

# A combinatorial genetic strategy for exploring complex genotype–phenotype associations in cancer

Received: 17 April 2023

Accepted: 25 January 2024

Published online: 29 February 2024

 Check for updates

Shan Li<sup>1,12</sup>, Alicia Wong<sup>1,12</sup>, Huiyun Sun<sup>1,2,12</sup>, Vipul Bhatia<sup>1</sup>, Gerardo Javier<sup>1</sup>, Sujata Jana<sup>1</sup>, Qian Wu<sup>3,4</sup>, Robert B. Montgomery<sup>5</sup>, Jonathan L. Wright<sup>6</sup>, Hung-Ming Lam<sup>6</sup>, Andrew C. Hsieh<sup>1,3,5</sup>, Bishoy M. Faltas<sup>7,8,9,10</sup>, Michael C. Haffner<sup>1,3,11</sup> & John K. Lee<sup>1,3,5,11</sup> ✉

Available genetically defined cancer models are limited in genotypic and phenotypic complexity and underrepresent the heterogeneity of human cancer. Here, we describe a combinatorial genetic strategy applied to an organoid transformation assay to rapidly generate diverse, clinically relevant bladder and prostate cancer models. Importantly, the clonal architecture of the resultant tumors can be resolved using single-cell or spatially resolved next-generation sequencing to uncover polygenic drivers of cancer phenotypes.

Most cancers are not driven by a single oncogenic driver but are instead the sum of multiple genetic perturbations that occur during tumor evolution<sup>1</sup>. However, the functional impact of most genomic abnormalities found in cancers remains largely unknown. Wrangling the catalog of recurrent genetic alterations in cancer and deriving meaningful insights into the functional and contextual contributions of these events is a major challenge in the field of cancer genomics. In vitro assays recapitulate only specific aspects of cancer behaviors such as cell proliferation, anchorage-independent colony formation or invasive migration. In vivo strategies such as the transplantation of cancer cell lines or chemical carcinogenesis are not genetically defined. Genetically engineered mouse models are a gold standard to define genetic drivers in cancer, but they are costly, slow and do not allow the facile manipulation of more than a few genes. Dissociated-cell tissue recombination and transplantation assays have also been applied to study the malignant transformation of primary epithelial cells but have been reliant on the introduction of discrete sets of candidate genes and limited by inefficient transgenesis. Collectively, existing cancer models

generated through these methods grossly underrepresent the diversity of human cancer. Furthermore, the use of these technologies to systematically investigate the permutations of genetic events associated with a single cancer would be incredibly challenging, if not impossible. To address these limitations of scale, throughput and economy, we developed a methodology incorporating barcoded lentiviral libraries encoding cancer-associated genetic events introduced efficiently and at random into primary epithelial cells, which are engrafted in mice for tumorigenic selection, at a high multiplicity of infection (MOI). This system enables the generation of genotypically and phenotypically diverse tumors and the massively parallel single-cell lentiviral barcode sequencing of tumors to identify cooperative oncogenic drivers of malignant transformation and specific cancer phenotypes.

Organotypic or organoid cultures permit the expansion of primary epithelial cells while maintaining their complex organization and tissue function<sup>2</sup>. A major barrier to higher-order genetic studies in this context has been inefficient transgenesis using available lentiviral transduction protocols. We proposed that enforced cell–virus contact

<sup>1</sup>Human Biology Division, Fred Hutchinson Cancer Center, Seattle, WA, USA. <sup>2</sup>Molecular Engineering and Sciences Institute, University of Washington, Seattle, WA, USA. <sup>3</sup>Clinical Research Division, Fred Hutchinson Cancer Center, Seattle, WA, USA. <sup>4</sup>Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, USA. <sup>5</sup>Division of Medical Oncology, University of Washington School of Medicine, Seattle, WA, USA. <sup>6</sup>Department of Urology, University of Washington School of Medicine, Seattle, WA, USA. <sup>7</sup>Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA. <sup>8</sup>Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY, USA. <sup>9</sup>Department of Medicine, Weill Cornell Medicine, New York, NY, USA. <sup>10</sup>Department of Cell and Developmental Biology, Weill Cornell Medicine, New York, NY, USA. <sup>11</sup>Department of Pathology and Laboratory Medicine, University of Washington School of Medicine, Seattle, WA, USA. <sup>12</sup>These authors contributed equally: Shan Li, Alicia Wong, Huiyun Sun. ✉e-mail: [jkleee5@fredhutch.org](mailto:jkleee5@fredhutch.org)

in a constrained volume of gel matrix could increase lentiviral transduction efficiency. Primary mouse bladder urothelial (mBU) and prostate epithelial (mPE) cells were isolated by fluorescence-activated cell sorting (FACS) on the basis of a lineage-negative ( $\text{Lin}^-$ ) ( $\text{CD45}^- \text{CD31}^- \text{Ter119}^-$ ),  $\text{EpCAM}^+ \text{CD49f}^{\text{high}}$  immunophenotype (Extended Data Fig. 1a), as these populations self-renew at high frequencies<sup>3</sup> and readily establish organoids in culture (Extended Data Fig. 1b). Cells were mixed into cold Matrigel containing concentrated lentivirus expressing GFP before the seeding and polymerization of organoid droplets<sup>4</sup>. Near complete transduction of mBU and mPE cells was achieved, delivering up to 10–20 copies per cell (Fig. 1a,b). We next developed a barcoding system to characterize the distribution of unique proviral copies per cell. Lentiviral constructs were barcoded with matching ten-nucleotide sequences distal to the 5' long terminal repeat (LTR) and proximal to the 3' LTR and produced as a pool. A custom single-cell amplicon panel was designed on the Mission Bio Tapestry platform to enable the sensitive enumeration of multiple uniquely barcoded lentiviruses per cell (Extended Data Fig. 1c). This approach was validated using a defined population of 3T3 cells engineered with lentiviruses to harbor up to four unique lentiviral barcodes per cell (Extended Data Fig. 1d). mPE were transduced with a diverse barcoded lentiviral pool at varying MOIs, and single-cell amplicon sequencing showed relatively normal distributions of proviral copies per cell (Fig. 1c and Extended Data Fig. 1e).

To determine the utility of this approach in understanding the initiation and progression of bladder and prostate cancer, we selected commonly mutated genes from cancer genome sequencing studies<sup>5–7</sup> (Extended Data Fig. 2a) and cloned these as open reading frames (ORFs) or short hairpin RNAs (shRNAs) into barcoded lentiviral constructs to mimic gain-of-function or loss-of-function events (Extended Data Fig. 2b and Supplementary Table 1). At least three shRNAs from The RNAi Consortium (TRC) targeting each gene were tested for knockdown in 3T3 cells by quantitative PCR. The shRNA demonstrating the most potent knockdown of target gene expression was incorporated into the lentiviral libraries (Extended Data Fig. 2c). A bladder urothelial lentiviral pool (BU-LVP) of 33 genes and a prostate epithelial lentiviral pool (PE-LVP) of 24 genes were produced in arrayed format to avoid lentiviral barcode recombination and concentrated by ultracentrifugation (Extended Data Fig. 3a). Infectivity (representation) of each lentivirus was evaluated by transducing either mBU or mPE cells with the respective lentiviral pool and performing bulk amplicon sequencing of lentiviral barcodes (Extended Data Fig. 3b). Initial lentivirus pools demonstrated over tenfold overrepresentation of shRNA vectors relative to ORF vectors (Extended Data Fig. 3c), presumably owing to more efficient viral packaging because of the reduced length between LTRs of the transfer plasmid<sup>8</sup>. These data were applied to adjust the cell surface area of producer cells for subsequent arrayed lentiviral library production, leading to near normalization of the representation of shRNA and ORF vectors (Extended Data Fig. 3d).

**Fig. 1 | Efficient lentiviral transduction of primary epithelial cells at high multiplicity of infection and transformation of urothelial cells to tumors with mixed cancer histologies.**

