SUPPLEMENTARY FIGURES

Jones*, Sun* et al. **Supplemental Figure 1**

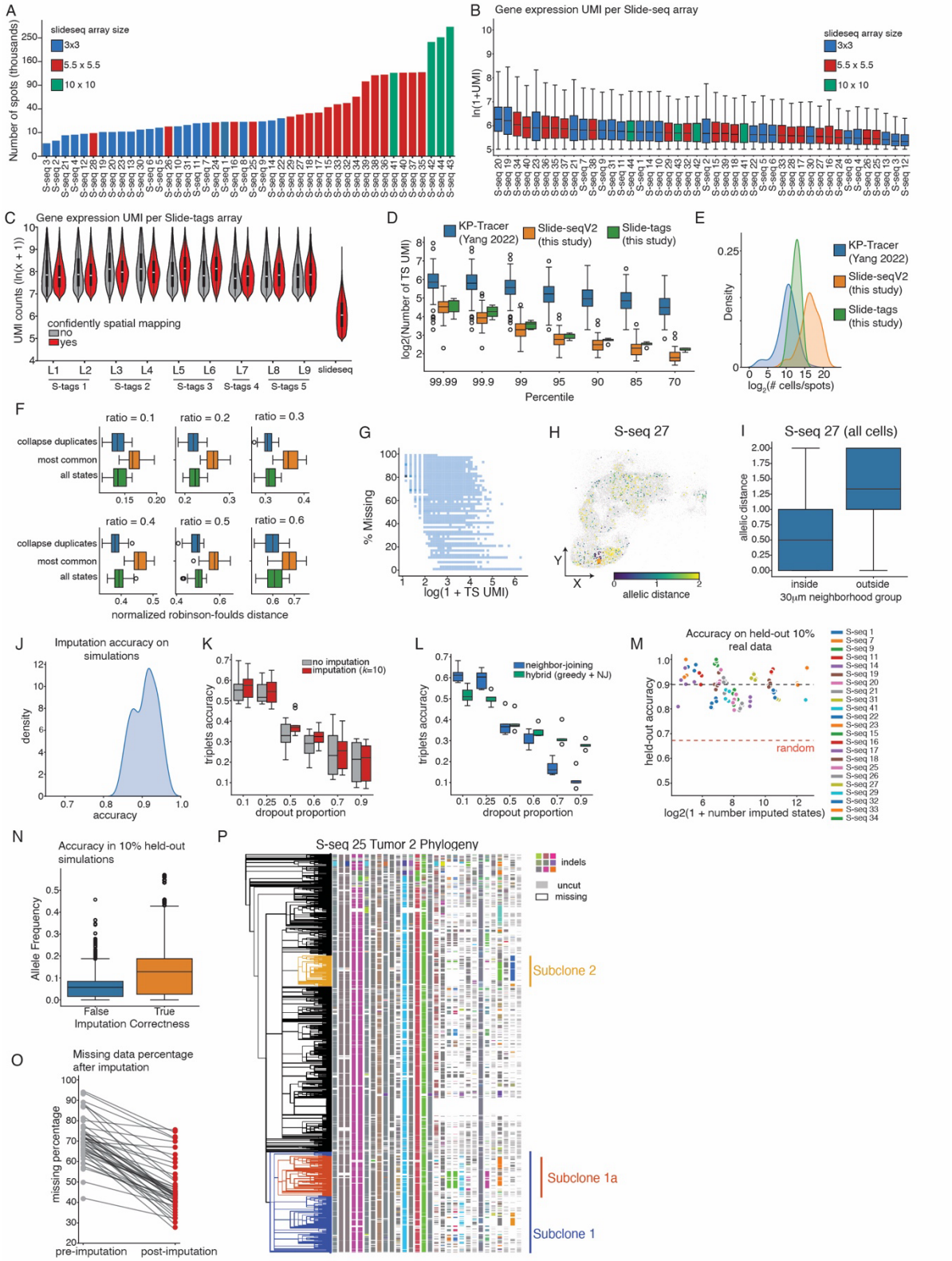

Figure S1. Characterization of spatial-lineage platform and benchmarking of computational approaches. Related to Figure 1.

(A) Number of spots that pass quality-control for all Slide-seq array. 3mm, 5mm, and 1cm arrays are uniquely colored.

