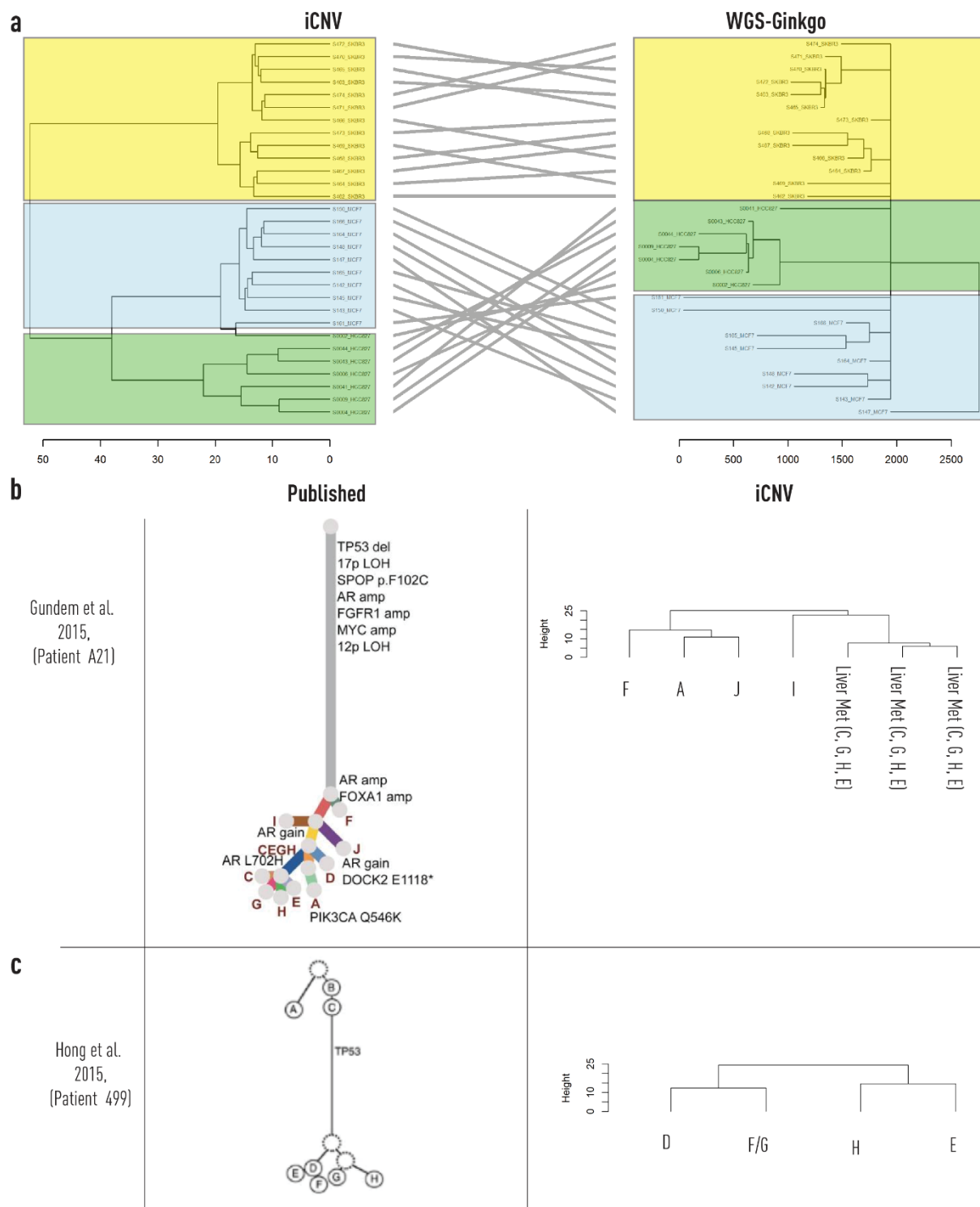
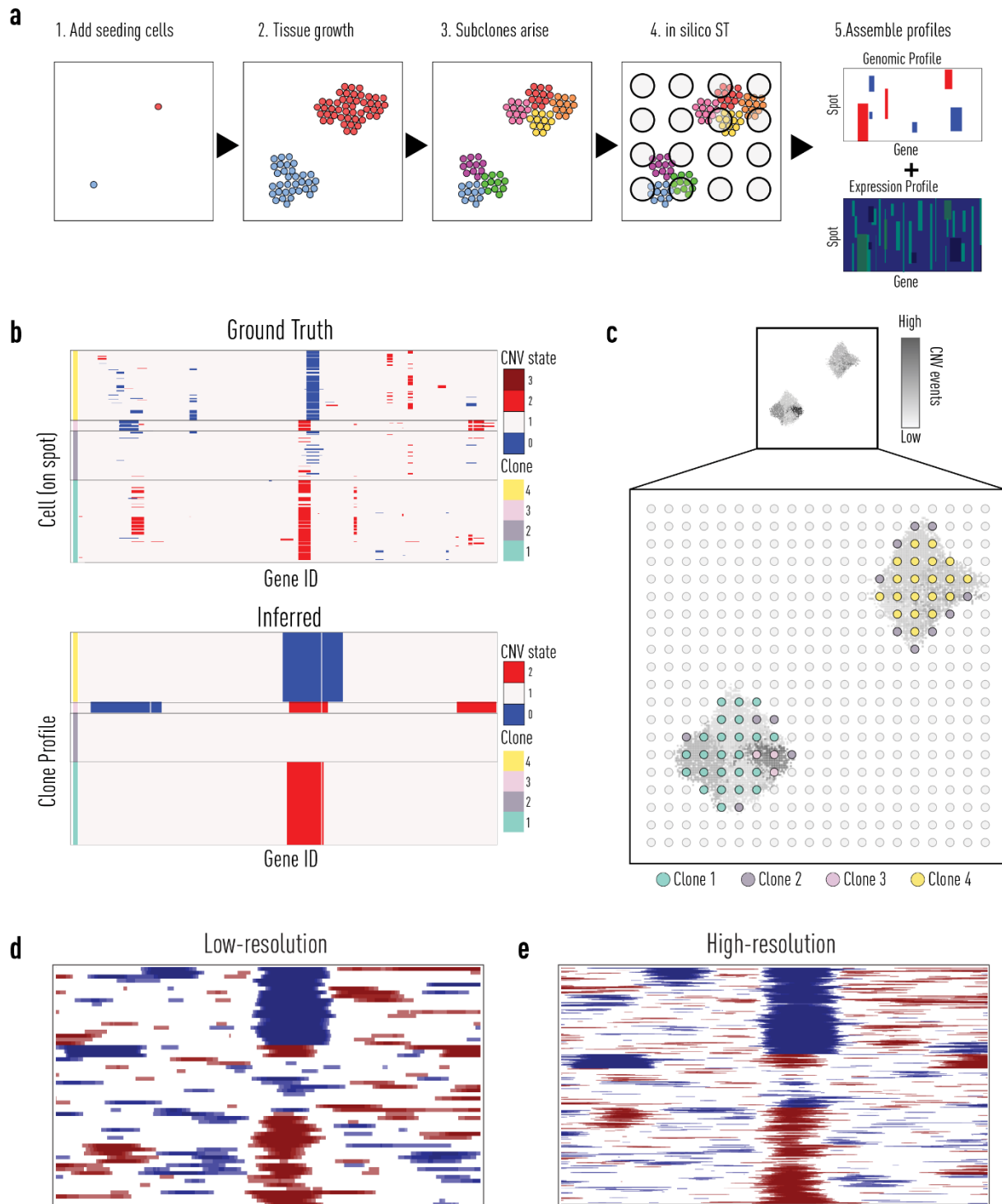