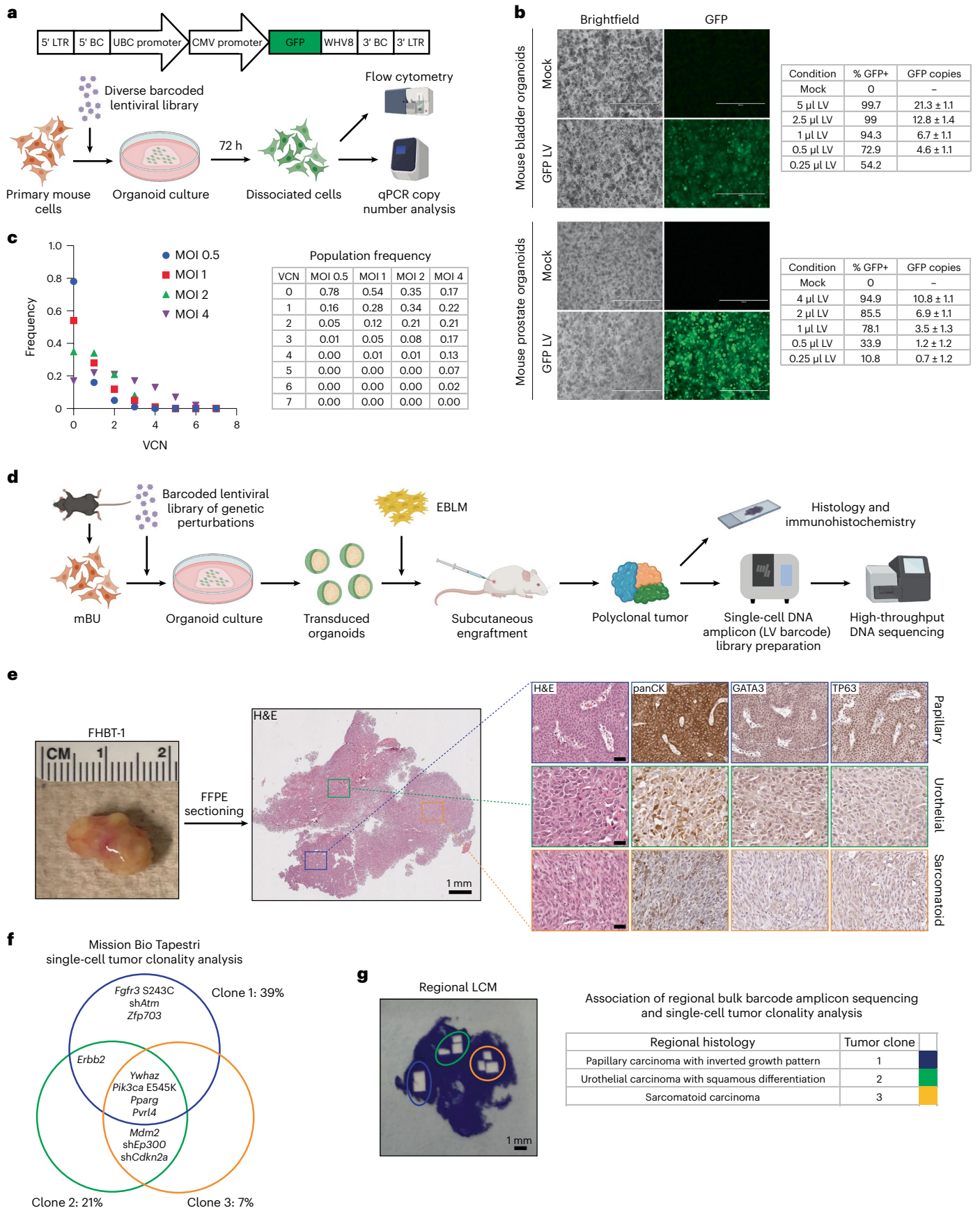
**a**, Top, schematic of a lentiviral (LV) construct with matching barcodes (BCs) at the 5' and 3' ends. Bottom, overview of experiments with LV infection of primary mouse cells in organoid culture and quantification of transduction. Created with BioRender.com. CMV, cytomegalovirus; UBC, ubiquitin C; WNVH8, Woodchuck hepatitis virus 8 post-transcriptional regulatory element. **b**, Left, brightfield and GFP images of mouse bladder or prostate organoids 72 h after mock or GFP LV transduction. Scale bar, 400  $\mu\text{m}$ . Right, tables summarizing quantification of LV transduction by flow cytometry and LV copies of GFP ( $\pm$  s.d.) by qPCR. **c**, Left, plot of the distribution of LV copies per mPE cell at different MOIs 72 h after transduction. Right, table summarizing viral copy number (VCN) population frequencies at varying MOIs. The experiment was independently repeated three times with similar results.

We adopted an approach in which primary mBU and mPE cells infected with BU-LVP or PE-LVP at high MOI in organoids were recombined with inductive mouse embryonic day 16 (E16) bladder mesenchyme (EBLM)<sup>9</sup> or urogenital sinus mesenchyme (UGSM)<sup>10</sup> and subsequently grafted subcutaneously in NOD scid gamma (NSG) mice to enable biological selection for tumorigenic clones (Fig. 1d). No tumors were appreciable from control grafts of untransduced mBU or mPE cells recombined with EBLM or UGSM. The efficiency of tumor formation (tumors formed per graft inoculated) was 80% (16 of 20) for mBU cells infected with BU-LVP and 38% (18 of 47) for mPE cells infected with PE-LVP (Supplementary Table 2). Tumor latency was measured as time from inoculation to achieving a maximal tumor diameter of 1 cm and ranged from 2.3 to 7.4 months (mean 4.2 months) for bladder tumors and 3.2 to 16 months (mean 8.9 months) for prostate tumors (Supplementary Table 2).

A representative tumor derived from primary mBU cells transduced with BU-LVP exhibited three morphologically distinct regions consistent with papillary urothelial carcinoma with an inverted growth pattern, urothelial carcinoma with squamous differentiation and sarcomatoid urothelial carcinoma, all three of which were also supported by GATA3, TP63 and pan-cytokeratin (panCK) immunostaining (Fig. 1e). Single-cell DNA amplicon sequencing was performed to enumerate the lentiviral barcodes for the determination of clonal architecture and deconvolution of lentivirus-delivered genetic events putatively involved in tumorigenesis. Three major clones harboring distinguishable sets of lentiviral barcodes were identified (Fig. 1f), but spatial resolution was lost owing to single-cell dissociation. To associate histology with clonality, we performed laser capture microdissection (LCM) of the three regions on stained tissue sections and performed bulk DNA amplicon sequencing (Fig. 1g). The papillary urothelial carcinoma was uniquely associated with *Fgfr3* S243C, *shAtm* and *Zfp703* mutations, in addition to the common *Ywhaz*, *Pik3ca* E545K, *Pparg* and *Pvrl4* mutations observed in all three dominant clones. Cancer genomics studies have shown that activating mutations in *FGFR3* are highly enriched in papillary urothelial carcinomas<sup>3,11</sup>. We further validated these findings in the mouse urothelial transformation assay in independent experiments using a defined lentiviral pool of *Fgfr3* S243C, *Ywhaz*, *Pik3ca* E545K, *Pparg* and *Pvrl4* (Extended Data Fig. 4a), which produced tumors with papillary urothelial carcinoma with an inverted growth pattern by histopathology and based on the endophytic proliferative pattern (Extended Data Fig. 4b). The co-occurrence of these genetic alterations was also evident in the human muscle-invasive bladder cancers from The Cancer Genome Atlas bladder cancer (TCGA-BLCA) cohort<sup>5</sup> (Extended Data Fig. 5).

Several tumors called the Fred Hutch Bladder Tumor (FHBT) series have been generated using this methodology, including those with pure urothelial carcinoma and others with mixtures of histologic subtypes (Fig. 2a and Extended Data Fig. 6). The urothelial origin of these tumors was supported by GFP staining (Fig. 2b–d), which was positive even

**d**, Scheme of the mBU organoid transformation assay to uncover functional genotype–phenotype associations in bladder cancer. Created with BioRender.com. **e**, Left, gross image of a tumor arising from mBU transformed with a BU-LVP. Middle, low-magnification image of the hematoxylin and eosin (H&E)-stained tumor section. Right, high-magnification images of H&E-stained and immunohistochemically stained sections of regions with distinct histologies. Scale bars, 50  $\mu\text{m}$ . Each FHBT model is a unique tumor that is the result of an independent experiment. **f**, Clonal architecture of the three dominant clones in the tumor as determined by Mission Bio Tapestry single-cell analysis of LV barcodes. **g**, Left, tumor tissue section after LCM of the histologically distinct regions. Right, table showing the associations between regional tumor histologies and clones in **f** based on LCM and bulk DNA amplicon sequencing of LV barcodes.









in regions of sarcomatoid carcinoma with low or absent panCK staining (Fig. 2b). We conducted molecular profiling of these tumors and their regional tumor histologies by LCM and RNA-seq analysis. Principal component analysis (PCA) of the gene expression data showed that squamous and sarcomatoid subtypes clustered together and were separate from urothelial and papillary urothelial carcinomas (Fig. 2e). The BASE47 subtype predictor<sup>12</sup>, a gene expression classifier used to distinguish luminal and basal subtypes of urothelial carcinoma, generally classified the tumors with papillary and papillary squamous subtypes as luminal and the squamous and sarcomatoid histologies as basal, consistent with an established relationship between sarcomatoid differentiation and the basal subtype<sup>13</sup> (Fig. 2f). The Consensus Molecular Classifier<sup>14</sup> revealed that the non-papillary urothelial histologies showed neuroendocrine-like gene expression with low or absent luminal and basal gene signatures (Fig. 2g and Extended Data Fig. 7a). Gene set enrichment analysis (GSEA) was used to compare these tumor histologies in a pairwise manner and revealed the enrichment of genes associated with epithelial-to-mesenchymal transition in sarcomatoid carcinoma, as expected from prior molecular analyses of human tumors<sup>13</sup> (Fig. 2h). We further confirmed the relevance of our FHBT models by projecting their RNA expression patterns onto principle component analysis plots of tumors from the TCGA-BLCA cohort (Fig. 2i) and established *N*-butyl-*N*-(4-hydroxybutyl)-nitrosamine (BBN)-induced mouse bladder cancer models<sup>15</sup> (Extended Data Fig. 7b) to show that they occupy overlapping space on the basis of histologic classification, indicating that the transcriptional features with the greatest variance between tumor subtypes are also conserved with FHBT models.

mPE cells transduced with PE-LVP and engrafted in mice (Extended Data Fig. 8a) also gave rise to mixed cancer morphologies. One tumor showed high-grade prostate adenocarcinoma with focal pleomorphic giant cells (Extended Data Fig. 8b), a rare histologic subtype associated with poor prognosis<sup>16</sup> that may contribute to therapeutic resistance and lethality<sup>17</sup>. Immunostaining revealed HOXB13 and AR expression in both histologies with pronounced nuclear TP53 expression in the pleomorphic giant cells (Extended Data Fig. 8b). We isolated large (pleomorphic giant cell carcinoma) and small (adenocarcinoma) cells from dissociated tumors using a flow cytometry-based strategy, propagated these cells briefly (one passage) in organoid cultures, then dissociated the cells and stained with the nuclear dye Hoechst 33342 to further isolate cells on the basis of DNA content for downstream single-cell lentiviral barcode enumeration (Extended Data Fig. 8c). This single-cell clonality analysis revealed striking enrichment of *shKmt2c* in the putative pleomorphic giant cell clones (Extended Data Fig. 8d). Recent studies have established that KMT2C mediates the DNA damage response in cancer<sup>18,19</sup>, and DNA damage repair alterations are common in human prostate adenocarcinoma with pleomorphic giant cell features<sup>20</sup>.

