Article

Spatial genomics enables multi-modal study of clonal heterogeneity in tissues

https://doi.org/10.1038/s41586-021-04217-4

Received: 1 February 2021

Accepted: 8 November 2021

Published online: 15 December 2021

Check for updates

Tongtong Zhao[1,2,7], Zachary D. Chiang[1,2,3,7], Julia W. Morriss[1,2], Lindsay M. LaFave[2,4,5], Evan M. Murray[1,2], Isabella Del Priore[4,5], Kevin Meli[4,5], Caleb A. Lareau[1,2], Naeem M. Nadaf[1], Jilong Li[1], Andrew S. Earl[1,2,3], Evan Z. Macosko[1,6], Tyler Jacks[1,4,5], Jason D. Buenrostro[1,2,3,8]✉ & Fei Chen[1,2,3,8]✉

The state and behaviour of a cell can be influenced by both genetic and environmental factors. In particular, tumour progression is determined by underlying genetic aberrations[1–4] as well as the makeup of the tumour microenvironment[5,6]. Quantifying the contributions of these factors requires new technologies that can accurately measure the spatial location of genomic sequence together with phenotypic readouts. Here we developed slide-DNA-seq, a method for capturing spatially resolved DNA sequences from intact tissue sections. We demonstrate that this method accurately preserves local tumour architecture and enables the de novo discovery of distinct tumour clones and their copy number alterations. We then apply slide-DNA-seq to a mouse model of metastasis and a primary human cancer, revealing that clonal populations are confined to distinct spatial regions. Moreover, through integration with spatial transcriptomics, we uncover distinct sets of genes that are associated with clone-specific genetic aberrations, the local tumour microenvironment, or both. Together, this multi-modal spatial genomics approach provides a versatile platform for quantifying how cell-intrinsic and cell-extrinsic factors contribute to gene expression, protein abundance and other cellular phenotypes.

Tissue function requires precise spatial organization of cell types, whose states are influenced by cell-intrinsic genetic factors and extrinsic environmental cues. In cancer, clonal populations of tumour cells evolve a diverse repertoire of DNA mutations, copy number alterations (CNAs), and large chromosomal rearrangements[1,2], resulting in intra-tumour genetic heterogeneity, which is associated with an increased risk of drug resistance, metastasis and relapse[3,4]. Concomitantly, surrounding normal cells that make up the tumour microenvironment communicate to form spatial neighbourhoods with distinct biochemical and biomechanical properties[5,6] that influence cell migration and invasion[7,8], as well as drug permeability[9]. Decoupling and quantifying these genetic aberrations and environmental cues within a tumour is critical to understanding cancer progression and improving treatments.

Current methods for delineating intratumour genetic heterogeneity include deep sequencing to quantify mutant allele frequencies[10,11] and single-cell whole-genome sequencing[12–14]. These methods leverage genetic alterations that occur during the evolution of the tumour to reconstruct phylogenetic cell lineages[10–14], but do not measure spatial organization. By contrast, multi-region sequencing methods[15–17] such as laser-capture microdissection (LCM) preserve spatial context, but are mostly limited to clearly observable late-stage cancers and require manual selection of cells, constraining throughput and de novo discovery. The recently developed in situ genome sequencing[18] enables untargeted spatial measurements of DNA, but focuses on high-resolution imaging of chromosome structure, precluding analysis of tissue sections. It therefore remains poorly understood how tumour clones are organized within a tissue, and to what extent cancer progression is driven by clone-specific genetic aberrations or environmental cues, highlighting a need for new methods that can integrate genomic, transcriptomic and spatial measurements at scale.

Spatially resolved DNA sequencing

Slide-seq[19,20] (hereafter referred to as slide-RNA-seq) is a scalable technology that uses barcoded bead arrays to capture spatially resolved genome-wide expression. Here we developed slide-DNA-seq, a method that enables spatially resolved DNA sequencing from intact tissues.

We first generate a spatially indexed array of 3-mm beads, as developed for slide-RNA-seq[19,20]. Each 10-μm polystyrene bead contains a unique DNA barcode that corresponds to a spatial location and is read out using sequencing by ligation chemistry[19,20]. We then cryosection tissues and transfer a single 10-μm-thick fresh-frozen section onto the sequenced bead array (Fig. 1a). To enable unbiased capture of DNA, the tissue section is treated with HCl to remove histones and transposed with Tn5 to create genomic fragments flanked by custom adapter

[1]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [2]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. [3]Gene Regulation Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]David H. Koch Institute, Massachusetts Institute of Technology, Cambridge, MA, USA. [6]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [7]These authors contributed equally: Tongtong Zhao, Zachary D. Chiang. [8]These authors jointly supervised this work: Jason D. Buenrostro, Fei Chen. ✉e-mail: jason_buenrostro@harvard.edu; chenf@broadinstitute.org

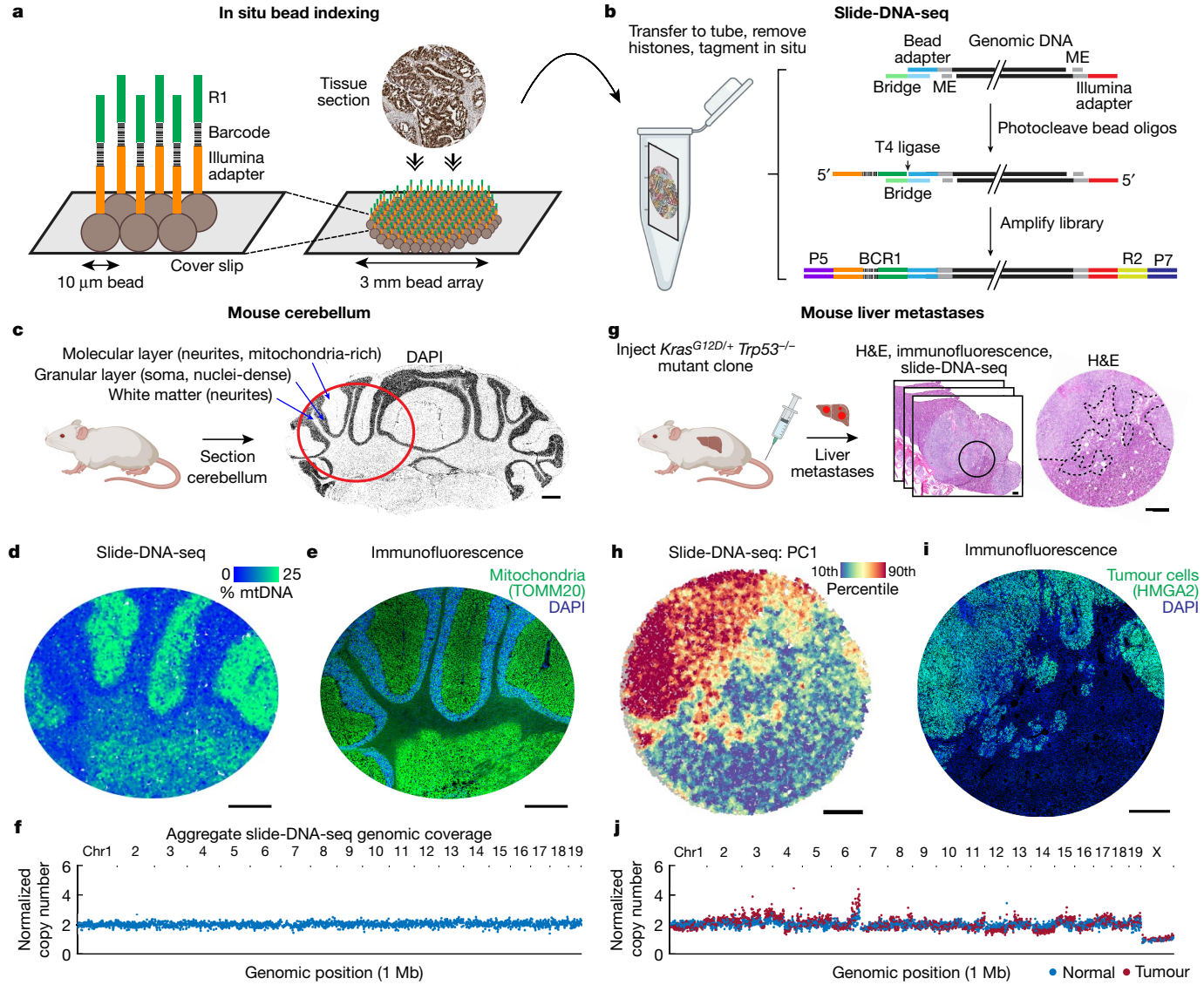


Fig. 1 | Slide-DNA-seq enables spatially resolved DNA sequencing. **a**, Schematic of in situ bead indexing. An array of randomly deposited beads is spatially indexed by in situ sequencing of DNA barcodes. Fresh-frozen tissue is cryosectioned onto the array. **b**, Schematic of slide-DNA-seq library construction. Genomic DNA is transposed with Tn5. Hybridization of a bridge oligonucleotide enables ligation of photocleaved, spatially indexed bead oligonucleotides to genomic fragments. BC, barcode; ME, mosaic ends; P5/P7, Illumina adaptor; R1, Illumina read 1; R2, Illumina read 2. **c**, A DAPI-stained cryosection of a mouse cerebellum. The red circle indicates the approximate region shown in **d**, **e**. **d**, Slide-DNA-seq of a cerebellar section with beads coloured by percentage of fragments aligned to mitochondrial genome. mtDNA, mitochondrial DNA. **e**, Adjacent section to the one in **d**, stained with DAPI and antibody against the mitochondrial protein TOMM20. **f**, Normalized copy number per 1-Mb genomic bin for aggregated beads from **d**. **g**, Serial sections from *Kras*$^{G12D/+}$;*Trp53*$^{-/-}$ liver metastases were processed for H&E staining. Centre, circle indicates the region in **h**, **i**. Right, dotted lines indicate tumour boundary. **h**, Slide-DNA-seq of mouse liver section with beads coloured by principal component 1 scores (PC1). For visualization, scores for each bead are smoothed by 50 PC neighbours and 10 spatial neighbours (36 µm diameter). **i**, Adjacent section to the one in **h**, stained with DAPI and antibody against the tumour marker HMGA2. **j**, Normalized copy number per 1-Mb genomic bin for aggregated normal and tumour beads from the liver section in **h**. Scale bars, 500 µm. Grey beads are shown for spatial context but are excluded from the analysis.

sequences[21,22] (Supplementary Table 1). We then photocleave spatial barcodes from the beads, ligate them to proximal genomic fragments, and PCR amplify the resulting DNA sequencing library (Fig. 1b). Following library construction, we perform high-throughput paired-end sequencing and use DNA barcodes to associate each genomic fragment with a spatial location on the bead array. These associations enable us to reconstruct the spatial organization of DNA in a tissue without imaging the sample under a microscope. We developed optimizations for tissue fixation, histone removal and bridge oligonucleotide hybridization that collectively maximize library size, make chromatin uniformly accessible to Tn5 (Extended Data Fig. 1), and preserve tissue architecture (Extended Data Fig. 2). Following our initial optimizations, each array contains 20,000 to 40,000 beads with a median 165 to 421 fragments per bead (tumour tissues; Extended Data Fig. 3). Furthermore, we developed a proof-of-concept protocol variant that uses repeated Tn5 tagmentation to improve yield, resulting in a tenfold increase in genomic fragments (Extended Data Fig. 3, Supplementary Methods). Detailed metrics for all tissues analysed in this study are listed in Supplementary Table 2.

To determine the spatial and genomic resolution of this approach, we first applied slide-DNA-seq to the mouse cerebellum, which contains distinct nuclei-dense (soma) and mitochondria-rich (neurites) regions (Fig. 1c). We reasoned that these patterns should be reflected in the spatial distribution of nuclear versus mitochondrial DNA fragments.

Indeed, striations of nuclear versus mitochondrial DNA content were apparent from slide-DNA-seq data (Fig. 1d, Extended Data Fig. 3). We then used these patterns to measure our spatial resolution by performing immunofluorescence and DAPI staining on serial tissue sections of the same cerebellum, resulting in a lateral diffusion estimate of approximately 25 μm (Fig. 1e, Extended Data Fig. 4, Supplementary Methods). To measure genomic resolution, we corrected the data for sequence biases and normalized coverage using bulk sequencing of the same tissue (Extended Data Fig. 5, Supplementary Methods, Supplementary Discussion). Using this approach, 99.78% of non-overlapping 1-Mb genomic bins had a normalized copy number between 1.5 and 2.5 (Fig. 1f, Extended Data Fig. 6). Together, these data show that slide-DNA-seq can spatially localize genomic information within normal tissues.

Detecting spatial distribution of CNAs

We next applied slide-DNA-seq to measure the spatial distribution of copy number alterations (CNAs) in a tumour section using genetically engineered mouse models of lung adenocarcinoma that are known to harbour chromosomal amplifications and deletions[23]. First, we isolated and expanded a single tumour clone from a *Kras*$^{G12D/+}$*Trp53*$^{-/-}$ (KP) mouse lung tumour[24,25] and injected this clone into the tail vein of a mouse, giving rise to large metastases in the liver (Fig. 1g). We then collected multiple serial sections of liver metastases to perform slide-DNA-seq, along with haematoxylin and eosin (H&E) staining and immunofluorescence for HMGA2, a late-stage tumour marker. To characterize tumour heterogeneity within the tissue, we developed a slide-DNA-seq analysis workflow comprising two main tasks: (1) de novo identification and spatial localization of clonal populations, and (2) characterization of genomic CNAs for each clone.

First, to detect and localize tumour clones within a tissue, we smoothed bead data on the basis of spatial proximity (k = 50 nearest beads, approximately 110 μm diameter; Extended Data Fig. 3; median 18,587 ± 5,300 fragments) and performed principal component analysis (PCA) to find co-associated genomic regions with variable coverage across the tissue. We then used these regions to assign a clonal identity to each bead on the slide-DNA-seq array by k-means clustering (Extended Data Fig. 7, Supplementary Methods). When we applied this approach to the slide-DNA-seq array from the liver metastases, principal component 1 (PC1, 2.89% variance explained) showed spatial patterning (Fig. 1h) that was visually concordant with immunofluorescence on a serial section against the late stage tumour marker HMGA2[26–28] (Fig. 1i). To validate whether this approach could be used to identify genetically distinct tumour clones, we performed downsampling on bulk sequencing of 4 tumour cell lines and found robust accuracy (99.38%) with as few as 1,000 fragments per sample (Extended Data Fig. 8, Supplementary Methods), suggesting that this strategy is sufficient for slide-DNA-seq data.

The second task in the analysis workflow is to characterize the CNAs present in each tumour clone. To do this, we aggregated data from hundreds to thousands of raw beads on the basis of the cluster assignments from the first task and visualized the genomic coverage of each cluster at 1-Mb resolution. When applied to the liver metastases array, the tumour-associated cluster displayed significant CNAs, including the amplification of chromosome (chr)6 that is characteristic for *Kras*-induced lung tumours[23], whereas the normal cluster showed comparatively uniform coverage (Fig. 1j). Further comparisons to a biological replicate performed on a serial section revealed visually concordant tissue architecture, as well as high correlation between tumour copy number profiles (Pearson's r = 0.986; Extended Data Fig. 9). To quantify the accuracy of the copy number analysis, we used the diploid mouse cerebellum data to systematically evaluate coverage at a range of bin sizes and spatial resolutions (Extended Data Fig. 10, Supplementary Methods). Together, these results demonstrate that our slide-DNA-seq analysis workflow enables de novo discovery and localization of tumour regions at approximately 1-Mb genomic resolution (Supplementary Discussion).

