospective co-variation and do not establish distinct subtypes, baseline prediction, or causal targets



3,855 signature/control genes excluded; patient-level logCPM changes adjusted for change in mean log UMI.
293 / 627 show same-direction corrected CARE association; 33 also pass baseline-state adjustment in both cohorts.
Axes are retrospective and programs are associative: these counts are not new validated treatment targets or baseline predictors.

Figure 6

Internal and external classifier performance. a, AUROC and patient-bootstrap 95% intervals for four fixed classifiers in 21 discovery and 33 CARE AC-state pairs. Discovery predictions use patient-grouped held-out folds; CARE uses frozen full-discovery fits without refitting. b, CARE ROC curves with primary defined as the positive class. The same sampled cells and patient folds were used across classifiers. Intervals condition on saved predictions. None of the four external patient-label-swap tests passes BH correction; no label reversal or outcome-selected threshold was applied

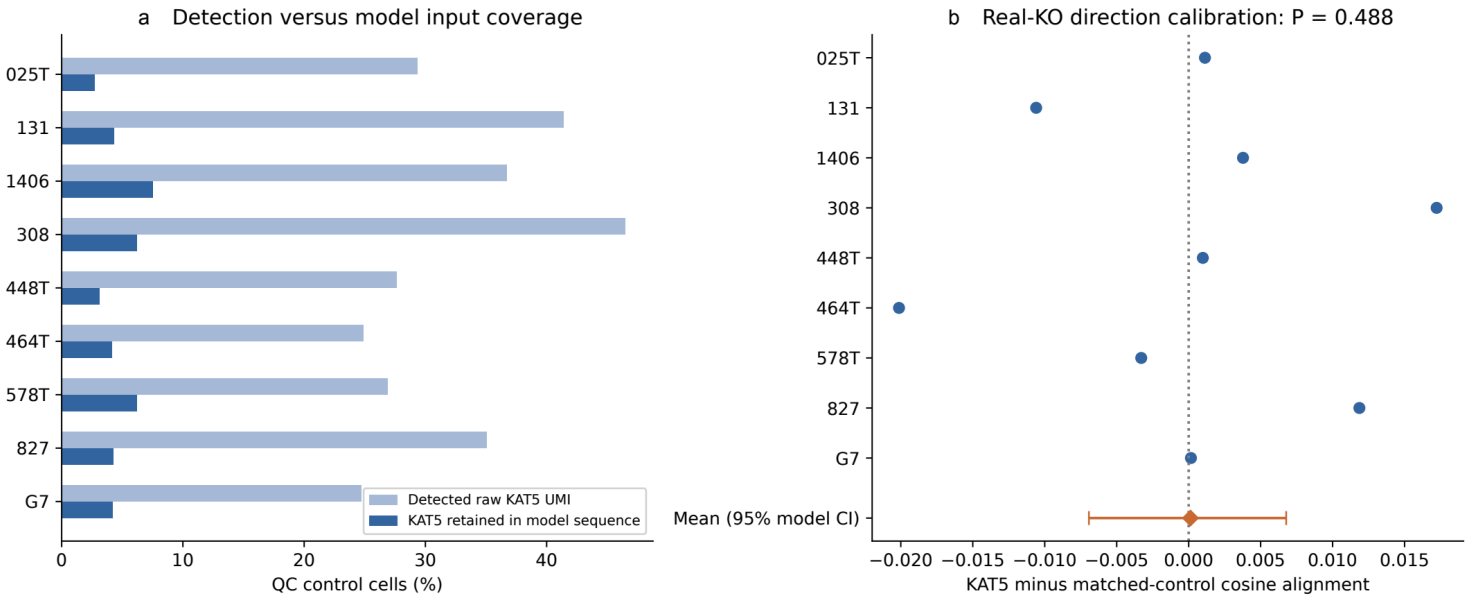


Figure 7

Calibration against real KAT5 knockout. a, Raw KAT5 UMI detection compared with retention in the Geneformer rank-truncated sequence, across nine matched in-vitro cell models. b, Model-level alignment of virtual KAT5 deletion with real KO-minus-control embedding direction, after subtracting matched-token deletion-control alignment. The orange point is the mean and its 95% model-bootstrap interval; the exact model sign-flip test gives P = 0.488. Expanded target-positive sampling is a labeled coverage sensitivity, with at most 64 eligible control cells per model. These are nine cell models, not nine clinical recurrences, and the endpoint does not establish therapeutic benefit