Extended Data Figure 1 | Comparison of phylograms created from whole genome sequencing CNVs and iCNV's. a, Comparison of single tumour cells with co-isolated DNA and RNA (Han et al., Genome Res 2018). Colours correspond to individual cell lines (yellow: SKBR3, green: HCC827, and light blue: MCF7). Entanglement of the phylograms was 0.11 (an entanglement value of 1 corresponds with full entanglement of two phylograms, whereas an entanglement value of 0 corresponds with no entanglement). **b** Phylogeny from patient A21, as published and reproduced from Gundem et al., Nature, 2015. Transcript data were available only for a subset of specimens. **c,** Phylogeny from patient 499, as published and reproduced from Hong et al., Nat.

Comms, 2015. Transcript data available for a subset of specimens (used to reproduce phylogenetic tree by iCNV).

iCNV = inferred Copy-Number Variant

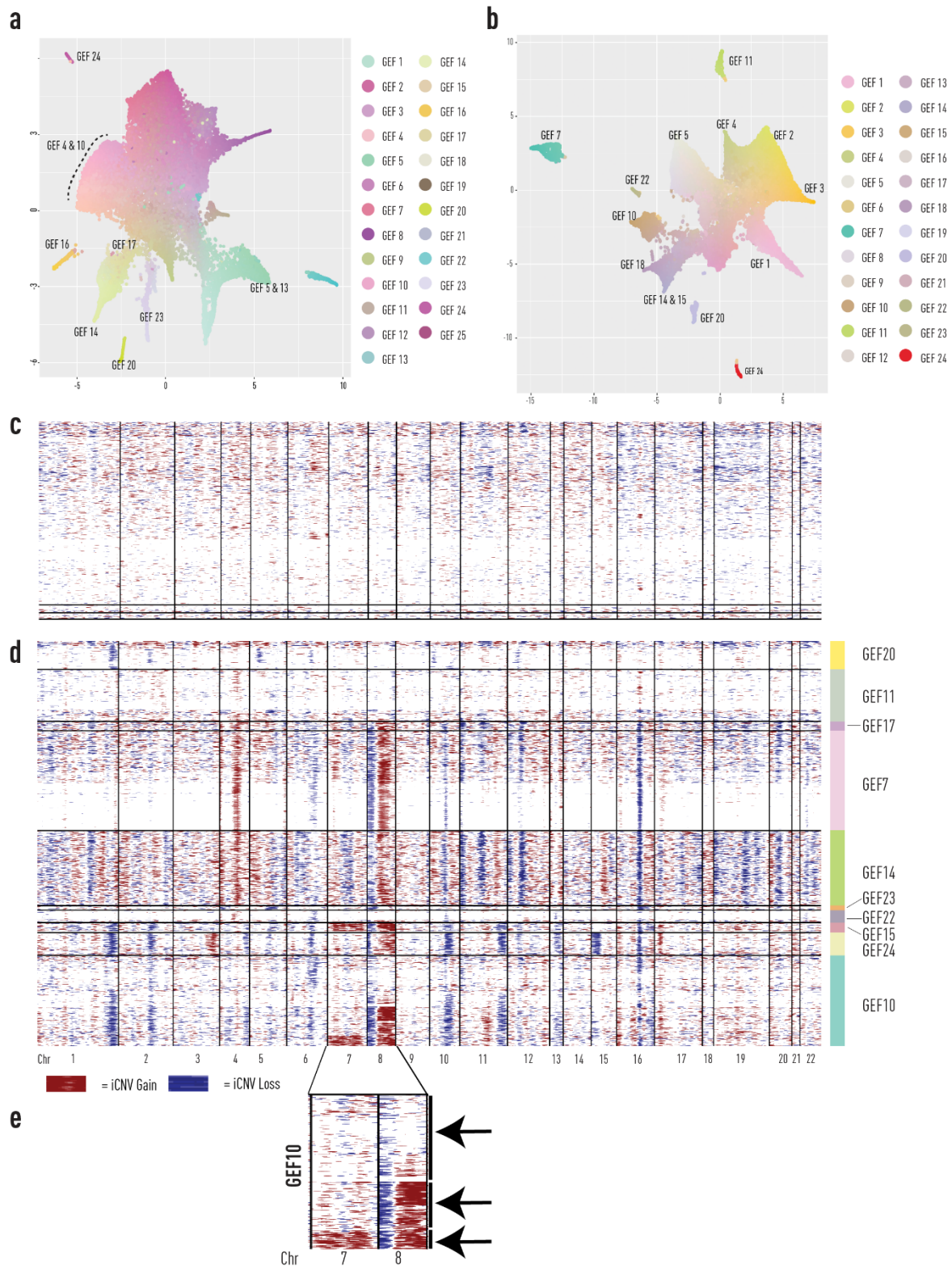


Extended Data Figure 2 | Generating and running iCNV on synthetic data. **a**, Schematic overview of the generative process used to produce artificial spatial data. 1) First a set of seeding cells (red and blue circles) are placed in a defined tissue domain (square), every seeding cell hosts one unique copy number event. 2) The cells are allowed to “grow” within the tissue domain until the number of cells in the domain exceeds a predetermined number. 3) Mutations in the genome occur stochastically during growth and as a result, subpopulations (indicated by colour) of cells with similar genomic profiles arise. 4) Unoccupied space in the tissue domain is filled with benign cells (no copy number

