nature genetics

Supplementary information

https://doi.org/10.1038/s41588-026-02779-5

Clonal heterogeneity in human pancreatic ductal adenocarcinoma and its impact on tumor progression

In the format provided by the authors and unedited

Supplemental results

Clonal CNV inference and validation

For each tumor with ≥100 epithelial cells (n=50), inferCNV[1] was used to infer CNVs and define malignant clones, followed by hierarchical clustering. We next constructed a phylogenetic tree for each patient sample using maximum parsimony, and overlaid inferred ancestral relationships based on the principle that descendant clones accrue additional CNVs absent in their predecessors. When two or more clones possessed mutually exclusive CNVs, we inferred the presence of an unobserved common ancestor from which these clones diverged prior to acquiring their unique CNVs (Fig. 2b,g, Suppl. Fig. 1a-b,e-f, Suppl. Fig. 2). Finally, we validated our trees by assessing Whole Genome Duplication (WGD) patterns in each sample by comparing the distribution of total UMI counts in malignant clones to those of non-malignant epithelial cells (Methods) (Suppl. Fig. 1c-d,g-h). Across all tumors, WGD was identified in 32.6% of samples, consistent with reported frequencies in the literature[2,3]. In sample P349, all clones except C02 lie along the same branch and display elevated UMI levels consistent with a shared WGD event (Suppl. Fig. 1g,h). In all other WGD-positive tumors, the WGD clone appeared as a leaf with no detectable descendants.

To validate the inferred CNVs, we performed orthogonal assessments at both the tissue and genome-wide levels. Fluorescence in situ hybridization (FISH) for MYC on matched fresh-frozen tumor sections confirmed *MYC* amplification in the two representative samples identified as *MYC*-amplified by our single-cell and spatial analyses (P332, P466.2), while a sample lacking *MYC* amplification in our dataset (P432) showed no FISH evidence of gain, supporting the specificity of our calls (Ext. Fig. 3a-d). In addition, for the 15 samples with available low-pass whole-genome sequencing (WGS), we compared pseudobulk CNVs inferred from inferCNV, CopyKAT[4], and Numbat[5] to WGS-derived CNVs. In our analysis, inferCNV showed the highest concordance with WGS, recovering 78% of WGS-called CNVs overall, including all WGS-detected deletions, outperforming CopyKAT (73%) and Numbat (46%) (Ext. Fig. 3e,f).

Clonal burden per tumor in single-cell and spatial cohorts

Across the scRNA-seq dataset, patients harbored on average 5.7 clones (range: 2–11; standard deviation (SD): 2.0; standard error of mean (SEM): 0. 30), with a mean of 409 cells per clone (range: 17–3101; SD: 505; SEM: 29.93) (Fig. 4a). In the spatial cohort, the average number of clones was slightly lower (4.4; range: 2-9; SD: 2.6; SEM: 0.80), consistent with spot level resolution capturing dominant but not all subclonal populations.

Gene set enrichment analysis in the malignant epithelial compartment

In order to investigate whether there is a link between the genetic and transcriptional profiles of the clones, we performed pathway analysis. First, we identified transcriptional clusters in the malignant epithelial cells of each sample using unsupervised, graph-based clustering. These clusters were then used for pathway enrichment analysis with gene set enrichment analysis (GSEA) (Fig. 4c). We then assessed pathway activity at the clone level to determine whether transcriptional programs align with genetic divergence. For each pathway, we tested whether any clone showed significantly elevated activity relative to all other clones in the dataset. This analysis revealed substantial heterogeneity across tumors. Pathways related to protein degradation/SUMOylation and chromatin organization were activated in the majority of tumors (45/50, and 44/50 tumors, respectively), whereas inflammatory signaling pathways were activated in only 9/50 tumors (Suppl. Fig. 4e). Within individual tumors, immune activation and cell cycle pathways were most frequently activated in more than one clone (87.5% and 86.5% of tumors, respectively), while small GTPase-related pathways were multiclonal in only 38% of tumors with pathway activation (Fig. 4d). We next examined whether pathway divergence between clones could be explained by CNV divergence. Although a small number of comparisons reached statistical significance, effect sizes were modest and no consistent trend was observed across pathways (Suppl. Fig. 4f). These results indicate that while pathway activity can differ between clones, such differences are not systematically driven by CNV divergence, and many transcriptional programs remain shared across clones.

To explore broader relationships among pathway programs, CNV burden, and transcriptomic subtype, we computed pairwise correlations across all variables (Fig. 4e). CNV burden correlated positively with the basal subtype and negatively with the classical subtype, consistent with greater genomic instability in basal tumors. Pathways related to cell cycle, translation, and protein degradation formed a cluster associated with basal identity, whereas chromatin organization, apoptosis, and small GTPase/cytoskeletal pathways showed inverse correlations with this cluster. Oncogenic signaling was inversely correlated with CNV burden, consistent with the early and ubiquitous acquisition of *KRAS* mutations, while EMT displayed a distinct pattern, correlating positively with the basal subtype and negatively with the classical subtype (Fig. 4e).

Clonal diversity shapes the interactions of clones with the TME

To quantify interaction heterogeneity, we categorized clones into three groups based on the number of interactions they exhibited with each cell type: low, intermediate, or high. Next, each patient sample was classified based on the overall interaction profile of its clones with other TME cell types. Samples were categorized as having low TME interactions if all clones were in the low or intermediate interaction groups, intermediate

TME interactions if all clones were in the intermediate group, and high TME interactions if all clones were in the high or intermediate groups. In cases where at least one clone exhibited low interactions and another clone displayed high interactions with the same cell type, the sample was classified as displaying heterogeneous interactions with this cell type (Fig. 5b). Of the 50 tumors analyzed, 40 (80%) displayed heterogeneous interactions with at least one cell type in the TME, a pattern that appeared independent of treatment or even the number of clones within a sample (Fig. 5b). Among the remaining samples, only one showed consistently high interactions across all its clones with all the cell types. The rest of the samples exhibited varying numbers of interactions between their clones and different cell types.

Heatmaps for representative six samples (Suppl. Fig. 3a-f) illustrate the interaction counts between clones and key TME cell types, along with clone sizes and cell type frequencies. These heatmaps highlight how samples can uniformly fall into the low (Suppl. Fig. 3a), intermediate (Suppl. Fig. 3b) or high (Suppl. Fig. 3c) interaction groups or exhibit distinct heterogeneity patterns (Suppl. Fig. 3d-f). For instance, some clones show strong interactions with T cells or fibroblasts, while others within the same sample interact minimally with these cell types, revealing variability in interaction profiles.

MYC is required to maintain basal identity in mouse and human PDAC models

To test our hypothesis that MYC amplification is necessary to drive evolutionary transition from the classical to the basal subtype, we knocked down MYC expression (Suppl. Fig. 7d) in two independently derived tumor cell lines from a PDAC mouse model with Ptf1a-CreERT2-induced expression of constitutive deregulated MYC and mutant KRAS (KMCE)[6], using doxycycline-inducible lentiviral shMyc (Methods).

To confirm whether *Myc* knockdown induces a reversal from a basal to a classical phenotype, we conducted GSEA using all published subtype signatures for PDAC, rather than only the Moffitt signature, as there is no direct correspondence between these gene signatures in mouse and human. In the KMCE PDAC cell lines basal subtype signatures were predominantly enriched, with the Bailey squamous and Puleo Pure-Basal signatures showing the strongest enrichment (FDR < 0.05) (Fig. 7e, red bars). In contrast, following *Myc* knockdown, the cells shifted to a classical subtype, with the Moffitt Classical, Puleo Pure-Classical, and Bailey progenitor signatures being most prominent (Fig. 7e, blue bars, Suppl. Fig. 7e). These results support that *Myc* knockdown drives a shift from a basal to a classical phenotype, and reinforce the role of *MYC* in driving the transition from the classical to a more aggressive basal phenotype in human PDAC clones.

To investigate whether the immune escape observed in human PDAC clones that underwent classical-to-basal transition is recapitulated in murine PDAC cells, we performed pathway enrichment analysis in KMCE cell lines with and without *Myc*

knockdown. Consistent with the findings from human PDAC clones, *Myc* knockdown resulted in the enrichment of immune-response related pathways (Fig. 7f). This finding suggests that MYC amplification is a prerequisite for the transition from a classical to a basal phenotype. Additionally, the restoration of immune pathway activity upon MYC loss suggests that MYC plays an immunosuppressive role.

To functionally validate the role of MYC in maintaining the basal phenotype in human PDAC too, we performed shRNA-mediated MYC knockdown in two independently derived primary human PDAC cell lines harboring MYC amplification (Suppl. Fig. 7f, Methods). In both lines, MYC suppression reduced basal markers (e.g., TP63, KRT5, S100A2, VIM) and induced expression of classical epithelial genes, including GATA6, CDH1, FOXA1, MUC1, and EPCAM (Suppl. Fig. 7g). These results are consistent with a MYC-dependent maintenance of the basal state and demonstrate that MYC loss is sufficient to shift human PDAC cells toward a classical transcriptional program. Moreover, MYC knockdown restored expression of HLA-A, B2M, HLA-DPA, and HLA-DPB (Suppl. Fig. 7h) in one of the two cell lines, mirroring the immune-pathway recovery observed in murine KMCE cells. Immune-gene analysis was not extended to the second cell line due to extremely low baseline expression of HLA class I-II genes, which precluded meaningful assessment of MYC-dependent regulation. Together, these data demonstrate that MYC is required to maintain the basal transcriptional state in both mouse and human PDAC and that MYC suppression reinstates classical identity and immune-response programs, consistent with the MYC-dependent classical-to-basal transition inferred from human clonal evolution.

Supplemental Methods

Single-cell library preparation, sequencing, and data processing

Samples were processed to isolate single cells as previously described by our group[7] before the single cell suspensions were processed by the Genome Technology Center at NYU School of Medicine using 10x Genomics Chromium single-cell 3' Reagent Kits v3, or Chromium single-cell 5' Reagent Kits v2 following manufacturer guidelines (samples sequenced after August 2021), and run on the Illumina NovaSeq 6000 system (Illumina, San Diego, CA). Following conversion of output BCL files via Illumina's bcl2fastq, the FastQ files were processed with 10x Genomics Cell Ranger v7.0.1[8] and aligned to the reference hg38/GRCh38 reference genome. Non-PCR duplicated reads that confidently mapped to the reference genome with valid unique molecular identifiers (UMIs) and barcodes were used to generate a counts matrix. Raw data were initially filtered to remove ambient RNA using SoupX v1.5.258. Seurat v5[9] and scooter[10] packages were used to perform downstream data analysis, including the identification of highly variable genes, dimensionality reduction, standard unsupervised clustering algorithms, visualization, and the discovery of differentially expressed genes. Cells with >1200 UMIs, >500 genes, and <15% mitochondrial genes, were retained. Counts data was normalized by total gene expression followed by log-transformation with a scaling factor of 10,000. scDblFinder v1.6.0[11] was used to identify and remove putative doublets and multiplets present in the data. Clusters were calculated using the first 30 PC dimensions and 30 nearest neighbors and visualized with UMAP projection.

The dimensionality of the scaled non-batch corrected data matrix of the combined data of the samples was reduced to project the cells in two-dimensional space, using principal component analysis followed by UMAP. A combination of different clustering resolutions (0.1 to 1.9) and marker gene expression were used to identify cells grouped together by classic marker genes. Markers used for both broad and granular cell types (Ext. Fig. 1a,b) were the same as those used in previous reports from our group[7]. The rest of the analysis was performed on data from each individual sample, using the cell type annotation from the combined data matrix.

Spatial transcriptomics library preparation, sequencing, and data processing

Samples were obtained via surgical resection and confirmed by a pathologist to contain sufficient tumor content. ST profiling was performed using the 10x Genomics Visium CytAssist platform on FFPE samples, following the manufacturer's protocol. After tissue sectioning and placement on Visium slides, spatially barcoded cDNA libraries were

generated using 10x Genomics Visium for FFPE Library Preparation Kits. Libraries were sequenced on the Illumina NovaSeq 6000 system.

Raw sequencing data were processed with 10x Genomics SpaceRanger v2.1.1, aligning reads to the GRCh38/hg38 human reference genome. Spatial gene expression matrices were generated by mapping transcripts to their corresponding tissue locations. Deconvolution was performed using the R package BayesPrism[58], with scRNA-seq data serving as the reference for cell type identification. Spots were filtered to include only those with >1000 UMIs, >500 detected genes, and <10% mitochondrial gene content. Downstream processing, including normalization, dimensionality reduction, clustering, and visualization, was performed using Seurat v5.

To assess the concordance between automatic cell-type annotations generated with BayesPrism and manual annotations derived from transcriptional clustering, we constructed a contingency table comparing the two annotation sources and performed a chi-square test with permutation. Specifically, statistical significance was evaluated using a chi-square test with 100,000 permutations to account for potential sparsity and unequal category sizes, and the strength of association was quantified using Cramér's V.