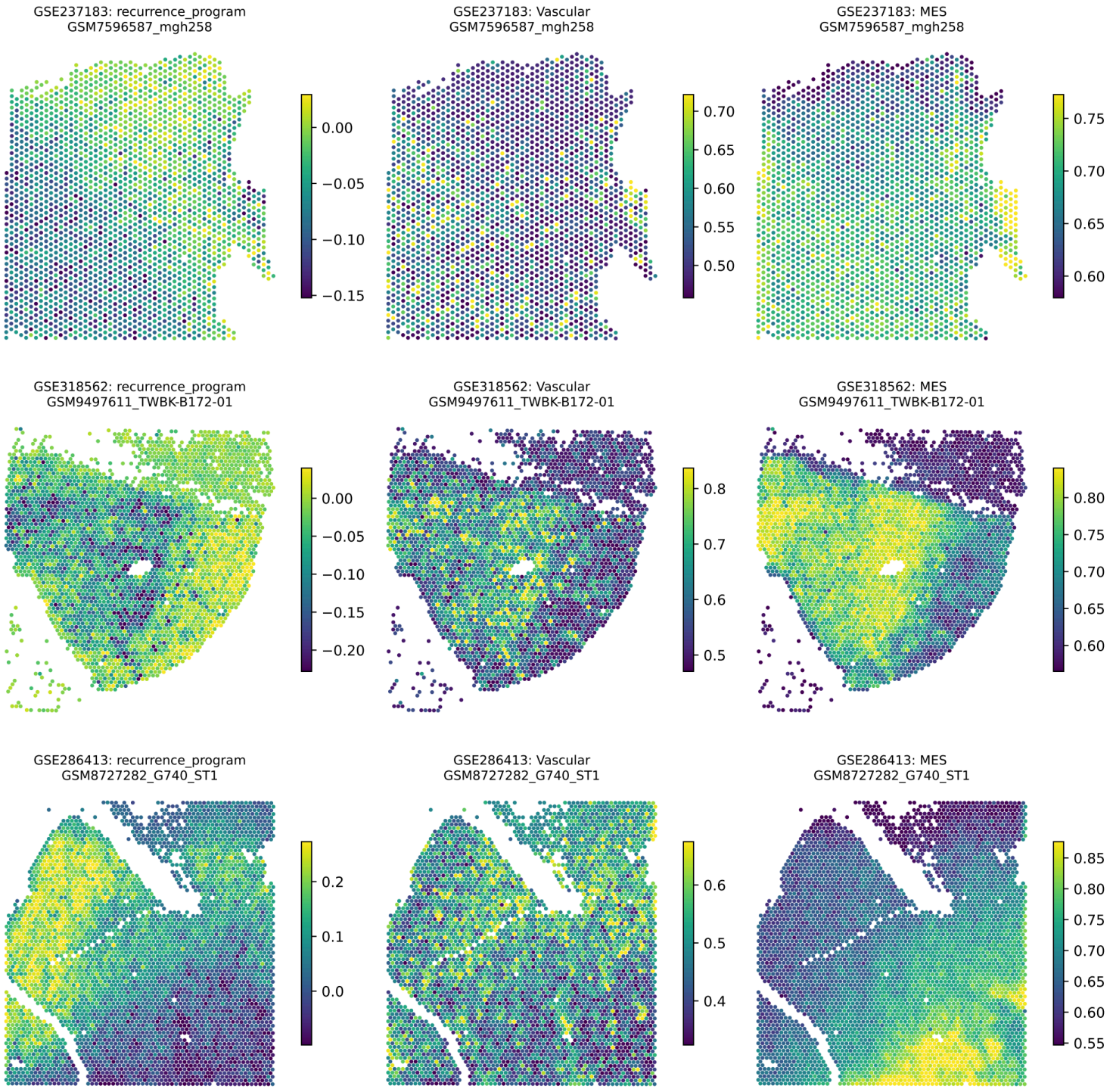


Figure 8

Spatial context of the frozen program. Deposited physical-coordinate maps for the frozen program, vascular marker score, and MES score in one illustrative section from each spatial cohort. Source section identifiers and separate color scales are shown. The analyzed cohorts contain five, nine, and two patients, respectively; inference uses patient summaries, with the two-patient cohort descriptive only. Maps are not histological segmentations or paired longitudinal spatial measurements. Program genes were removed from comparison marker sets before association testing

Supplementary Files

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

- SupplementaryCode1.zip
- SupplementaryData6.zip
- SupplementaryData1.zip
- SupplementaryData3.zip
- SupplementaryData2.zip
- SupplementaryData4.zip
- SupplementaryData5.zip
- Supplementaryinformation.pdf
- SupplementaryData7.zip