variations), spatial capture locations are placed in a grid over the grown tissue and transcripts are “captured” from the cells overlying each spot. 5) Synthetic spatial expression data is produced together with associated ground truth genomic data (both on spot and cell level). **b**, Results from applying spatial inferCNV (bottom) to a set of synthetic data together with ground truth information (top), only cells residing at spots being annotated as non-benign are shown. Blue indicates a deletion event while red indicates an amplification event. The ground truth shows the genomic profiles for all cells contributing to the spots assigned to a given clone. Comparing the inferred state with the

ground truth on a clone 19 level, the average accuracy across genes was 0.90 (standard deviation 0.10) **c**, Spatial organization of the synthetic data analysed in (b), with thumbnail of the complete cell population in the artificial tissue, each pixel corresponding to a cell. The cells' intensity levels are proportional to their total number of associated copy number events. Circles represent the spots used to "capture" transcripts. Spots are coloured by their inferred clone identity. Note how Clone 2,

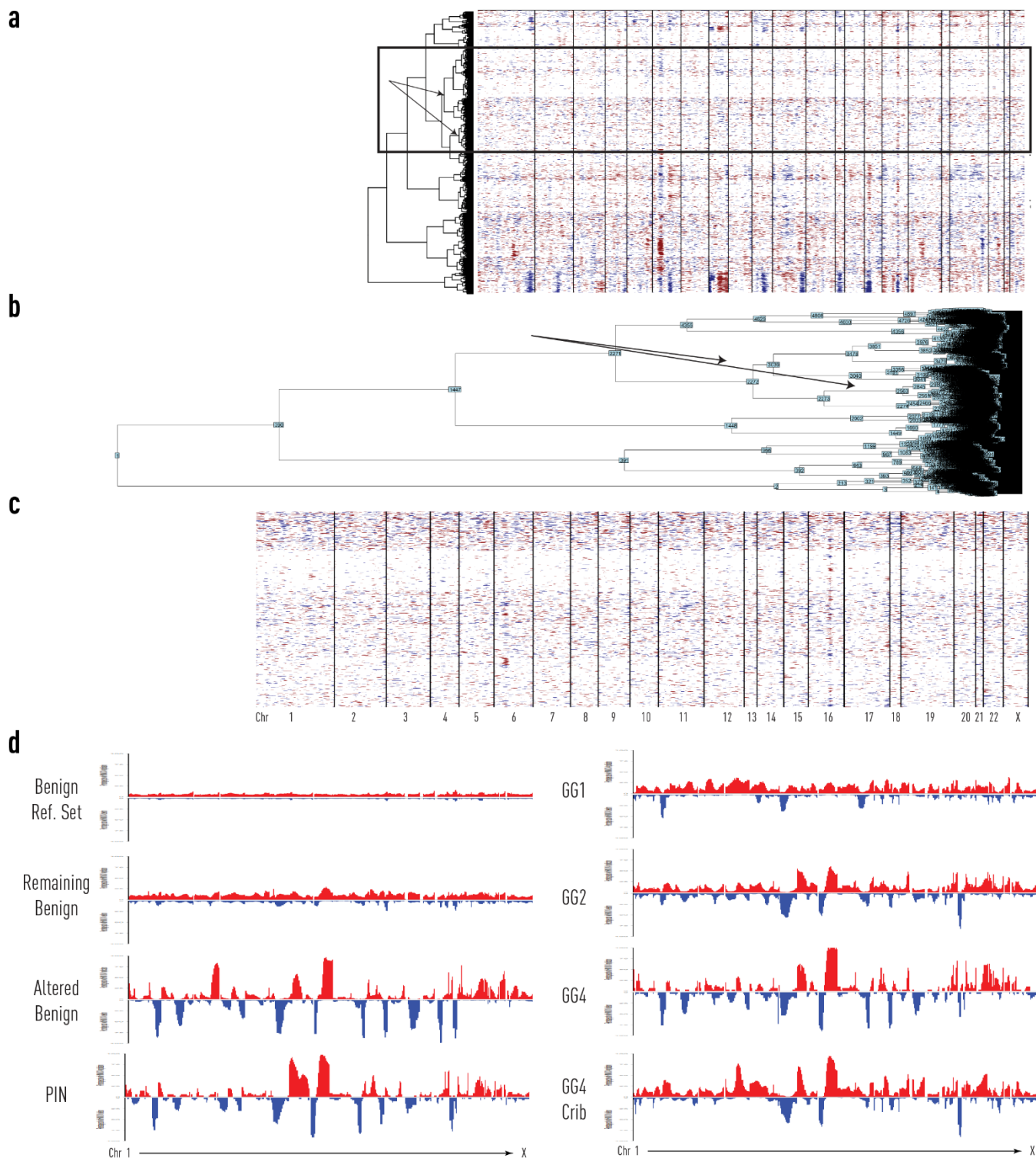
predicted to have zero copy number events, is found along the borders of both foci, where there's a mixture of benign and non-benign cells. **d**, iCNV outputs from simulated synthetic data of spots simulating ST 1k array (low-resolution) with 100 μm spot diameter and centre-to-centre distance of 200 μm . **e**, Visium (high-resolution). High resolution spots were 0.55x size of low resolution and had 5x more spots per area. The synthetic ground truth data were identical for both.



Extended Data Figure 3 | GEF directed iCNV analysis from prostate patient 1 (high resolution Visium analysis). **a**, UMAP summary of GEFs from 1k spatial transcriptomics experiments of prostate samples from patient 1. **b**, UMAP summary of GEFs from high resolution Visium experiments of prostate samples from patient 1. Top marker genes for each GEF are available in Supplementary Table 3, 4. **c**, Benign GEFs from b (high resolution) were used as a reference set for analysis of

d, Tumour GEFs from b (high resolution). **e**, Snapshot of iCNV profiles for chr 7 and 8 from GEF10. GEF iCNV heterogeneity is highlighted by 3 subclones: the first harbouring no changes to chr 7 and 8, the second having a deletion and amplification in chr 8, and the last having alterations in both chr 7 and chr 8.