Spatial genomics of metastatic clones

To demonstrate that our experimental and computational approach can distinguish between clones within a tissue, we injected multiple clones originating from two independently derived metastatic KP tumours into the tail vein of a mouse, which gave rise to large metastases in the liver. We then performed H&E staining and identified a region of the tissue that appeared to have two spatially distinct metastases (Fig. 2a). Immunohistochemistry on the same region of a serial section revealed that the two varied in protein levels of tumour marker HMGA2[26], suggesting that they may originate from different metastatic clones (Extended Data Fig. 11a).

We then applied slide-DNA-seq to a third serial section of the same liver tissue. Using the PCA approach described above, we found that both PC1 and PC2 explained substantial variance (4.21% and 2.50%, respectively), allowing the beads to be assigned to 3 distinct clusters on the basis of their genomic profiles (Fig. 2b). One of these clusters was visually concordant with H&E staining of normal tissue, whereas the other two appeared to correspond to the different metastases. We developed a permutation test to spatially localize statistically significant CNA gains or losses present in one or both of the metastases, and detected differential regions on chr6, chr15 and chr19 (Fig. 2c, Supplementary Methods). We then tested the aggregate genomic coverage in selected regions for statistical significance (two-sided Wilcoxon rank sum test; P-values in Fig. 2d), providing further evidence that they were seeded by different clones. Additionally, we observed that one clone was probably triploid, which we confirmed independently using flow cytometry (Extended Data Fig. 12).

To test whether genetic differences between the two clones were reflected in cell state, we performed slide-RNA-seqV2[20] on a fourth serial section and collected paired single-nucleus RNA sequencing (snRNA-seq). Unsupervised clustering of the snRNA-seq data and spatial projection[29] onto slide-RNA-seq beads (Methods, Supplementary Table 3) revealed that the two metastases were transcriptionally distinct (Fig. 2e, Extended Data Fig. 11b, c); there were 3,732 genes differentially expressed between the two clones (Fig 2f, Supplementary Table 4; two-sided z-test, false discovery rate < 0.01, $\log_2$(fold change) > 1, minimum of 100 transcripts). Clone A had higher expression of late-stage tumour markers, including *Hmga2* (lung metastases), *Tm4sf1* (JAK–STAT signalling) and *Vim* (cell motility), whereas the top hits for clone B included *Aqp5* (loss of lineage identity) and epithelial-to-mesenchymal transition markers *S100a4* and *Vcan*[25] (Fig. 2f, g). Although both clones exhibited epithelial-to-mesenchymal transition and metastasis expression signatures, these differentially expressed genes may reflect divergent paths of tumour evolution. Furthermore, we found differential monocyte localization (P = 0.0002; permutation test) into clone B, reflecting a higher degree of immune infiltration (Extended Data Fig. 11e, f, Supplementary Methods). Altogether, these data demonstrate that paired slide-DNA-seq and slide-RNA-seq enable spatial characterization of genetically distinct metastatic tumour clones and their associated cell states.

Subclone detection in human colon cancer

We then sought to determine whether slide-DNA-seq could detect clonal heterogeneity de novo in a primary human tumour. We selected a sample from a stage IIIB colorectal tumour, because colorectal cancer is one of the most common causes of cancer-related deaths worldwide and 84% of such tumours display chromosome instability[30,31]. As before, we performed H&E staining, multiplexed immunohistochemistry and slide-DNA-seq on serial sections (Fig. 3a). First, we examined H&E staining, observing many localized aggregates of tumour cells about

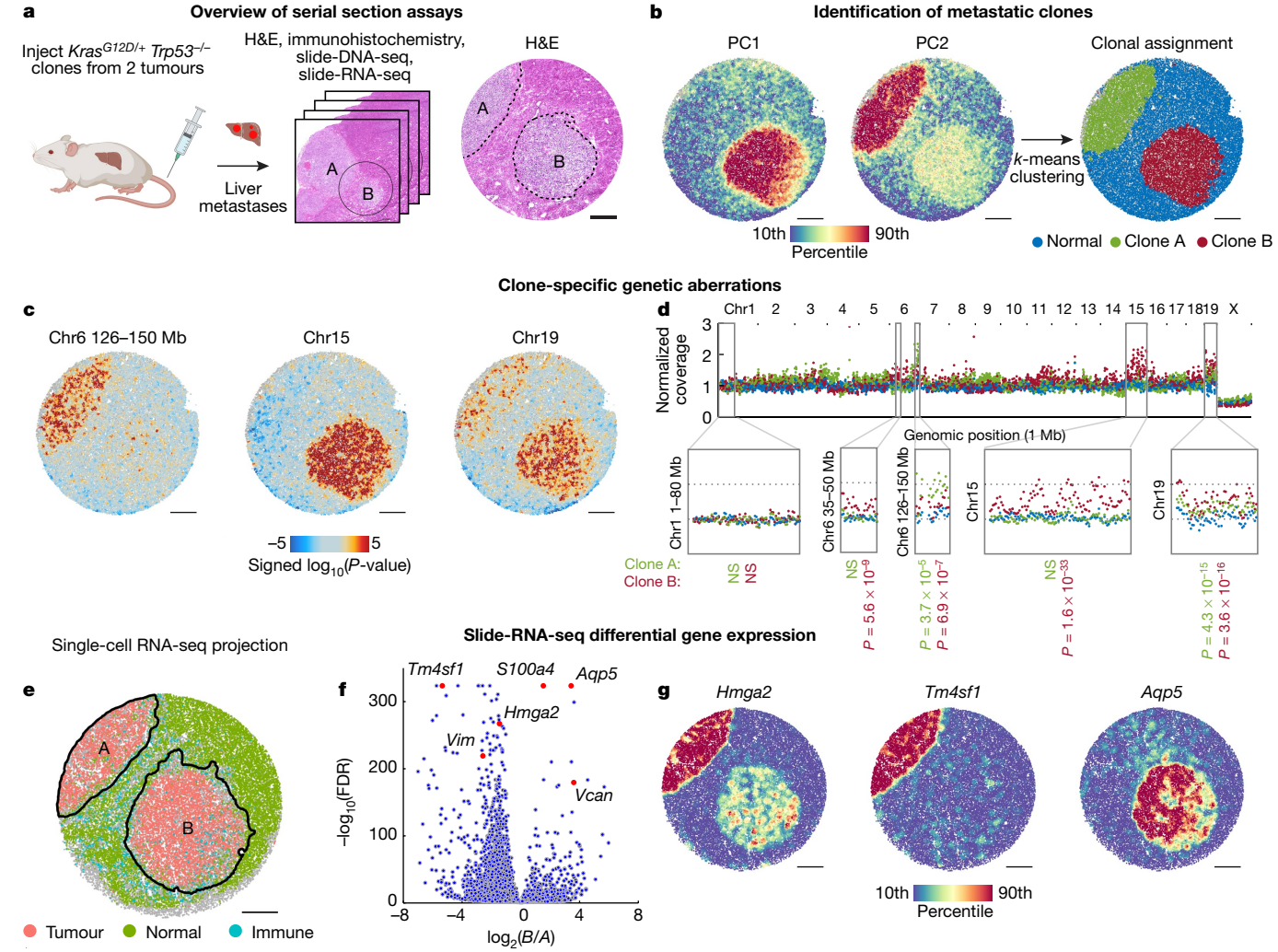


Fig. 2 | Paired slide-DNA-seq and slide-RNA-seq characterize the genetics and transcriptomes of distinct metastatic clones. a–**g**, Serial sections from *Kras*$^{G12D/+}$*Trp53*$^{-/-}$ liver metastases were processed for H&E staining (centre, circle indicates analysed region; right, dotted lines indicate tumour boundaries), immunohistochemistry (Extended Data Fig. 5), slide-DNA-seq (**b**–**d**) and slide-RNA-seq (**e**–**g**). **b**, PC1 (left) and PC2 (middle) of slide-DNA-seq genomic coverage. Beads clustered using *k*-means ($k = 3$) and annotated as normal, clone A or clone B (right). **c**, Genomic region enrichment signed *P*-values for chromosomes 6 (126–150 Mb), 15 and 19 (two-sided permutation test, not adjusted for multiple comparisons). Amplifications, red; deletions, blue. **d**, Genomic coverage profiles of aggregate normal (blue), clone A (green), and clone B (red) beads from **b**. Genomic coverage normalized to 1 to compare profiles of different ploidy (Extended Data Fig. 12). *P*-values calculated using two-sided Wilcoxon rank sum test to compare clone and normal coverage. NS, not significant. **e**, Slide-RNA-seq of mouse liver serial section coloured by tumour, normal and immune cell classes, as assigned from single-cell projection. Detailed cell-type labels are shown in Extended Data Fig. 13. **f**, Differentially expressed genes between clones A and B. Genes shown in **g** or referred to in text are labelled and indicated by red dots. **g**, Normalized expression of selected genes from **f**. Scale bars, 500 µm. Grey beads are shown for spatial context but are excluded from the analysis.

100–500 µm in diameter. We hypothesized that each of these aggregates could arise from a single clonal lineage, suggesting constraints on migration or, alternatively, that each aggregate could contain a mixture of cells from different lineages, indicating cell intermixing.

To distinguish between these two possibilities, we performed PCA and unsupervised clustering on the slide-DNA-seq data as described above, which resulted in three distinct clusters of genomic profiles (Fig. 3b). One of these clusters had a spatial distribution that was visually concordant with normal tissue in the H&E staining (Fig. 3b right, blue), but also included regions of moderate PC1 scores, suggesting a low abundance of cancer cells harbouring CNAs. By contrast, the other two clusters displayed high PC1 scores and were spatially restricted to distinct tumour aggregates, supporting the hypothesis that each aggregate originates from a single lineage. This finding is consistent with reports suggesting that individual colorectal tumour cells seed a glandular organization in which neighbouring cells share a recent common ancestor[32,33]. We validated the tumour architecture detected by slide-DNA-seq through co-registration of the slide-DNA-seq array, H&E staining and immunohistochemistry for the tumour marker MKI67 and the immune marker CD45 (Fig. 3c).

We then set out to characterize the genetic aberrations of the identified subclones. We found several genetic aberrations, including chr8q amplification and loss of chr15 and chr18, that were shared across all tumour regions (Fig. 3d, e, Supplementary Methods), indicating that they arose early in tumour evolution and may have had an important role in tumour initiation. The chr8q amplification contains genes known to promote tumour progression, including the proto-oncogenes *MYC* and *MYBL1*[34], whereas deletion of chr15 results in loss of multiple genes required for genome stability, including *TP53BP1*[35], *RAD51*[36] and *FAN1*[37]. Supporting these observations, chr8q gain and chr18 loss were identified as typical early events in an evolutionary history of 60 colorectal tumours[10]. In contrast to these shared aberrations, we observed

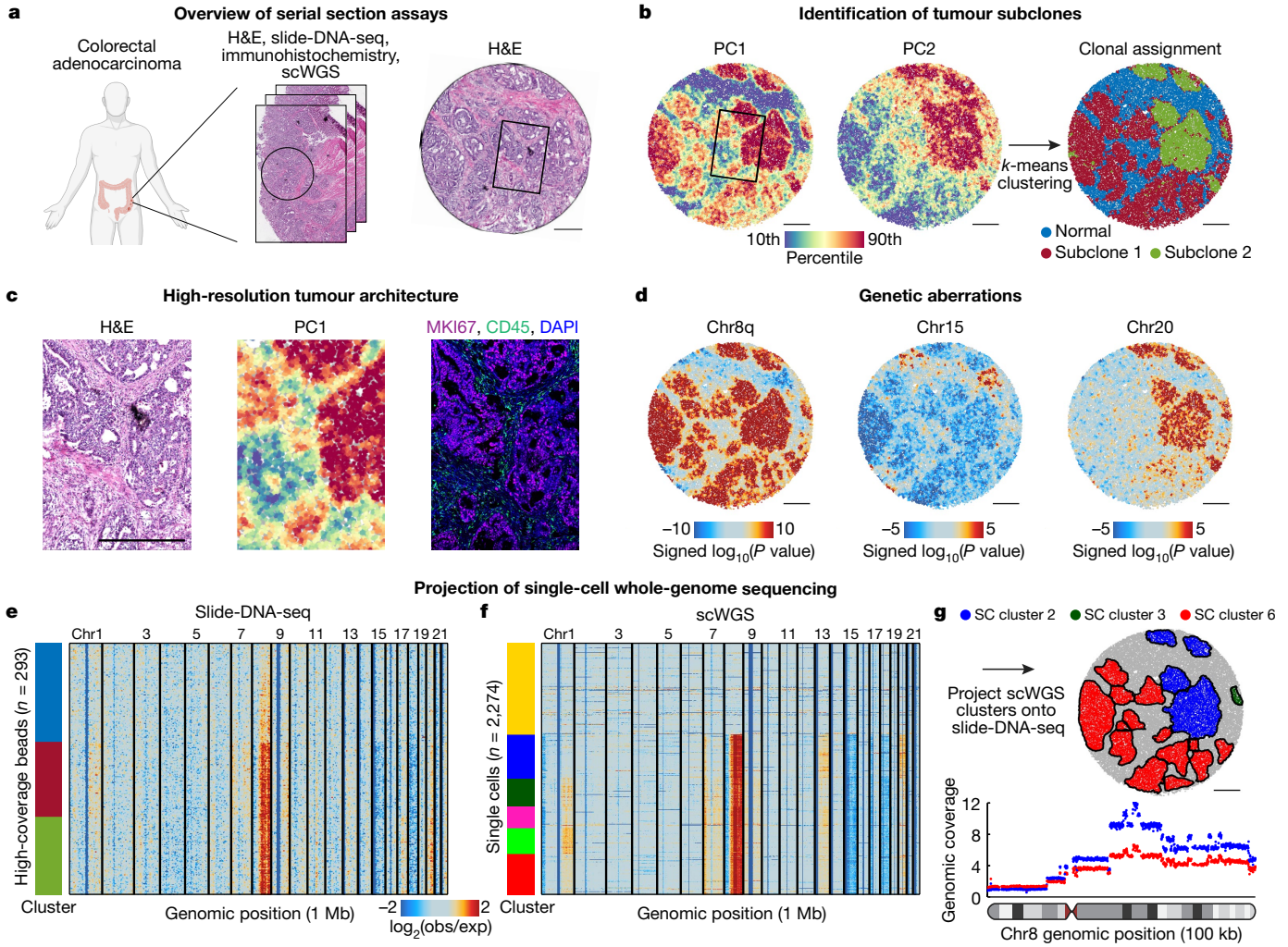


Fig. 3 | De novo identification of spatial tumour clones in primary human colorectal cancer. a–**g**, Serial sections of primary human colorectal tumour were processed for H&E staining (right, **c**), slide-DNA-seq (**b**–**e**), and multiplexed immunohistochemistry (**c**)—scWGS was performed on the same sample (**f**, **g**). **b**, PC1 (left) and PC2 (middle) of slide-DNA-seq genomic coverage. Beads clustered using *k*-means ($k = 3$) and annotated as normal, subclone 1 or subclone 2 (right). **c**, Magnified view of boxed regions in **a**, **b**. Right, serial section stained with antibodies against MKI67, CD45 and DAPI. **d**, Genomic region enrichment signed *P*-values for chromosomes 8q, 15 and 20 (two-sided permutation test, not adjusted for multiple comparisons). **e**, Copy number profiles for 293 high-coverage slide-DNA-seq beads. **f**, Copy number profiles for 2,274 single cells profiled via scWGS. Profiles within each cluster are ordered by the PC1 score in **e**, **f**. **g**, Matched single-cell clusters projected onto slide-DNA-seq array (top). Genomic coverage of chromosome 8 at 100 kb resolution for single-cell clusters 2 and 6 (bottom). Scale bars, 500 µm. Amplifications, red; deletions, blue (**d**–**f**). Grey beads are shown for spatial context but are excluded from the analysis.

subclonal amplifications of chr1q, chr7, and chr20, which presumably occurred at a later stage of evolution (Fig. 3d, e). Notably, previous analyses of colorectal cancers classified chr7p amplification as a typically clonal (rather than subclonal) event, whereas both loss and gain of chr20p were identified as frequent subclonal aberrations[10,38]. The detection and temporal classification of these events demonstrate the utility of slide-DNA-seq for studying the evolution of clonal heterogeneity.