In summary, we describe a set of technologies that form a functional in vivo cancer genomics assay with efficient delivery of compound genetic perturbations from barcoded lentiviral libraries and single-cell sequencing to rapidly investigate genotype–phenotype relationships in cancer initiation and progression using primary epithelial cells. We leveraged this strategy to develop a series of mouse bladder cancers that recapitulate the phenotypic diversity of human bladder cancer and a mouse prostate cancer with pleomorphic giant cell carcinoma, representing cancer subtypes that have not previously been modeled in a genetically defined fashion. Importantly, single-cell lentiviral barcode deconvolution associated mutant active *Fgfr3* with the luminal papillary differentiation of urothelial carcinoma and the loss of *Kmt2c* with pleomorphic giant cell carcinoma in prostate cancer. These initial studies provide proof of principle that this approach can be deployed to investigate higher-order genetic interactions to explore complex genotype-to-phenotype relationships in cancer.

## Online content

Any methods, additional references, Nature Portfolio 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/s41588-024-01674-1>.

## References

- Loeb, L. A., Loeb, K. R. & Anderson, J. P. Multiple mutations and cancer. *Proc. Natl Acad. Sci. USA* **100**, 776–781 (2003).
- Kretschmar, K. & Clevers, H. Organoids: modeling development and the stem cell niche in a dish. *Dev. Cell* **38**, 590–600 (2016).
- Santos, C. P. et al. Urothelial organoids originating from Cd49high mouse stem cells display Notch-dependent differentiation capacity. *Nat. Commun.* **10**, 4407 (2019).
- Drost, J. et al. Organoid culture systems for prostate epithelial and cancer tissue. *Nat. Protoc.* **11**, 347–358 (2016).
- Robertson, A. G. et al. Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* **171**, 540–556.e525 (2017).
- Winters, B. R. et al. Genomic distinctions between metastatic lower and upper tract urothelial carcinoma revealed through rapid autopsy. *JCI Insight* **5**, e128728 (2019).
- Robinson, D. et al. Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015).
- Kumar, M., Keller, B., Makalou, N. & Sutton, R. E. Systematic determination of the packaging limit of lentiviral vectors. *Hum. Gene Ther.* **12**, 1893–1905 (2001).
- Oottamasathien, S. et al. Bladder tissue formation from cultured bladder urothelium. *Dev. Dyn.* **235**, 2795–2801 (2006).
- Zong, Y., Goldstein, A. S. & Witte, O. N. Preparation of urogenital sinus mesenchymal cells for prostate tissue recombination models. *Cold Spring Harb. Protoc.* **2015**, 988–990 (2015).
- Marzouka, N. A. et al. The Lund molecular taxonomy applied to non-muscle-invasive urothelial carcinoma. *J. Mol. Diagn.* **24**, 992–1008 (2022).
- Damrauer, J. S. et al. Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *Proc. Natl Acad. Sci. USA* **111**, 3110–3115 (2014).
- Guo, C. C. et al. Dysregulation of EMT drives the progression to clinically aggressive sarcomatoid bladder cancer. *Cell Rep.* **27**, 1781–1793.e1784 (2019).
- Kamoun, A. et al. A consensus molecular classification of muscle-invasive bladder cancer. *Eur. Urol.* **77**, 420–433 (2020).
- Shah, S. D. et al. Syngeneic model of carcinogen-induced tumor mimics basal/squamous, stromal-rich, and neuroendocrine molecular and immunological features of muscle-invasive bladder cancer. *Front Oncol.* **13**, 1120329 (2023).
- Alharbi, A. M., De Marzo, A. M., Hicks, J. L., Lotan, T. L. & Epstein, J. I. Prostatic adenocarcinoma with focal pleomorphic giant cell features: a series of 30 cases. *Am. J. Surg. Pathol.* **42**, 1286–1296 (2018).
- Pienta, K. J., Hammarlund, E. U., Brown, J. S., Amend, S. R. & Axelrod, R. M. Cancer recurrence and lethality are enabled by enhanced survival and reversible cell cycle arrest of polyaneploid cells. *Proc. Natl Acad. Sci. USA* **118**, e2020838118 (2021).
- Rampias, T. et al. The lysine-specific methyltransferase KMT2C/MLL3 regulates DNA repair components in cancer. *EMBO Rep.* **20**, e46821 (2019).
- Chang, A. et al. Recruitment of KMT2C/MLL3 to DNA damage sites mediates DNA damage responses and regulates PARP inhibitor sensitivity in cancer. *Cancer Res.* **81**, 3358–3373 (2021).
- Lotan, T. L., Kaur, H. B., Alharbi, A. M., Pritchard, C. C. & Epstein, J. I. DNA damage repair alterations are frequent in prostatic adenocarcinomas with focal pleomorphic giant-cell features. *Histopathology* **74**, 836–843 (2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this

article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

## Methods

### Lentiviral constructs and lentiviral library production

Double-barcoded lentiviral vectors were generated from the parental vector FU-CGW<sup>21</sup> by sequentially inserting matched ten-nucleotide barcodes into the PacI site distal to the HIV FLAP using the Quick Ligation Kit (New England Biolabs) and PCR amplification of the WPRE sequence and barcode with insertion into the ClaI sites proximal to the 3' LTR by HiFi DNA Assembly (New England Biolabs). ORFs were cloned into the EcoRI site of the double-barcoded lentiviral vectors by HiFi DNA Assembly. To generate shRNA lentiviral vectors, the ubiquitin C promoter sequence was excised from the double-barcoded plasmid by digesting with PspXI and EcoRI. U6 promoter and shRNA cassettes were isolated by digesting pLKO.1 TRC shRNA clones with PspXI and EcoRI and were inserted into the digested double-barcoded plasmid using the Quick Ligation Kit. Individual lentiviruses were generated in arrayed format in 293T cells (CRL-11268, ATCC) by co-transfection of each double-barcoded lentiviral ORF or shRNA plasmid with the helper plasmids pVSV-G, pMDL and pRev using FuGENE HD Transfection Reagent (Promega). Lentiviral supernatants were collected 36 h after transfection, pooled and concentrated by ultracentrifugation in V-bottom polypropylene centrifuge tubes on a SW 32 Ti in an Optima XE 90 (Beckman Coulter) at 82,520g at 4 °C for 2 h. Supernatants were aspirated, and lentiviral pellets were resuspended in residual media and cryopreserved.

### shRNA screening

The top three to five shRNA sequences identified from The RNAi Consortium for each target gene were identified from the Genetic Perturbation Platform Web Portal at the Broad Institute. shRNA sequences were cloned into pLKO.1. pLKO.1-TRC control and pLKO.1-shRNA lentiviruses were generated and used to transduce 3T3 cells (gift from V. Vasioukhin, Fred Hutchinson Cancer Center). Seventy-two hours after lentiviral transduction, 3T3 cells were collected, and RNA was collected using an RNeasy Mini Kit (Qiagen). Reverse transcription of RNA was performed using SuperScript IV Reverse Transcriptase (Invitrogen) as per the manufacturer's instructions. qPCR was performed on a QuantStudio 6 using SYBR Green qPCR Master Mix (ThermoFisher Scientific), and primers specific to each target gene and *Ubc* as a control. All primers used for these studies are listed in Supplementary Table 3. Relative expression was calculated using ddCT analysis.

### Embryonic bladder mesenchyme and urogenital sinus mesenchyme preparation

All animal care and studies were performed in accordance with an approved Fred Hutchinson Cancer Center Institutional Animal Care and Use Committee (IACUC) protocol (PROTO000051048) and Comparative Medicine regulations. All animals were housed in an Association for Assessment and Accreditation of Laboratory Animal Care (AALAC)-accredited facility and subjected to a 12-h light/dark cycle with the temperature maintained between 18 °C and 24 °C and 40–60% humidity. UGSM was isolated and propagated as previously described<sup>21</sup>. E16 fetal bladders were also collected at the same time as the urogenital sinus and subjected to similar steps for preparation of EBLM. UGSM and EBLM were passaged less than five times before use in engraftment studies.

### Mouse bladder and prostate dissociation and organoid culture

Bladder and prostates from 8- to 12-week-old male C57BL/6 mice (The Jackson Laboratory) were dissected and mechanically and enzymatically dissociated as previously described<sup>21</sup>. Cells were stained with antibodies for FACS on a Sony SH800 Cell Sorter with collection of Lin<sup>-</sup>CD49<sup>high</sup>EpCAM<sup>high</sup> cells. Between  $1 \times 10^4$  and  $2 \times 10^4$  bladder urothelial and prostate epithelial cells were resuspended in a total of 15  $\mu$ l of growth factor-reduced Matrigel (Corning) with or without

concentrated lentivirus and seeded as droplets in each 48-well tissue culture plate well. Cells were cultured as previously described<sup>22</sup>. Mouse bladder organoid culture media consisted of Advanced DMEM-F12, 10 mM HEPES, 2 mM GlutaMAX, B27 supplement, 1.25 mM N-acetylcysteine, 50 ng ml<sup>-1</sup> hEGF, 100 ng ml<sup>-1</sup> hNoggin and 500 ng ml<sup>-1</sup> hR-spondin, 200 nM A83-01 and 10  $\mu$ M Y-27632. Mouse prostate organoid culture media consisted of mouse bladder organoid culture media with the addition of 1 nM dihydrotestosterone.

