Supplementary Methods and Reporting Notes

This document provides the detailed methods and interpretation notes for the frozen S09 analyses.

Supplementary Methods

S1. Data provenance and cohort roles

All analyses used the frozen S05, S08 and S09 contracts included in the release package. GSE294300 was the primary independent validation cohort for the adjacent-normal to primary-tumor paired endpoint. GSE315534, GSE178318 and Moorman 2024 were retained as separate progression-concordance cohorts. Cohort roles were assigned before interpretation of the corresponding validation results and were not changed during figure or manuscript preparation.

The patient or donor was the inferential unit. Cell-level records were used for matrix structure checks, coverage, scoreability and low-cell flags. They were not treated as independent biological replicates. The package retains accession-level manifests, barcode maps, frozen contracts, handoff reports, execution logs and SHA-256 checksums. Large public matrices and raw archives are not duplicated in the package; the source-data index identifies the original repositories and available checksums.

S2. Input layers and matrix handling

The GSE294300 analysis used a provisional barcode-rank-derived candidate layer. The candidate input contained 62,700 feature rows and 36 sample matrices. All matrices passed the declared structural checks. The layer was retained as a reproducible candidate input and was not described as equivalent to an author-provided filtered matrix. Raw counts remained separate from normalized and summarized layers.

GSE315534 used deposited filtered 10x triplets. GSE178318 used the deposited aggregate 10x matrix after recovery of barcode boundaries. Moorman 2024 used the public epithelial object and its raw count layer, with the published malignant-cell label and nonmalignant epithelial reference retained for the selected donors. These input layers were not harmonized by post hoc filtering after the results were inspected.

For GSE178318 and GSE315534, the epithelial marker gate included EPCAM, KRT8, KRT18, KRT19, KRT20, MUC1, MUC13, CEACAM5, CEACAM6 and SOX9. A candidate cell required at least two marker hits and a minimum of 1,000 UMI under the relevant contract. The GSE178318 COL17 liver-metastasis block contained 60 marker-gated cells and was retained as a prespecified low-count sensitivity observation.

S3. State-program scoring

The two state programs were frozen before S09 validation: ABSORPTIVE_BEST4_CA4 and SECRETORY_MUC2_TFF3. For each sample, the program score was the arithmetic mean of log1p counts-per-10,000 values across frozen genes present in the input matrix. A score was reported only when at least three genes from the relevant program were present. This minimum was applied before patient-level aggregation and was not modified after validation.

The programs were analyzed in parallel. No composite score was formed, and one program was not used to substitute for a missing or low score in the other. Where cell-level mapping was required, a cell was assigned to the highest-scoring module only when its margin over the

next-highest module reached 0.05. Cells below this margin were recorded as abstentions. This rule was used to preserve ambiguity rather than force a categorical state label.

S4. GSE294300 paired endpoint

The primary GSE294300 endpoint included 16 complete paired patients with adjacent-normal and primary-tumor samples. Patients 1 and 17 formed a prespecified sensitivity layer because their sample-level candidate floor was not met; they remained outside the primary denominator. For each patient and module, the primary-tumor score minus the adjacent-normal score was calculated. The estimand was the mean of these patient-specific paired differences.

Uncertainty was estimated by complete-patient bootstrap resampling. Each replicate resampled the 16 patients with replacement and recomputed the mean paired difference. The 2.5th and 97.5th percentiles defined the 95% interval. Cell counts and scoreable-cell counts were reported as coverage audits and were not used as weights. No pooled-cell p-value or cell-level significance test was calculated.

S5. Trajectory execution audit

The formal trajectory audit tested PAGA and Palantir in Python and Slingshot and Monocle3 in R. The frozen resource layer specified complete-patient resampling, a per-sample cap of 2,000 cells, 2,000 highly variable genes, 15 principal components and no dense count conversion. The formal audit used 10 patient-bootstrap replicates plus an initial execution record, yielding 11 records in the handoff tables. Normal-cell anchors were used under the frozen root rule. Root selection was not changed after validation results were inspected.

Execution outcomes were classified before reviewing the results. A missing root anchor or fewer than 50 anchor cells produced an abstention. Non-finite pseudotime produced an invalid result. Missing branch output produced an abstention, and software errors were retained as hard failures. Failed records were not removed or repaired. The audit therefore reports structural pass, abstention, invalid-result and hard-failure counts rather than a software accuracy ranking.