To validate these genetic aberrations, we performed single-cell whole-genome sequencing (scWGS) on the same colorectal tumour. This approach sampled cells from the entirety of the tumour (100-fold more material than the slide-DNA-seq tissue section), so we expected to potentially identify additional subclones. In line with this expectation, analysis of 2,274 high-coverage single-cell CNA profiles resulted in one normal cluster and five tumour clusters, some of which resembled the slide-DNA-seq CNA profiles (Fig. 3f). We then sought to project the high-coverage sequencing onto the slide-DNA-seq array to identify CNAs at enhanced resolution (Supplementary Methods). The spatial regions predominantly matched two separate scWGS clusters, supporting the analysis using slide-DNA-seq only, but we also found a small region with distinct genetic aberrations that was revealed only with the higher coverage of the scWGS data (Fig. 3g, top, Extended Data Fig. 13). Having demonstrated improved spatial resolution, we then re-analysed the matched scWGS clusters at 100 kb genomic resolution, revealing a complex CNA landscape in chromosome 8 (Fig. 3g, bottom). Together, these analyses validate that slide-DNA-seq alone is sufficient for de novo discovery and localization of distinct tumour clones within a tissue and show that CNA characterization can be enhanced through integration with scWGS.

Multi-modal analysis of clonal heterogeneity

Finally, to demonstrate the unique capabilities of a multi-modal spatial sequencing approach, we sought to quantify how tumour transcriptional programs are controlled by both genetics and environmental cues. We first performed H&E staining, slide-DNA-seq and slide-RNA-seqV2 on serial sections from a nearby region of the

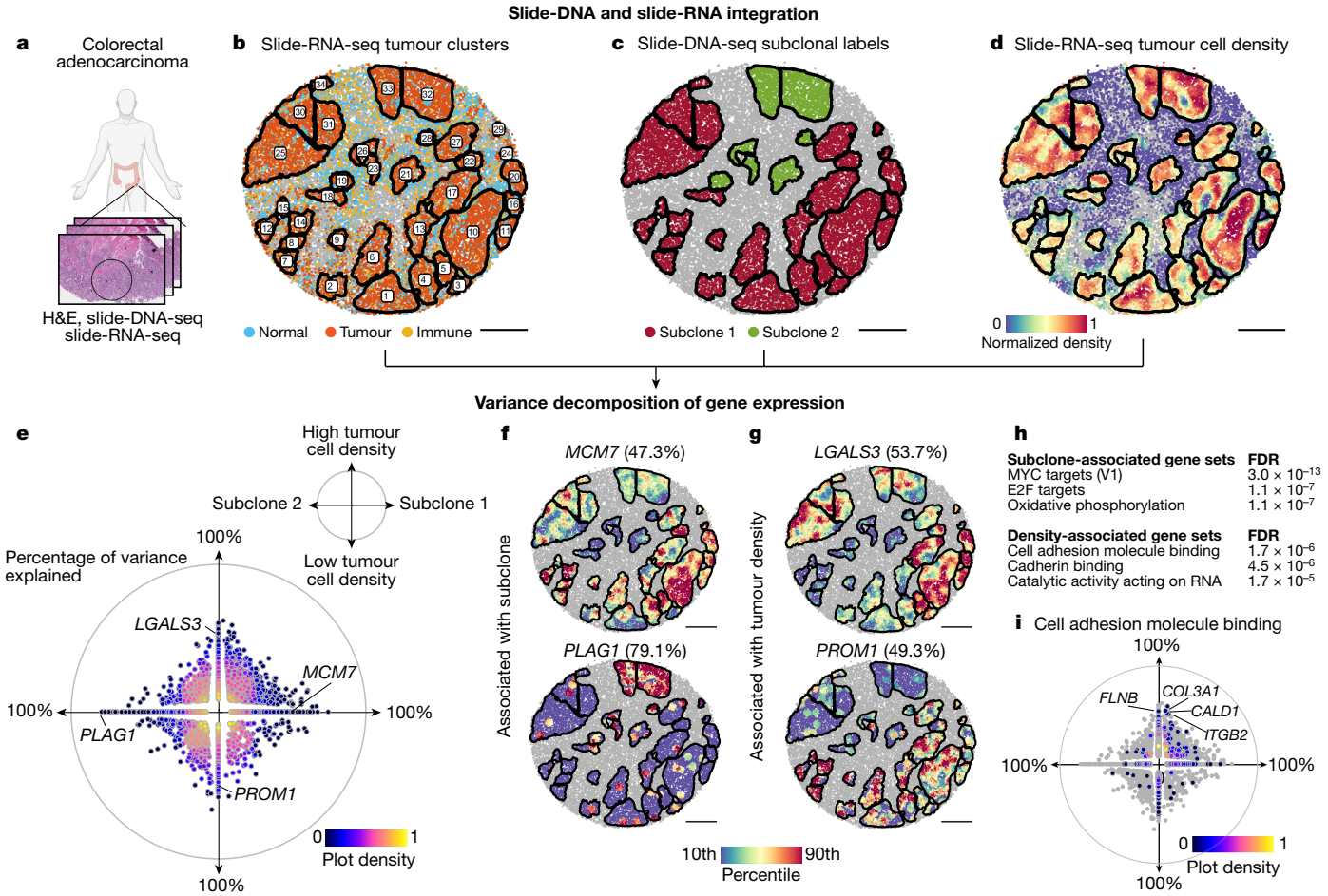


Fig. 4 | Decomposition of transcriptional programs driven by genetic aberrations and tumour density. a–**i**, Serial sections from nearby region of human colorectal tumour from Fig. 3 were processed for H&E staining, slide-DNA-seq (Extended Data Fig. 14) and slide-RNA-seq (**b**–**i**). **b**, Slide-RNA-seq of a colon tumour section with beads coloured by assignment to normal, tumour or immune clusters. Black lines denote boundaries of spatially distinct tumour regions. **c**, Subclone labels for spatial tumour regions (defined via co-registration with slide-DNA-seq serial section) plotted on slide-RNA-seq array from **b**. **d**, Tumour density plotted on the slide-RNA-seq array from **b**. **e**, Genes plotted by percentage of variance explained by subclone (*x*-axis) and/or tumour density (*y*-axis), coloured by plot density ($n = 2{,}148$; stepwise regression, $P < 0.05$). **f**, Top subclone-associated genes, with expression plotted for spatial tumour regions. **g**, Same as **f** but for top tumour density-associated genes. **h**, Selected gene sets significantly associated with either subclone or tumour density. **i**, Cell adhesion molecule-binding genes ($n = 544$) plotted by percent variance explained by subclone (*x*-axis) and tumour density (*y*-axis), coloured by plot density. All other genes from **e** are shown in grey. Scale bars, 500 µm. Grey beads shown for spatial context but excluded from analysis.

colorectal tumour (Fig. 4a) and co-registered the arrays to integrate pathological, genomic and transcriptomic information. We then identified spatially distinct regions of tumour cells (Fig. 4b, Supplementary Methods) and proceeded to assign each one with a subclonal identity (Fig. 4c) and quantify the local tumour density (Fig. 4d, Supplementary Methods). Comparison with the H&E staining validated the spatial architecture of the subclones identified by slide-DNA-seq as well as the tumour density quantified by slide-RNA-seq (Extended Data Fig. 14).

Given both subclonal identity (cell-intrinsic) and tumour density (cell-extrinsic) measurements, we set out to deconvolve how these factors contribute to the transcriptional programs of the colorectal tumour. To this end, we used a variance-decomposition approach that, for each gene, calculates the percentage of gene expression variance explained by subclonal identity, tumour density and unexplained variance (Supplementary Methods). Of the 25,074 genes detected by slide-RNA-seq 412 genes were significantly associated with subclonal identity, 638 genes were associated with tumour density, and 1,098 genes were associated with a combination of both ($P < 0.05$, variance explained > 30%, Fig. 4e, Supplementary Table 5). Genes associated with subclonal identity included known cancer genes located in amplified regions, such as *PLAG1*, an oncogene on chr8q[39], and *MCM7*, a MYC target gene on chr7q that is involved in DNA replication initiation[40] (Fig. 4f). Notable tumour density-associated genes included *LGALS3*[41] (also known as galectin-3), which contributes to immunosuppression in the tumour microenvironment, and *PROM1* (also known as CD133), which is important for intestinal homeostasis, regeneration and tumour initiation[42] (Fig. 4g).

As well as characterizing individual genes, we also performed gene set enrichment analysis to determine which molecular pathways were associated with subclonal identity or tumour density (Fig. 4h, Supplementary Methods). This analysis showed that subclonal identity primarily altered the expression of genes involved in cell growth and proliferation, with MYC- and E2F-target genes representing the top hallmark gene sets for subclone 1 (Extended Data Fig. 15). By contrast, genes associated with high tumour density were most enriched for cell adhesion molecule- and cadherin-binding properties (Fig. 4i, Extended Data Fig. 15), including the extracellular matrix component gene *COL3A1*, actin modulator genes *FLNB* and *CALD1*, and the mechanotransduction regulator gene *ITGB2* (also known as CD18). Notably, extracellular matrix stiffness and remodelling are thought to promote

cell proliferation and tumour progression[43], which may contribute to high tumour cell density. Overall, these analyses demonstrate the utility of this multi-modal approach for decoupling and quantifying contributions of genetic and environmental factors to gene expression.

Discussion

This study demonstrates that slide-DNA-seq can detect clonal heterogeneity, characterize the copy number alterations of each clone and analyse their spatial distribution within a tissue. These capabilities, in combination with processing of serial sections for histopathology and slide-RNA-seq enable high-resolution multi-omic characterization of intratumoural heterogeneity[44]. Additionally, integration with single-cell whole-genome sequencing may enable spatial characterization of complex subclonal events, such as loss of heterozygosity or extrachromosomal DNA amplifications[45]. Going forward, we anticipate that slide-DNA-seq will be particularly useful for large-scale efforts to create atlases of tumour evolution[10], adding spatial information to studies of clonal heterogeneity. It may also drive new frontiers in clinical diagnoses as a complement to standard pathology assays such as H&E staining, karyotyping and DNA fluorescence in situ hybridization.

Spatially resolved DNA sequencing may also enable advances in many fields beyond cancer genomics, including spatially resolved metagenomics[46], evaluation of gene therapy delivery[47], synthetic DNA data storage[48] and lineage tracing in healthy tissues[49]. The core of this technology—that is, fragmenting and barcoding DNA in situ to preserve spatial information for next-generation sequencing—is compatible with other sequencing-based assays. For example, direct tagmentation of the DNA without HCl treatment, or converting methylated cytosines to dihydrouracil before amplification, would enable spatially resolved measurements of chromatin accessibility and DNA methylation, respectively[22,50]. In summary, slide-DNA-seq enables new opportunities to chart the spatial organization of cell states in human development, homeostasis and disease.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-04217-4.

1. McGranahan, N. & Swanton, C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell* **168**, 613–628 (2017).
2. Turajlic, S., Sottoriva, A., Graham, T. & Swanton, C. Resolving genetic heterogeneity in cancer. *Nat. Rev. Genet.* **20**, 404–416 (2019).
3. Ramón y Cajal, S. et al. Clinical implications of intratumor heterogeneity: challenges and opportunities. *J. Mol. Med.* **98**, 161–177 (2020).
4. Pogrebniak, K. L. & Curtis, C. Harnessing tumor evolution to circumvent resistance. *Trends Genet.* **34**, 639–651 (2018).
5. Duan, Q., Zhang, H., Zheng, J. & Zhang, L. Turning cold into hot: firing up the tumor microenvironment. *Trends Cancer Res.* **6**, 605–618 (2020).
6. Jin, M.-Z. & Jin, W.-L. The updated landscape of tumor microenvironment and drug repurposing. *Signal Transduct. Target. Ther.* **5**, 166 (2020).
7. Tse, J. M. et al. Mechanical compression drives cancer cells toward invasive phenotype. *Proc. Natl Acad. Sci. USA* **109**, 911–916 (2012).
8. Zhao, Y. et al. Selection of metastasis competent subclones in the tumour interior. *Nat. Ecol. Evol.* **5**, 1033–1045 (2021).
9. Stylianopoulos, T., Munn, L. L. & Jain, R. K. Reengineering the physical microenvironment of tumors to improve drug delivery and efficacy: from mathematical modeling to bench to bedside. *Trends Cancer Res.* **4**, 292–319 (2018).
10. Gerstung, M. et al. The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128 (2020).
11. Landau, D. A. et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714–726 (2013).
12. Wang, Y. et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512**, 155–160 (2014).
13. Xu, X. et al. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* **148**, 886–895 (2012).
14. Minussi, D. C. et al. Breast tumours maintain a reservoir of subclonal diversity during expansion. *Nature* **592**, 302–308 (2021).
15. Casasent, A. K. et al. Multiclonal invasion in breast tumors identified by topographic single cell sequencing. *Cell* **172**, 205–217.e12 (2018).
16. McPherson, A. et al. Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nat. Genet.* **48**, 758–767 (2016).
17. Jamal-Hanjani, M. et al. Tracking the evolution of non-small-cell lung cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).
18. Payne, A. C. et al. In situ genome sequencing resolves DNA sequence and structure in intact biological samples. *Science* **371**, eaay3446 (2020).
19. Rodriques, S. G. et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
20. Stickels, R. R. et al. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2021).
21. Adey, A. et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.* **11**, R119 (2010).
22. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
23. McFadden, D. G. et al. Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. *Cell* **156**, 1298–1311 (2014).
24. Johnson, L. et al. Somatic activation of the K-*ras* oncogene causes early onset lung cancer in mice. *Nature* **410**, 1111–1116 (2001).
25. LaFave, L. M. et al. Epigenomic state transitions characterize tumor progression in mouse lung adenocarcinoma. *Cancer Cell* **38**, 212–228.e13 (2020).
26. Morishita, A. et al. HMGA2 is a driver of tumor metastasis. *Cancer Res.* **73**, 4289–4299 (2013).
27. Thuault, S. et al. Transforming growth factor-β employs HMGA2 to elicit epithelial–mesenchymal transition. *J. Cell Biol.* **174**, 175–183 (2006).
28. Kumar, M. S. et al. HMGA2 functions as a competing endogenous RNA to promote lung cancer progression. *Nature* **505**, 212–217 (2014).
29. Cable, D. M., Murray, E., Zou, L. S., Goeva, A. & Macosko, E. Z. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-021-00830-w (2021).
30. Dienstmann, R. et al. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat. Rev. Cancer* **17**, 268 (2017).
31. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
32. Kang, H. et al. Many private mutations originate from the first few divisions of a human colorectal adenoma. *J. Pathol.* **237**, 355–362 (2015).
33. Humphries, A. et al. Lineage tracing reveals multipotent stem cells maintain human adenomas and the pattern of clonal expansion in tumor evolution. *Proc. Natl Acad. Sci. USA* **110**, E2490–E2499 (2013).
34. Brayer, K. J., Frerich, C. A., Kang, H. & Ness, S. A. Recurrent fusions in MYB and MYBL1 define a common, transcription factor-driven oncogenic pathway in salivary gland adenoid cystic carcinoma. *Cancer Discov.* **6**, 176–187 (2016).
35. Wang, B., Matsuoka, S., Carpenter, P. B. & Elledge, S. J. 53BP1, a mediator of the DNA damage checkpoint. *Science* **298**, 1435–1438 (2002).
36. Long, D. T., Räschle, M., Joukov, V. & Walter, J. C. Mechanism of RAD51-dependent DNA interstrand cross-link repair. *Science* **333**, 84–87 (2011).
37. Lachaud, C. et al. Ubiquitinated Fancd2 recruits Fan1 to stalled replication forks to prevent genome instability. *Science* **351**, 846–849 (2016).
38. Sottoriva, A. et al. A Big Bang model of human colorectal tumor growth. *Nat. Genet.* **47**, 209–216 (2015).
39. Landrette, S. F. et al. *Plag1* and *Plagl2* are oncogenes that induce acute myeloid leukemia in cooperation with *Cbfb-MYH11*. *Blood* **105**, 2900–2907 (2005).
40. Ren, B. et al. *MCM7* amplification and overexpression are associated with prostate cancer progression. *Oncogene* **25**, 1090–1098 (2006).
41. Farhad, M., Rolig, A. S. & Redmond, W. L. The role of Galectin-3 in modulating tumor growth and immunosuppression within the tumor microenvironment. *Oncoimmunology* **7**, e1434467 (2018).
42. Karim, B. O., Rhee, K.-J., Liu, G., Yun, K. & Brant, S. R. Prom1 function in development, intestinal inflammation, and intestinal tumorigenesis. *Front. Oncol.* **4**, 323 (2014).
43. Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K. J. & Werb, Z. Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nat. Commun.* **11**, 5120 (2020).
44. Wu, C.-Y. et al. Integrative single-cell analysis of allele-specific copy number alterations and chromatin accessibility in cancer. *Nat. Biotechnol.* **39**, 1259–1269 (2021).
45. Wu, S. et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature* **575**, 699–703 (2019).
46. Integrative HMP (iHMP) Research Network Consortium. The Integrative Human Microbiome Project. *Nature* **569**, 641–648 (2019).
47. van Haasteren, J., Li, J., Scheideler, O. J., Murthy, N. & Schaffer, D. V. The delivery challenge: fulfilling the promise of therapeutic genome editing. *Nat. Biotechnol.* **38**, 845–855 (2020).
48. Goldman, N. et al. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* **494**, 77–80 (2013).
49. Ludwig, L. S. et al. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* **176**, 1325–1339.e22 (2019).
50. Liu, Y. et al. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat. Biotechnol.* **37**, 424–429 (2019).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Raw sequencing data are available from the Sequence Read Archive under accession PRJNA768453. Spatial barcode locations and counts matrices are available from the Broad Institute Single Cell Portal (https://singlecell.broadinstitute.org/single_cell/study/SCP1278). GC-content tracks for hg19 and mm10 were downloaded from the UC Santa Cruz Genome Browser. k36 mappability tracks for both genomes were downloaded from https://bismap.hoffmanlab.org/. Replication timing data were downloaded from Gene Expression Omnibus accession GSM923451 for hg19 and GSE137764 for mm10. Tn5 insertion bias tracks for both genomes were generated using the bias command from pyatac (https://nucleoatac.readthedocs.io/en/latest/pyatac/). Gene sets were downloaded from the Molecular Signatures Database Collections (http://www.gsea-msigdb.org/gsea/msigdb/collections.jsp).