(B) Number of gene expression UMIs for each Slide-seq array. Ln(1+UMI) is reported for each dataset. 3mm, 5mm, and 1cm arrays are uniquely colored.

(C) Number of gene expression UMIs for each Slide-tags array, and one representative Slide-seq array. Each array is sequenced across multiple 10X libraries; assignment of 10X library to array is annotated. Distributions are split between cells that are confidently mapped and those that are not. Ln(1+UMI) is reported.

(D) Distribution of number of target-site UMIs marking the top *X* percentile for whole-cell (KP-Tracer), Slide-seq, or Slide-tags datasets. Ln(1+UMI) is reported.

(E) Distribution of number of observations (cells or spots) that pass target-site quality-control in whole-cell (KP-Tracer), Slide-seq, or Slide-tags datasets. Log_2 of the number of observations is reported.

(F) Normalized Robinson-Foulds reconstruction error for simulated trees with increasing ratios of pooled cells and different pre-processing techniques. A ratio of *p* indicates that simulated lineage-tracing data of *p*% of cells are combined into a single observation to simulate multiple-cell capture in spatial transcriptomics (**Methods**).

(G) Relationship between percentage of missing lineage-tracing data in a cell or spot and the log-number of UMIs (ln(1+x)) for Slide-seq and Slide-tags data.

(H) Representative example of spatial coherence of lineage-tracing data on S-seq 27. For a selected spot (shown as a star), normalized allelic distance is reported for all spots with confident lineage-tracing data. Allelic distance is normalized between 0 and 2.

(I) Distribution of allelic distances to spots within a $30\mu m$ neighborhood of a spot versus outside this neighborhood. Distribution over all spots in S-seq 27 is reported.

(J) Distribution of spatial imputation accuracy in lineage-tracing data simulated on a two-dimensional array.

(K) Triplets-correct accuracy of reconstructed phylogenies simulated on a spatial array for various amounts of missing data rates, with and without spatial imputation.

(L) Triplets-correct accuracy of reconstructions with modified Neighbor-Joining and hybrid Cassiopeia-Greedy / Neighbor Joining algorithms for data simulated on a spatial array with various amounts of missing data, after spatial imputation.

(M) Accuracy of spatial imputation and number of imputed states after holding-out 10% of all lineage-tracing data in Slide-seq datasets. Datasets where at least 10 imputations are made are shown. Median accuracy of random predictions is reported in a red dashed line.

(N) Allele frequency of held-out data in a given tumor binned by imputation correctness.

(O) Overview of missing data reduction across all Slide-seq datasets after five rounds of spatial-imputation.

(P) Phylogeny and lineage tracing heatmap of tree reconstructed in **Figure 1E**. Subclones of interest are annotated in the same colors as in **Figure 1E**. Unique colors of the heatmap indicate unique insertions or deletions (“indels”), white indicates missing data, and gray colors indicates no indel detected.