Root-sensitivity analyses varied the declared resource layer and retained rank-agreement summaries only when a valid root and finite output were available. These analyses assess computational dependence and reproducibility of execution. They do not identify a biologically correct root or establish a lineage sequence.

S6. Progression-concordance analyses

GSE315534 included three patients with complete normal, primary-tumor and liver-metastasis chains. Both adjacent edges were summarized at the patient level. GSE178318 included six primary-tumor and liver-metastasis pairs. The COL17 low-count pair remained in the all-pair descriptive analysis and was highlighted in the sensitivity figure. Moorman 2024 included six complete donors with normal, primary and metastasis arms; the primary-to-metastasis edge was the progression comparison and the normal arm supplied descriptive context.

Each progression cohort used 10,000 complete-patient or complete-donor bootstrap resamples with the seed specified in its frozen contract. Intervals were calculated within cohort. No pooled estimate, meta-analysis, direction vote or cross-cohort composite was generated. Treatment strata, low-cell flags and malignancy-evidence tiers were retained as context rather than used to post hoc explain or adjust the observed estimates.

S7. Evidence grading

CNV, marker and malignancy evidence were graded separately from expression-state evidence. GSE294300 had auxiliary genomic-consistency evidence; GSE315534 had supportive proxy evidence; GSE178318 had restricted or uncertain evidence with a PBMC proxy reference; and Moorman 2024 provided a stronger public malignant-cell label for the selected donors. Limited evidence was not treated as negative evidence. None of these evidence tiers was used to define a direction, lineage or conversion probability.

S8. Reproducible execution

Formal analysis and rendering used WSL Ubuntu-24.04 with the project cast-py and cast-r environments on CPU. Run identifiers, frozen configuration paths, seeds, worker settings, logs and output locations are preserved in the package provenance. The package checksum manifest is regenerated after each release edit. A user can reconstruct the reporting environment by retrieving the public source inputs listed in source_data/, installing the recorded environment, and executing the scripts in code/ without changing the frozen contracts.

Supplementary Reporting Notes

Interpretation of the primary endpoint

The GSE294300 endpoint is a paired comparison of two tissue contexts using frozen expression-state summaries. It estimates an average patient-level score difference. It does not estimate the probability that an individual cell changes state, and it does not require the two tissues to represent a developmental sequence.

The absorptive-program estimate was -0.20474 with a complete-patient bootstrap 95% interval of -0.26994 to -0.14409. The secretory-program estimate was +0.05494 with a 95% interval of -0.02083 to +0.15804. These values are reported at the patient level and should not be replaced by cell-level summaries.

Interpretation of trajectory outputs

Trajectory figures report execution readiness, finite-output coverage, graph summaries and failure handling. A finite pseudotime value indicates that a method returned a numerically usable output under a specified root and resource layer. It does not establish a patient history, a biological lineage or a conversion probability. Abstentions, invalid results and hard failures are part of the reproducibility record.

Interpretation of progression cohorts

GSE315534, GSE178318 and Moorman 2024 were analyzed as independent descriptive cohorts. Their tissue edges, patient or donor denominators, treatment contexts, cell coverage and malignancy evidence differ. Cohort-specific deltas and intervals are therefore reported without pooling or a direction vote. The absence of a common direction is an external-generalization result that identifies where stronger independent evidence is needed.

Figure and table crosswalk

Figure 1 and Supplementary Figure 1 document cohort roles and denominators. Figure 2 documents the frozen state-program foundation. Figure 3 reports the GSE294300 paired endpoint. Figure 4 reports trajectory execution and failure categories. Figure 5 and

Supplementary Figure 6 report separate progression estimates. Supplementary Figures 2–5 provide input, malignancy-evidence, root-sensitivity and low-count audits. Supplementary Tables 1–6 provide machine-readable denominators, quality-control counts, evidence tiers, execution outcomes and patient- or donor-level deltas.

Release scope

This supplement describes the frozen reporting package and is not a new analysis. No feature, threshold, cohort role, root rule or parameter was changed in response to the validation results. The package is intended to support independent review, reproducible rerendering and extension with future cohorts.