Quantification of sample specificity of clusters using Shannon entropy

To assess the degree of sample specificity within epithelial transcriptional clusters, we calculated normalized Shannon entropy based on the sample composition of each cluster. Following epithelial cell subsetting, Leiden clustering was performed at a resolution selected to yield at least as many clusters as patient samples (64 clusters for 62 samples), providing sufficient granularity to detect sample-specific structure. For each cluster, we computed the proportion of epithelial cells originating from each patient and used these proportions to calculate the normalized Shannon entropy H, where values close to 0 indicate enrichment for a single patient and values closer to 1 indicate a more uniform distribution across patients. Clusters were classified as patient-specific if $H<0.3$.

Whole genome sequencing and analysis

In cases where fresh frozen tissue was available, the inferred CNVs were further validated with low passage WGS (median read depth 1x). FastQ files were processed using the Seq-N-Slide analysis workflow for WES/WGS[12]. Briefly, FastQ files were trimmed with Trimmomatic v0.36 (trailing=5, sliding window=4:15, minlen=35). Next the reads were

aligned to the reference hg38 genome with BWA-MEM alignment v0.7.17, filtering for a mapping quality of at least 10. Duplicates were removed by sambamba v0.6.8. Filtered bam files were then realigned and recalibrated with GATK v3.8.1. CNVs were predicted by Control-FREEC v11.6[13] using default parameters for WGS. From this, the BedGraph formatted tracks of CNVs were plotted in a heatmap, aligned to the inferCNV profile.

Inferring whole genome duplication in tumor clones

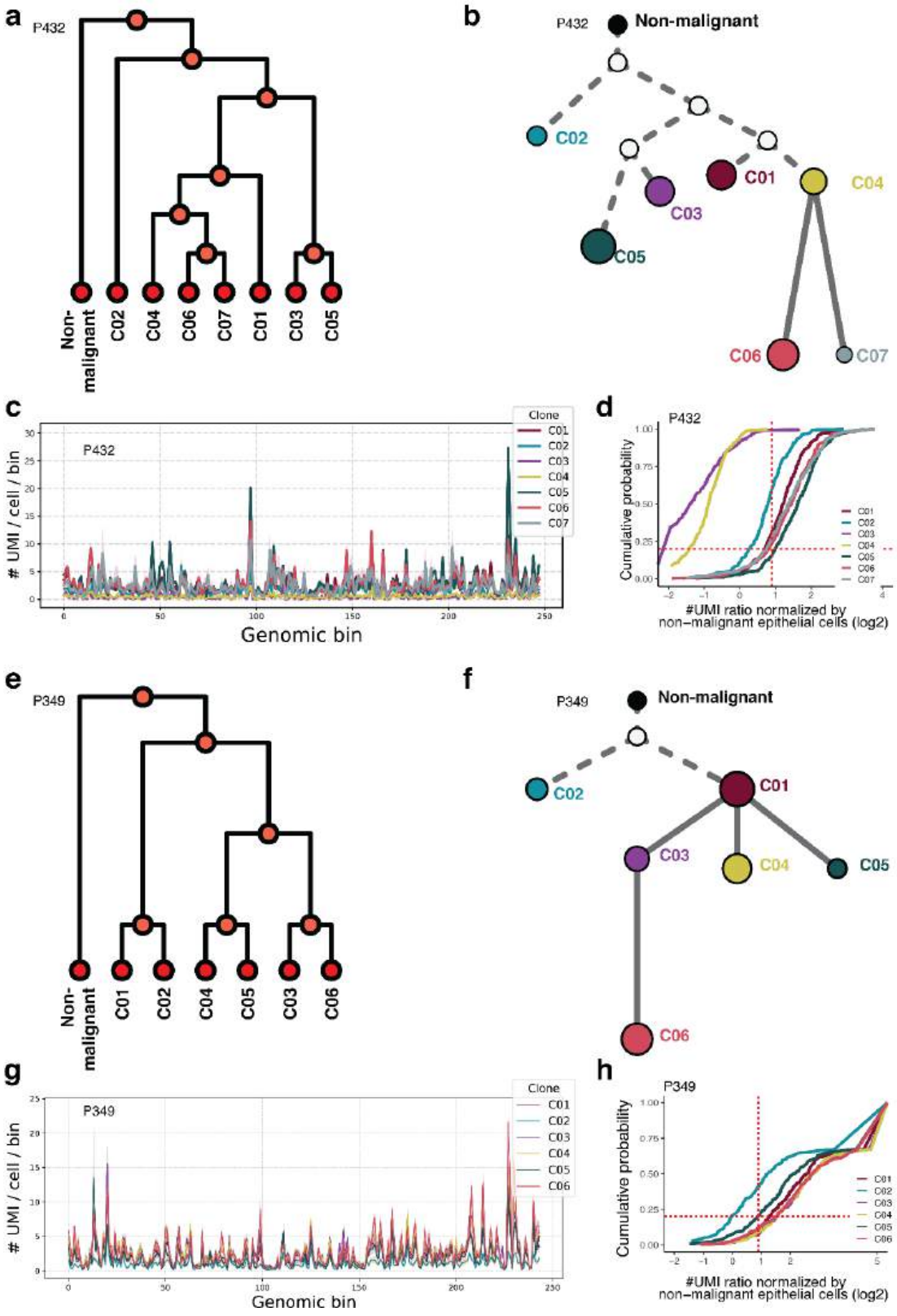
Raw scRNA-seq reads (FASTQ files) were aligned to the human reference genome (hg38) using 10x Genomics Cell Ranger count. Genomic positions were segmented into 10-Mb bins, yielding ~300 bins across the genome. For each bin, UMI counts per cell were calculated based on the paired cell barcode associated with each read. Cells were grouped by their clone assignment, and for each clone we computed the average UMI count per cell per bin. During preprocessing, we excluded (i) cells lacking aligned reads in more than 20% of bins, and (ii) bins with aligned reads in fewer than 20% of cells.

To estimate relative ploidy across tumor clones, UMI counts were normalized using non-malignant epithelial cells as a reference. For each bin, the average UMI count per cell for a tumor clone was divided by the corresponding value in the reference group. The distribution of these ratios across all bins was then used to infer relative ploidy. A clone was classified as whole-genome doubled if at least 80% of its cells exhibited a log-ratio higher than or equal to 0.9.

DNA-fluorescent in situ hybridization (FISH)

Fresh frozen (FF) pancreatic cancer patient tissues P332 and P466.2 harboring MYC amplification, and negative control P432 (without MYC amplification) were cryo-sectioned to 10 µm by the biorepository core. Sections were fixed with 4% PFA, pre-treated with 0.1M citrate buffer (pH = 6.0), and protein digested by pepsin (10 mg/ml in 0.01N HCl, pH = 2). DNA FISH was then performed using the ZytoLignt SPEC MYC/CEN8 Dual Color FISH probe (ZytoVision) following the manufacturer's protocol in a humidified chamber (ThermobirteSystem). After post-hybridization washes, slides were counterstained with DAPI and mounted with Vectashield (Vector Laboratories). Images were acquired with a Nikon CSU-W1 SoRA Spinning Disc confocal microscope equipped with a CFI Plan Apochromat Lambda D 100X NA-1.45 oil-immersion objective and Kinetix sCMOS camera. Nuclei, MYC foci, and centromere 8 (CEN8) foci were imaged by 405, 488, and 561 nm channels, respectively. The numbers of MYC and CEN8 foci in each nucleus were quantified with ImageJ.

Supplemental Figures

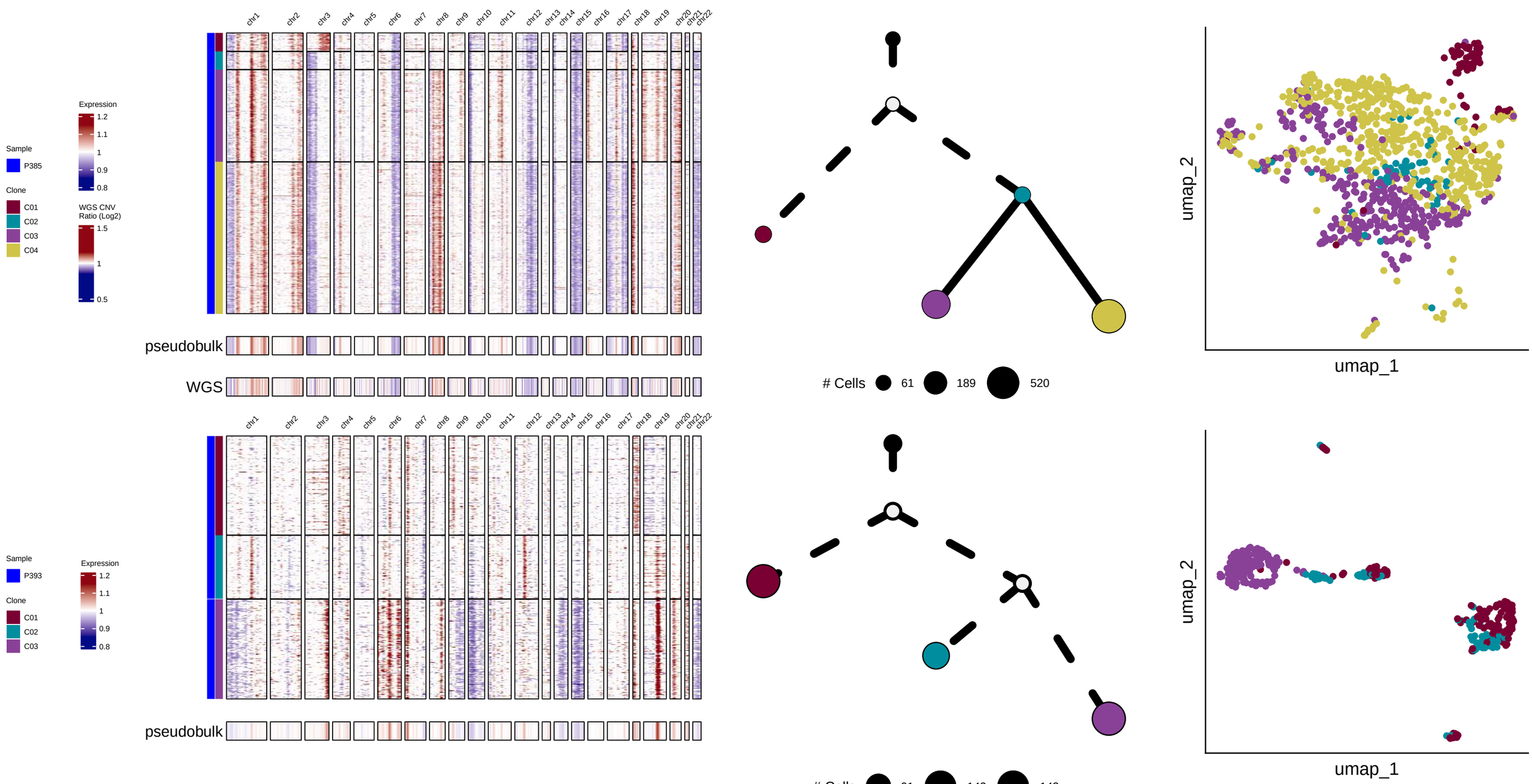


Suppl. Figure 1. Phylogenetic reconstruction. a-d, Results for sample P432. a, Maximum-parsimony output showing inferred evolutionary relationships among clones. b, Phylogenetic tree incorporating ancestral relationships between clones. c, Number of UMIs per clone across genomic bins. d, Cumulative distribution of log2-transformed UMI ratios of malignant clones relative to non-malignant epithelial cells, used to assess whole-genome doubling (WGD). e-h, Results for sample P349. e, Maximum-parsimony output showing inferred evolutionary relationships among clones. f, Phylogenetic tree incorporating ancestral relationships between clones. g, Number of UMIs per clone across genomic bins. h, Cumulative distribution of log2-transformed UMI ratios of malignant clones relative to non-malignant epithelial cells, used to assess WGD.