### Organoid transformation assay

After 5–7 days of culture, transduced mouse bladder urothelial or prostate epithelial organoids were liberated by dissociating the Matrigel matrix with 5 U ml<sup>-1</sup> dispase (STEMCELL Technologies). Organoids were washed with PBS and resuspended in ice-cold Matrigel with either 10<sup>5</sup> EBLM or UGSM and subcutaneously injected into the flanks of 6- to 8-week-old male NSG (NOD-SCID-IL2Ry-null) mice (The Jackson Laboratory). For prostate epithelial transformation studies, mice were supplemented with testosterone through the subcutaneous implantation of 90-day release testosterone pellets (Innovative Research of America). Tumors were collected when they reached 1 cm in maximal diameter. The maximum tumor size permitted by the Fred Hutchinson Cancer Center IACUC is 2 cm in diameter, which was not exceeded during these studies.

### Copy number assay

DNA was extracted from organoids using a GeneJET Genomic DNA Purification Kit (ThermoFisher Scientific). Copy number analysis was performed by TaqMan Real-Time PCR Assay (ThermoFisher Scientific) using the TaqMan Copy Number Reference Assay, mouse, Tfrc (4458366) and EGFP TaqMan Copy Number Assay (Mr00660654\_cn) on a QuantStudio 6. Genomic DNA extracted from the tails of transgenic C57BL/6 mice with one or two copies of GFP was used as a calibrator sample. GFP copy number was determined using ddCT analysis, where sample copy number = calibrator copy number  $\times 2^{-\text{ddCT}}$ .

### Single-cell DNA amplicon sequencing library preparation and sequencing

A custom panel was designed for the Mission Bio Tapestri to amplify segments of ten mouse genes at two exons each, the 5' and 3' lentiviral barcodes and lentiviral GFP. Libraries were generated either from cryopreserved or freshly dissociated tumor cells using the Mission Bio Tapestri Single-cell DNA Custom Kit according to the manufacturer's recommendations. Single cells (3,000 to 4,000 cells per  $\mu$ l) were resuspended in Tapestri cell buffer and encapsulated using a Tapestri microfluidics cartridge, lysed and barcoded. Barcoded samples were subjected to targeted PCR amplification, and PCR products were removed from individual droplets, purified with KAPA Pure Beads (Roche Molecular Systems) and used as a template for PCR to incorporate Illumina P7 indices. PCR products were purified by KAPA Pure Beads and quantified by Qubit dsDNA High Sensitivity Assay (ThermoFisher Scientific). Sample quality was assessed by Agilent TapeStation analysis. Libraries were pooled and sequenced on an Illumina MiSeq or HiSeq 2500 with 150 bp paired-end reads in the Fred Hutchinson Cancer Center Genomics Shared Resource.

### Laser capture microdissection and DNA and RNA isolation for high-throughput sequencing

Sections 10  $\mu$ m thick were cut from formalin-fixed paraffin-embedded (FFPE) tumor tissue blocks and mounted onto PEN Membrane Frame Slides (ThermoFisher Scientific). Sections were fixed with 95% ethanol for 1 min, stained with 3% cresyl violet and dehydrated through graded alcohols and xylene. Histology review and annotation were performed by a pathologist. Laser capture microdissection was performed on an Arcturus XT Laser Capture Microdissection System (ThermoFisher Scientific). Microdissected specimens were collected for DNA and



RNA extraction. DNA was extracted using a GeneRead DNA FFPE Kit (Qiagen), and RNA was extracted using an RNeasy FFPE Kit (Qiagen) according to the manufacturer's protocols. Two-step PCR for lentiviral barcode amplification and sequencing library adaptor ligation was performed. The first PCR reaction consisted of 2x KAPA HiFi HotStart ReadyMix, 100 nM of 1° FWD primer (5'-TCGTCGGCAGCGTCAGATGTGATAAGAGACAGCAAATTTTCGGGTTTATTACAGG-3'), 100 nM of 1° REV primer (5'-GTCTCGTGGGCTCGGAGATGTGATAAGAGACAGGCCGCTCGAGGACTATTAAG-3') and 80 ng of genomic DNA. Thermal cycling conditions were 95 °C for 3 min; (95 °C for 30 s, 64 °C for 30 s, 72 °C for 30 s) × 25 cycles; 72 °C for 5 min; and hold at 4 °C. PCR cleanup was conducted using the Wizard SV Gel and PCR Clean-Up System (Promega), with elution in 30 µl of double distilled water. The second PCR reaction consisted of 2x KAPA HiFi HotStart ReadyMix, 140 nM of 2° i7 primer, 140 nM of 2° i5 primer and 5 µl of elution from the PCR cleanup of the 1° PCR. Thermal cycling conditions were 95 °C for 3 min, (95 °C for 30 s, 61 °C for 30 s, 72 °C for 30 s) × 8 cycles; 72 °C for 5 min; and hold at 4 °C. The sequences of 2° primers used to incorporate dual-indexed Illumina sequencing adaptors are displayed in Supplementary Table 4. PCR cleanup was conducted using the Wizard SV Gel and PCR Clean-Up System, with elution in 30 µl of double distilled water. Sample quality was assessed by Agilent TapeStation analysis. Sequencing was performed on an Illumina MiSeq or HiSeq 2500 instrument using 150 bp single-end reads. PhiX sequences were excluded from the sequencing reads by Bowtie 2 v2.4.4 (ref. 23). Cutadapt v4.1 (ref. 24) was used to trim the reads to the barcode region. Then the trimmed reads were aligned to custom DNA references containing all barcodes using Bowtie 2. Samtools v1.11 (ref. 25) was used to extract read counts for each barcode. The RNA-seq libraries were prepared using a SMARTer Stranded Total RNA-Seq Kit v3 - Pico Input Mammalian (Takara Bio) and sequenced on an Illumina NovaSeq 6000 using a NovaSeq S4 flow cell with 100 bp paired-end reads by MedGenome. Sequencing reads were mapped to mouse genome reference GRCm39, and gene expression was quantified and normalized using the UC Santa Cruz Computational Genomics Lab Toil RNA-seq pipeline v4.1.2 (ref. 26).

### Transcriptional subtype analysis and PCA projections

All computational analyses were carried out in RStudio v4.1.0. Mouse Ensembl genes were converted to Mouse Genome Informatics (MGI) gene symbols using the biomaRt package v2.24.1 (<https://bioconductor.org/packages/release/bioc/html/biomaRt.html>). MGI gene symbols were then converted to their human orthologs by referencing the mouse–human ortholog database available from The Jackson Laboratory ([http://www.informatics.jax.org/downloads/reports/HOM\\_MouseHumanSequence.rpt](http://www.informatics.jax.org/downloads/reports/HOM_MouseHumanSequence.rpt)). The human ortholog matrix was used for downstream analysis in transcriptional subtype analysis. FHBT samples were classified using the BASE47 subtype predictor gene list and the ConsensusMIBC package v1.1 (<https://github.com/cit-bioinfo/consensusMIBC>). Z-score means of genes and signature scores were calculated for each sample. Heatmaps of both the BASE47 and ConsensusMIBC results were generated using the pheatmap package v1.0.12 (<https://www.rdocumentation.org/packages/pheatmap/versions/1.0.12/topics/pheatmap>). For PCA analysis, the FPKM human ortholog matrix was normalized by  $\log_2 + 1$  transformation before performing mean-centered PCA using the prcomp package v3.6.2 (<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/prcomp>). Visualization of the PCA plot was performed using the factoextra package v1.0.7 (<https://cran.r-project.org/web/packages/factoextra/index.html>) and ggpubr package v0.6.0 (<https://www.rdocumentation.org/packages/ggpubr/versions/0.6.0>).

For PCA projections, RNA-seq count data from the FHBT, GSE220999 and TCGA-BLCA datasets were transformed to counts per million, normalized and batch corrected using ComBat-seq<sup>27</sup> to compare across each dataset using the DGEobj.utils package v1.0.6

(<https://rdrr.io/cran/DGEobj.utils>). PCA projection of the FHBT data onto the TCGA-BLCA space was done by first generating a PCA of the TCGA-BLCA samples from the common genes between the FHBT and GSE220999 data. A PCA for both the FHBT and GSE220999 samples was then scaled by the eigenvalues of the TCGA-BLCA using the base package v3.6.2 (<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/scale>). A plot was constructed overlaying the reference TCGA-BLCA samples with either FHBT or GSE220999 tumor projections using ggplot2 v3.4.1 (<https://cran.r-project.org/web/packages/ggplot2/index.html>). TCGA-BLCA samples were colored by their Consensus Molecular Classifier subtype. Differential gene expression analysis was performed pairwise between FHBT histologies using the DESeq2 package v1.38.3 (ref. 28). P values were generated via the Wald test and P-adjusted using the Benjamini–Hochberg correction. Pre-ranked GSEA (Broad Institute) was conducted by inputting a ranked list of differentially expressed genes based on  $\log_{10}$ -transformed P values from the DESeq2 analysis for each pairwise comparison. Dot plots were generated by plotting the normalized enrichment score and log-transformed false discovery rate for each pre-ranked GSEA output using ggplot2.

### Single-cell lentiviral barcode enumeration and clonality analysis

Raw sequencing reads were trimmed to the amplicon regions using the awk command. Barcode sequences in the reads were filtered and extracted using UMI-tools v1.0.0 (ref. 29). Processed reads were aligned to custom references containing all amplicon sequences using bwa-mem v0.7.17-r1188 (ref. 30). Samtools was used to extract amplicon counts for each barcode. Mouse cells with no GFP amplicon counts were removed. Counts per cell were normalized to total counts for each barcode. A minimum threshold normalized count of 1% of total counts was used to define the presence of a barcode in a cell. The clonal architecture of cells was determined by enumerating all cells containing each distinct combination of barcodes.

### Immunohistochemistry

Tumor samples were formalin-fixed and paraffin-embedded, sectioned to a 5-µm thickness and placed on positively charged glass slides. For each tumor, slides were stained with a standard hematoxylin and eosin protocol. Immunohistochemical staining was performed according to an established protocol<sup>31</sup>. Stained slides were digitally scanned on a VENTANA DP 200 (Roche) and analyzed using QuPath 0.2.3 (ref. 32).