Code availability

Code for the in situ bead indexing is available from https://github.com/broadchenf/Slideseq. Code for all analyses is available from https://github.com/buenrostrolab/slide_dna_seq_analysis and archived at https://doi.org/10.5281/zenodo.5553305.

Acknowledgements J.D.B. and F.C. acknowledge funding from the Allen Institute Distinguished Investigator award and funding from the NIH R21HG009749. F.C. also acknowledges funding from NIH DP50D024583, and R33CA246455. F.C. and E.Z.M acknowledge funding from NIH R01HG010647. F.C. also acknowledges the Searle Scholars Award and the Burroughs Wellcome CASI award. J.D.B. acknowledges support from the NIH New Innovator Award (DP2HL151353). Z.D.C. acknowledges funding from NHGRI training grant T32HG002295 and the Harvard Quantitative Biology Initiative. We thank S. Nagaraja for assistance in making figures. Components of our figures were created with BioRender.com. We thank J. Strecker for the gift of Tn5 enzyme, and the Buenrostro and Chen laboratories for helpful discussions. We thank the cancer patients and their families for their invaluable donations to science, making this work possible.

Author contributions T.Z. and J.W.M. developed the protocol and performed experiments. Z.D.C. developed the computational processing pipeline. T.Z., Z.D.C., J.D.B., and F.C. performed analyses. L.M.L., I.D.P. and K.M. assisted with the mouse experiments under the supervision of T.J. E.M.M. performed the in situ bead indexing and slide-RNA-seq experiments under the supervision of E.Z.M. and F.C. J.L. wrote the in situ bead-indexing pipeline. N.M.N. assisted with the 10x experiment. C.A.L. and A.S.E. assisted with the computational processing pipeline. T.Z., Z.D.C., J.D.B. and F.C. wrote the manuscript with input from all authors. E.Z.M., T.J., J.D.B. and F.C. supervised this work.

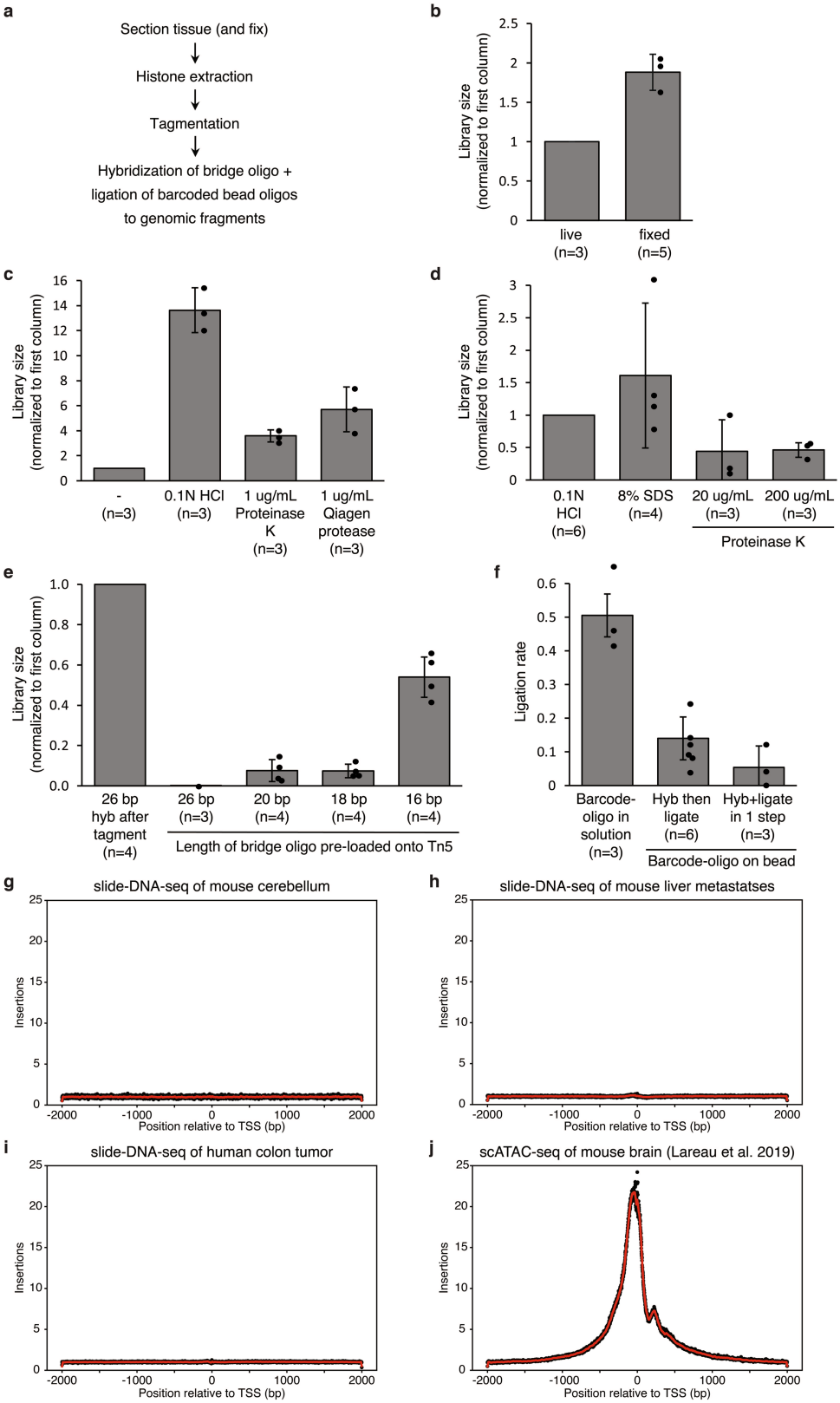
Competing interests E.Z.M. and F.C. are listed as inventors on a patent application related to slide-seq. T.J. is a member of the board of directors of Amgen and Thermo Fisher Scientific. He is also a co-founder of Dragonfly Therapeutics and T2 Biosystems. T.J. serves on the scientific advisory board of Dragonfly Therapeutics, SQZ Biotech and Skyhawk Therapeutics, and is the president of Break Through Cancer. J.D.B. holds patents related to ATAC-seq and is on the scientific advisory board for Camp4, Seqwell and Celsee. F.C. is a paid consultant for Celsius Therapeutics and Atlas Bio. E.Z.M is a paid consultant for Atlas Bio. T.Z., E.Z.M., J.D.B. and F.C. have filed a patent application based on this work.

Additional information

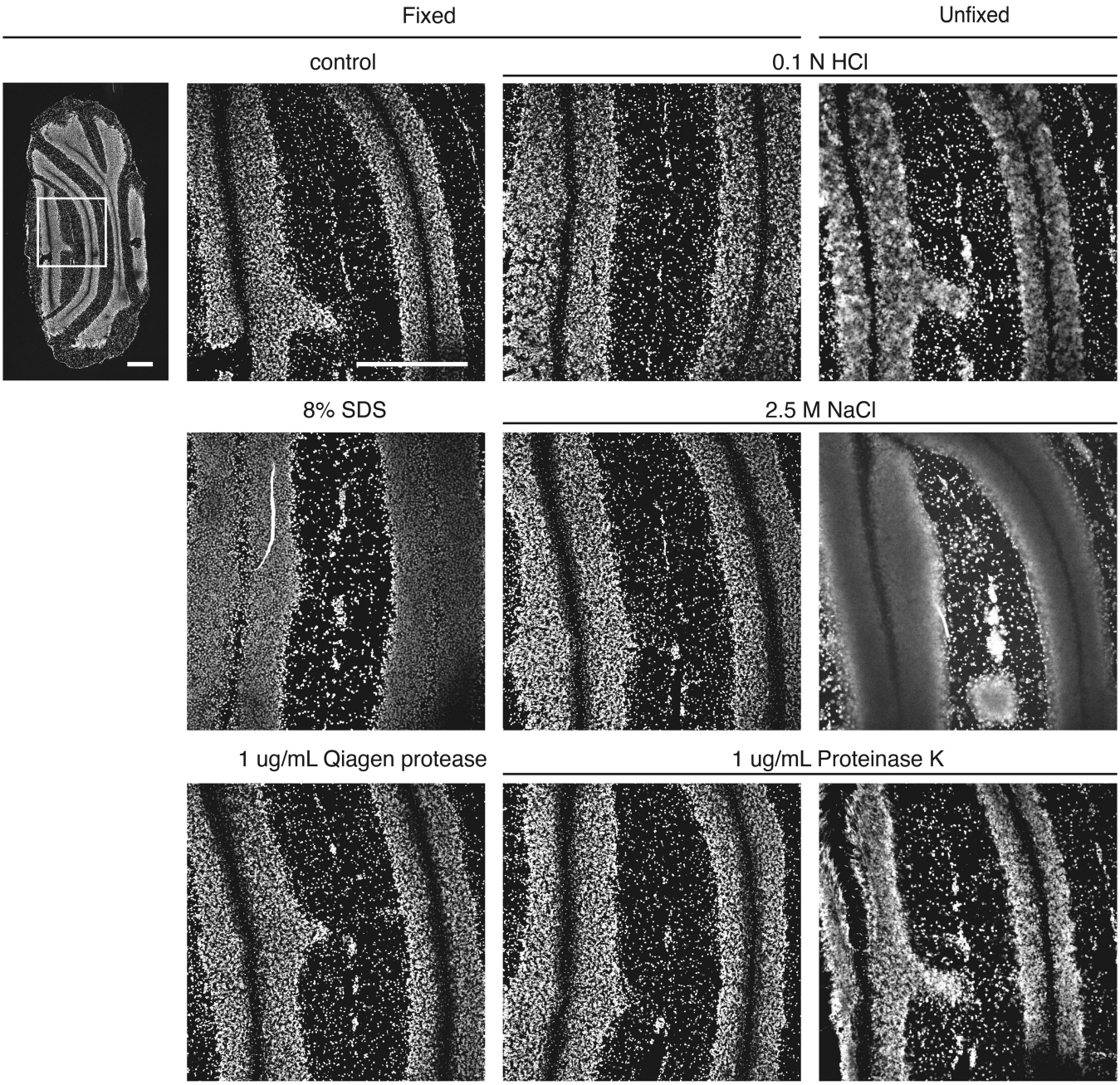
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-04217-4.

Correspondence and requests for materials should be addressed to Jason D. Buenrostro or Fei Chen.

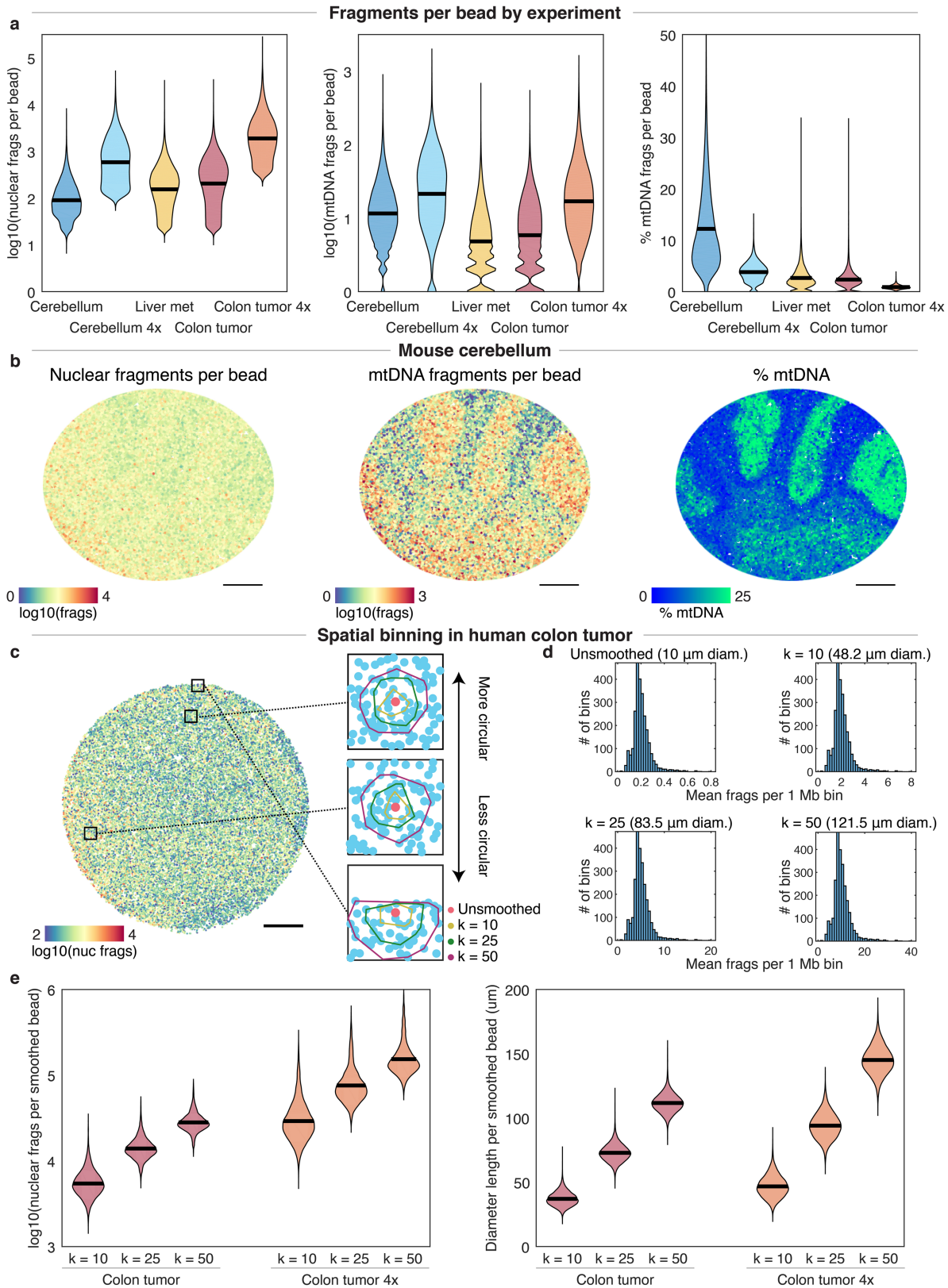
Peer review information *Nature* thanks Andrew Adey and Naveed Ishaque for their contribution to the peer review of this work.