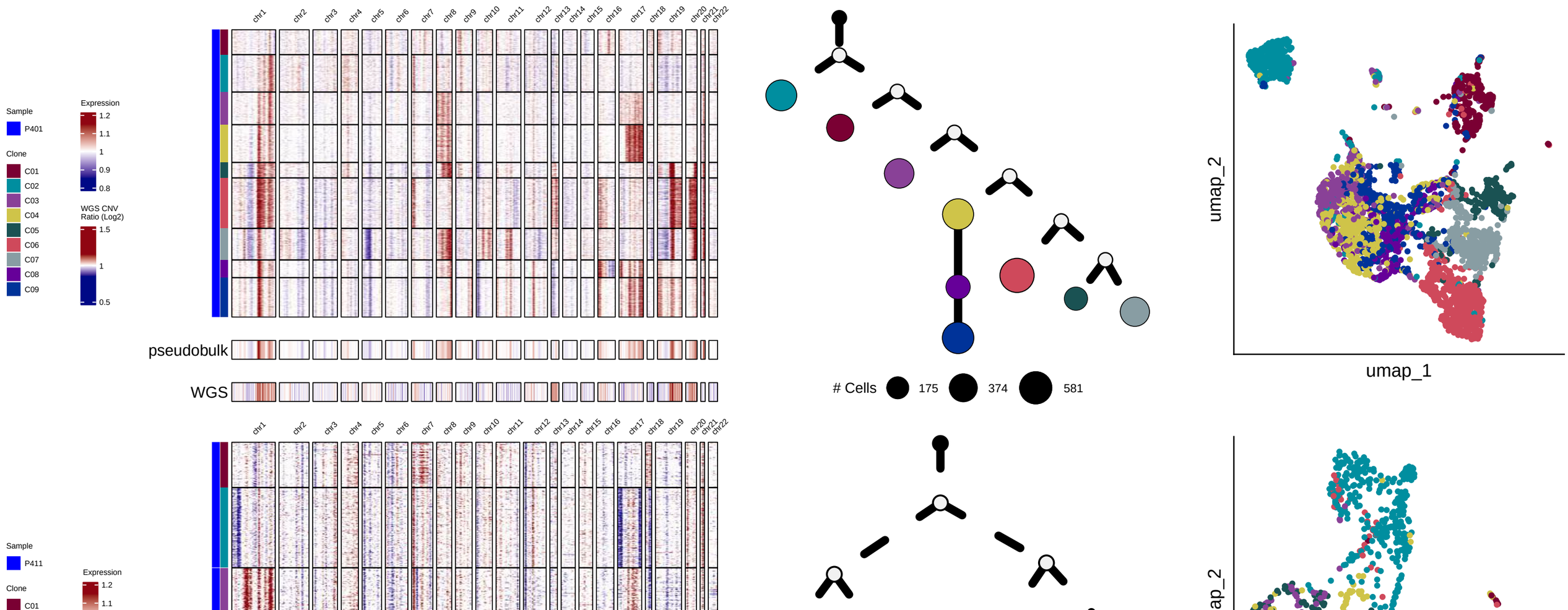
Sample
P323
Clone
C01
C02
C03
C04
C05
C06
Expression
1.2
1.1
1
0.9
0.8
WGS CNV Ratio (Log2)
1.5
1
0.5
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
WGS
Cells 169 342 487
umap_2
umap_1
Sample
P332
Clone
C01
C02
C03
C04
C05
C06
C07
Expression
1.2
1.1
1
0.9
0.8
WGS CNV Ratio (Log2)
1.5
1
0.5
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
WGS
Cells 168 456 655
umap_2
umap_1
Sample
P334
Clone
C01
C02
C03
C04
C05
Expression
1.2
1.1
1
0.9
0.8
WGS CNV Ratio (Log2)
1.5
1
0.5
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
WGS
Cells 44 130 184
umap_2
umap_1
Sample
P339
Clone
C01
C02
C03
C04
C05
C06
Expression
1.2
1.1
1
0.9
0.8
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
Cells 96 264 488
umap_2
umap_1
Sample
P342.1
Clone
C01
C02
C03
C04
C05
C06
C07
Expression
1.2
1.1
1
0.9
0.8
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
Cells 125 653 1828
umap_2
umap_1

Sample
P342.2
Clone
C01
C02
C04
C05
C06
C07
Expression
1.2
1.1
1
0.9
0.8
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
Cells 31 89 1524
umap_2
umap_1
Sample
P345
Clone
C01
C02
C03
C04
C05
C06
C07
Expression
1.2
1.1
1
0.9
0.8
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
Cells 67 245 382
umap_2
umap_1
Sample
P349
Clone
C01
C02
C03
C04
C05
C06
Expression
1.2
1.1
1
0.9
0.8
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
Cells 57 113 207
umap_2
umap_1
Sample
P357
Clone
C01
C02
C03
C04
C05
C06
C07
C08
Expression
1.2
1.1
1
0.9
0.8
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
Cells 134 1246 1657
umap_2
umap_1
Sample
P358
Clone
C01
C02
C03
C04
C05
C06
C07
C08
Expression
1.2
1.1
1
0.9
0.8
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
Cells 172 463 1039
umap_2
umap_1

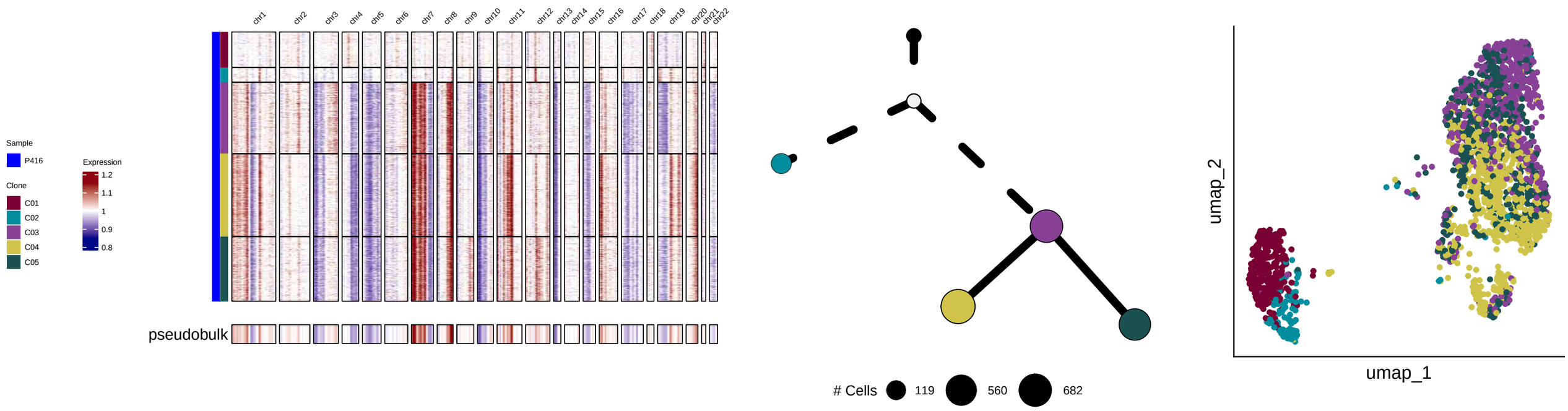
Sample
P360
Clone
C01
C02
C03
C04
C05
C06
C07
C08
Expression
1.2
1.1
1
0.9
0.8
WGS CNV Ratio (Log2)
1.5
1
0.5
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
WGS
Cells 44 428 843
umap_2
umap_1
Sample
P367.1
Clone
C01
C02
C03
C04
C05
Expression
1.2
1.1
1
0.9
0.8
pseudobulk
Cells 273 571 868
umap_2
umap_1
Sample
P371.1
Clone
C02
C03
Expression
1.2
1.1
1
0.9
0.8
pseudobulk
Cells 50 192 726
umap_2
umap_1
Sample
P371.2
Clone
C01
C02
C03
C04
C05
C06
C07
C08
Expression
1.2
1.1
1
0.9
0.8
pseudobulk
Cells 38 153 1402
umap_2
umap_1
Sample
P375
Clone
C01
C02
C03
C04
C05
C06
C07
C08
Expression
1.2
1.1
1
0.9
0.8
WGS CNV Ratio (Log2)
1.5
1
0.5
pseudobulk
WGS
Cells 20 1809 3101
umap_2
umap_1


chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
Sample
P385
Clone
C01
C02
C03
C04
Expression
1.2
1.1
1
0.9
0.8
WGS CNV Ratio (Log2)
1.5
1
0.5
pseudobulk
WGS
Cells 61 189 520
umap_2
umap_1

chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
Sample
P393
Clone
C01
C02
C03
Expression
1.2
1.1
1
0.9
0.8
pseudobulk
Cells 91 142 143
umap_2
umap_1


chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
Sample
P401
Clone
C01
C02
C03
C04
C05
C06
C07
C08
C09
Expression
1.2
1.1
1
0.9
0.8
WGS CNV Ratio (Log2)
1.5
1
0.5
pseudobulk
WGS
Cells 175 374 581
umap_2
umap_1

chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
Sample
P411
Clone
C01
C02
C03
C04
C05
C06
Expression
1.2
1.1
1
0.9
0.8
pseudobulk
Cells 61 162 317
umap_2
umap_1


chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
Sample
P416
Clone
C01
C02
C03
C04
C05
Expression
1.2
1.1
1
0.9
0.8
pseudobulk
Cells 119 560 682
umap_2
umap_1

Sample
P427
Clone
C01
C02
C03
C04
C05
C06
Expression
1.2
1.1
1
0.9
0.8
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
Cells 50 425 2889
umap_2
umap_1
Sample
P428
Clone
C01
C02
C03
C04
C05
C06
C07
C08
C09
C10
C11
Expression
WGS CNV Ratio (Log2)
1.5
1
0.5
pseudobulk
WGS
Cells 90 354 1123
Sample
P432
Clone
C01
C02
C03
C04
C05
C06
C07
Cells 35 205 337
Sample
P436.1
Clone
C02
C03
Cells 50 148 288
Sample
P438
Clone
C01
C02
C03
C04
C05
C06
Cells 24 117 221

Sample
P443
Clone
C01
C02
C03
C04
C05
C06
C07
C08
Expression
1.2
1.1
1
0.9
0.8
WGS CNV Ratio (Log2)
1.5
1
0.5
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
WGS
Cells 37 290 679
umap_2
umap_1
Sample
P444
Clone
C01
C02
C03
C04
C05
C06
Expression
1.2
1.1
1
0.9
0.8
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
Cells 120 254 381
umap_2
umap_1
Sample
P455
Clone
C01
C02
C03
C04
C05
Expression
1.2
1.1
1
0.9
0.8
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
Cells 67 163 261
umap_2
umap_1
Sample
P459
Clone
C01
C02
C03
C04
C05
C06
C07
Expression
1.2
1.1
1
0.9
0.8
WGS CNV Ratio (Log2)
1.5
1
0.5
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
WGS
Cells 132 736 2073
umap_2
umap_1
Sample
P463
Clone
C01
C02
C03
C04
C05
C06
C07
Expression
1.2
1.1
1
0.9
0.8
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
Cells 21 101 948
umap_2
umap_1