### Antibodies

Antibodies used for FACS: Human/mouse/bovine integrin alpha 6/CD49f PE-conjugated antibody (FAB13501P, R&D Systems, 1:40); PE/Cyanine7 anti-mouse CD326 (Ep-CAM) antibody (118216, BioLegend, 1:40); CD31 (PECAM-1) monoclonal antibody (390), FITC (11-0311-82, eBioscience, 1:100); CD45 monoclonal antibody (30-F11), FITC (11-0451-85, eBioscience, 1:100); TER-119 monoclonal antibody (TER-119), FITC (11-5921-82, eBioscience, 1:100). Antibodies used for immunohistochemistry: Anti-wide spectrum Cytokeratin antibody (ab9377, Abcam, 1:100); rabbit monoclonal GFP antibody (clone D5.1, Cell Signaling, 1:100); rabbit polyclonal p63 antibody (12143-1-AP, Proteintech, 1:200); mouse monoclonal p53 antibody (clone 1C12, Cell Signaling, 1:500); rabbit monoclonal HOXB13 antibody (clone D7N8O, Cell Signaling, 1:50); rabbit polyclonal AR antibody (06-680, Millipore, 1:2,000); rabbit monoclonal GATA3 antibody (clone D13C9, Cell Signaling, 1:200); rabbit monoclonal CD44 antibody (clone E7K2Y, Cell Signaling, 1:100).

### Statistical analyses

Data analysis was performed on GraphPad Prism 9 (GraphPad Software). qPCR results were analyzed in Excel. Statistical significance was determined using the unpaired two-tailed Student's *t*-test. Results are

depicted as mean + s.d. unless stated otherwise. For all statistical tests, *P* values of <0.05 were considered significant.

### Reporting summary

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

### Data availability

Sequencing data pertaining to this study are available from the Gene Expression Omnibus (GEO) as SuperSeries [GSE229783](#). RNA-seq data from FHBT models are available from accession number [GSE229780](#). Bulk DNA amplicon sequencing data from lentiviral library representation studies and from FHBT models are available from accession numbers [GSE231542](#) and [GSE229781](#), respectively. Single-cell DNA amplicon sequencing data related to determining the unique proviral copies per cell after lentiviral transduction across a range of MOIs are available from accession number [GSE231543](#). Single-cell DNA amplicon sequencing data from FHBT models and enriched cells from the tumor model with prostate adenocarcinoma and focal pleomorphic giant cell carcinoma are available from accession number [GSE229782](#).

### Code availability

The study did not use any custom code or software. All code or software used for all data processing and analysis has been described in the Methods section. No custom code was used to generate figures.

### References

- Xin, L., Ide, H., Kim, Y., Dubey, P. & Witte, O. N. In vivo regeneration of murine prostate from dissociated cell populations of postnatal epithelia and urogenital sinus mesenchyme. *Proc. Natl Acad. Sci. USA* **100**, 11896–11903 (2003).
- Karthauss, W. R. et al. Identification of multipotent luminal progenitor cells in human prostate organoid cultures. *Cell* **159**, 163–175 (2014).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
- Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
- Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
- Vivian, J. et al. Toil enables reproducible, open source, big biomedical data analyses. *Nat. Biotechnol.* **35**, 314–316 (2017).
- Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom. Bioinform.* **2**, lqaa078 (2020).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- DeLucia, D. C. et al. Regulation of CEACAM5 and therapeutic efficacy of an anti-CEACAM5-SN38 antibody-drug conjugate in neuroendocrine prostate cancer. *Clin. Cancer Res.* **27**, 759–774 (2021).
- Bankhead, P. et al. QuPath: open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).

### Acknowledgements

We thank C. Morrissey, L. Xin and P. Nelson for critical discussion and review of this work. This work was supported by National Institutes of Health (NIH) grants DP2 CA271301 (J.K.L.), R01 CA276308 (A.C.H.), R37 CA230617 (A.C.H.), a Bladder Cancer Advocacy Network Research Innovator Award (J.K.L.), a Department of Defense Peer Reviewed Cancer Research Program Career Development Award W81XWH-19-1-0569 (J.K.L.), a Department of Defense Prostate Cancer Research Program Early Investigator Research Award W81XWH-20-1-0083 (S.L.) and a Department of Defense Peer Reviewed Cancer Research Program Horizon Award W81XWH-19-1-0658 (S.J.). We acknowledge support from the Seattle Translational Tumor Research Program in Bladder Cancer and the University of Washington Medicine Urethral Cancer Research Fund provided by donor D.L. Rich. This research was also supported by the Flow Cytometry, Experimental Histopathology and Genomics Shared Resources of the Fred Hutch/University of Washington Cancer Consortium funded by NIH grant P30 CA015704.

### Author contributions

S.L., A.W., H.S. and J.K.L. designed experiments. S.L., A.W., H.S., V.B., G.J. and S.J. performed the experiments. H.S. and G.J. analyzed the RNA-seq and bulk and single-cell DNA amplicon sequencing data. Q.W. provided statistical oversight. R.B.M., J.L.W., H.-M.L., A.C.H., B.M.F., M.C.H. and J.K.L. supervised the study. S.L. and J.K.L. wrote the manuscript with contributions from all authors.

### Competing interests

J.K.L. served on the Speaker's Bureau for Mission Bio. B.M.F. has consulting or advisory roles with QED Therapeutics, Boston Gene, Astrin Biosciences Merck, Immunomedics/Gilead, Guardant and Janssen and receives patent royalties from Immunomedics/Gilead and research support from Eli Lilly. All other authors declare no competing interests.

### Additional information

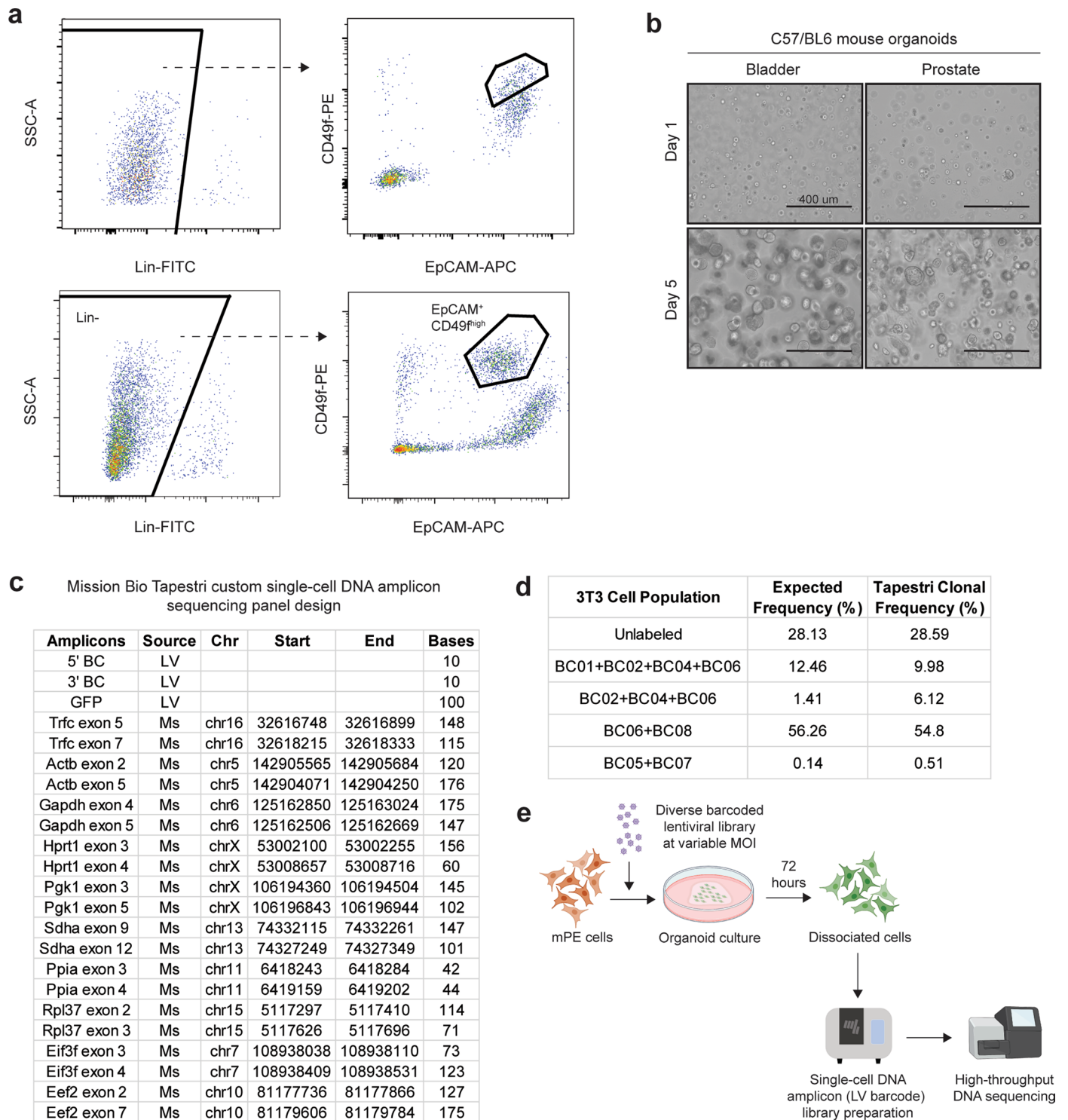
**Extended data** is available for this paper at <https://doi.org/10.1038/s41588-024-01674-1>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41588-024-01674-1>.