Jones*, Sun* *et al.* **Supplemental Figure 2**

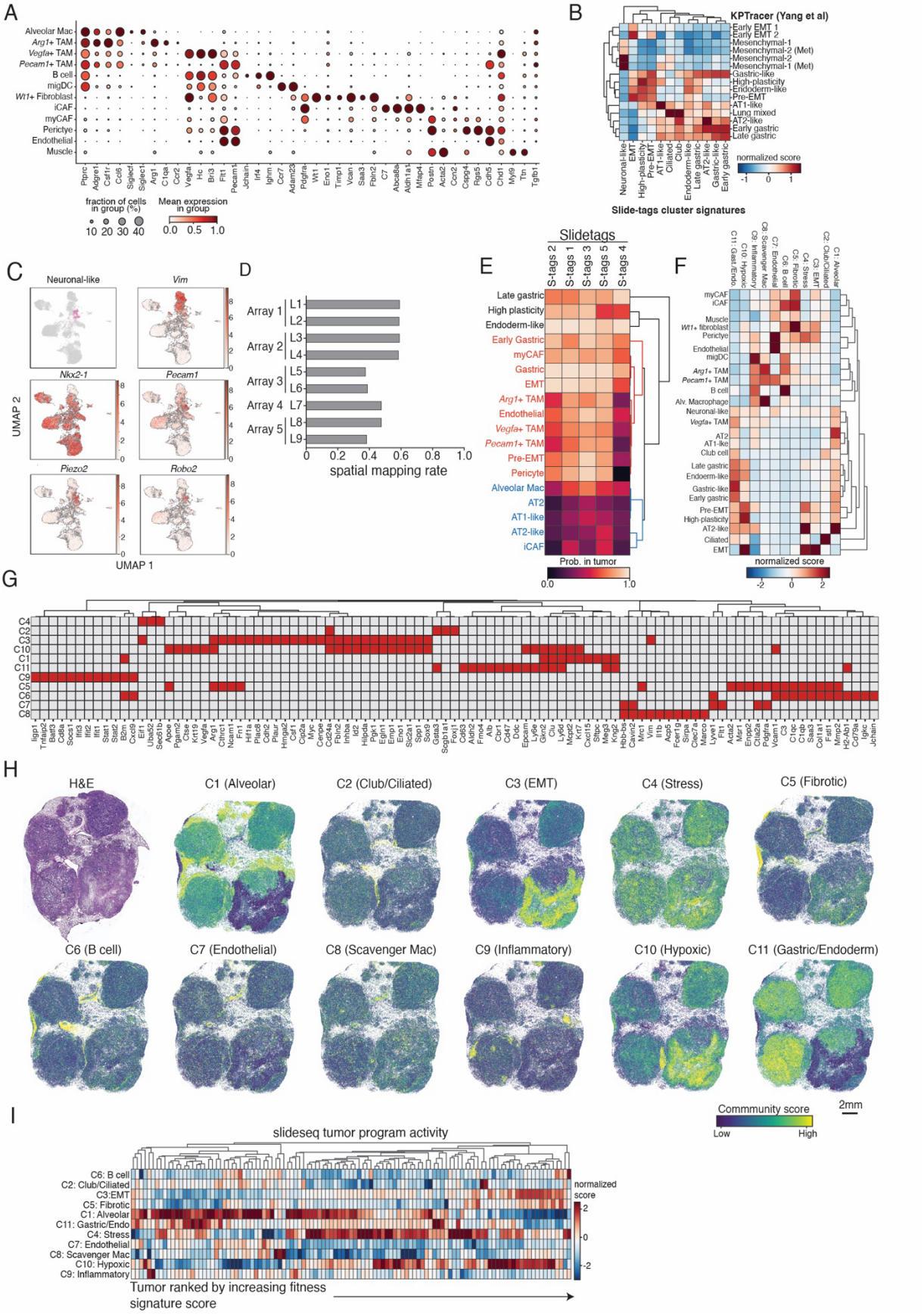

Figure S2. Profiling of cell types and spatial communities underlying tumor progression. Related to Figure 2.

(A) Summary of gene markers for each stromal cell population identified in Slide-tags. Each row corresponds to a stromal or immune cell-type cluster and each column corresponds to a marker gene. Dot size indicates the proportion of cells expression that gene, and color indicates the average gene expression value (unit scaled between 0 and 1).

(B) Clustered heatmap of transcriptional score of marker genes identified from Slide-tags data of tumor and epithelial cell types applied to previous KP-Tracer data. Scores are Z-normalized.

(C) Annotation of Slide-tags tumor and epithelial UMAP projection with the Neuronal-like cell-type, and log-normalized gene expression patterns of selected genes: *Vim*, *Nkx2-1, Pecam1, Piezo2,* and *Robo2.*

(D) Proportion of cells that are confidently mapped in each Slide-tags array.

(E) Proportion of cells for each cell type that are found within the tumor boundary across Slide-tags arrays.

(F) Clustered heatmap of transcriptional scores for each spatial community, identified from Hotspot analysis of Slide-seq data, for each Slide-tags cell type cluster. Scores are Z-normalized.

(G) Clustered heatmap showing selected genes for each spatial community. Red colors indicate that a gene is found within that module.

(H) Community scores for each spatial community and paired H&E for a representative Slide-seq community.

(I) Clustered heatmap of community scores for each tumor in the Slide-seq dataset ordered by increasing fitness signature scores. Scores are Z-normalized.