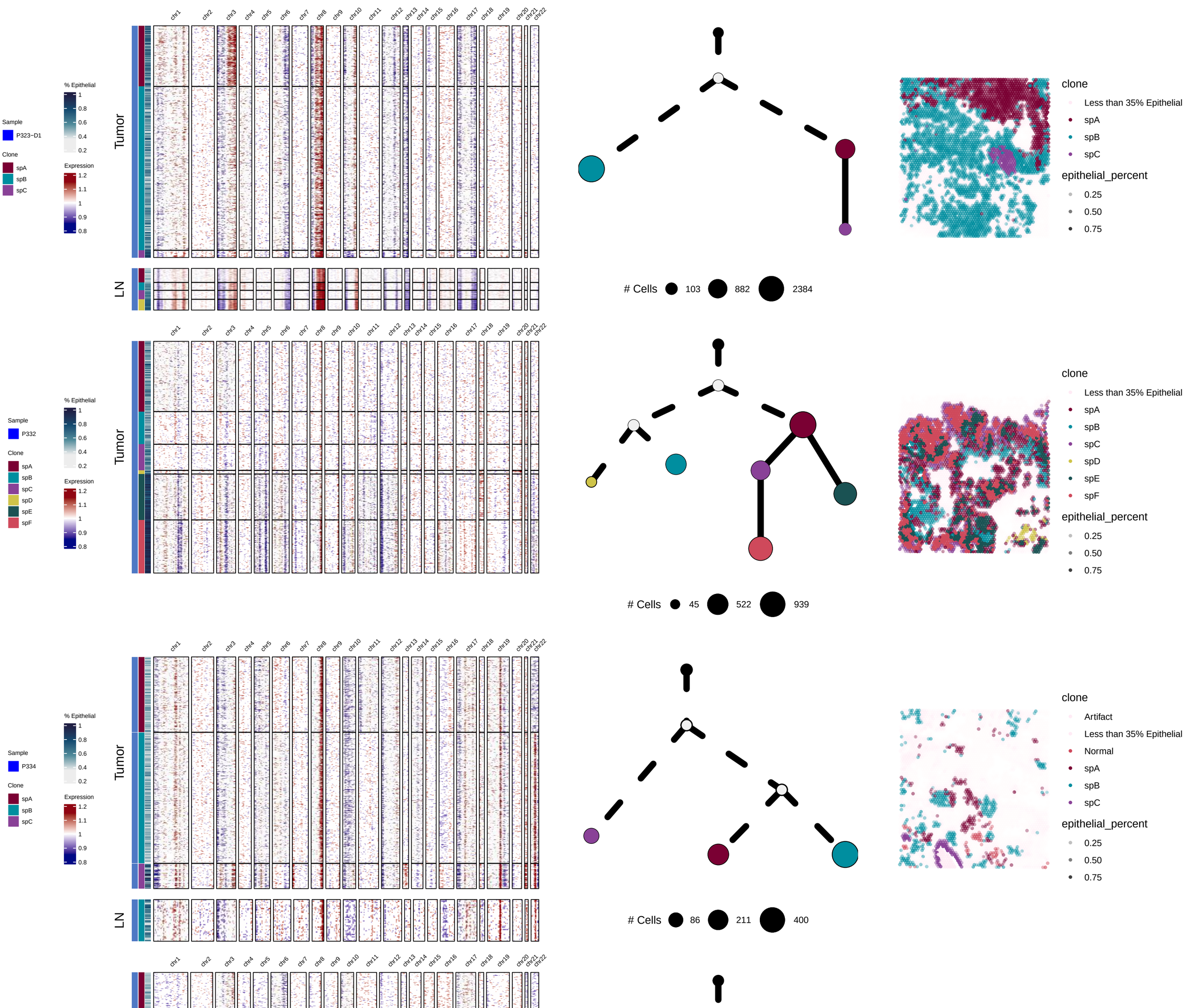
Sample
P465
Clone
C01
C02
C03
C04
C05
C06
C07
Expression
1.2
1.1
1
0.9
0.8
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
Cells 54 295 869
umap_2
umap_1
Sample
P466.1
Clone
C01
C02
C05
C06
C07
C08
C09
C10
Expression
1.2
1.1
1
0.9
0.8
pseudobulk
Cells 130 355 1071
umap_2
umap_1
Sample
P470
Clone
C01
C02
C03
C04
C05
Expression
1.2
1.1
1
0.9
0.8
pseudobulk
Cells 62 110 203
umap_2
umap_1
Sample
P436.2
Clone
C01
C03
C04
C05
Expression
1.2
1.1
1
0.9
0.8
WGS CNV Ratio (Log2)
1.5
1
0.5
pseudobulk
WGS
Cells 118 1017 2358
umap_2
umap_1
Sample
P477
Clone
C01
C02
C03
C04
C05
Expression
1.2
1.1
1
0.9
0.8
pseudobulk
Cells 27 138 222
umap_2
umap_1

chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
Sample
P480.1
Clone
C01
C02
C03
C04
Expression
1.2
1.1
1
0.9
0.8
Cells 57 118 238
umap_2
umap_1
Sample
P487
Clone
C01
C02
C03
C04
Cells 75 112 313
Sample
P518.1
Clone
C01
C02
C03
Cells 83 142 163
Sample
P480.2
Clone
C01
C04
Cells 122 134 147
Sample
P524
Clone
C01
C02
C03
C04
C05
Cells 40 70 94

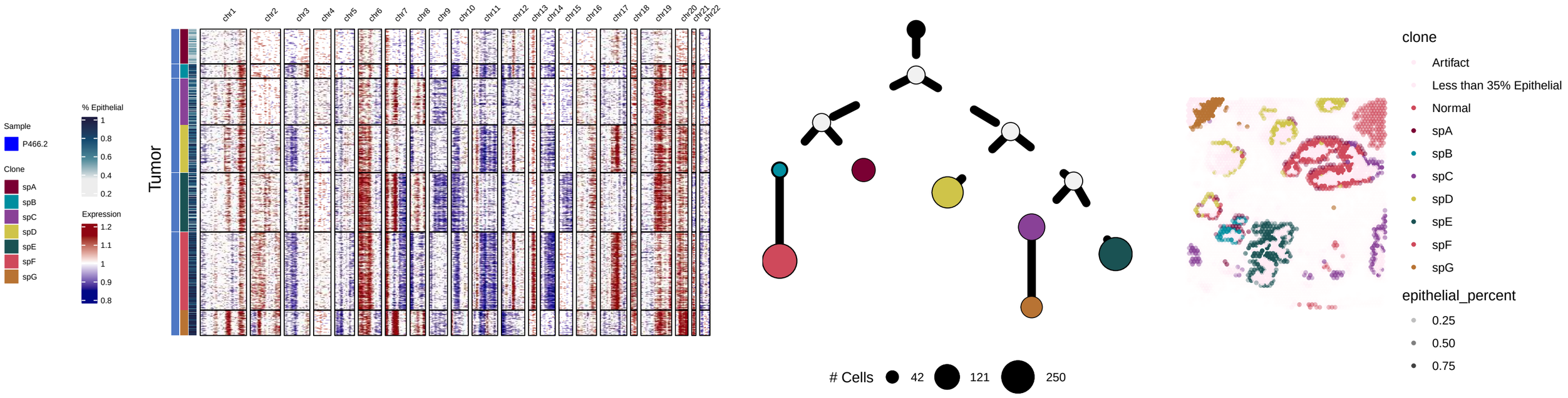
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
Sample
P532
Clone
C01
C02
C03
Expression
1.2
1.1
1
0.9
0.8
Cells 70 217 313
umap_2
umap_1
Sample
P466.2
Clone
C02
C03
C04
C06
C09
C11
Cells 43 309 1311
Sample
P559.1
Clone
C01
C02
C03
C04
C05
C06
Cells 17 86 458
Sample
P567.1
Clone
C01
C05
Cells 47 124 202
Sample
P586
Clone
C01
C02
C03
C04
C05
C06
C07
Cells 54 123 466

Sample
P567.2
Clone
C01
C02
C03
C04
C05
Expression
1.2
1.1
1
0.9
0.8
WGS CNV Ratio (Log2)
1.5
1
0.5
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
WGS
Cells 67 128 757
umap_2
umap_1
Sample
P615
Clone
C01
C02
C03
C04
C05
C06
Expression
1.2
1.1
1
0.9
0.8
WGS CNV Ratio (Log2)
1.5
1
0.5
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
WGS
Cells 50 595 2044
umap_2
umap_1
Sample
P601.2
Clone
C01
C02
C03
C04
C05
C06
C07
C08
C09
Expression
1.2
1.1
1
0.9
0.8
WGS CNV Ratio (Log2)
1.5
1
0.5
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
WGS
Cells 35 242 564
umap_2
umap_1
Sample
P643
Clone
C01
C02
C03
Expression
1.2
1.1
1
0.9
0.8
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
Cells 51 74 94
umap_2
umap_1
Sample
P645
Clone
C01
C02
Expression
1.2
1.1
1
0.9
0.8
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
pseudobulk
Cells 64 80 97
umap_2
umap_1

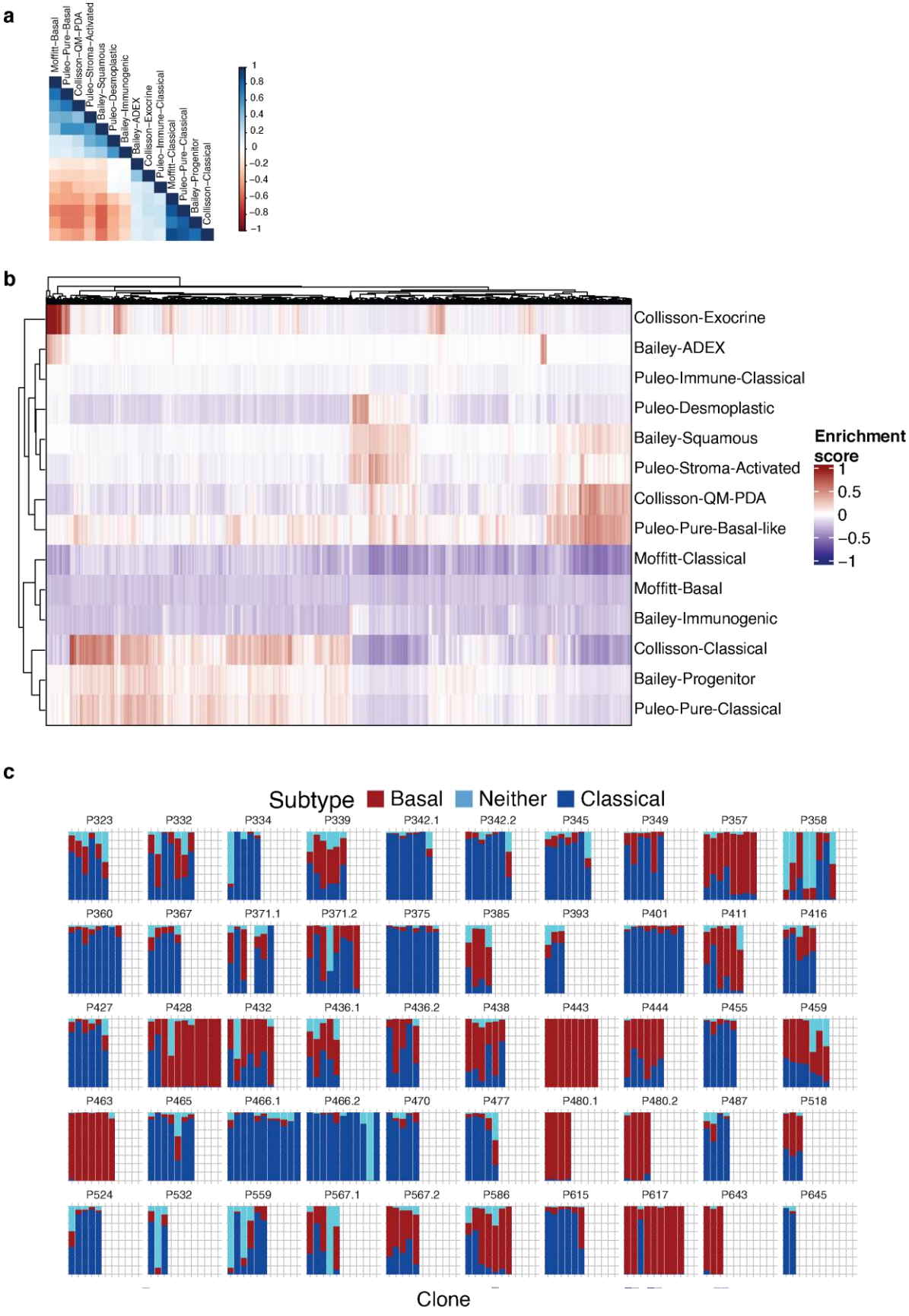
Suppl. Figure 2. Clonal composition across all samples profiled with single-cell RNA-sequencing