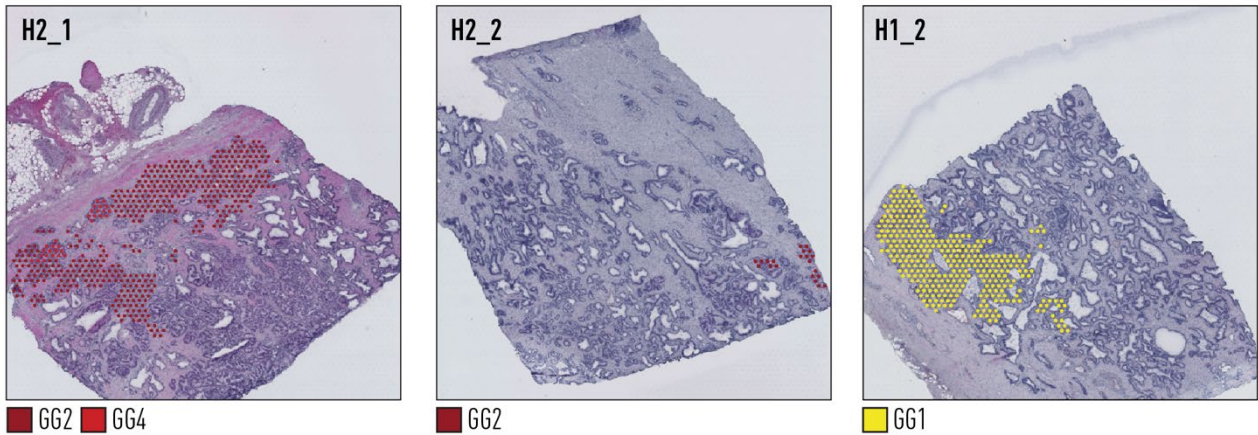
GEF = Gene Expression Factor, chr = Chromosome, iCNV = inferred Copy-Number Variant.



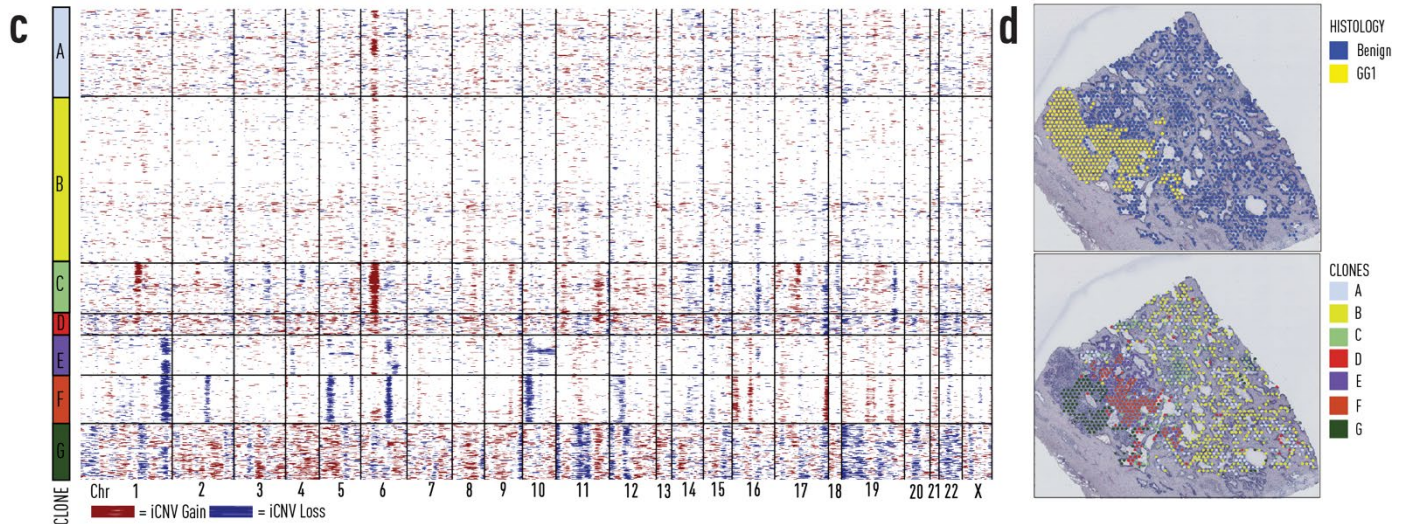
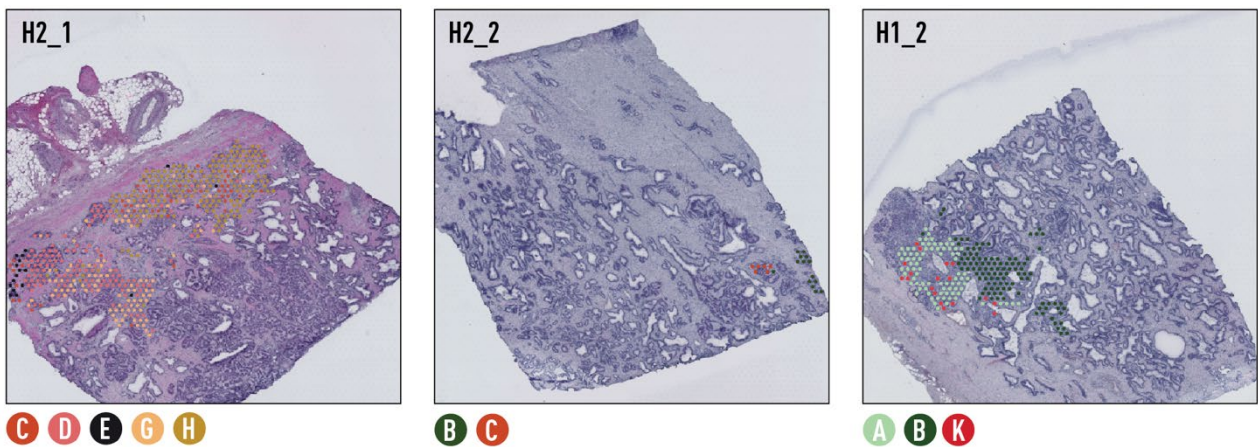
Extended Data Figure 4 | Identification of a histologically benign reference set from prostate patient 1. **a**, Visual selection of benign epithelial spots harbouring the least amount of inferred copy number variations (iCNV), as outlined by the black box bounding box. Arrows identify dendrogram nodes corresponding to barcoded spots within the box. **b**, SpatialInferCNV output of the dendrogram nodes with numerical identifiers for selection corresponding to Panel a. **c**, Finalized benign reference set from analysis of epithelial cells in prostate patient 1, section H2_1 (Fig. 3). **d**, Global iCNV profiles of the

selected benign reference set from panel a, the remainder of the benign not included in the reference set, altered benign (Clone C, Fig. 3), and the other Visium spots with luminal epithelial annotations (PIN, GG1, GG2, GG4, GG4 Cribiform).