Jones*, Sun* *et al.* **Supplemental Figure 3**

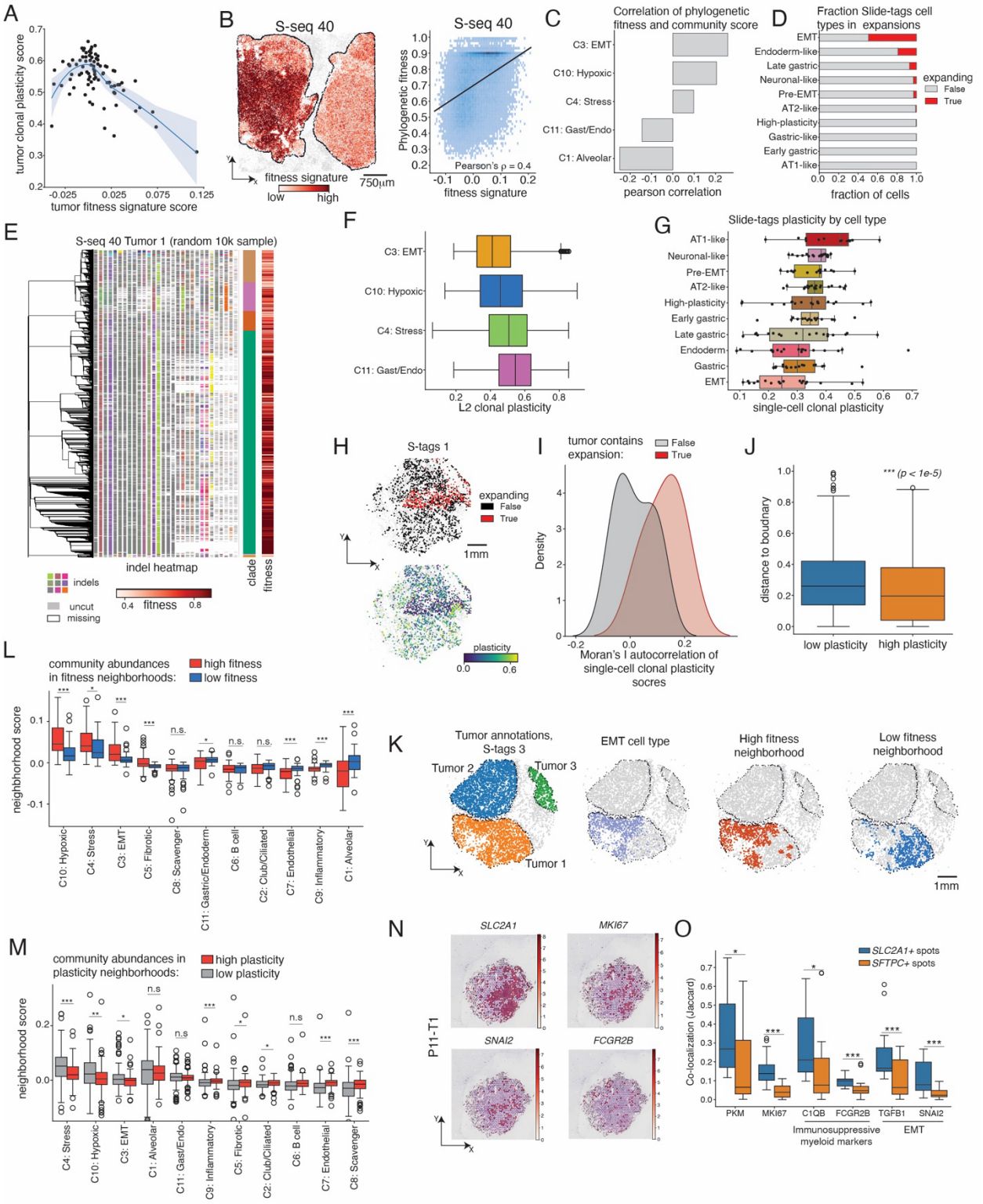

Figure S3. Characterization of subclonal tumor and microenvironmental dynamics. Related to Figure 3.

(A) Joint distribution of mean tumor clonal plasticity and fitness signatures across Slide-seq datasets.

(B) Relationship between phylogenetic fitness, estimated from inferred trees, and transcriptional fitness signature score (Pearson's correlation = 0.4)

(C) Correlation of phylogenetic fitness, estimated from inferred trees, and community scores for cancer-associated communities (C1: Alveolar; C3: EMT; C4: Stress; C10: Hypoxic; and C11: Gastric/Endoderm). Correlations are ordered in decreasing order.

(D) Fraction of cells found in expanding regions of Slide-tags phylogenies, summarized for each cancer cell-type.