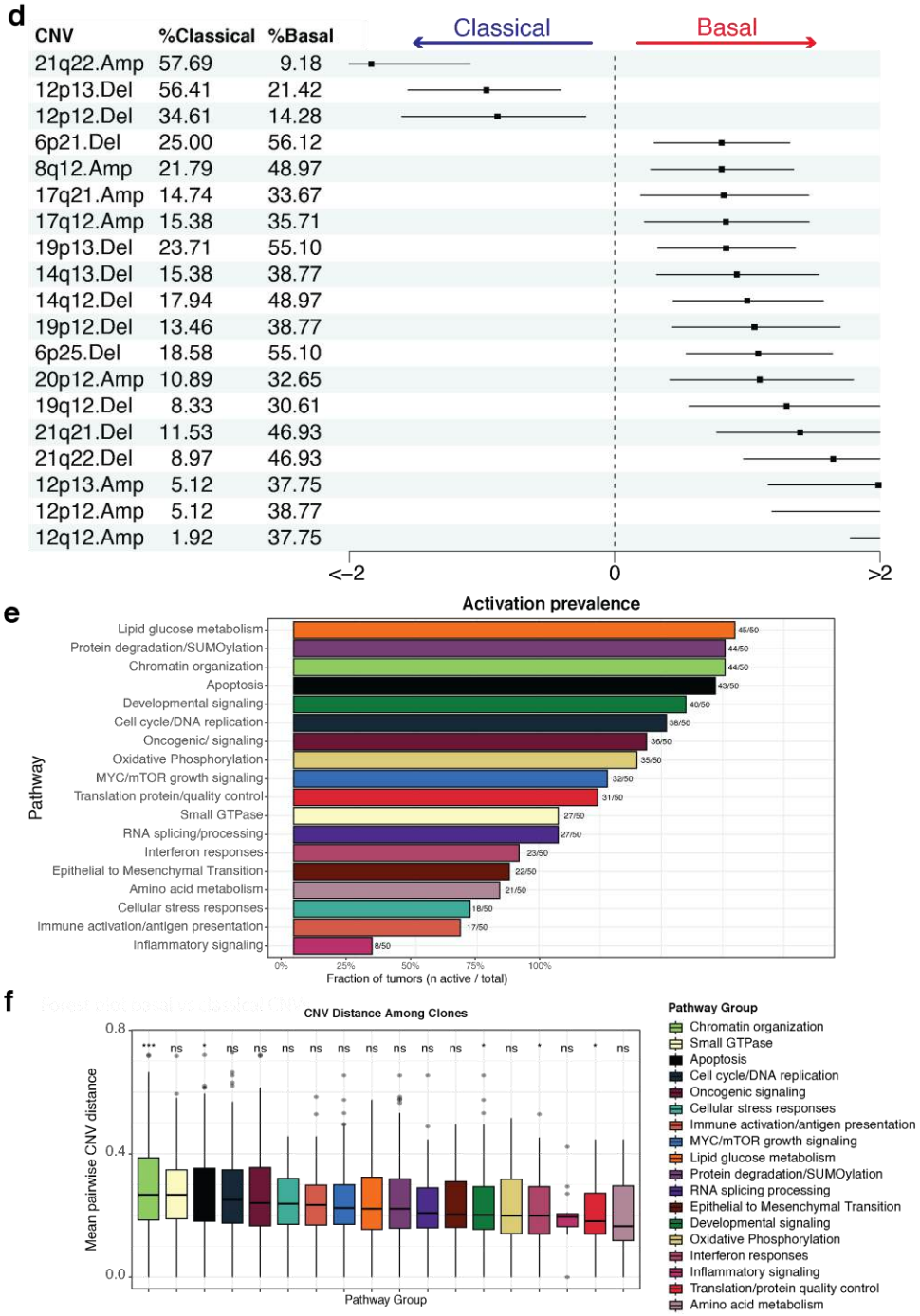

chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
Tumor
LN
% Epithelial
1
0.8
0.6
0.4
0.2
Expression
1.2
1.1
1
0.9
0.8
Sample
P323−D1
Clone
spA
spB
spC
Cells 103 882 2384
clone
Less than 35% Epithelial
spA
spB
spC
epithelial_percent
0.25
0.50
0.75
Sample
P332
Clone
spA
spB
spC
spD
spE
spF
Cells 45 522 939
clone
Less than 35% Epithelial
spA
spB
spC
spD
spE
spF
epithelial_percent
0.25
0.50
0.75
Sample
P334
Clone
spA
spB
spC
Cells 86 211 400
clone
Artifact
Less than 35% Epithelial
Normal
spA
spB
spC
epithelial_percent
0.25
0.50
0.75
Sample
P357
Clone
spA
spB
spC
Cells 336 452 889
clone
Less than 35% Epithelial
spA
spB
spC
epithelial_percent
0.25
0.50
0.75
Sample
P358
Clone
spA
spB
spC
spD
spE
spF
spG
spH
spI
Cells 41 77 499
clone
Less than 35% Epithelial
spA
spB
spC
spD
spE
spF
spG
spH
spI
epithelial_percent
0.25
0.50
0.75

Sample
P401
Clone
spA
spB
spC
spD
spE
spF
spG
spH
% Epithelial
1
0.8
0.6
0.4
0.2
Expression
1.2
1.1
1
0.9
0.8
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
Tumor
Cells 47 252 703
clone
Less than 35% Epithelial
spA
spB
spC
spD
spE
spF
spG
spH
epithelial_percent
0.25
0.50
0.75
Sample
P416
Clone
spA
spB
Cells 123 282 442
epithelial_percent
0.2
0.4
0.6
0.8
clone
Less than 35% Epithelial
Normal
spA
spB
Sample
P427
Clone
spA
spB
spC
Cells 45 128 333
epithelial_percent
0.2
0.4
0.6
0.8
clone
Less than 35% Epithelial
Normal
spA
spB
spC
Sample
P432
Clone
spA
spB
Cells 16 25 35
clone
Less than 35% Epithelial
Normal
spA
spB
epithelial_percent
0.25
0.50
0.75
Sample
P436.2
Clone
spA
spB
Cells 183 961 1739
clone
Artifact
Less than 35% Epithelial
Normal
spA
spB
epithelial_percent
0.25
0.50
0.75

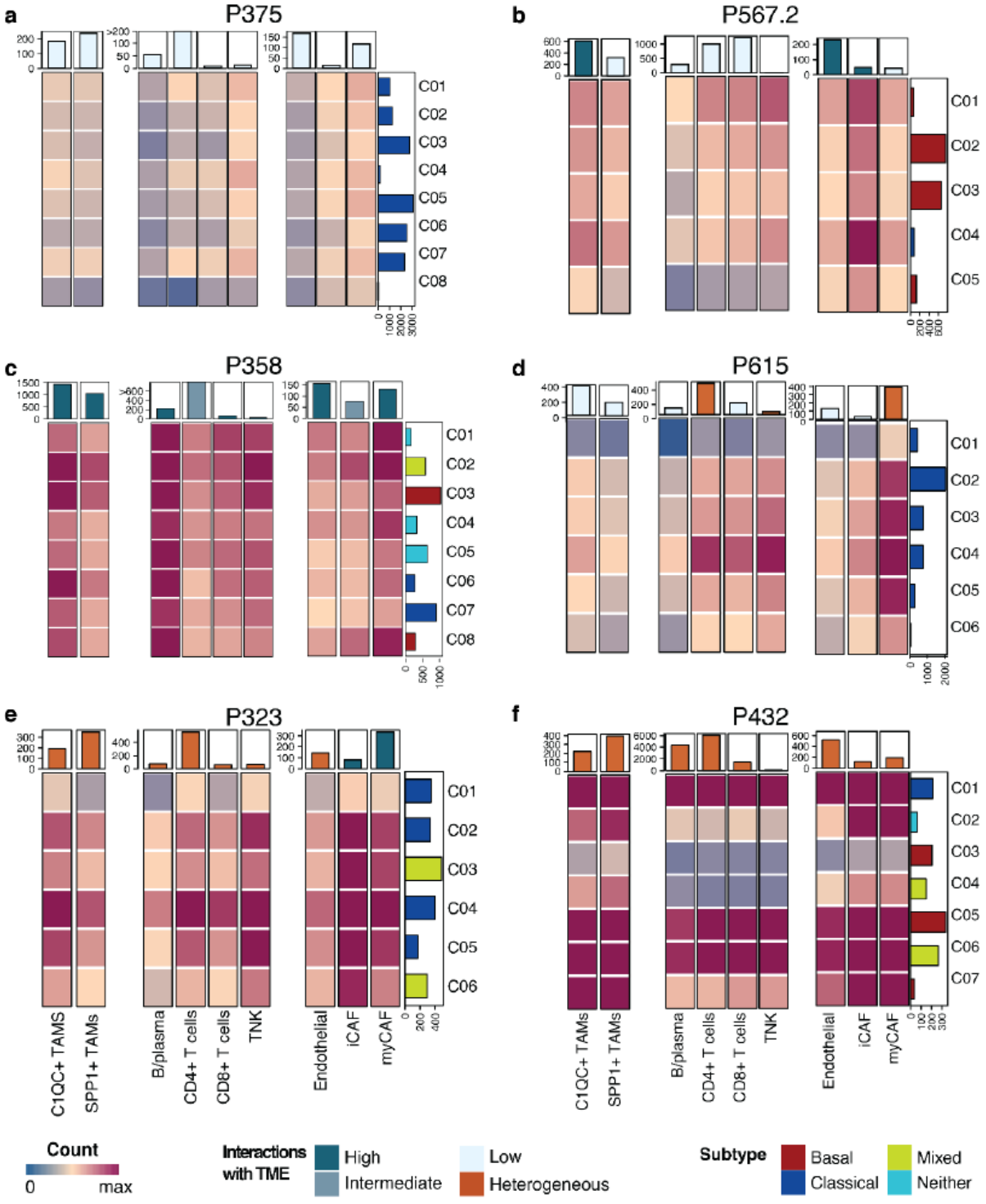

Sample
P466.2
Clone
spA
spB
spC
spD
spE
spF
spG
% Epithelial
1
0.8
0.6
0.4
0.2
Expression
1.2
1.1
1
0.9
0.8
Tumor
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
Cells
42
121
250
clone
Artifact
Less than 35% Epithelial
Normal
spA
spB
spC
spD
spE
spF
spG
epithelial_percent
0.25
0.50
0.75

Suppl. Figure 3. Clonal composition and distribution across space for all samples profiled with spatial transcriptomics

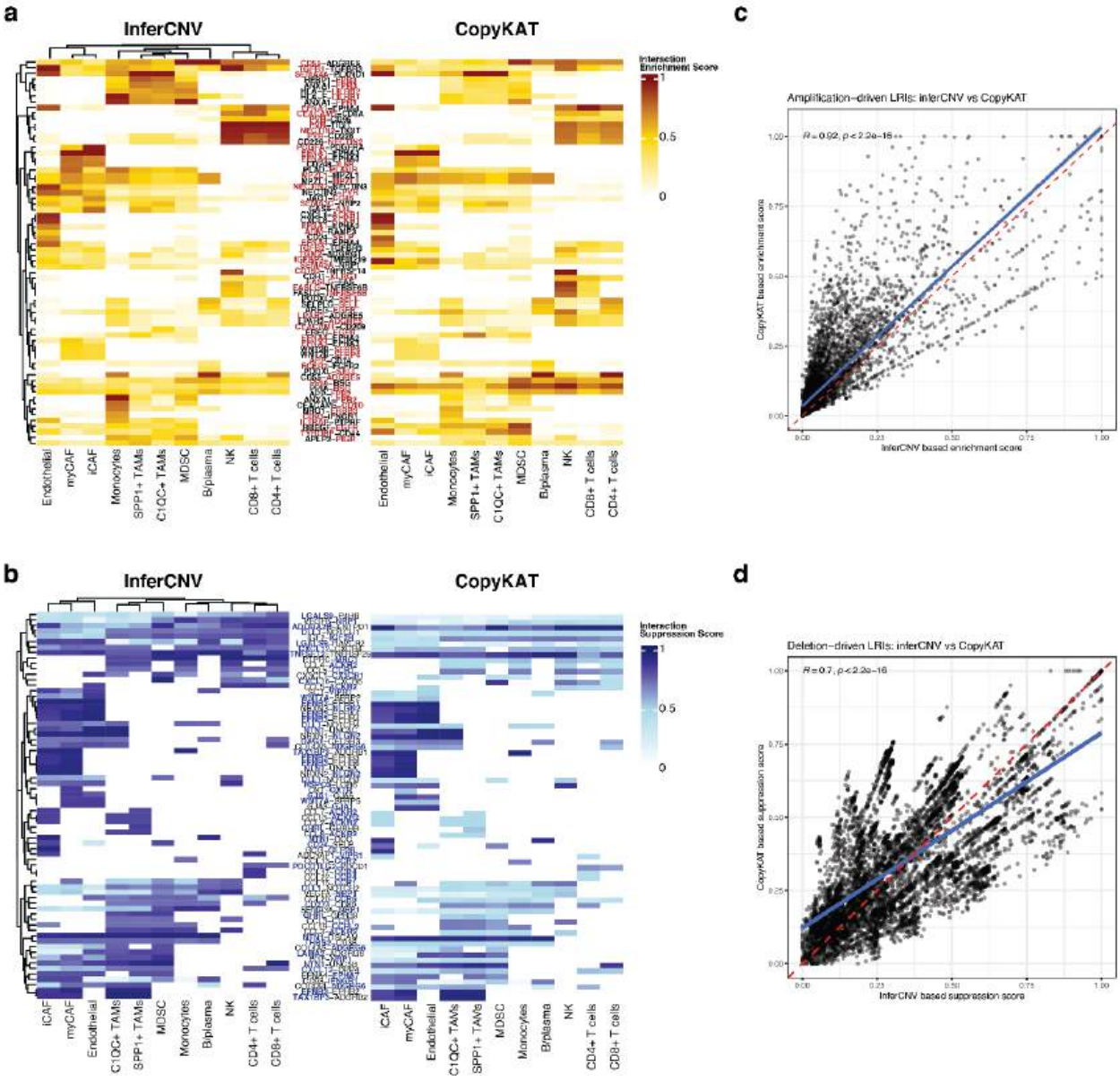

a
Moffitt–Basal
Puleo–Pure–Basal
Collisson–QM–PDA
Puleo–Stroma–Activated
Bailey–Squamous
Puleo–Desmoplastic
Bailey–Immunogenic
Bailey–ADEX
Collisson–Exocrine
Puleo–Immune–Classical
Moffitt–Classical
Puleo–Pure–Classical
Bailey–Progenitor
Collisson–Classical
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
b
Collisson-Exocrine
Bailey-ADEX
Puleo-Immune-Classical
Puleo-Desmoplastic
Bailey-Squamous
Puleo-Stroma-Activated
Collisson-QM-PDA
Puleo-Pure-Basal-like
Moffitt-Classical
Moffitt-Basal
Bailey-Immunogenic
Collisson-Classical
Bailey-Progenitor
Puleo-Pure-Classical
Enrichment score
1
0.5
0
−0.5
−1
c
Subtype
Basal
Neither
Classical
P323
P332
P334
P339
P342.1
P342.2
P345
P349
P357
P358
P360
P367
P371.1
P371.2
P375
P385
P393
P401
P411
P416
P427
P428
P432
P436.1
P436.2
P438
P443
P444
P455
P459
P463
P465
P466.1
P466.2
P470
P477
P480.1
P480.2
P487
P518
P524
P532
P559
P567.1
P567.2
P586
P615
P617
P643
P645
Clone