**Correspondence and requests for materials** should be addressed to John K. Lee.

**Peer review information** *Nature Genetics* thanks David Fernandez-Antoran and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



**Extended Data Fig. 1 | Isolation of mouse bladder urothelial and prostate epithelial cells for organoid culture and design/validation of a custom Mission Bio Tapestri single-cell DNA amplicon sequencing panel.**

(a) Representative flow cytometry plot for the isolation of mouse bladder urothelial (top) and prostate epithelial (bottom) from dissociated tissues based on a Lin<sup>-</sup> (CD45<sup>-</sup> CD31<sup>-</sup> Ter119<sup>-</sup>) EpCAM<sup>+</sup> CD49f<sup>high</sup> immunophenotype. (b) Images of organoid cultures of mouse bladder urothelial and prostate epithelial cells on day 1 and day 5 after seeding. (c) Table showing the amplicons represented in a custom Mission Bio Tapestri single-cell DNA amplicon sequencing panel.

(d) Table showing results of a validation study where a defined mixture of 3T3 cells with an unlabeled population and others labeled with combinations of lentiviruses encoding distinct barcodes were analyzed using the Mission Bio Tapestri single-cell DNA amplicon sequencing panel to determine clonality. ~2,000 cells were analyzed. (e) Overview of experiments with infection of mouse prostate epithelial (mPE) cells with a diverse barcoded lentiviral library in organoid culture across a range of multiplicity-of-infection (MOI) and quantification of viral copy number per cell across the population by single-cell amplicon sequencing. Created with [BioRender.com](https://www.biorender.com).

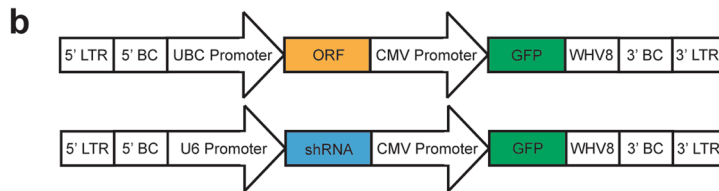


**a** Genetic alterations associated with bladder cancer

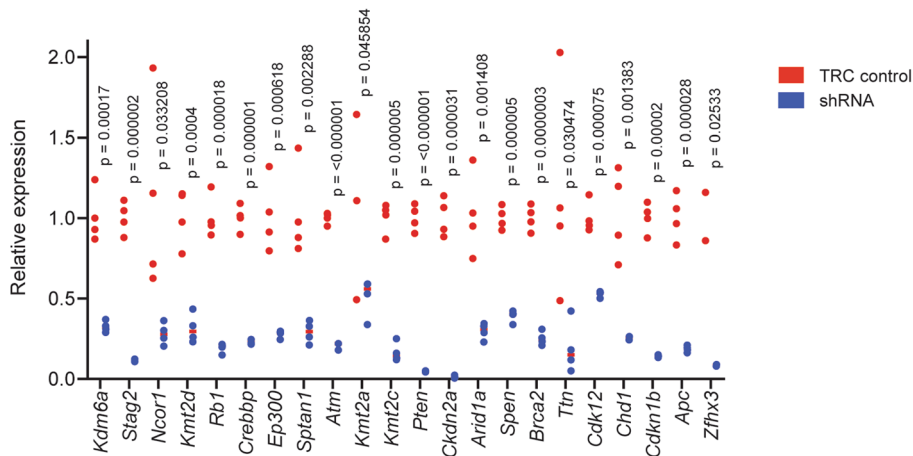
Gain-of-function	Loss-of-function
E2f3	Kmt2d
Pparg	Kdm6a
Mdm2	Arid1a
Ccnd1	Kmt2c
Egfr	Ep300
Pvrl4	Stag2
Ywhaz	Atm
Yap1	Fat1
Ccne1	Sptan1
Myc	Kmt2a
Zfp703	Cdkn2a
Trp53 R245Q	Rb1
Pik3ca E545K	Crebbp
Fgfr3 S243C	Ncor1
ErbB3 V104L	Pten
ErbB2 S311Y	

Genetic alterations associated with prostate cancer

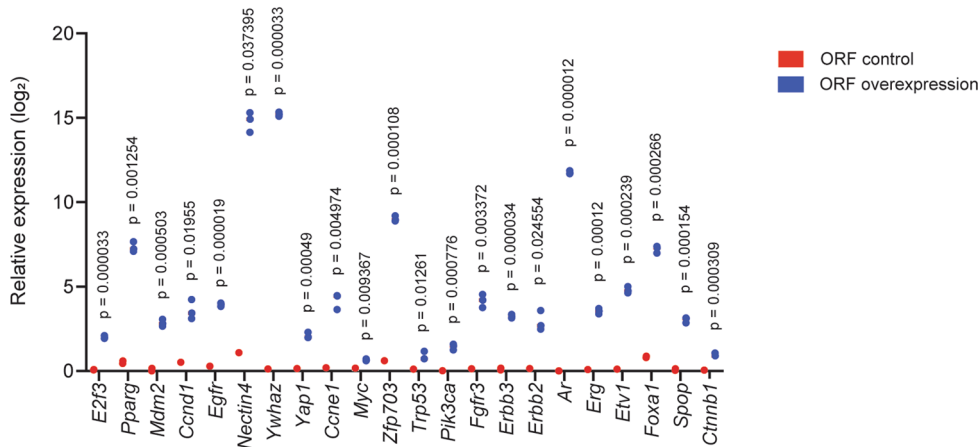
Gain-of-function	Loss-of-function
Ar	Pten
Erg	Ncor1
Etv1	Spn
Pik3ca E545K	Apc
Trp53 R245Q	Brca2
Spop F133C	Atm
Foxa1 R261C	Cdk12
Ctnnb1 D32A	Cdkn1b
Myc	Kmt2c
	Kmt2d
	Kdm6a
	Chd1
	Ttn
	Zfx3



**c** shRNA validation in 3T3s by qPCR



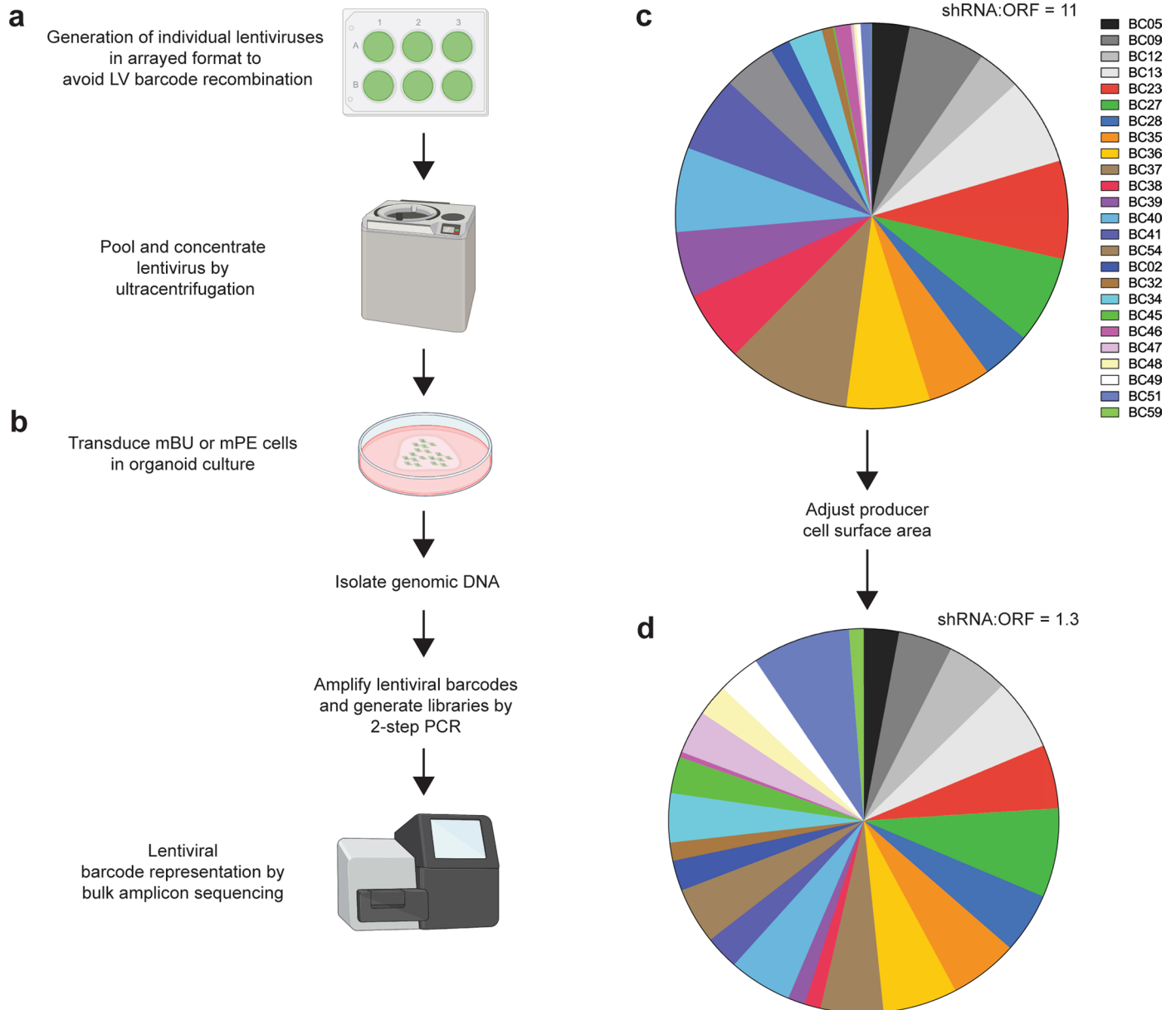
**d** ORF validation in 3T3s by qPCR



Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | Recurrent genetic alterations associated with bladder and prostate cancer encoded in barcoded lentiviral libraries. (a)** Tables showing gain-of-function and loss-of-function genetic alterations associated with bladder and prostate cancer selected for representation in cancer-specific barcoded lentiviral libraries. **(b)** Schematics of barcoded lentiviral vectors expressing open reading frames (ORF) or short-hairpin RNA (shRNA). LTR = long terminal repeat; BC = barcode; UBC = Ubiquitin C; CMV = cytomegalovirus; GFP = green fluorescent protein; WHV8 = Woodchuck hepatitis virus 8 post-transcriptional regulatory element. **(c)** Plot showing relative expression of target genes as determined by quantitative polymerase chain reaction (qPCR) in 3T3