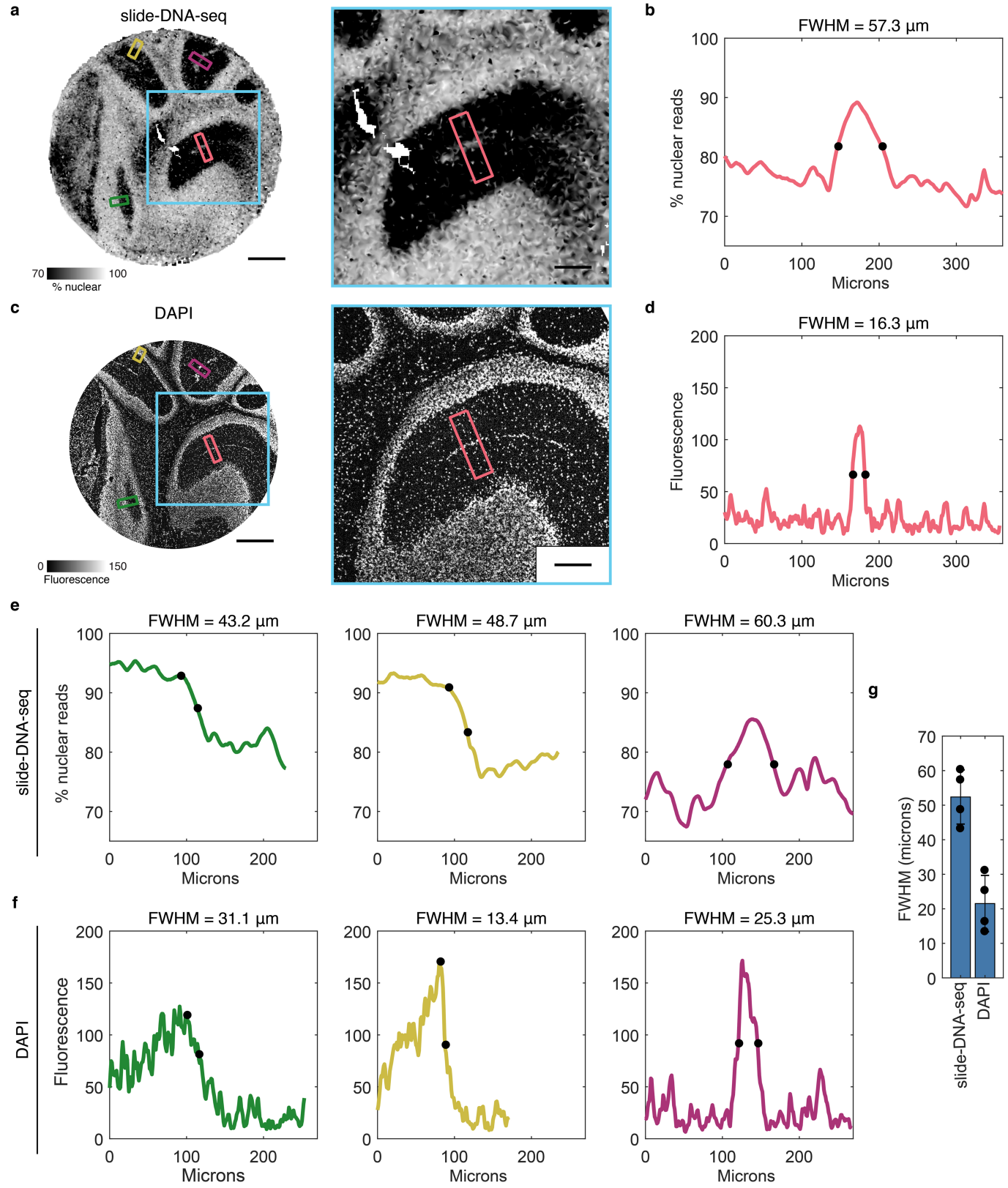
Extended Data Fig. 1 | Optimization of slide-DNA-seq protocol. a, Library preparation steps. **b**–**e**, Library size comparisons for live vs fixed tissue (**b**); histone extraction protocols (**c**, **d**); and varying lengths of a bridge oligo used to connect the barcoded bead oligo to genomic fragments **e**, either hybridized after tagmentation (left bar) or pre-loaded onto the Tn5 transposase prior to tagmentation (rest). All values are normalized to control condition (first column). **f**, Rate of ligation of genomic fragments to barcoded oligo either ordered in solution from IDT (left) or cleaved off from beads (center, right). **g**–**j**, Frequency of Tn5 insertions in the genome relative to the nearest transcription start site (TSS) for slide-DNA-seq of mouse cerebellum (**g**), mouse liver metastases (**h**), human colon tumor (**i**), and for single-cell ATAC-seq of a mouse brain (**j**). Error bars, mean ± s.d; n, number of replicate comparisons (generated from 4 biological samples); dots represent values of each replicate.



Extended Data Fig. 2 | Comparison of fixation conditions during histone extraction. Cerebellar sections are exposed to treatment as stated (with or without prior fixation) and stained with DAPI. Scale bars, 500 µm.

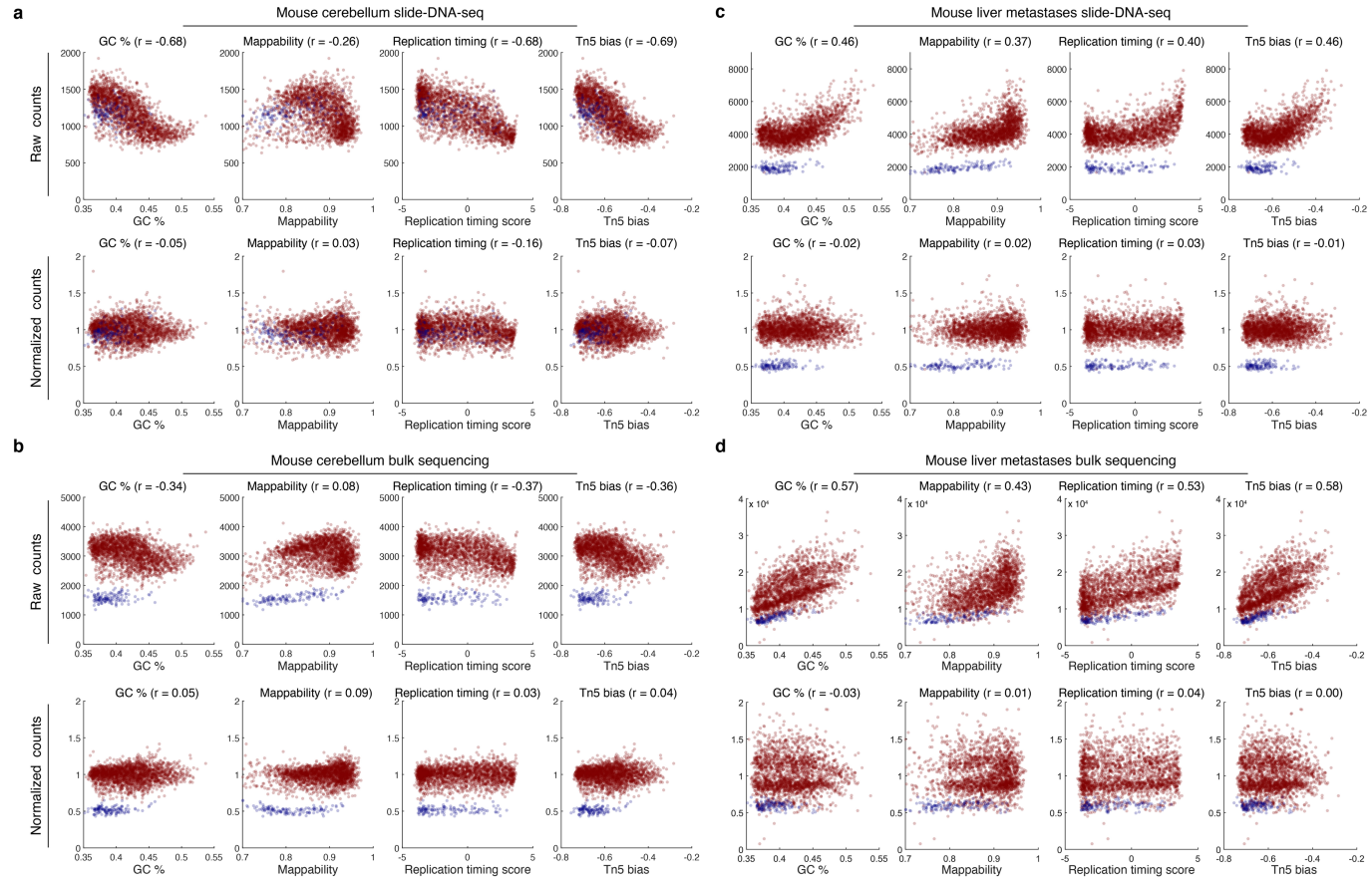


Extended Data Fig. 3 | Quantification of DNA fragments per slide-DNA-seq array. a, Nuclear (left) and mitochondrial (center) DNA fragments per bead obtained for tissues used in this study. Right, mitochondrial fraction of fragments. 4x, protocol variant with 4x tagmentation. Black lines on violin plots indicate the mean. **b**, slide-DNA-seq of the mouse cerebellum experiment in Fig. 1. Beads are colored by the number of nuclear fragments (left), mitochondrial DNA fragments (center), and fraction of mitochondrial DNA fragments (right). **c**, Visualization of representative convex hulls for different spatial bin values of k for k-nearest neighbor smoothing. Beads are colored by raw counts, insets show convex hulls for k = 1, 10, 25, and 50, centered on salmon colored beads. Hulls are generally circular except at the edge of the array. **d**, Distribution of mean fragments per 1 Mb genomic bin for different spatial bin values of k. The median diameter of the smoothed features is indicated in parentheses. **e**, Comparison of nuclear fragments (left) and effective diameter (right) per bead for different spatial bin values of k. Scale bars, 500 µm.

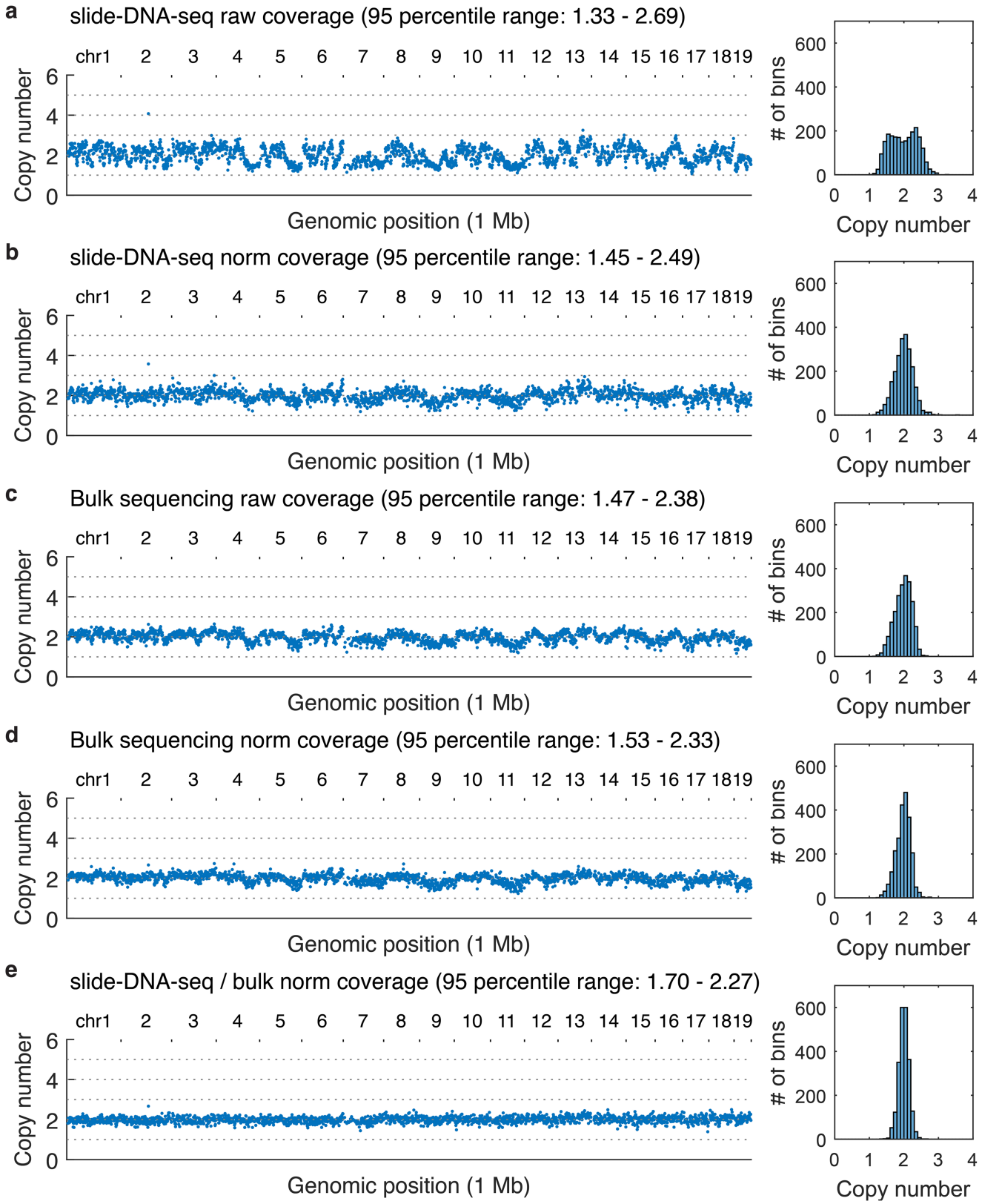


Extended Data Fig. 4 | See next page for caption.