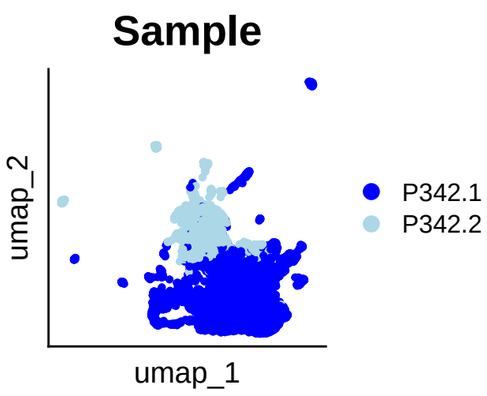
Suppl. Figure 4. a, Correlation of published PDAC subtype signature scores in our dataset. b, Enrichment of published PDAC subtypes per cell in epithelial cells that have a negative score for the Moffitt-Classical and Moffitt-Basal scores. c, Percentage of cells expressing each subtype within each clone. d, Forest plot of significant CNVs between basal and classical clones (Fisher's exact test; see exact metrics in raw data). e, Fraction of clones in each sample that are active for a pathway. f, CNV distance between clones in the same tumor that are active for the same pathway (n=285 clones; Wilcoxon two-sided rank sum test; see exact p-values in raw data).

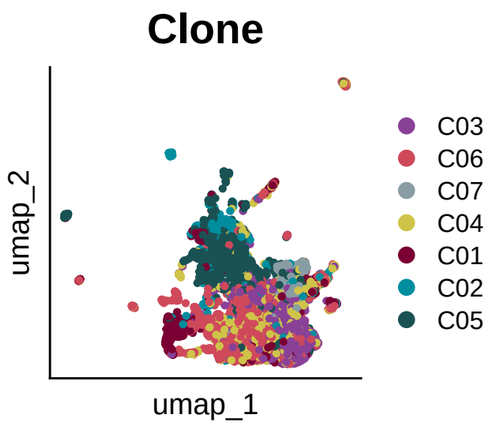


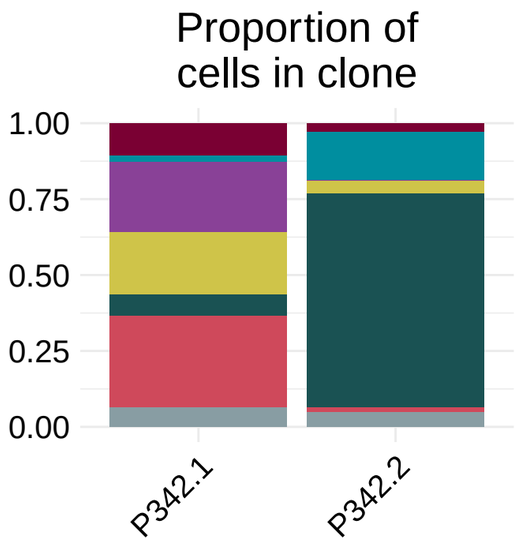
Suppl. Figure 5. a-f. Examples of samples with clones with comparable number of interactions with the TME (a-c), with clones interacting differently with some cell types (d-e), and with clones interacting differently with all TME cell types (f).

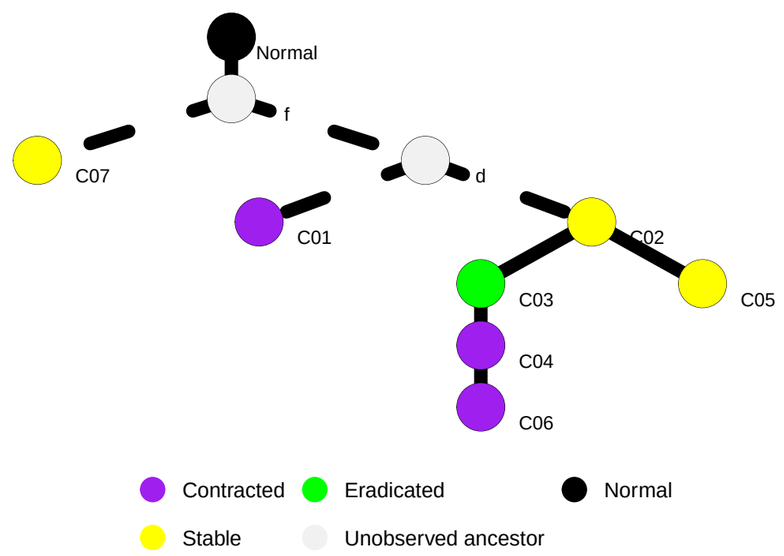


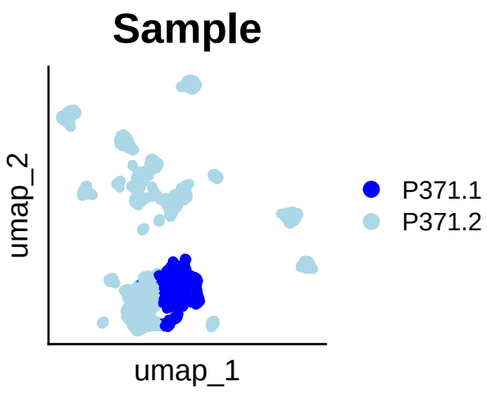
Suppl. Figure 6. a,b, Ligand Receptor Interaction (LRI) scores based on amplifications (a) and deletions (b) detected by inferCNV (left) and CopyKAT (right). c,d, Correlation of scores based on amplifications (c) or deletions (d) detected by inferCNV vs CopyKAT (spearman correlation).

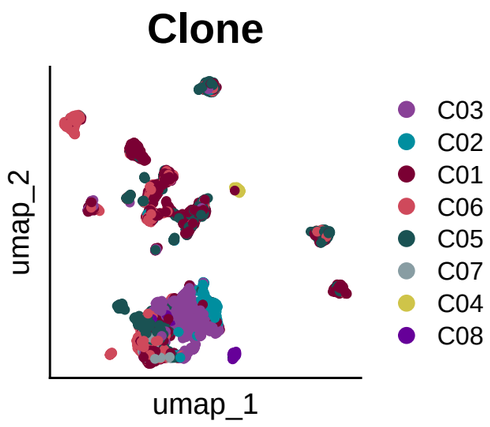

Sample
umap_2
umap_1
P342.1
P342.2

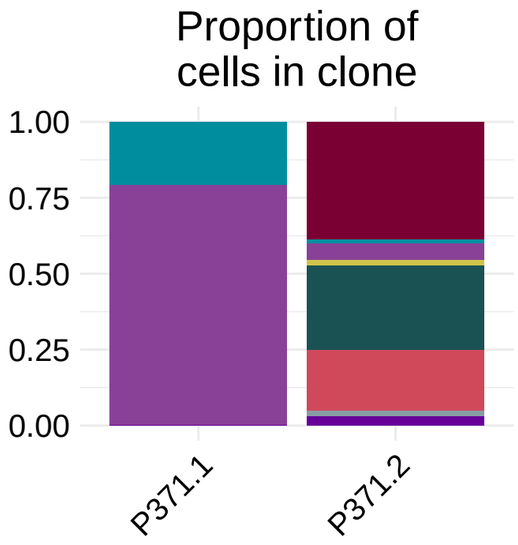

Clone
umap_2
umap_1
C03
C06
C07
C04
C01
C02
C05

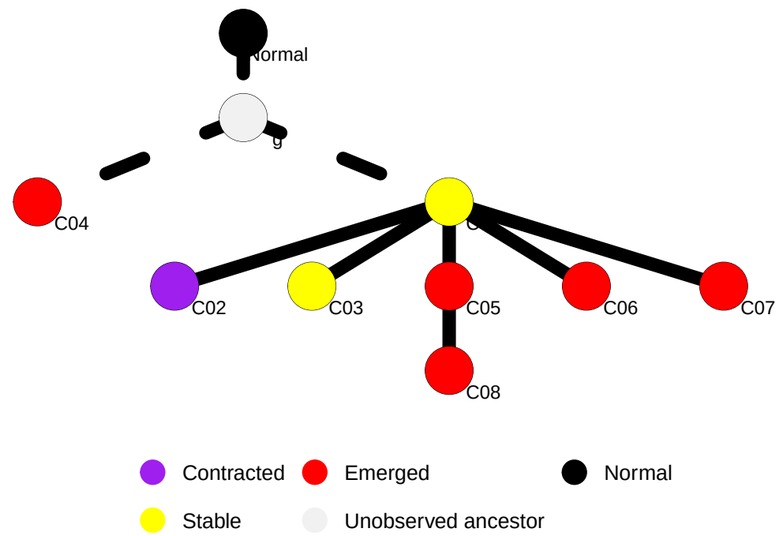

Proportion of cells in clone
1.00
0.75
0.50
0.25
0.00
P342.1
P342.2

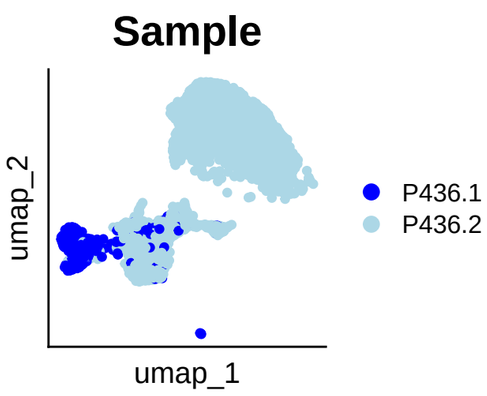

Normal
f
C07
d
C01
C02
C03
C05
C04
C06
Contracted
Eradicated
Normal
Stable
Unobserved ancestor

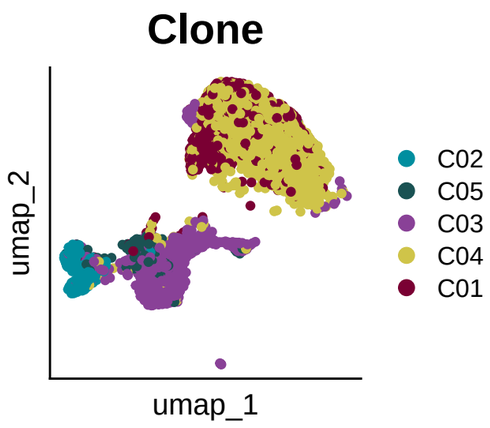

Sample
umap_2
umap_1
P371.1
P371.2

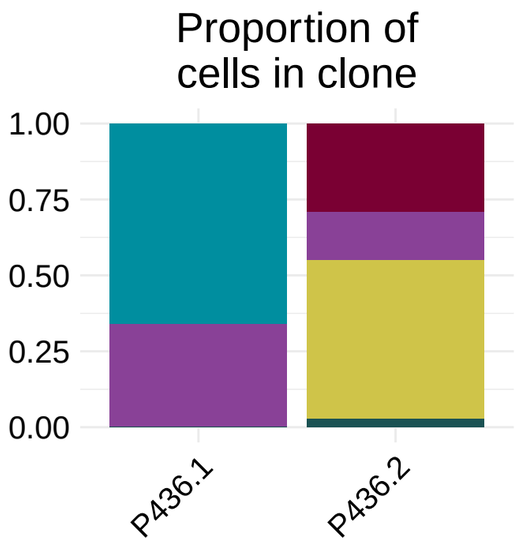

Clone
umap_2
umap_1
C03
C02
C01
C06
C05
C07
C04
C08

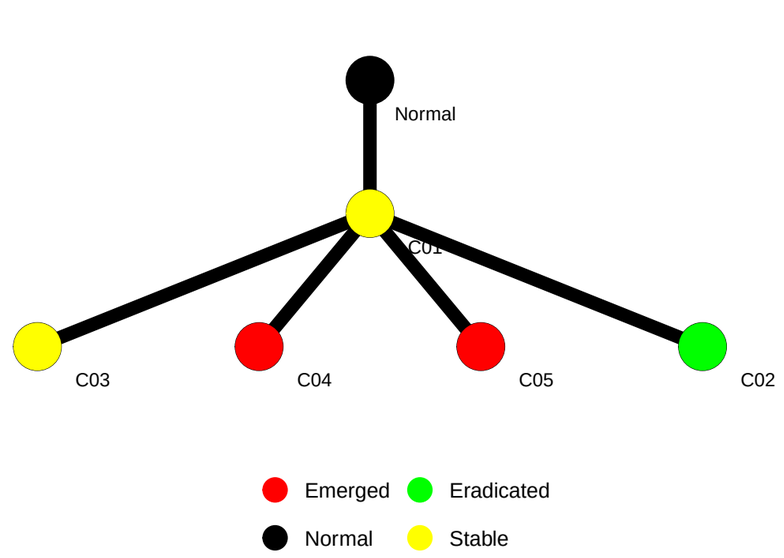

Proportion of cells in clone
1.00
0.75
0.50
0.25
0.00
P371.1
P371.2

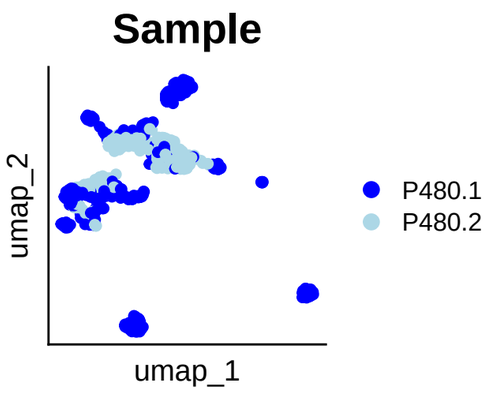

Normal
g
C04
C02
C03
C05
C06
C07
C08
Contracted
Emerged
Normal
Stable
Unobserved ancestor