(E) Reconstructed phylogeny and lineage tracing heatmap of representative tumor presented in **Figure 3A-B**. Unique colors of the heatmap indicate unique insertions or deletions ("indels"), white indicates missing data, and gray colors indicates no indel detected. Color bars indicate the subclonal clade and fitness, identical to those reported in **Figure 3A-B**.

(F) Distribution of L2 clonal plasticity (**Methods**) quantified in Slide-seq phylogenies summarized across spots annotated by cancer-dominated communities.

(G) Distribution of single-cell clonal plasticity scores computed in Slide-tags phylogenies, stratified by cancer cell-types, and reported across tumor-array combinations.

(H) Representative spatial localization of phylogenetic expansion (top) and single-cell clonal plasticity scores (bottom) in a single Slide-tags array (S-tags 3). Scale bar indicates 1mm.

(I) Distribution of autocorrelation values, computed by Moran's I, of single-cell clonal plasticity scores for tumors with or without expansions. Higher autocorrelation values indicate that values have higher spatial coherence. Autocorrelations are reported across all Slide-tags datasets.

(J) Distance to nearest non-tumor cell (i.e., tumor boundary) for high- and low-plasticity cells across all Slide-tags arrays. Cells with high-plasticity are closer to the tumor boundary ($p < 1e\text{-}5$, wilcoxon rank-sums test).

(K) Representative example demonstrating the stratification of neighborhoods of high- and low-fitness cells in Slide-tags data, and comparison to spatial localization of the EMT state. Scale bar indicates 1mm.

(L) Distribution of average community scores in 30μm neighborhoods of high- or low-fitness spots in Slide-seq data. Each observation corresponds to a tumor. Significance is

indicated above each comparison (*n.s.* = not significant; * = *p<0.1*; ** = *p<0.05;* *** = *p < 0.01*).

(M) Distribution of average community scores in 30μm neighborhoods of high- or low-plasticity spots in Slide-seq data. Each observation corresponds to a tumor. Significance is indicated above each comparison (*n.s.* = not significant; * = *p<0.1*; ** = *p<0.05;* *** = *p < 0.01*).

(N) Representative example of spatial log-normalized gene expression values for selected genes in a human lung adenocarcinoma (LUAD) spatial transcriptomics dataset (see **Methods**).

(O) Overall distribution of log-normalized gene expression values of selected genes co-expressed in hypoxic (*SLC2A1+*) or epithelial-like (*SFTPC+*) tumor spots across all LUAD samples in dataset shown in (M). Ontologies are indicated underneath genes. Hypoxia+ spots have higher expression of proliferation (*MKI67)*, immunosuppressive myeloid (*FCGR2B* and *C1QB)* and EMT (*SNAI2* and *TGFB1*) markers. Statistical significance between gene expression distributions is shown for each comparison (*n.s.* = not significant; * = *p<0.1*; ** = *p<0.05;* *** = *p < 0.01*).