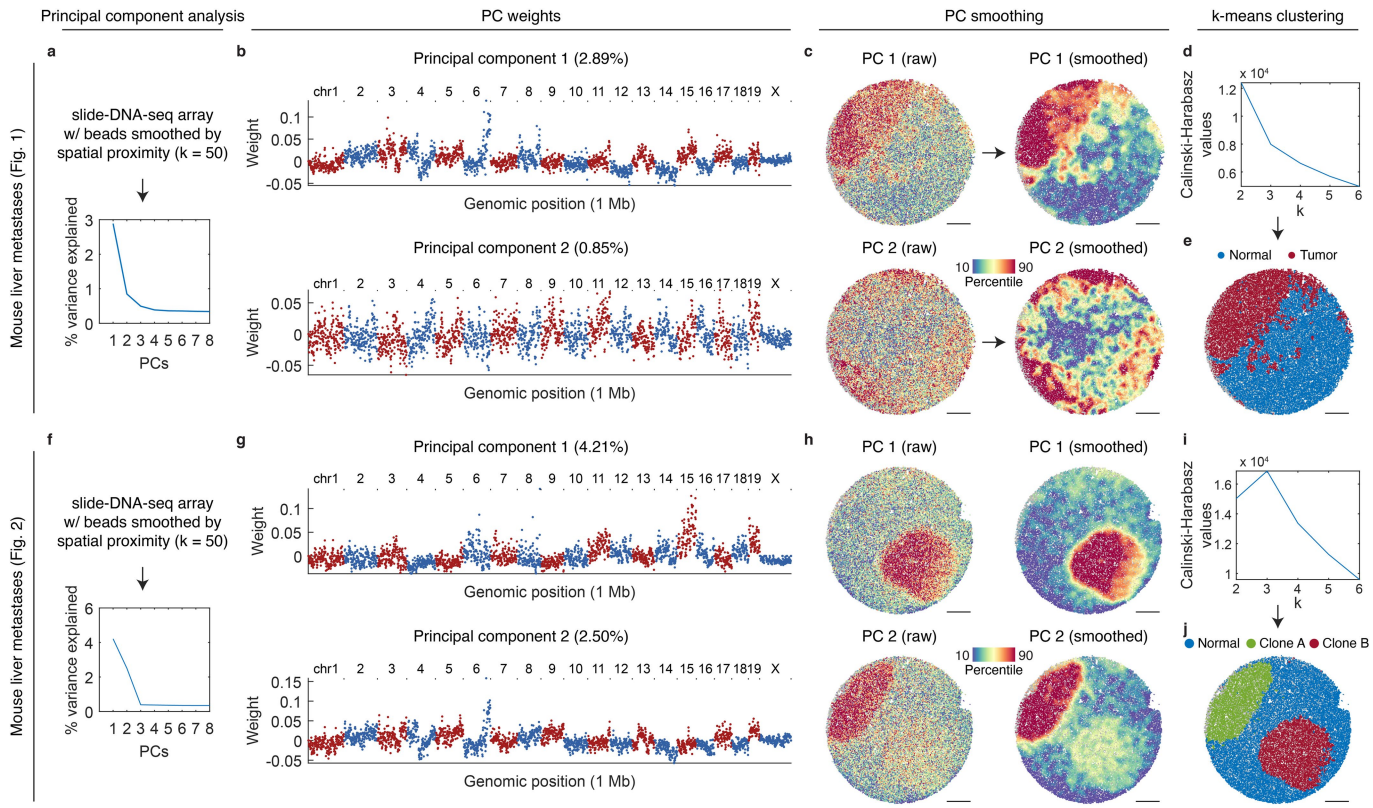
Extended Data Fig. 4 | Estimation of slide-DNA-seq lateral diffusion. **a**, Interpolated image showing the nuclear fraction of fragments of a 10 µm mouse cerebellar section processed for slide-DNA-seq. Cyan box indicates magnified area (right). Smaller boxes indicate regions taken for linescans in **b** and **e**. **b**, Pseudo-intensity (representing nuclear fraction of fragments) of linescan as indicated by red box in **a**. Black dots, halfmax. Full Width at Half Maximum (FWHM) = 57.3 µm. **c**, 10 µm serial section of the same mouse cerebellum stained with DAPI. Blue box indicates magnified area (right). Smaller boxes indicate regions taken for linescans in **d** and **f**. **d**, Linescan of DAPI intensity as indicated in **c**. Black dots, halfmax. FWHM = 16.4 µm. **e**, Same as b, but for 3 different regions as indicated by the smaller non-red boxes in **a**. For the left (green) and middle (yellow) panel, FWHM is calculated as twice the distance between the peak and the halfmax (marked by black dots). **f**, Same as **d**, but for 3 different regions as indicated by the smaller non-red boxes in **b**. For the left (green) and middle (yellow) panel, FWHM is calculated as twice the distance between the peak and the halfmax (marked by black dots). **g**, Bar graph of average FWHM (n=4 regions). Error bars, mean ± s.d. Upper bound for the diffusion measurement is half of largest FWHM (not taking into account the DAPI measurement). Scale bars, 500 µm (**a**, **c**), 200 µm (**b**, **d**).



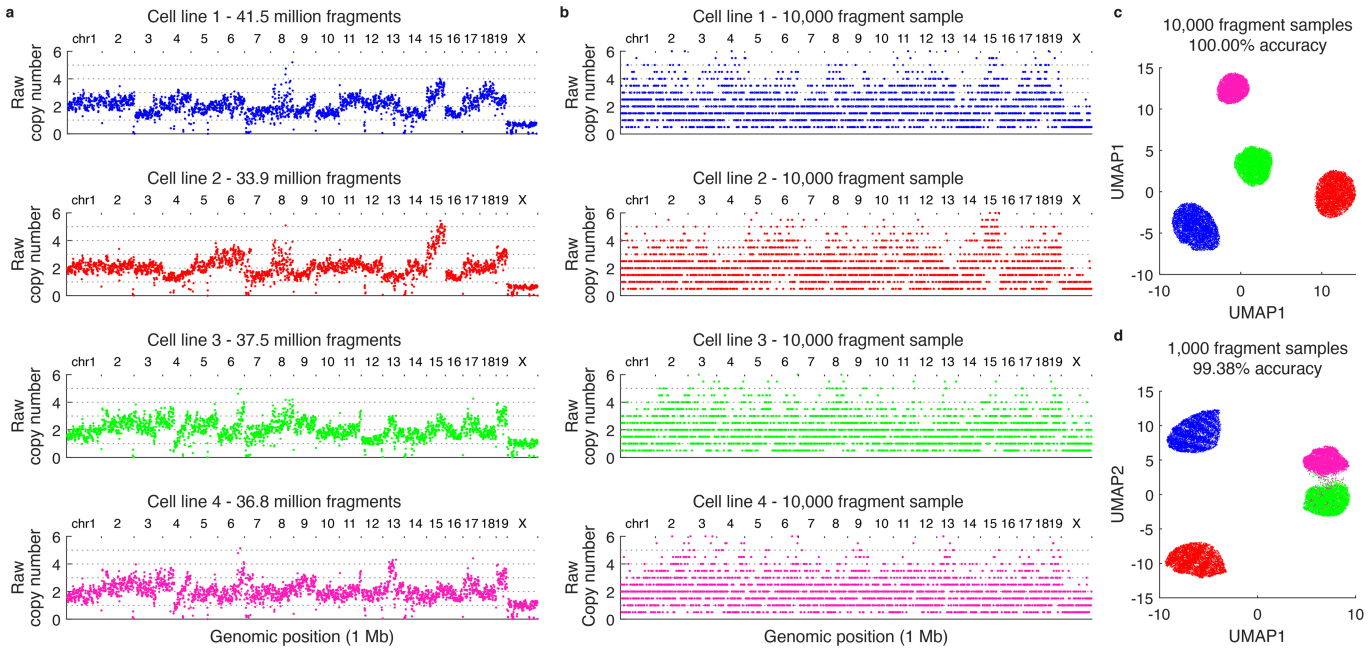
Extended Data Fig. 5 | Normalization of slide-DNA-seq sequencing biases. **a**, Top, raw sequencing reads per 1 Mb bin for mouse cerebellum slide-DNA-seq are plotted for GC-content, mappability, replication timing score, and Tn5 bias. Pearson's r values are shown for each. Bottom, bias corrected coverage and correlation values after normalization. **b**, Same as **a** but for tagmentation-based bulk sequencing of mouse cerebellum (**Methods**). **c**, Same as **a** but for slide-DNA seq of mouse liver metastases. **d**, Same as **b** but for tagmentation-based bulk sequencing of mouse liver metastases. Blue points, bins from chrX (not included in the calculation of the fit).



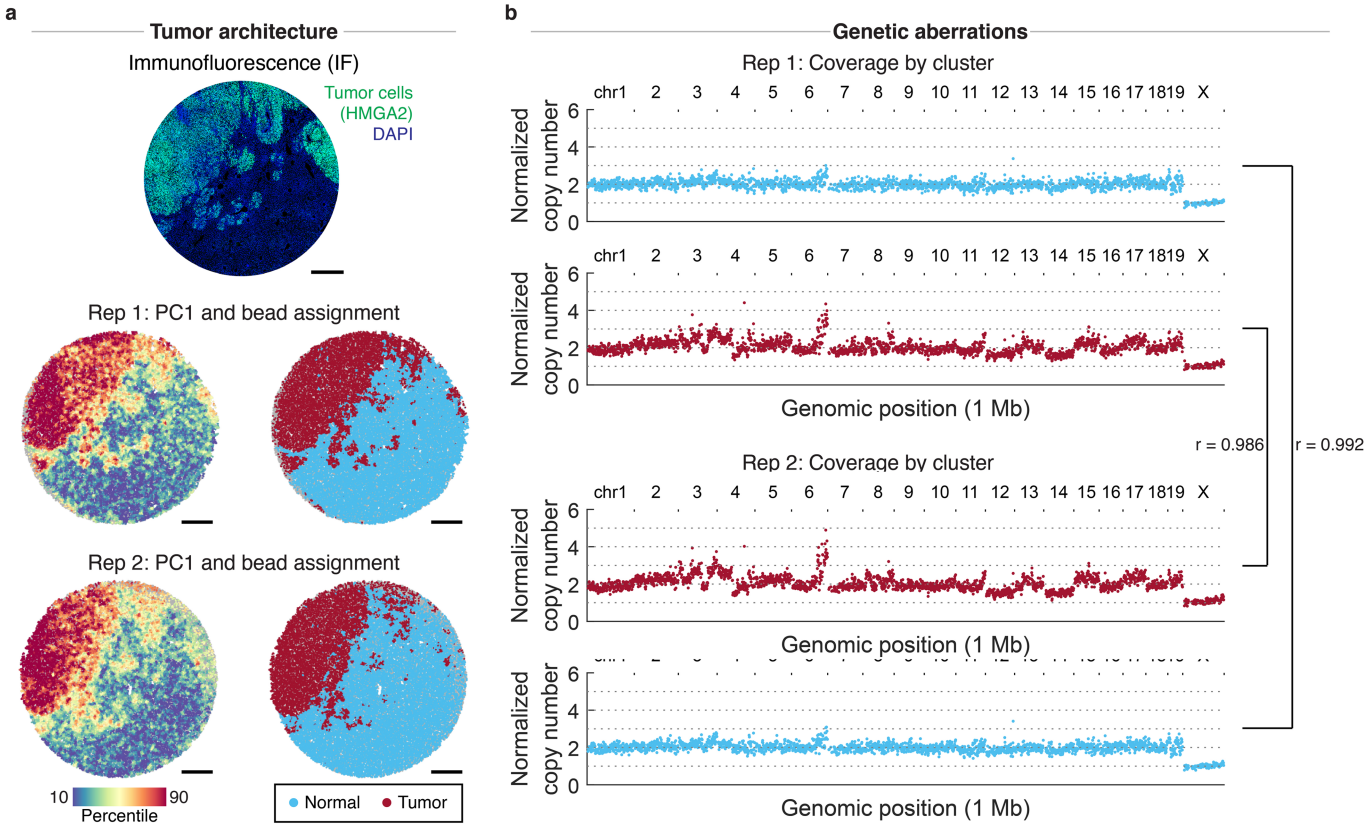
Extended Data Fig. 6 | Quantification of genomic coverage in a diploid sample. Left (all panels), copy number profiles at 1 Mb genomic resolution of the mouse cerebellum for the sequencing modality and processing indicated. For this diploid sample, each copy number distribution is normalized to a median of 2. Right (all panels), histogram of the number of bins per copy number. **a**, Raw coverage profile of slide-DNA-seq. **b**, Coverage profile of slide-DNA-seq normalized by GC-content and mappability. **c**, Coverage profile for bulk tagmentation-based sequencing. **d**, Coverage profile of bulk sequencing normalized by GC-content and mappability. **e**, Coverage profile of slide-DNA-seq normalized by the GC-content and mappability divided by bulk sequencing normalized by GC-content and mappability.



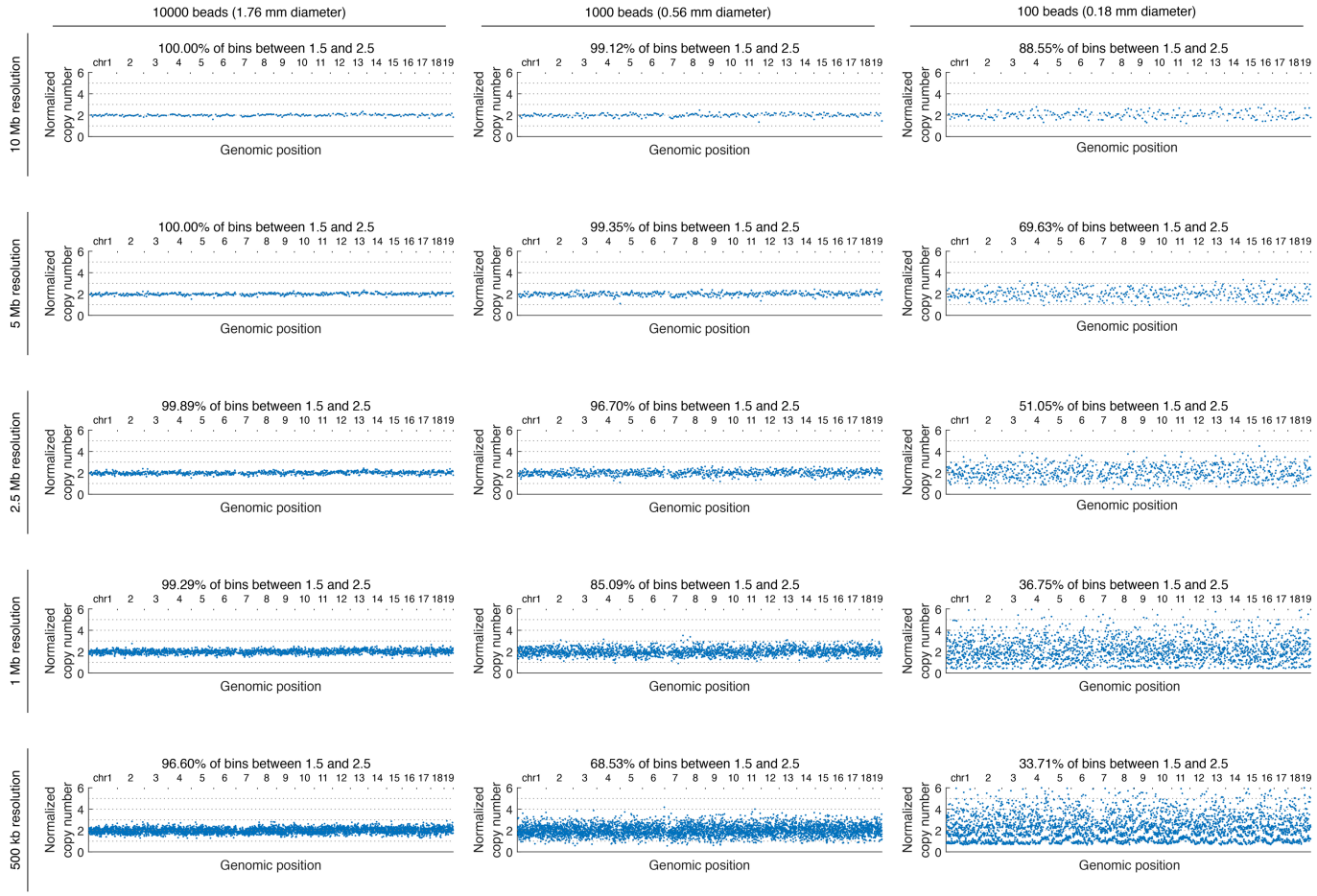
Extended Data Fig. 7 | slide-DNA-seq clonal analysis workflow. a, Principal components calculated from smoothed slide-DNA-seq beads, ordered by the percentage of variance explained for the mouse liver metastases array shown in Fig. 1. **b**, Weights per 1 Mb genomic bin for principal components 1 and 2. Red points indicate bins from chromosomes with an odd number, blue from chromosomes with an even number (and chrX). **c**, slide-DNA-seq array for the mouse liver metastases array shown in Fig. 1 with points colored by raw PC 1 scores (top left), smoothed PC 1 scores (top right), raw PC 2 scores (bottom left), smoothed PC 2 scores (bottom right). **d**, Calinski-Harabasz criterion values used to select the optimal value of k for k-means clustering. **e**, slide-DNA-seq array colored by cluster assignment using the value of k selected in **d**. **f**–**j**, Same as **a**–**e**, but for the mouse liver metastases array shown in Fig. 2.