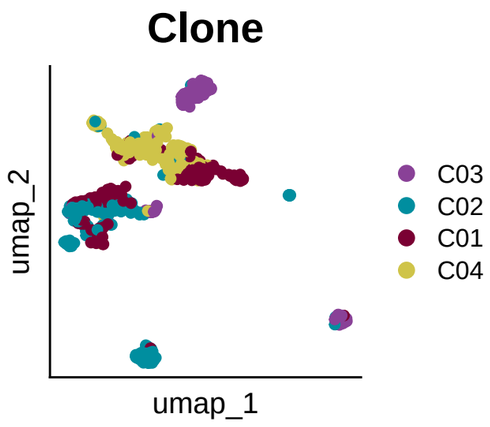

Sample
umap_2
umap_1
P436.1
P436.2

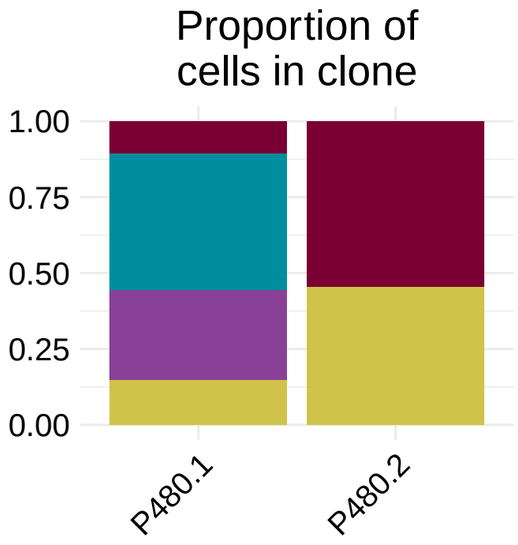

Clone
umap_2
umap_1
C02
C05
C03
C04
C01

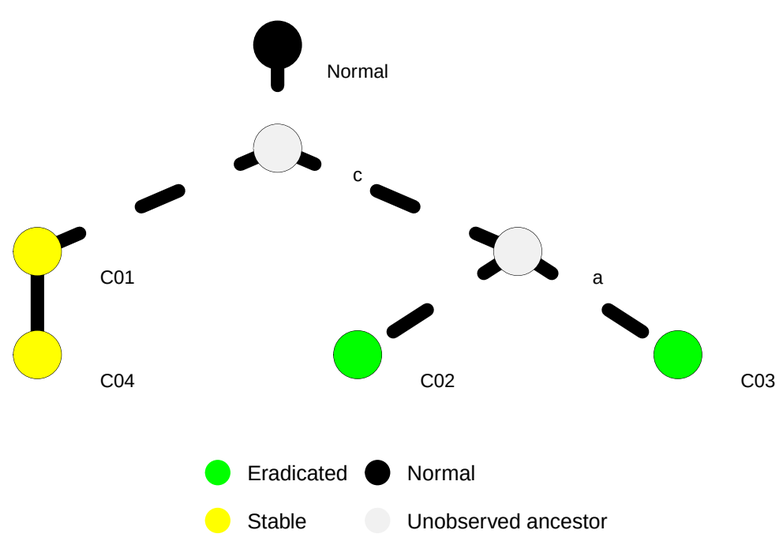

Proportion of cells in clone
1.00
0.75
0.50
0.25
0.00
P436.1
P436.2

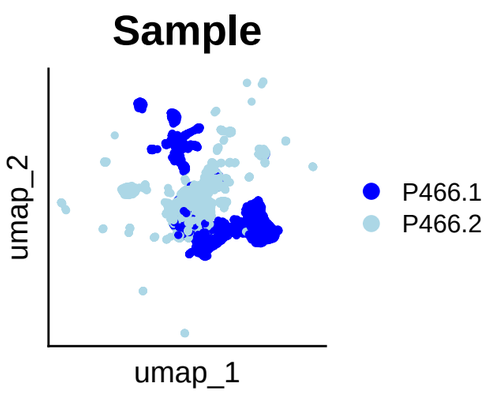

Normal
C01
C03
C04
C05
C02
Emerged
Eradicated
Normal
Stable

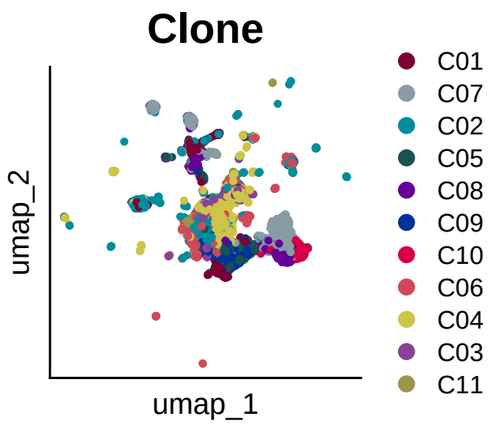

Sample
umap_2
umap_1
P480.1
P480.2

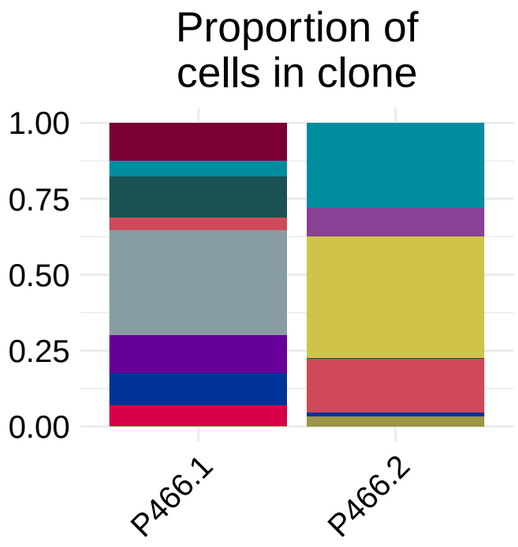

Clone
umap_2
umap_1
C03
C02
C01
C04

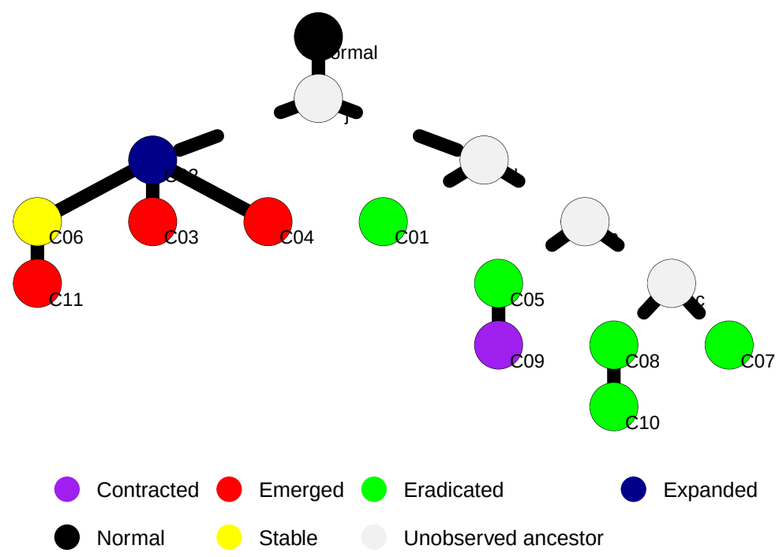

Proportion of cells in clone
1.00
0.75
0.50
0.25
0.00
P480.1
P480.2

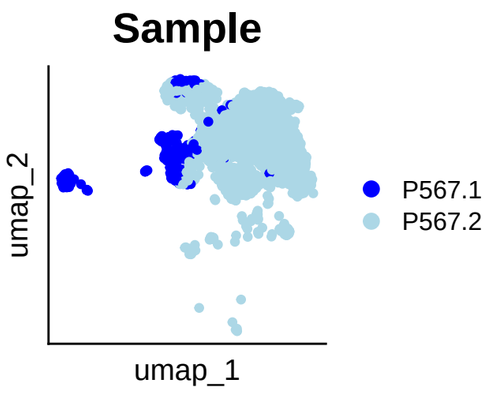

Normal
c
C01
a
C04
C02
C03
Eradicated
Normal
Stable
Unobserved ancestor

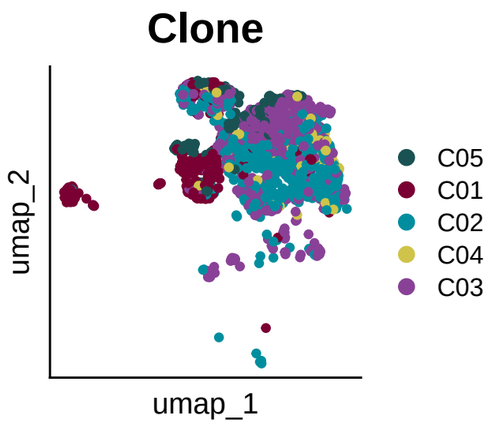

Sample
umap_2
umap_1
P466.1
P466.2

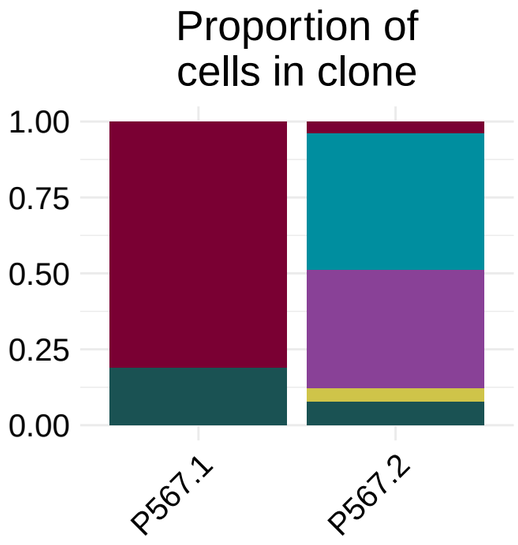

Clone
umap_2
umap_1
C01
C07
C02
C05
C08
C09
C10
C06
C04
C03
C11

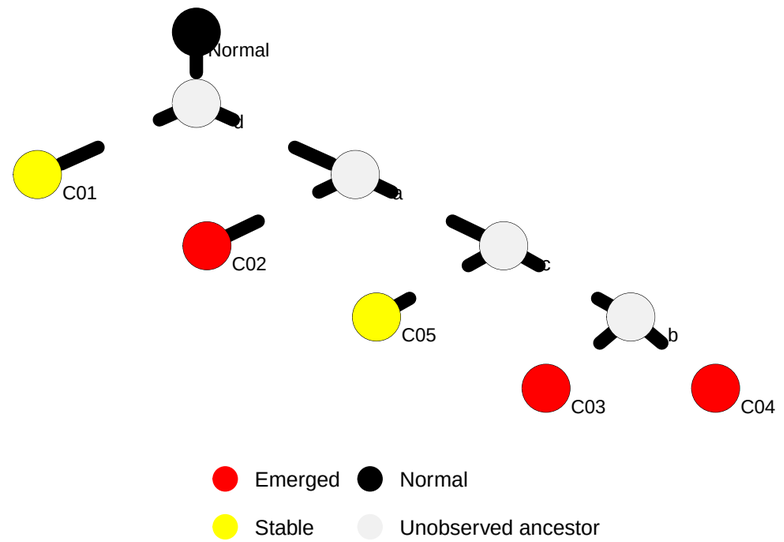

Proportion of cells in clone
1.00
0.75
0.50
0.25
0.00
P466.1
P466.2


C06
C03
C04
C01
C11
C05
C09
C08
C07
C10
Contracted
Emerged
Eradicated
Expanded
Normal
Stable
Unobserved ancestor