a HISTOLOGY

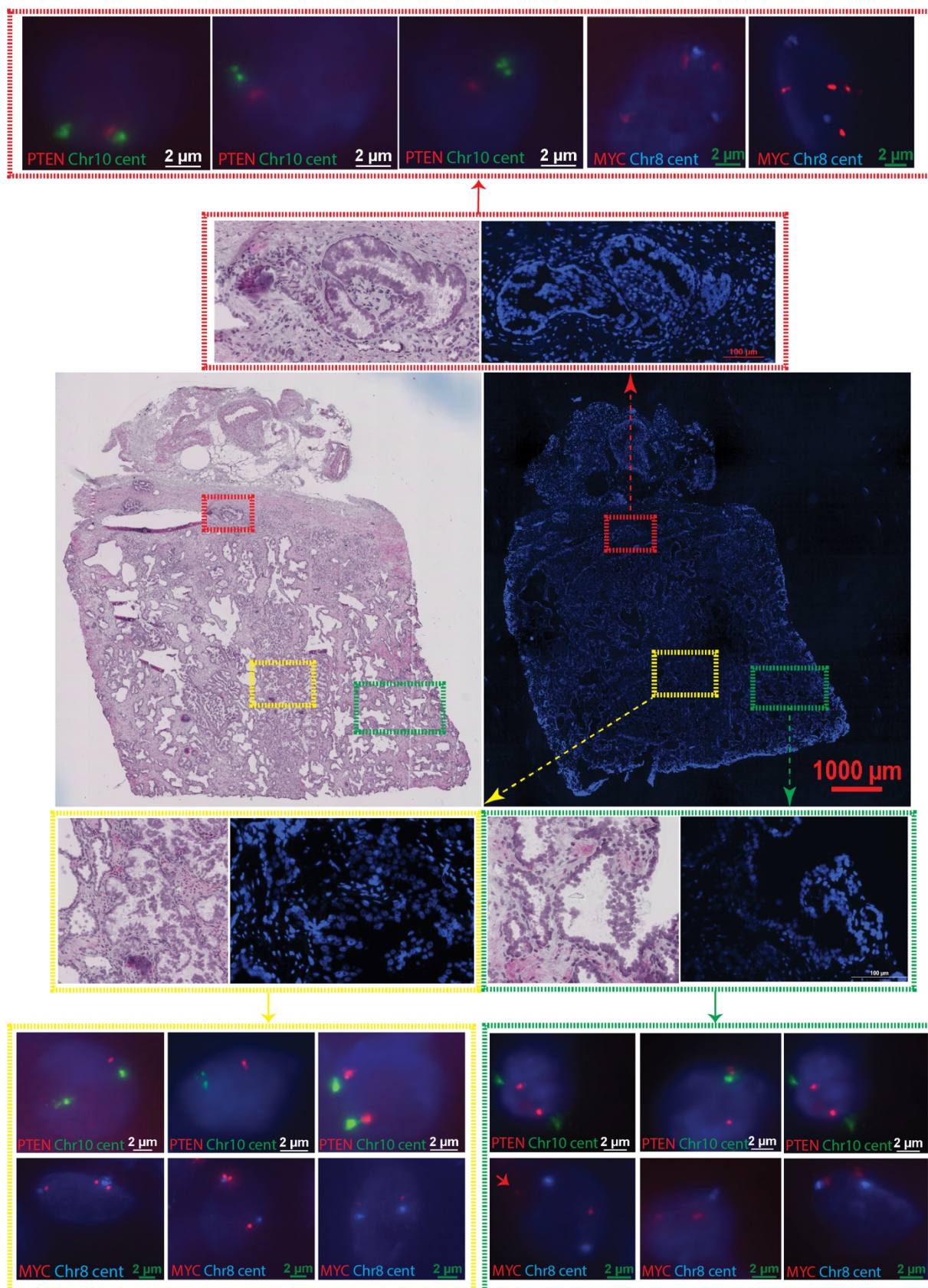


b CLONES



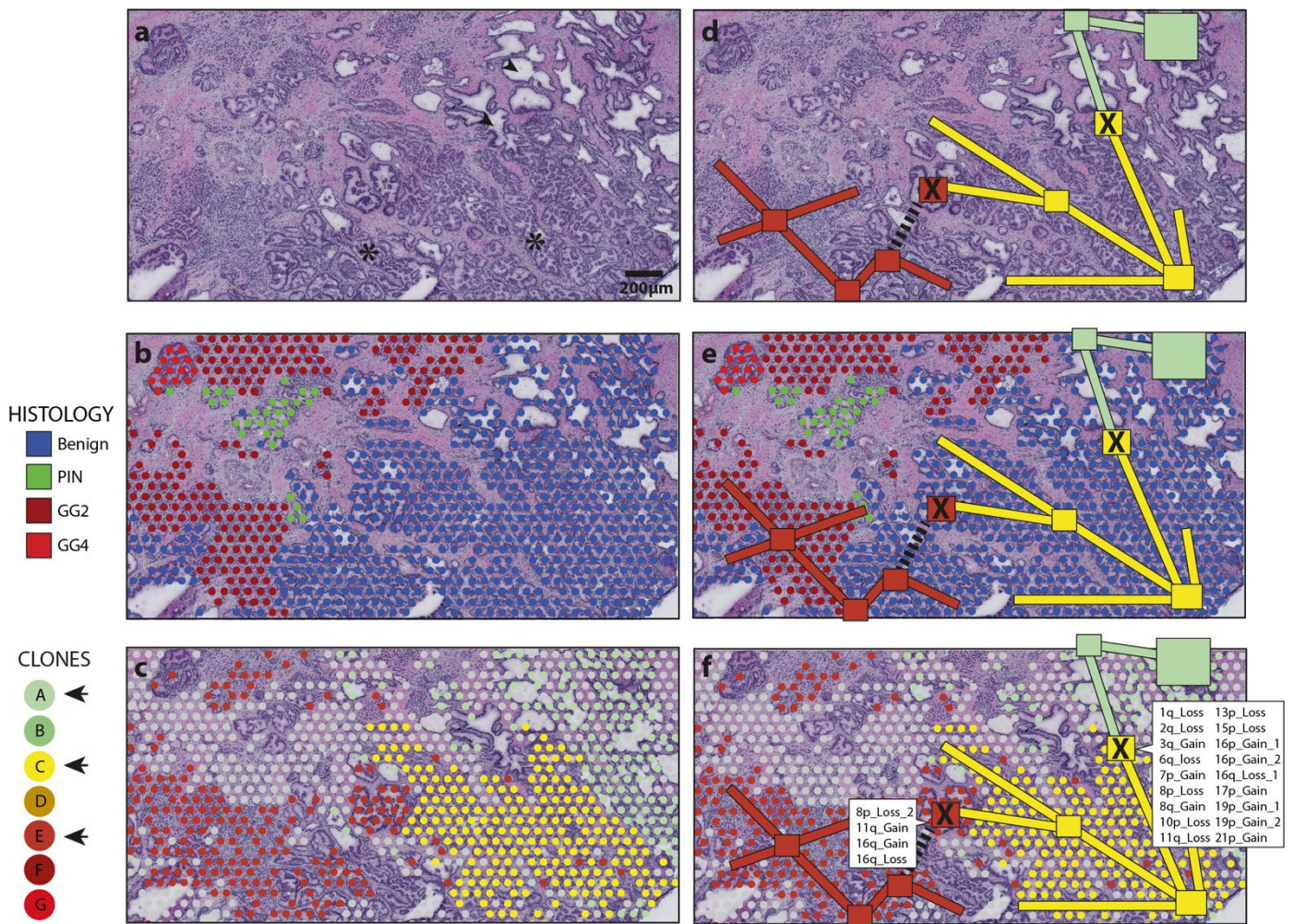
Extended Data Figure 5 | Histology and clones (from Fig. 2a) for prostate patient 1. **a**, Consensus pathology annotations for tumour spots from sections H2_1, H2_2, and H1_2. **b**, Clonal groupings of spots (approx. 10-15 cells each) determined by hierarchical clustering. **c**, Distinct iCNV profile of GG1 tumour focus from organscale prostate patient 