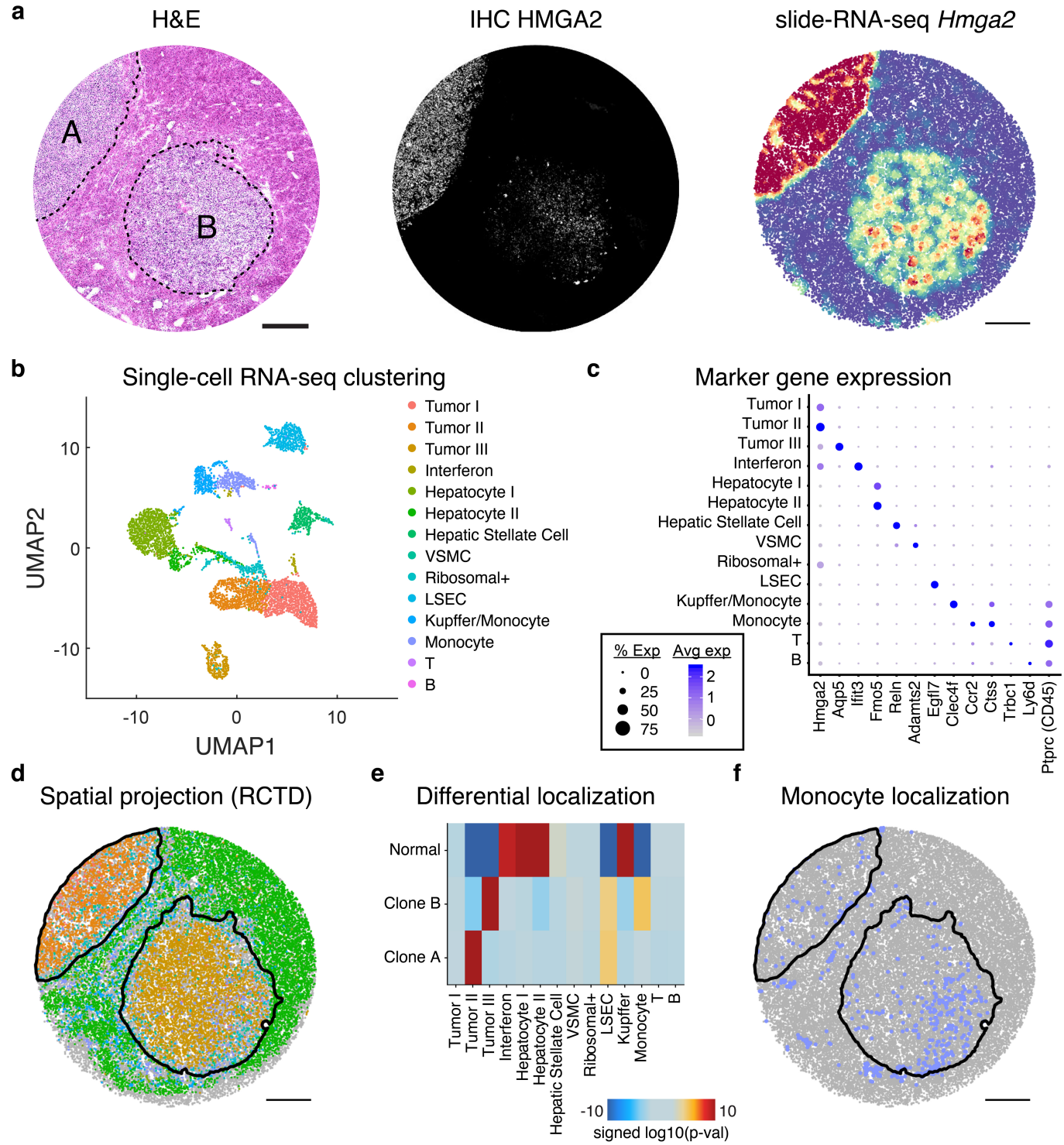
Extended Data Fig. 8 | Accuracy of clonal assignment via downsampling of bulk tumor cell lines. a, Raw copy number profiles for four tumor cell lines profiled using tagmentation-based bulk sequencing. **b**, Representative 10,000 fragments samples of the cell lines shown in **a**. **c**, Clonal assignment accuracy for 10,000 fragment samples (n=5,000 samples of each cell line) using the analysis workflow shown in Extended Data Fig. 7. **d**, Same as **c** but for 1,000 fragment samples.



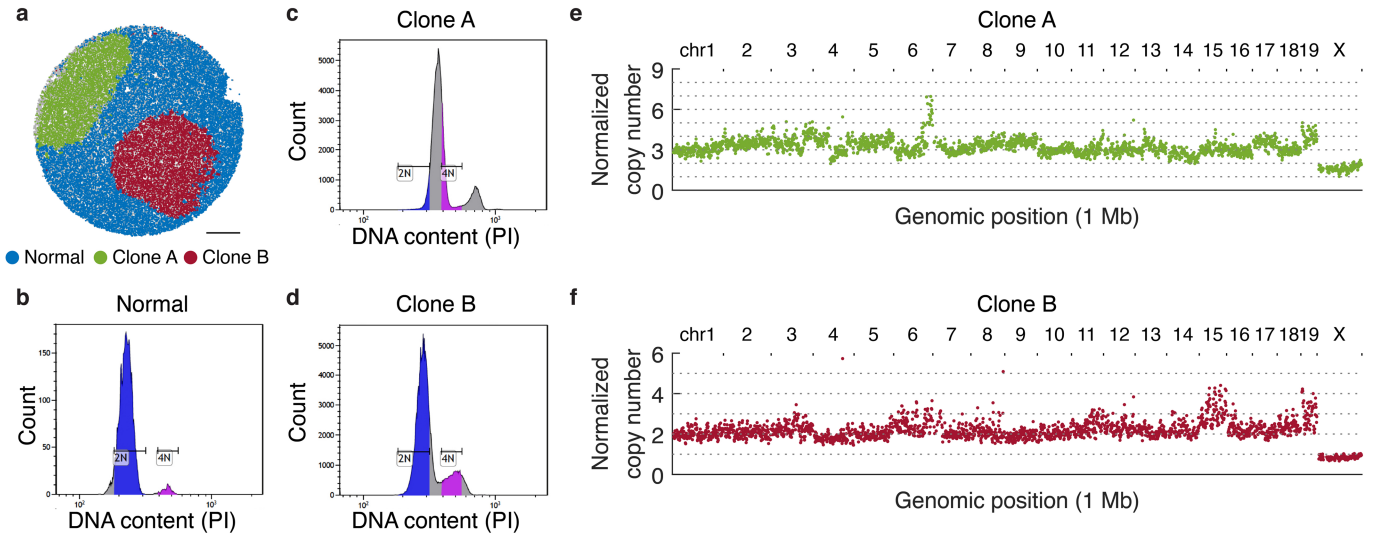
Extended Data Fig. 9 | Reproducibility of slide-DNA-seq across serial sections. a, Immunofluorescence (IF) against tumor marker HMGA2 (top) and two slide-DNA-seq replicates (center, bottom) were performed on two serial sections of a mouse liver metastasis. Beads colored by PC1 scores (left) and cluster assignment (right) show similar spatial architecture between replicates. Scale bars, 500 µm. **b**, Aggregate copy number profiles of normal and tumor beads show high correlation (Pearson's r = 0.986 and 0.992) between the two replicates.



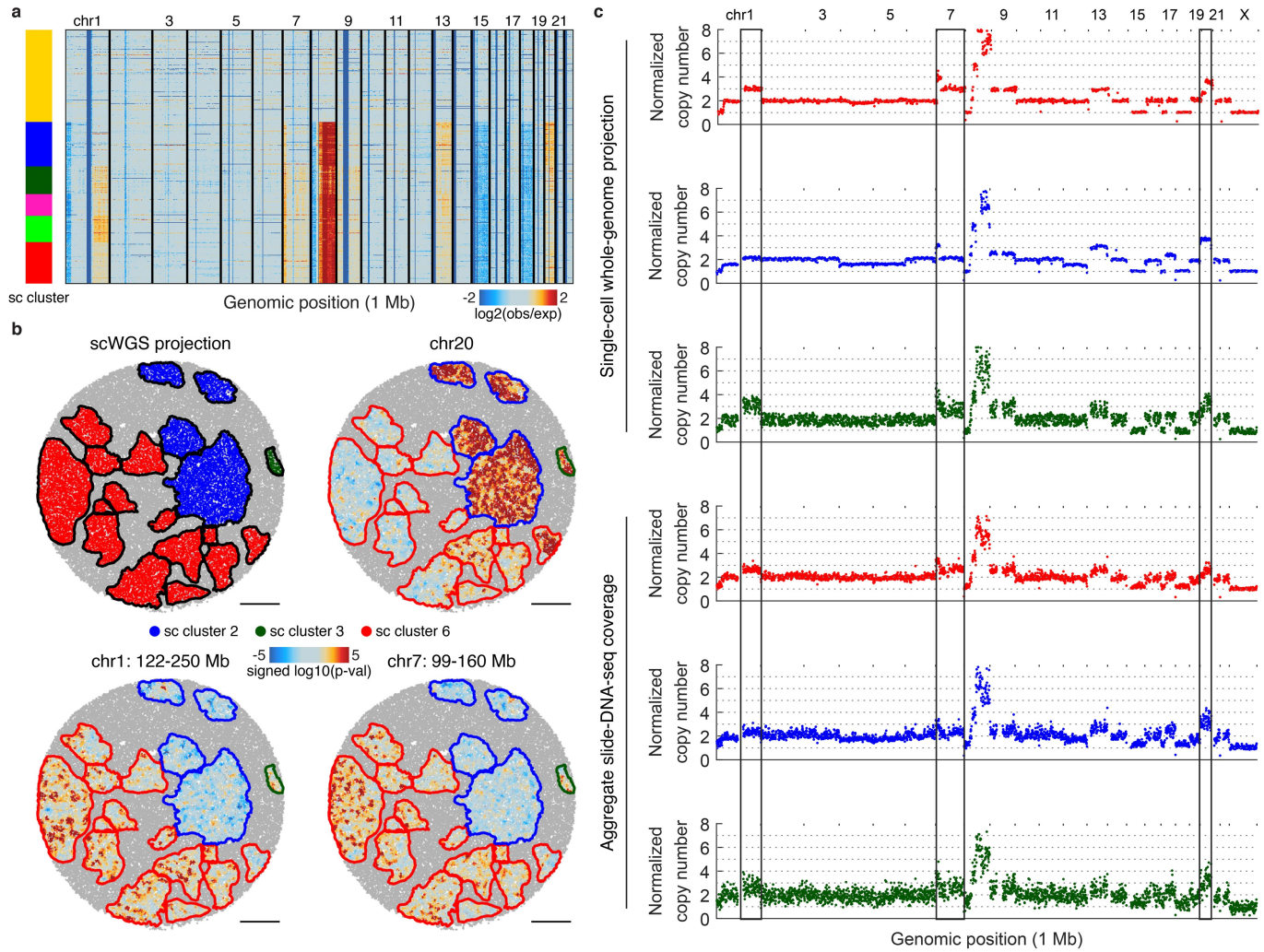
Extended Data Fig. 10 | Quantification of genomic coverage by bin size and number of beads. Each column represents normalized copy number profiles aggregated across the number of slide-DNA-seq beads indicated (10,000; 1,000; or 100), while each row indicates the genomic bin size (10 Mb, 5 Mb, 2.5 Mb, 1 Mb, and 500 kb) for the mouse cerebellum array.



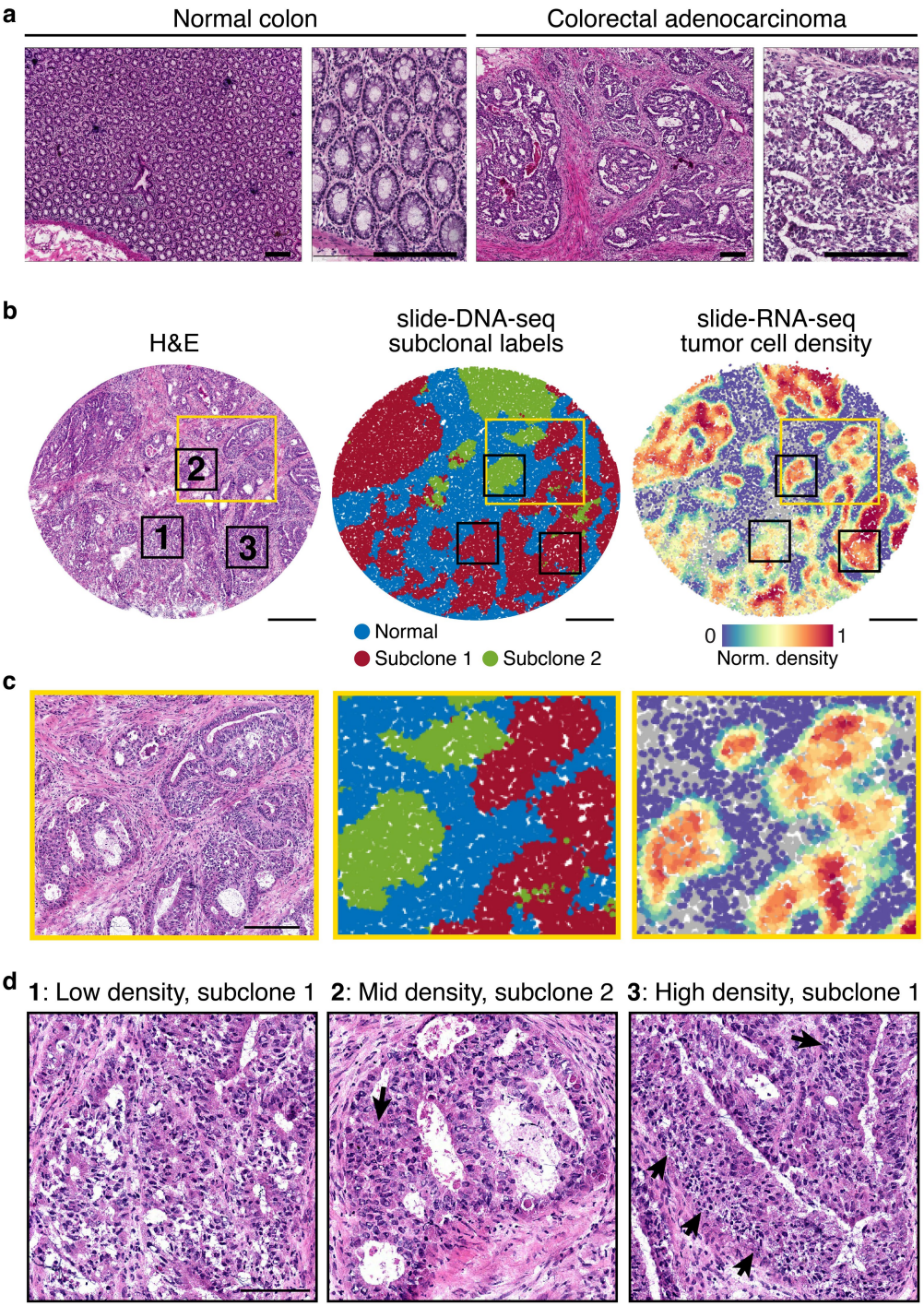
Extended Data Fig. 11 | Integrated slide-RNA-seq and single-nucleus RNA-seq analysis of clones. a, H&E stain (left), IHC against tumor marker HMGA2 (center), and *Hmga2* expression from slide-RNA-seq (right) of three serial sections of mouse liver metastases. **b**, UMAP of unsupervised clustering of single nucleus RNA-seq performed on nuclei from mouse liver metastasis sample. **c**, Dot plot showing the expression of marker genes used to annotate clusters in **b**. **d**, Spatial projection of cell types from **b** onto the slide-RNA-seq array, colored in the same fashion. Black lines indicate spatial tumor clusters. **e**, Differential localization of cell types between clone A, clone B and normal regions. Heatmap shows signed (positive, enrichment; negative, depletion) log10(p-value) from permutation testing (two-sided, not adjusted for multiple comparisons). **f**, Spatial plot of monocyte localization on the array, which is significantly enriched for clone B. Black lines indicate spatial tumor clusters. VSMC, vascular smooth muscle cell; LSEC, liver sinusoidal endothelial cells.