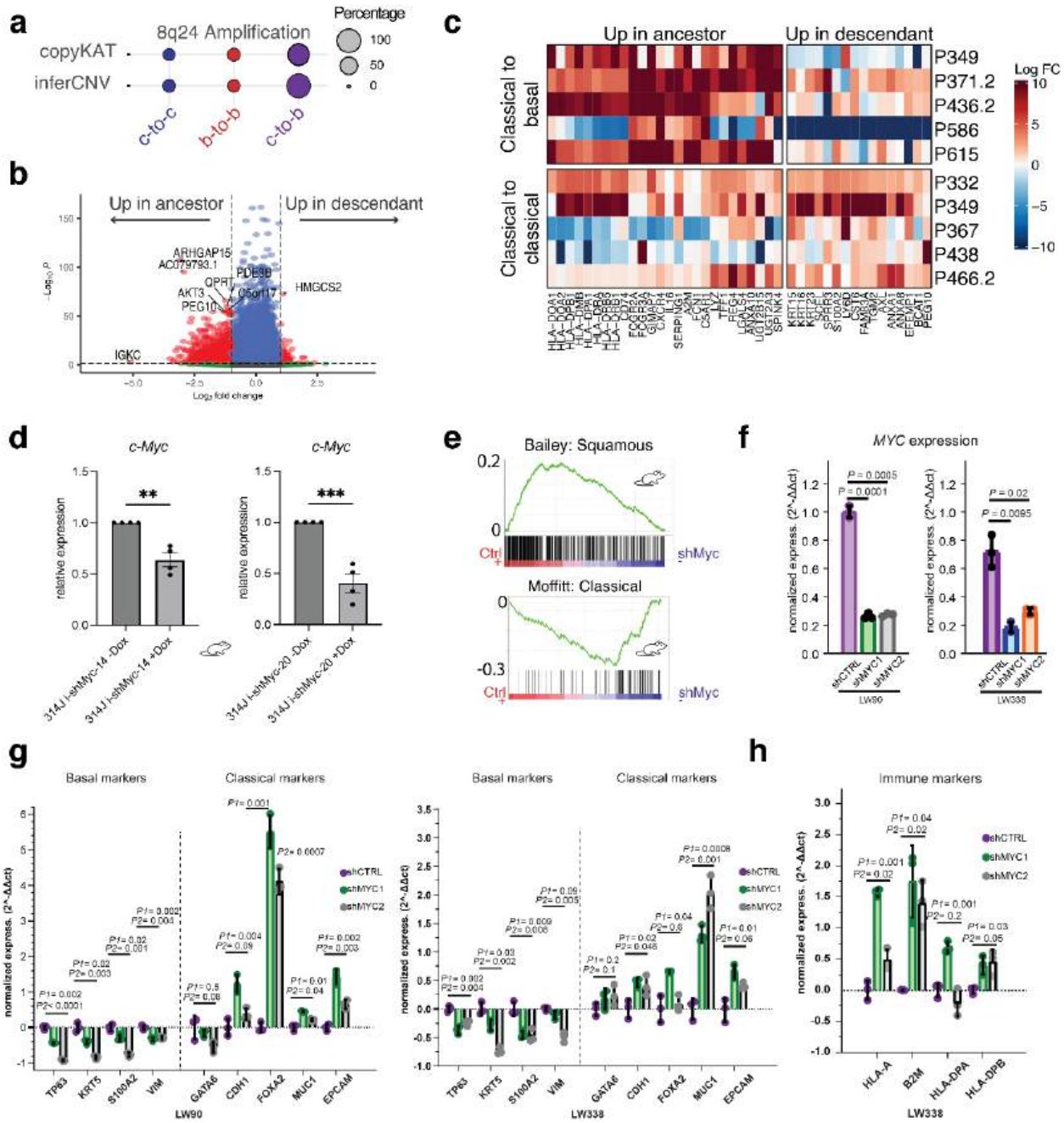

Sample
umap_2
umap_1
P567.1
P567.2


Clone
umap_2
umap_1
C05
C01
C02
C04
C03


Proportion of cells in clone
1.00
0.75
0.50
0.25
0.00
P567.1
P567.2

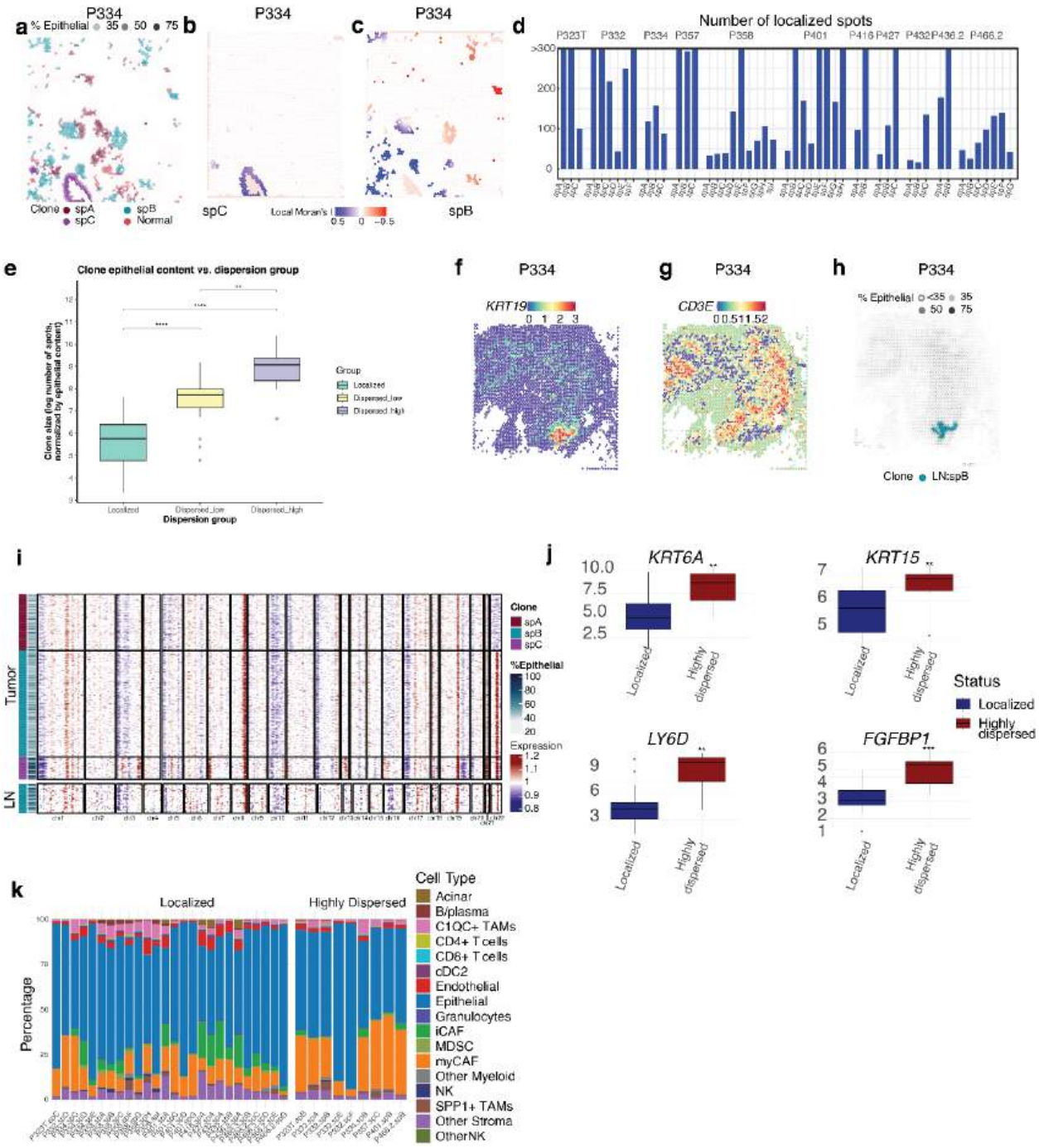

C01
C02
C05
C03
C04
Emerged
Normal
Stable
Unobserved ancestor

Suppl. Figure 7. Patterns of clonal dynamics and interactions following chemotherapy



Suppl. Figure 8. a, Frequency of 8q24 amplification in classical-to-basal transitions detected by inferCNV and CopyKAT. (c-to-c: classical ancestor clone and classical descendant; b-to-b: basal ancestor clone and basal descendant; c-to-b: classical ancestor clone and basal descendant) b, Genes differentially expressed upon the classical-to-classical evolutionary transition with 8q24 gain (two-sided Wald test; y-axis, −log10 Benjamini–Hochberg–adjusted p-value). c, Log fold change of selected gene expression between the pairs of classical to basal transition and classical to classical transition that include the 8q24 amplification. d, Myc knockdown in two *Myc*+ basal mouse cell lines (n=4 replicates each; **: p<0.01, ***: p<0.001, two-sided t-test). e. *Myc* knockdown in two

Myc+ basal mouse cell lines results in the depletion of the basal (top) and the enrichment of the classical subtype (bottom) (n=3 replicates each). f, MYC knockdown in two basal human PDAC cell lines positive for MYC amplification (n=3 replicates each, two-sided t-test). g, Downregulation of basal markers and upregulation of classical markers in the two human PDAC cell lines after successful MYC knockdown (n=3 replicates each, two-sided t-test). h, Upregulation of MHC I, II genes in one human PDAC cell lines after successful MYC knockdown (n=3 replicates, two-sided t-test).



Suppl. Fig. 9. Dispersed clonal populations facilitate lymph node invasion in PDAC. a. Spatial mapping of the clones in the profiled region of a representative sample (P334). b. Local Moran's I score for clone spCNV2 of P334. c. Local Moran's I score for clone spCNV3 of P334. d. Barplot showing the number of localized spots for each clone of each sample. e. Epithelial content of clones in each dispersion group ($n_{localized}$=26 clones, $n_{dispersed\_low}$=17 clones, $n_{dispersed\_high}$=9 clones; exact p-values in raw data; , two-sided Wilcoxon rank-sum test). f. *KRT19* expression in the lymph node of P334. g. *CD3E*

expression in the lymph node of P334. h. Spatial mapping of spCNV2 clone that has invaded P334's lymph node. i. CNV profile of spots of the tumor region (top) and the lymph node (bottom) of 334 with more than 35% epithelial content. Clustered spots are considered clones. j. Boxplots of basal marker gene expression in highly dispersed and localized clones ($n_{localized}$=26 clones, $n_{dispersed\_high}$=9 clones; $pval_{KRT6A}$=0.0013, $pval_{KRT15}$=0.0031, $pval_{LY6D}$=0.0001, $pval_{FGFBP1}$= 2.494e-06; two-sided Wilcoxon rank-sum test on pseudobulked expression). k. Cell type composition in localized and highly dispersed clones.

References

1. Tickle, T., Tirosh, I., Georgescu, C., Brown, M. & Haas, B. inferCNV of the Trinity CTAT Project. Klarman Cell Observatory, Broad Institute of MIT and Harvard (2019).
2. Bielski, C. M. *et al*. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nature Genetics* **50**, 1189–1195 (2018).
3. Notta, F. *et al*. A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* **538**, 378–382 (2016).
4. Gao, R. *et al*. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nature Biotechnology* **39**, 599–608 (2021).
5. Gao, T. *et al*. Haplotype-aware analysis of somatic copy number variations from single-cell transcriptomes. *Nature Biotechnology* **41**, 417–426 (2023).
6. Farrell, A. S. *et al*. MYC regulates ductal-neuroendocrine lineage plasticity in pancreatic ductal adenocarcinoma associated with poor outcome and chemoresistance. *Nat Commun* **8**, 1728 (2017).
7. Werba, G. *et al*. Single-cell RNA sequencing reveals the effects of chemotherapy on human pancreatic adenocarcinoma and its tumor microenvironment. *Nature Communications* **14**, 797 (2023).
8. Zheng, G. X. Y. *et al*. Massively parallel digital transcriptional profiling of single cells. *Nature Communications* **8**, 14049 (2017).
9. Hao, Y. *et al*. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology* **42**, 293–304 (2024).
10. Dolgalev, I. & Yeaton, A. scooter: Streamlined scRNA-Seq Analysis Pipeline. (2023).

11. Germain, P.-L., Lun, A., Meixide, C. G., Macnair, W. & Robinson, M. D. Doublet identification in single-cell sequencing data using *scDblFinder*. Preprint at https://doi.org/10.12688/f1000research.73600.2 (2022).
12. Dolgalev, I. Seq-N-Slide [Computer software].
13. Boeva, V. *et al*. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423–425 (2012).