Jones*, Sun* *et al.* **Supplemental Figure 4**

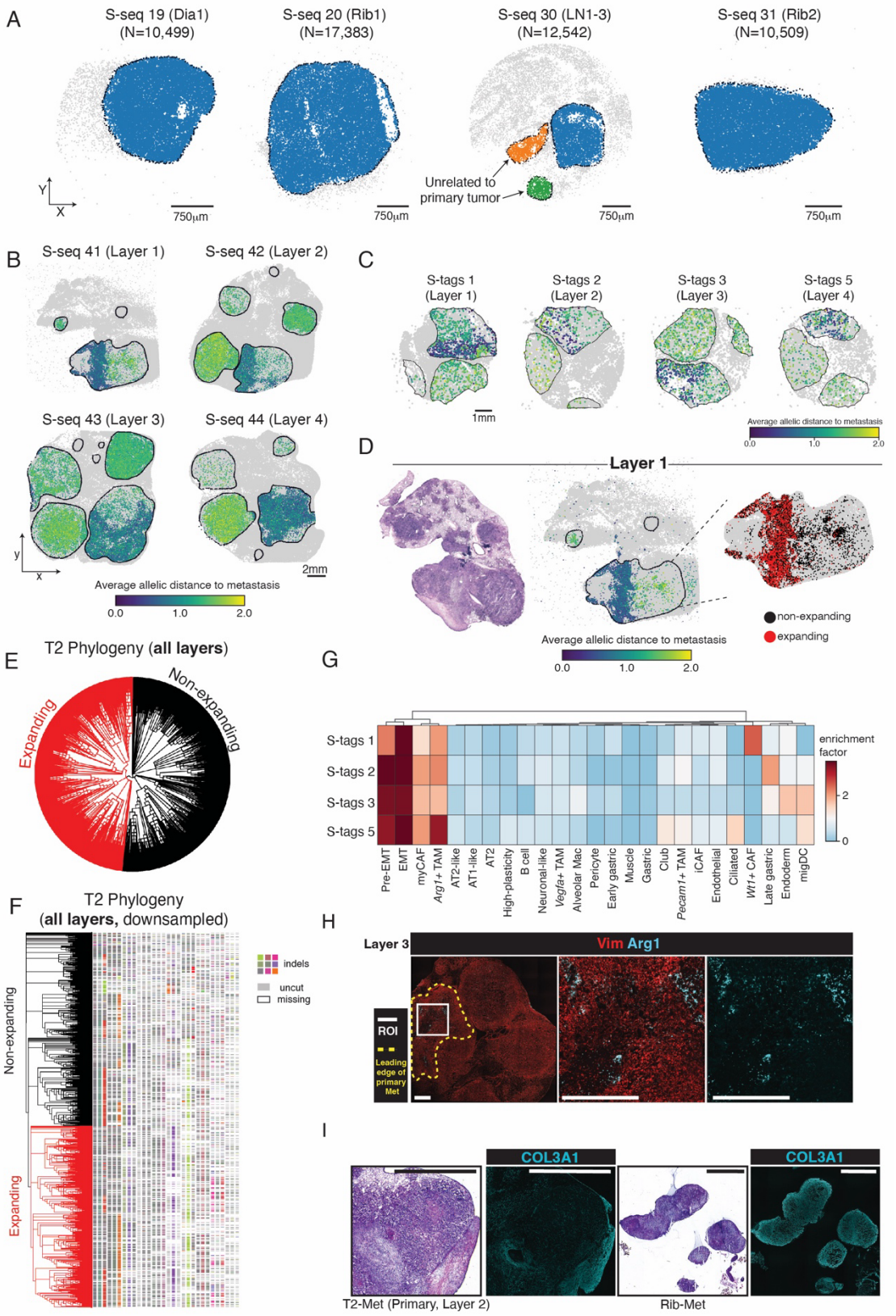

Figure S4. Profiling of metastases and microenvironmental evolution during metastasis. Related to Figure 4.

(A) Summary of metastases identified in Slide-seq spatial transcriptomics dataset. Each sample is annotated the metastatic site (LN: lymph node; Dia: Diaphragm). Two metastases in the lymph node (S-seq 30) were not found to be related to the primary tumor studies in **Figure 4** and thus removed from comparative analysis.

(B) Spatial projection of allelic distances for each spot with lineage-tracing data to consensus metastatic parental allele across all four layers profiled in Slide-seq. Allelic distances are normalized between 0 and 2.

(C) Spatial projection of allelic distances for each cell with lineage-tracing data to consensus metastatic parental allele across paired Slide-tags arrays. Allelic distances are normalized between 0 and 2.

(D) H&E staining, spatial mapping of allelic distances to consensus metastatic parental allele state, and spatial localization of phylogenetic expansion for T2 in representative dataset. Allelic distances are normalized between 0 and 2.

(E) Reconstructed phylogeny of T2 from all layers with phylogenetic expansion annotated in red.

(F) Reconstructed phylogeny and lineage tracing heatmap of T2 from all layers. Unique colors of the heatmap indicate unique insertions or deletions (“indels”), white indicates missing data, and gray colors indicates no indel detected. Clades participating in expansion shown in (E) are shown in red.

(G) Clustered heatmap of enrichments of cell type abundances in spatial neighborhoods of cells related to metastases in Slide-tags arrays.

(H) Immunofluorescence imaging of ARG1 and VIM in a section of the tumor-bearing lung close to Layer 3. Leading edge of the metastasis-initiating subclone is indicated with yellow dashed line. Scale bar indicates 1mm.

(I) H&E and immunofluorescence imaging of COL3A1 in a section of the metastasis-initiating primary tumor (Layer 2) and related metastasis. Scale bar indicates 1mm.