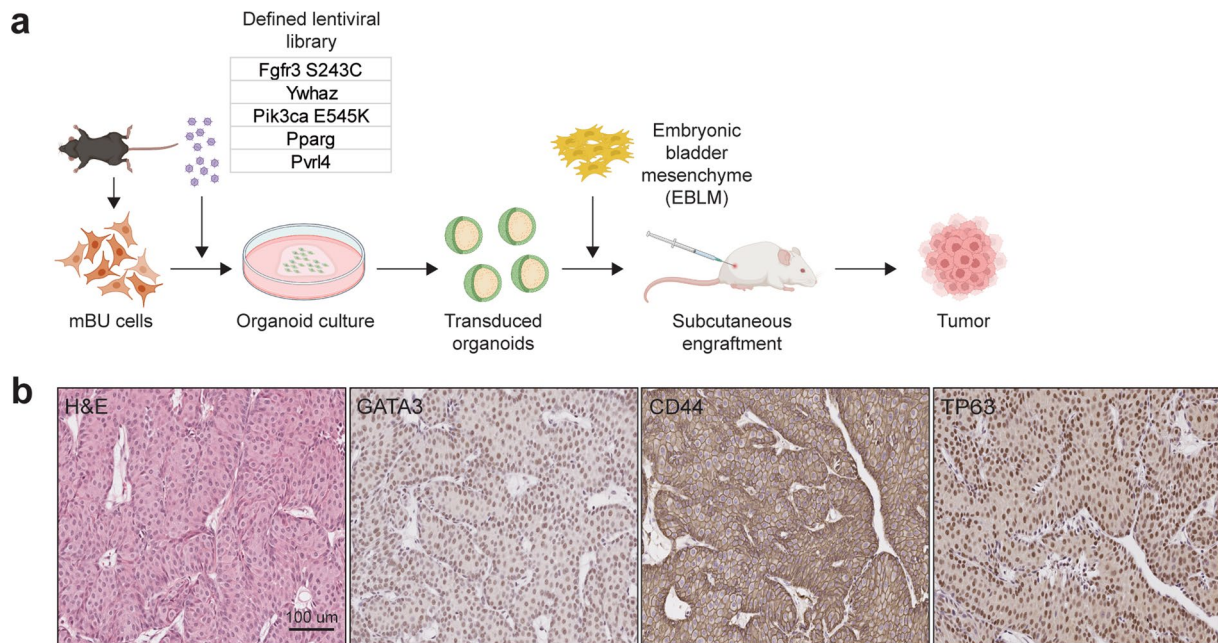
cells 72 hours after lentiviral transduction with pLKO.1-TRC control or pLKO.1 expressing select shRNA previously screened and selected for inclusion in the barcoded lentiviral libraries based on the extent of gene knockdown. qPCR reactions were performed on four biologically independent replicates. Statistical analysis was performed by two-tailed, unpaired *t*-test with p-values shown. **(d)** Plot showing relative overexpression of gain-of-function genes as determined by qPCR in 3T3 cells 72 hours after lentiviral transduction with barcoded ORF vectors. qPCR reactions were performed on three biologically independent replicates. Statistical analysis was performed by two-tailed, unpaired *t*-test with p-values shown.



**Extended Data Fig. 3 | Generation of barcoded lentiviral libraries and normalization of library representation.** Schema showing the (a) generation of individual lentiviruses from the library in arrayed format with subsequent pooling and concentration by ultracentrifugation and (b) transduction of respective mouse bladder urothelial (mBU) or prostate epithelial (mPE) cells in organoid culture with concentrated lentiviral libraries to determine lentiviral

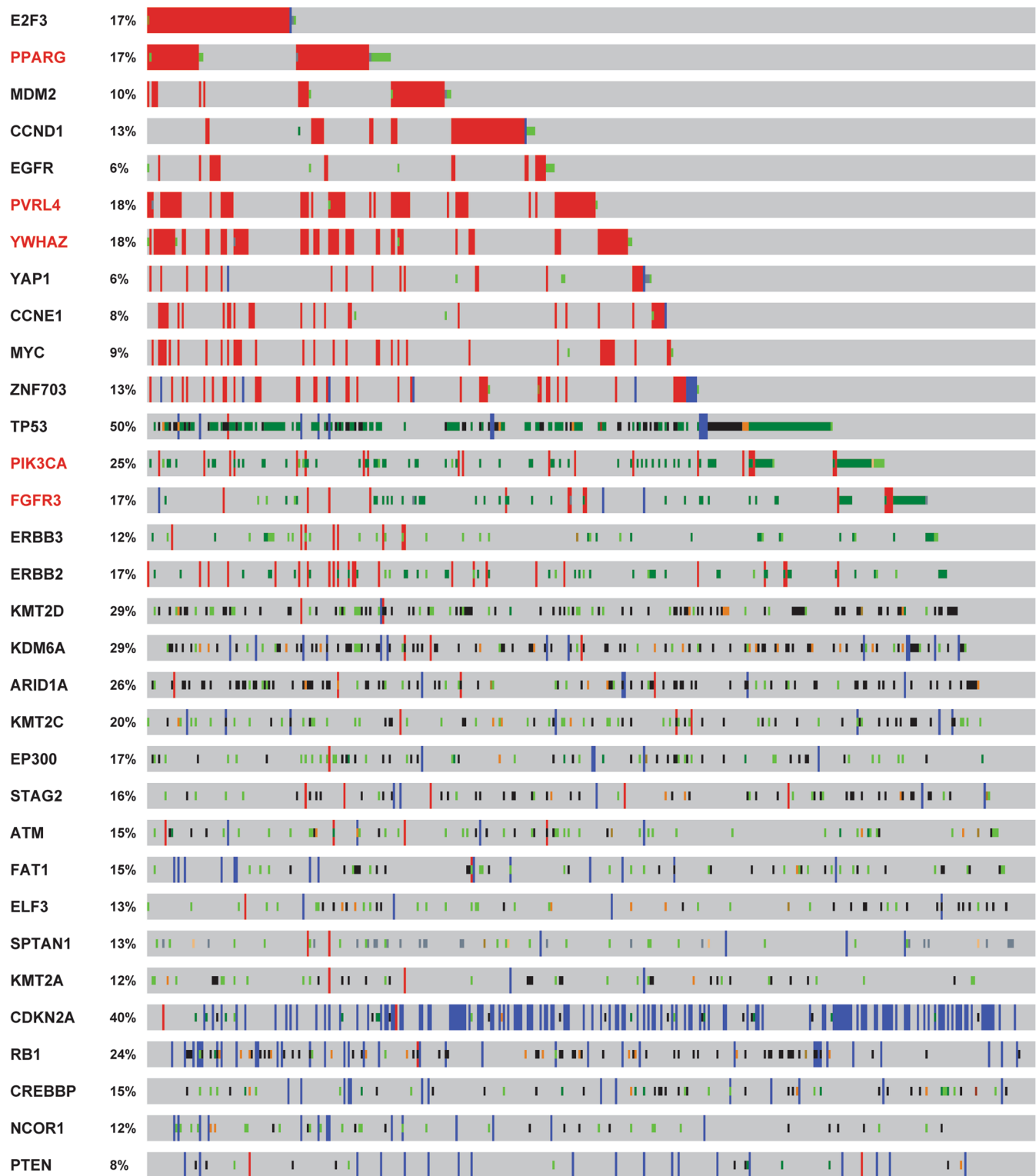
barcode representation by bulk amplicon sequencing of genomic DNA. Created with [BioRender.com](https://www.biorender.com). (c) Representative distribution of barcoded lentiviruses within a library with skewed enrichment of shRNA relative to ORF lentiviruses. (d) Representative distribution of barcoded lentiviruses within a library after applying information from c to adjust producer cell surface area in a for the generation of the lentiviral library.



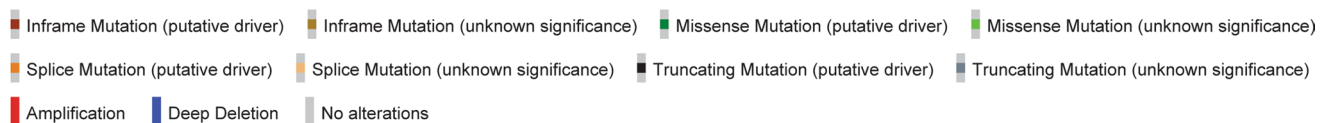


**Extended Data Fig. 4 | Active mutant Fgfr3 S243C cooperates with other oncogenic factors in mouse bladder urothelial cells to drive papillary urothelial carcinoma with inverted growth pattern. (a)** Scheme of the mBU organoid transformation assay using a defined lentiviral library to confirm

functional genotype-phenotype associations. Created with [BioRender.com](https://www.biorender.com). **(b)** High-magnification images of H&E- and IHC-stained sections of a resultant tumor of the experiment in **a** with histologic features consistent with papillary urothelial carcinoma with inverted growth pattern.