1. iCNV profiling of epithelial Visium spots from section H1_2. **d**, Spot level histology and iCNV clone calls.

GG = ISUP Gleason 'Grade Group', iCNV = inferred copy number variant,



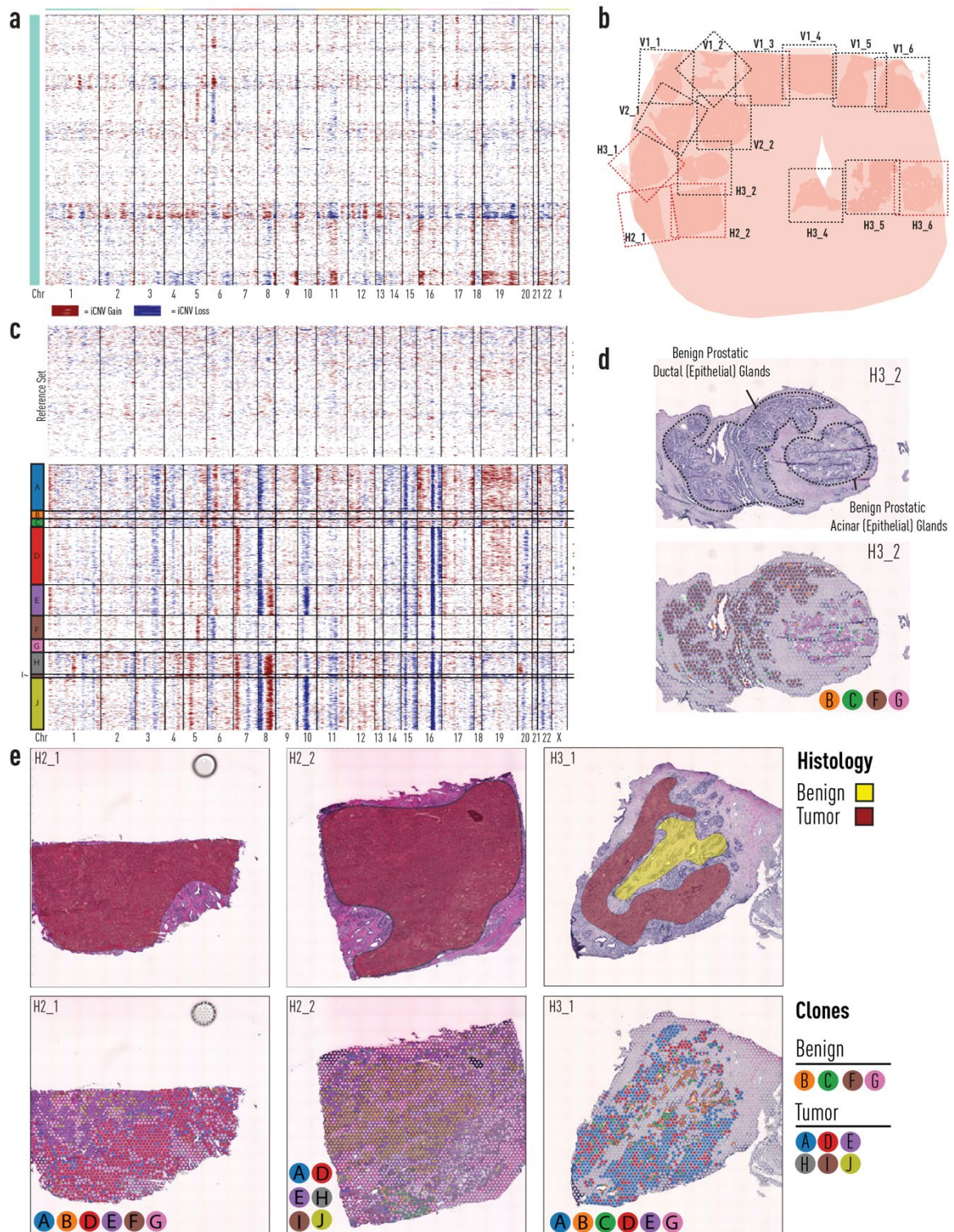
Extended Data Figure 6 | DNA FISH targeting MYC and PTEN loci. Representative images from fresh frozen prostate tissue sections obtained from patient H2.1 labelled with CytoCell MYC/8cent and PTEN/10cen probes. Three consecutive sections