Extended Data Fig. 12 | Validation of ploidy and copy number of metastatic clones. a, Assignment of beads to normal tissue, clone A, and clone B based on k-means clustering. **b**–**d**, Histogram of DNA content of single cells measured by propidium iodide (PI) fluorescence intensity through flow cytometry. (**b**) bone marrow cells (normal control); (**c**) clone A; (**d**) clone B. Diploid G1 (2N) and G2 (4N) gates are determined on bone marrow histogram and applied to clones A and B, revealing that the clone A genome is triploid; and the clone B genome is diploid with some amplifications (e.g. of chr. 15 and 19, see Fig. 2d). **e**, Aggregate copy number profiles of beads assigned to clone A. **f**, Aggregate copy number profiles of beads assigned to clone B.



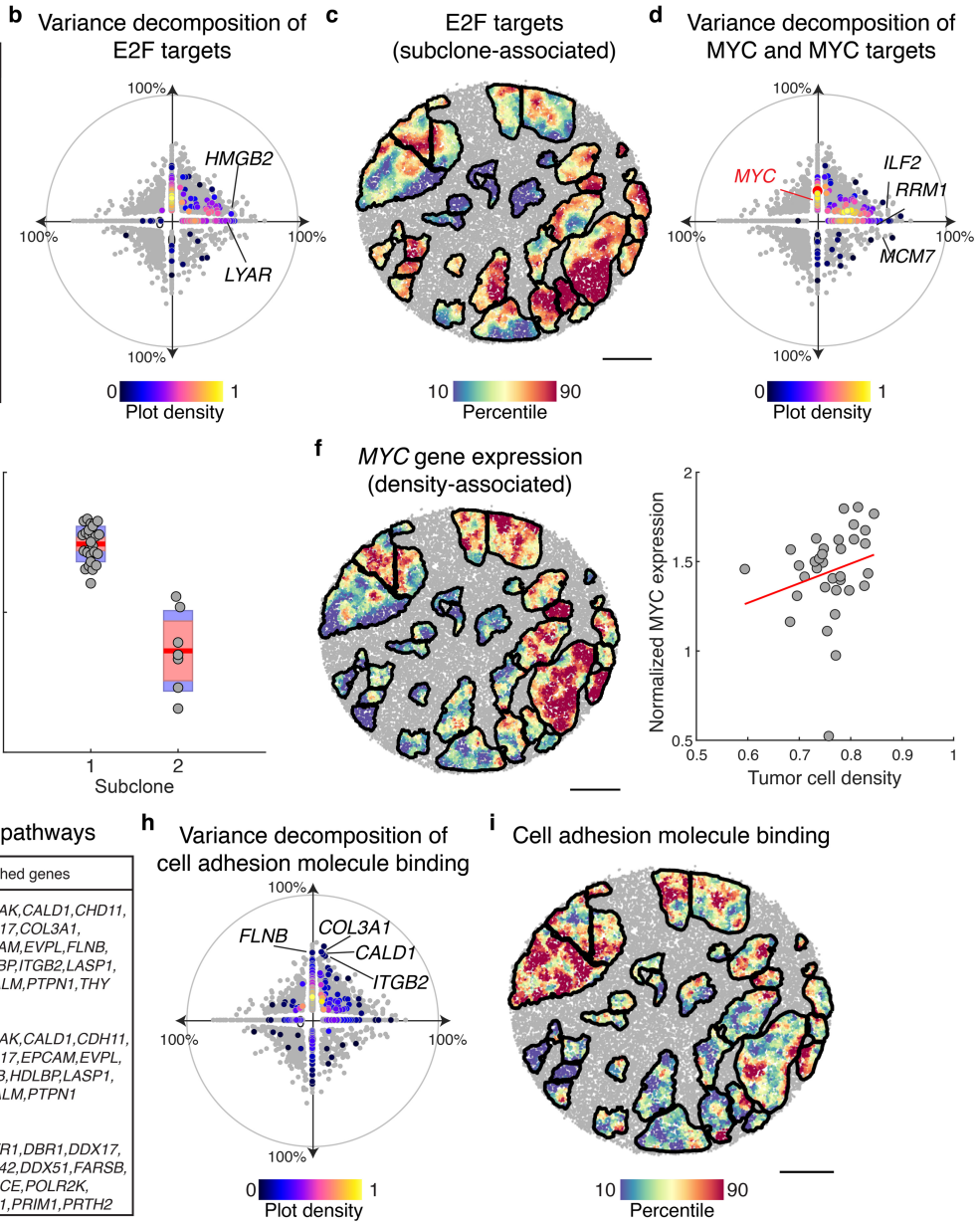
Extended Data Fig. 13 | Spatial projection of single-cell whole-genome sequencing (scWGS) clusters. a, Genomic copy number profiles for 2,274 single cells obtained using scWGS, with cluster annotations colored. **b**, Top left: projection of scWGS clusters onto slide-DNA-seq. All other: three genomic regions of differential CNA profiles between the three projected clusters, shown are spatial heatmaps of signed p-value differences from the average profile (two-sided permutation test, not adjusted for multiple comparisons). **c**, Normalized copy number profiles for the three scWGS clusters, and the corresponding spatial clusters. Vertical lines denote variable regions from **b**. Single-cell cluster 2 (blue) shows complex CNA patterns that obscure cluster ploidy, nevertheless, copy number values are normalized to 2 for easy comparison to other clusters.



Extended Data Fig. 14 | Tumor morphology of primary human colon cancer sample. a, H&E stain of normal colon (left) and colon tumor (right) tissue from the same patient. Scale bars, 200 µm. **b**, Serial sections processed for H&E stain (left), slide-DNA-seq (center), and slide-RNA-seq (right). Scale bars, 500 µm. Yellow and black boxes indicate magnified areas in **c**, **d**, respectively. **c**, Magnified views of H&E stain, slide-DNA-seq and slide-RNA-seq reconstructions show concordant spatial tissue architecture across three modalities; scale bar, 200 µm. **d**, Magnified view of H&E stain of three regions that are assigned low, medium, and high tumor density by slide-RNA-seq transcriptomic analysis (**b**, right). Arrows indicate regions of high tumor density identified through H&E stain. Scale bar, 100 µm.

a Subclone-associated pathways

Gene set	FDR	Enriched genes
MYC targets (V1)	3.0E-13	*CBX3,CCNA2,CCT3, FBL,HDGF,HNRNPU, ILF2,MCM4,MCM7, NAP1L1,PSMB3, PTGES3,RAN,RRM1, SLC25A3,USP1*
E2F targets	1.1E-07	*ANP32E,HMGB2, LMNB1,LYAR,MCM4, MCM7,NAA38,NAP1L1, RAN,TOP2A,USP1*
Oxidative phosphorylation	1.1E-07	*COX6B1,COX7C, HSPA9,LDHB,NDUFB1, NDUFB2,NDUFB7, SLC25A3,TIMM13, UQCRH,UQCRQ*



Gene set	FDR	Enriched genes
Cell adhesion molecule binding	1.7E-06	*AHNAK,CALD1,CHD11, CDH17,COL3A1, EPCAM,EVPL,FLNB, HDLBP,ITGB2,LASP1, PICALM,PTPN1,THY*
Cadherin binding	4.5E-06	*AHNAK,CALD1,CDH11, CDH17,EPCAM,EVPL, FLNB,HDLBP,LASP1, PICALM,PTPN1*
Catalytic activity acting on RNA	1.7E-05	*CMTR1,DBR1,DDX17, DDX42,DDX51,FARSB, MEPCE,POLR2K, POP1,PRIM1,PRTH2*

Extended Data Fig. 15 | Biological pathways explained by subclone or tumor density. a, Subclone-associated pathways identified through gene set enrichment analysis. **b**, Hallmark E2F target genes (n=200) plotted according to percent variance explained by clonal identity (x-axis) and tumor density (y-axis). Included genes colored by normalized density on the scatter plot, all other genes are shown in grey. **c**, Expression of highly subclone-associated E2F target genes (n=11, listed in **a**), plotted for spatial tumor regions of the slide-RNA-seq array from Fig. 4. **d**, MYC target genes (n=200) plotted according to percent variance explained by clonal identity (x-axis) and tumor density (y-axis). Included genes are colored by normalized density on the scatter plot, MYC is colored red, all other genes are shown in grey. **e**, Expression of highly subclone-associated MYC target genes (n=16, listed in **a**), plotted for spatial tumor regions (left). Box plot showing normalized MYC target gene expression by subclone assignment; each point represents a spatial tumor cluster (right). Red line, mean, red box, 95% confidence interval for mean, blue box, standard deviation. **f**, MYC expression plotted for spatial tumor regions (left). Scatter plot showing normalized MYC expression by tumor cell density; each point represents a spatial tumor cluster (right). **g**, Subclone-associated pathways identified through gene set enrichment analysis. **h**, Cell adhesion molecule binding genes (n=514) plotted according to percent variance explained by clonal identity (x-axis) and tumor density (y-axis). Included genes are colored by normalized density on the scatter plot, all other genes are shown in grey (reproduced from Fig. 4i). **i**, Expression of highly density-associated cell adhesion molecule binding genes (n=14, listed in **g**), plotted for spatial tumor regions. Scale bars, 500 µm.

nature research

Corresponding author(s): Jason Buenrostro, Fei Chen

Last updated by author(s): Oct 6, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	For the sequence alignment pipeline, fastp version 0.21.0, bowtie2 version 2.3.4.1, samtools version 1.7, and UMI-tools 0.5.5 were used. Code for the in situ bead indexing is available from https://github.com/broadchenf/Slideseq.
Data analysis	Code for the in situ bead indexing is available from https://github.com/broadchenf/Slideseq. Code for all analyses are available from https://github.com/buenrostrolab/slide_dna_seq_analysis and archived at https://doi.org/10.5281/zenodo.5553305.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw sequencing data is available from the Sequence Read Archive (SRA) at accession PRJNA768453. Spatial barcode locations and counts matrices are available from the Broad Institute Single Cell Portal (https://singlecell.broadinstitute.org/single_cell/study/SCP1278).

GC-content tracks for hg19 and mm10 were downloaded from the UC Santa Cruz Genome Browser. k36 mappability tracks for both genomes were downloaded from https://bismap.hoffmanlab.org/. Replication timing data was downloaded from GEO accession GSM923451 for hg19 and GSE137764 for mm10. Tn5 insertion bias tracks for both genomes were generated using the bias command from pyatac (https://nucleoatac.readthedocs.io/en/latest/pyatac/). Gene sets were downloaded from the Molecular Signatures Database Collections (MSigDB, http://www.gsea-msigdb.org/gsea/msigdb/collections.jsp).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. Samples sizes were chosen primarily based on experiment length, sample availablity, and sequencing costs. These sample sizes are sufficient because each sample serves as a proof-of-concept for the new technology.
Data exclusions	No data was excluded.
Replication	All replication attempts were successful. For slide-DNA-seq experiments, we performed 3 biological replicates of the mouse cerebellum, 4 biological replicates of the mouse liver metastases, and 4 biological replicates of the human colon cancer. The slide-RNA-seq and single-cell whole genome sequencing experiments were performed independently because the development of these technologies is not the primary focus of the paper.
Randomization	Randomization was not applicable because the focus of this paper is the development of a new genomic technology and did not involve allocating samples/organisms/participants into experimental groups.
Blinding	Blinding was not applicable because because the focus of this paper is the development of a new genomic technology and did not involve group allocation, and by extension, blinding.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	TOMM20 (EPR15581-540), Abcam ab186735; HMGA2 (D1A7), Cell signaling 8179S; Ki67 (polyclonal), Abcam ab15580; CD45 (polyclonal), Abcam ab10558; Alexa Fluor 488 goat anti-rabbit IgG (H+L), Life Technologies A11034; Opal Polymer anti-rabbit HRP, Opal 520 and Opal 570, Akoya Biosciences SKU NEL810001KT
Validation	TOMM20: validated according to manufacturer's website for Flow Cytometry 1/90, ICC/IF 1/250, WB 1/1000-1/10000 detecting a band of approximately 16 kDa (predicted molecular weight: 16 kDa), IHC-P 1/50 - 1/100 Perform heat mediated antigen retrieval with Tris/EDTA buffer pH 9.0 before commencing with IHC staining protocol, IHC-Fr 1/50 Heat mediated antigen retrieval using sodium citrate buffer (10mM citrate pH 6.0 + 0.05% Tween-20); HMGA2: validation data on manufacturer's website for WB, IHC and IF; Ki67: according to manufacturer's website knockout-validated, IHC-P Use a concentration of 0.1 - 5 µg/ml. Perform heat mediated antigen retrieval before commencing with IHC staining protocol, ICC Use a concentration of 0.5 - 1 µg/ml. If fixing cells in 4% PFA (20 min, room temp), it is recommended to permeabilized cells with 0.1% Triton-X for 5 min. Positive Control: HeLa and HAP1 cells; CD45: validated according to manufacturer's website for Flow Cyt Use 1µg for 10^6 cells, WB 1/500. Detects a band of approximately 190 kDa (predicted molecular weight: 147 kDa), IHC-P Use a concentration of 0.5 - 5 µg/ml. Perform heat mediated antigen retrieval before commencing with IHC staining protocol.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Cell lines derived were from KP primary tumors 1183T3 and 860T3 (LaFave et al., 2020).
Authentication	Cell lines were authenticated by genotyping for specified alleles.
Mycoplasma contamination	All cell lines tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	B6129SF1/J (Jackson lab, stock 101043) mice, male, 8-10 weeks. Facilities are AAALAC accredited animal facilities and under the supervision of the Division of Comparative Medicine at MIT. Animals are housed on hardwood bedding with a light dark cycle of 12/12. Room temp is 70 +/-2 with a humidity range of 30-70%
Wild animals	This study did not involve wild animals.
Field-collected samples	No field-collected samples were used.
Ethics oversight	All animal experiments described in this study were approved by the Massachusetts Institute of Technology Institutional Animal Care and Use Committee (IACUC) (institutional animal welfare assurance no. A-3125-01).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	This is a commercially-available sample from a 67 year old male.
Recruitment	The sample was obtained from a commercial source.
Ethics oversight	This was determined to be non-human subject research by Broad IBC.

Note that full information on the approval of the study protocol must also be provided in the manuscript.