were used for H&E staining and FISH. Control probes labelled chromosome 8 and 10 centromeres in (green & aqua) respectively, and MYC and PTEN shown in (red). Nuclei counterstained with DAPI (dark blue).



Extended Data Figure 7 | Branching morphogenesis and somatic mosaicism in prostate epithelium. **a**, Close up histology of Section H2_1 demonstrating clear ductal (e.g. arrow heads) and acinar (e.g. stars) branching patterns. **b**, Overlaid spot-level histology. **c**, Overlaid clone groupings (from Fig. 3).

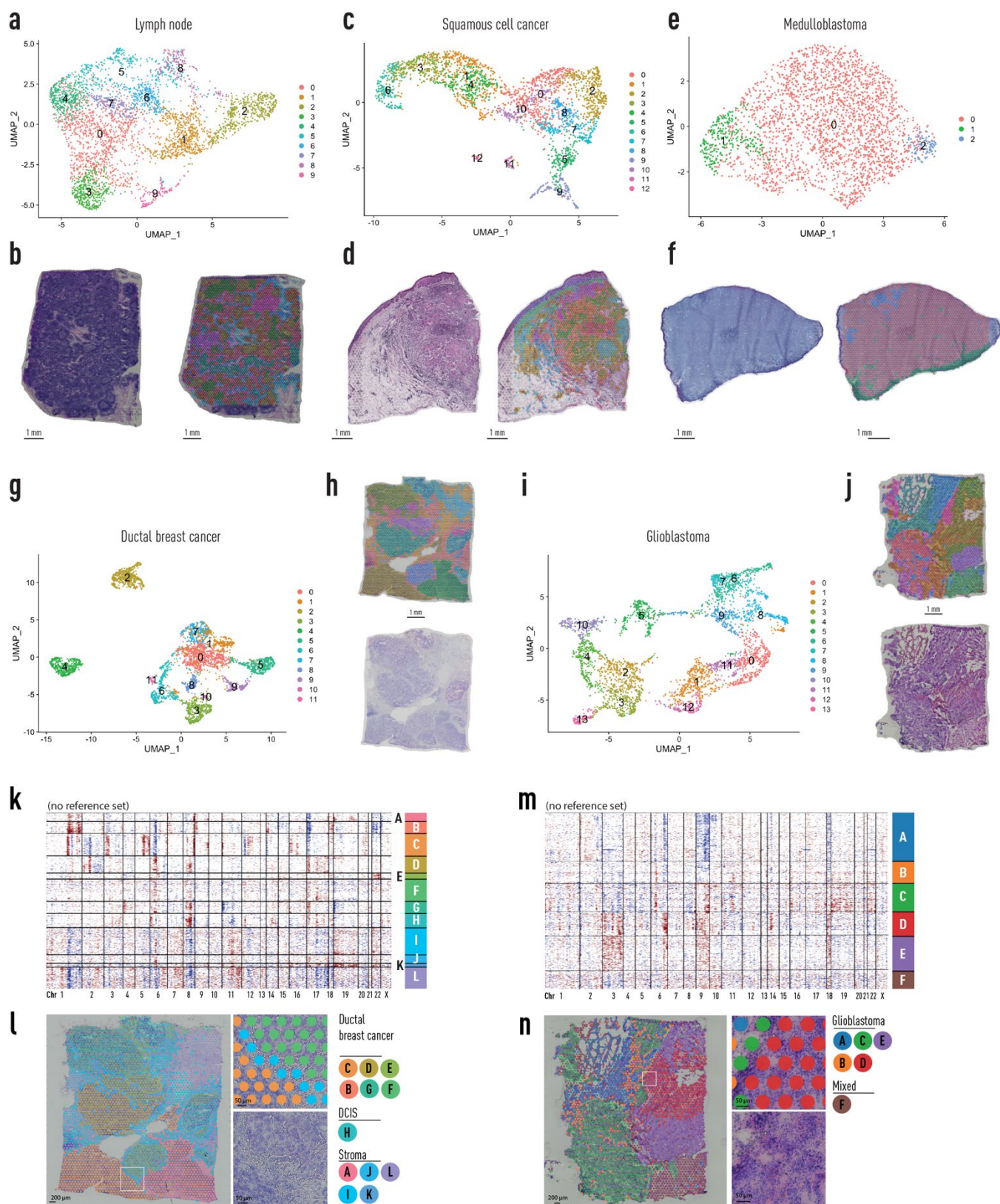
d-f, Possible arrangement of clonal expansion during branching morphogenesis with key mutational events (marked with X, iCNV events from Fig. 3) passed on to downstream branches. Dotted line represents presumed branch/duct not visible in two-dimensional plane.



Extended Data Figure 8 | Organscale prostate patient 2. a, iCNV profiles of histologically benign prostatic epithelial cells from 11 sections from prostate patient 2. **b,** Reference overview of 15 sections available for analysis: sections H2_1, H2_2, H3_1, and H3_6 harbour tumour (marked with red dotted lines). Black dotted lines represent the area covered by spatial transcriptomics array surface. **c,** Analysis of tumour foci in sections H3_1, H2_1 and H2_2. Analysis includes section H3_2, a non-tumour bearing section which included spatially co-localized benign spots harbouring iCNV alterations from panel a. **d,** Spatial histology and

clone distribution in section H3_2 (no-tumour). Benign ductal histology (Clone F) harbours distinct iCNVs (chr5 amplification, chr6 deletion), not harboured in neighbouring benign acinar glands (Clone G). **e,** Section histology (transparent red indicates tumour, and transparent yellow denotes benign) and clones from tumour-bearing sections H3_1, H2_1, and H2_2.

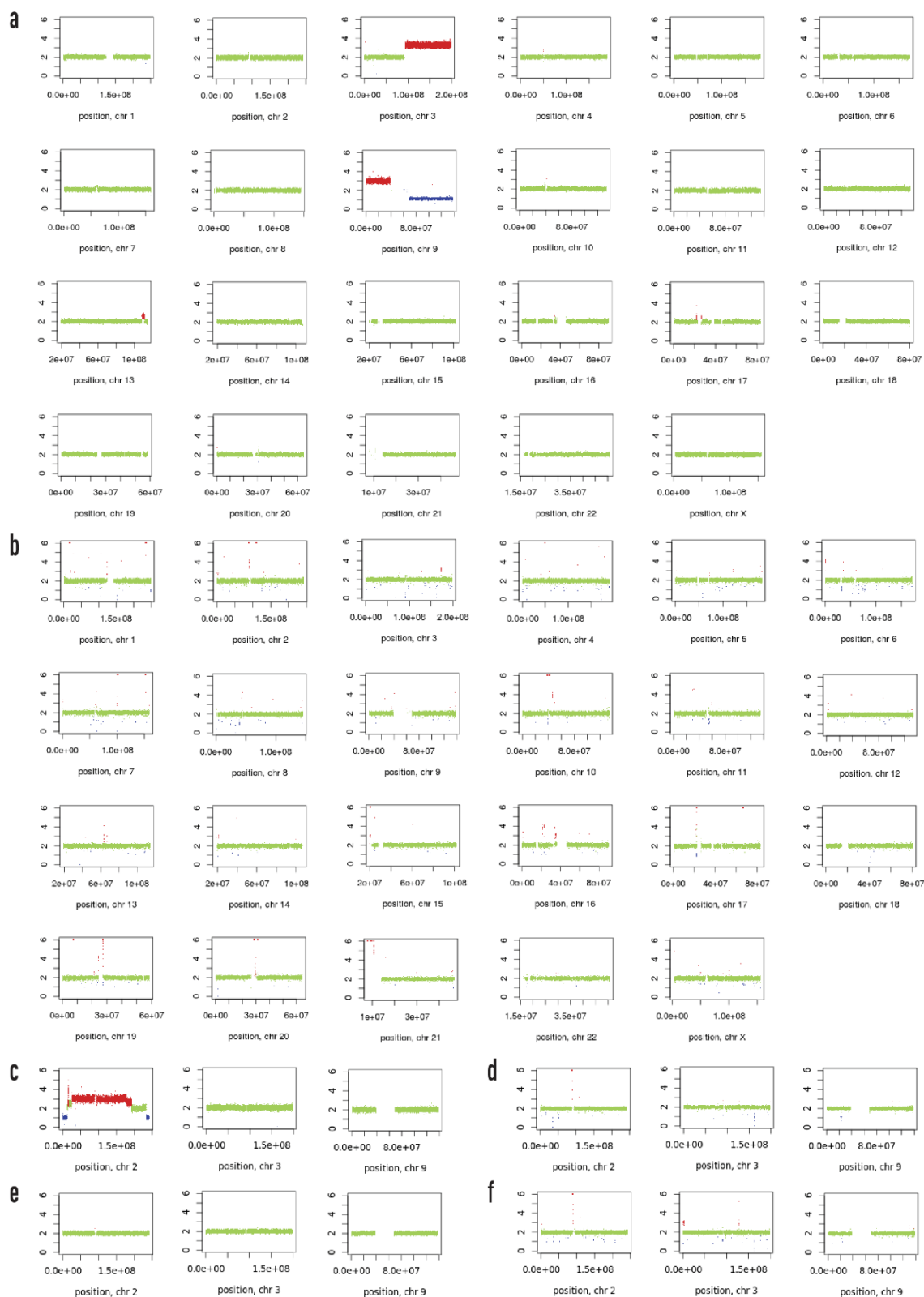
iCNV = inferred Copy-Number Variant, chr = chromosome



Extended Data Figure 9 | Spatial transcriptomics and iCNV analysis of multiple sample types. a, c, e, g, i, Transcript UMAPs of all spots labelled by cluster from human lymph node (a), human squamous cell carcinoma (c), malignant childhood brain tumour diagnosed as medulloblastoma (e) human invasive ductal breast carcinoma (g), malignant childhood brain tumour diagnosed as medulloblastoma SHH grade IV (i). **b, d, f, h, j,**

H&E stain and unbiased cluster spots visualized spatially on tissue from human lymph node (b), human squamous cell carcinoma (d), childhood medulloblastoma (f), human invasive ductal breast carcinoma (h), human glioblastoma multiforme (j). **k-n,** somatic copy number alterations in breast tissue containing ductal breast cancer and DCIS (k, l) and brain tissue containing glioblastoma (m, n). While some of the samples did not have an

annotated benign reference set, interestingly, unsupervised iCNV could still segment different histological clones. However, the lack of a reference set did reduce the ability to identify specific inferred CNVs.



Extended Data Figure 10 | Whole-genome sequencing-based copy number profiles for paediatric brain tumour patients. a, Somatic WGS CNV profile of patient 1 diagnosed with medulloblastoma (grade IV, desmoplastic/nodular, SHH-activated) with **b,** match normal blood. **c,** Somatic WGS CNV profile of Chr 2, 3 and 9 of patient 2 diagnosed with medulloblastoma (grade IV, classic morphology, SHH-activated) with **d,** match normal blood. Notably our iCNV analysis on Visium

data did not show any genomic variability in chr 2 but since Visium and WGS data were generated from different locations of each tumour, we speculate that the observed WGS CNV patterns in patient 2 could be due to the inherent spatial heterogeneity of DNA copy number alterations observed by others when sampling multiple sites of medulloblastoma tumours **e,** Somatic WGS CNV profile of Chr 2, 3 and 9 of patient 3 diagnosed with CNS embryonal tumour (grade IV, multi-layered rosettes, NOS) with **d,**

match normal blood. No CNV was detected by WGS in the chromosomes not displayed.

WGS = Whole-genome sequencing. Chr = Chromosome. SHH = Sonic hedgehog. CNS = Central nervous system. NOS = Not otherwise specified.