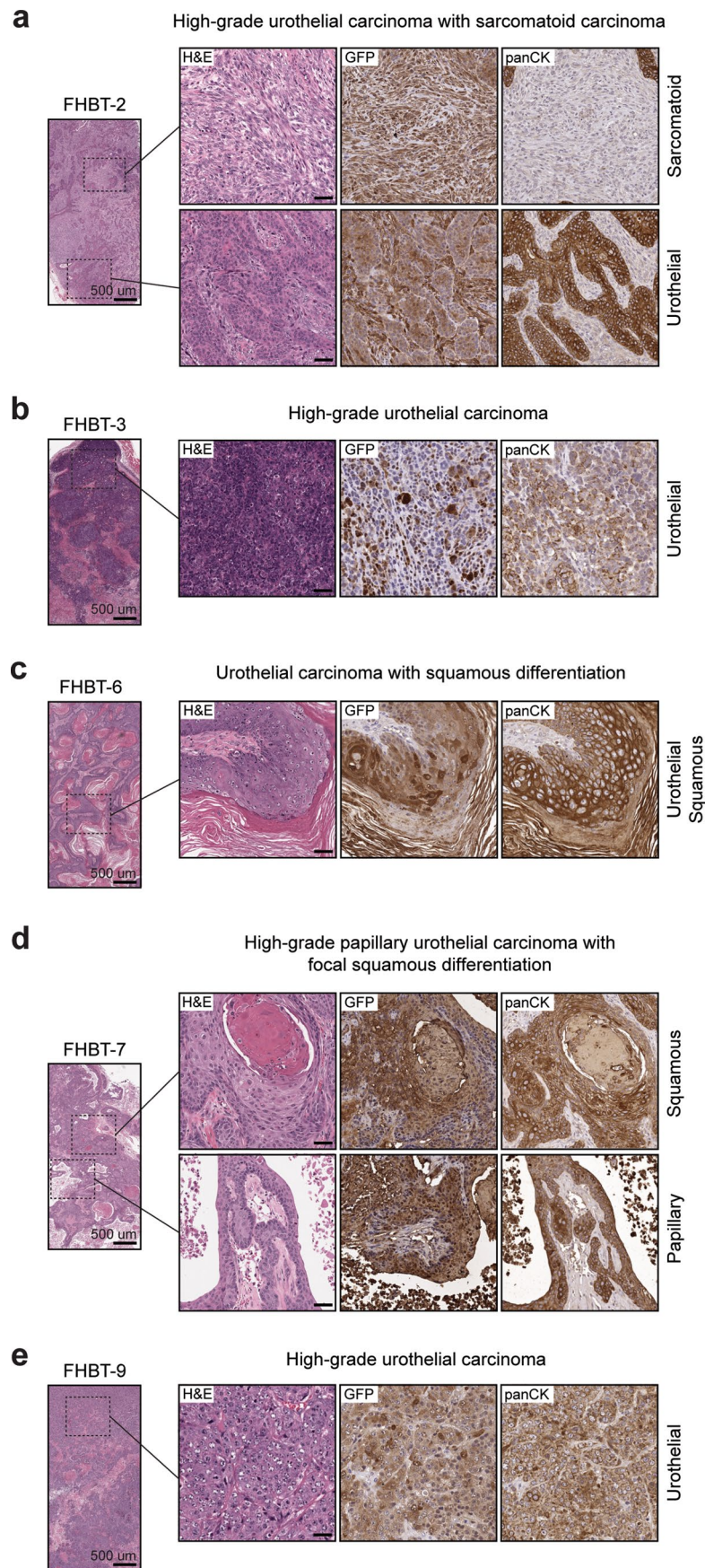


Genetic Alteration



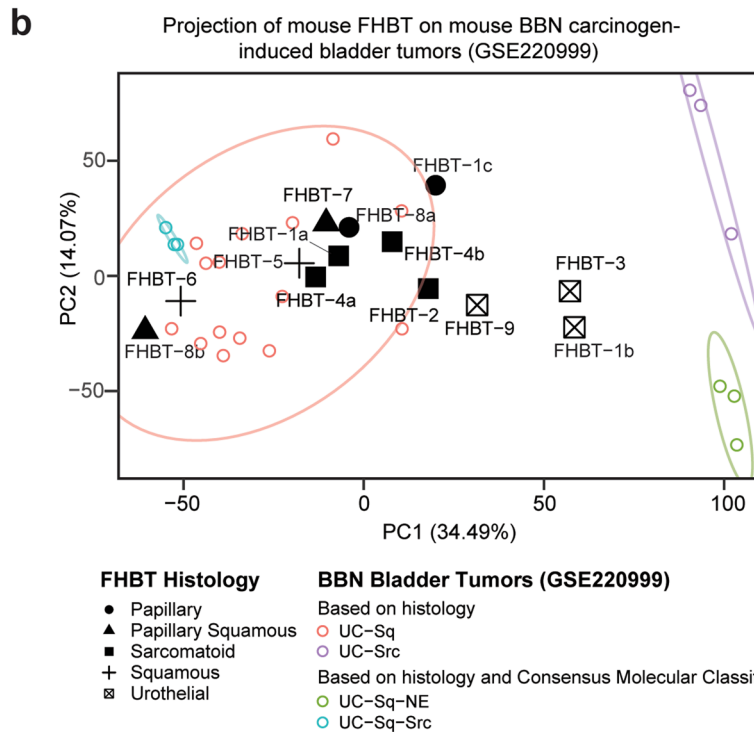
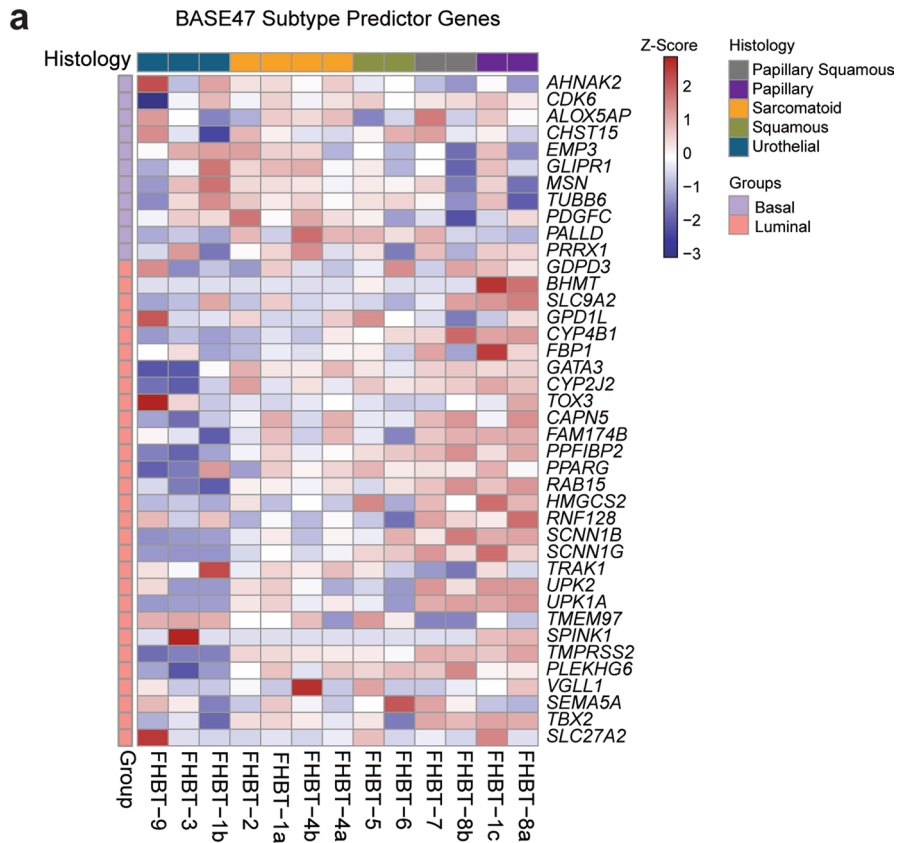
**Extended Data Fig. 5 | Frequencies of gene alterations represented in the bladder urothelial LV pool in human muscle-invasive bladder cancer.**  
 Oncoprint from cBioPortal analysis of human muscle-invasive bladder cancers

from The Cancer Genome Atlas bladder cancer (TCGA-BLCA) cohort showing select genes for which gain- or loss-of-function events were incorporated into the bladder urothelial LV pool.



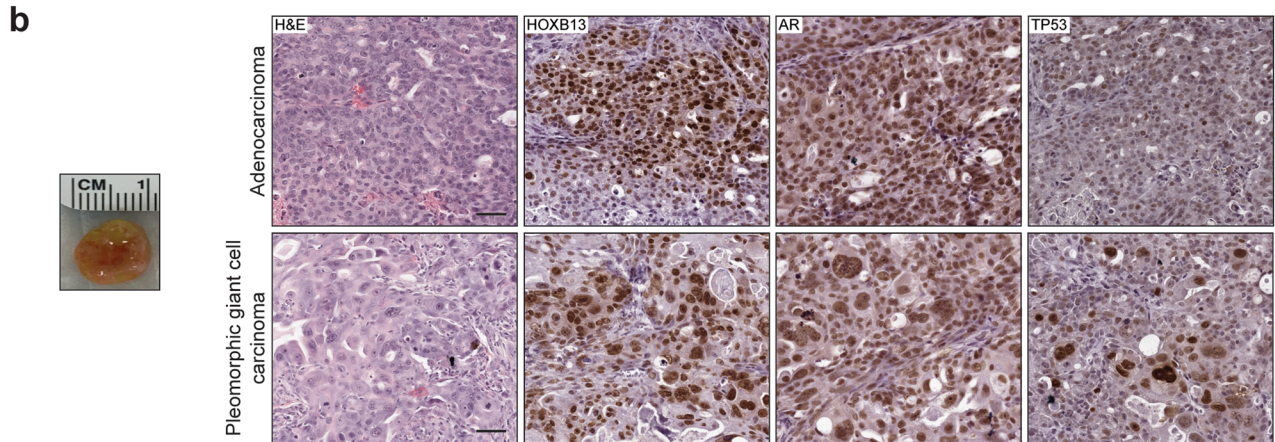
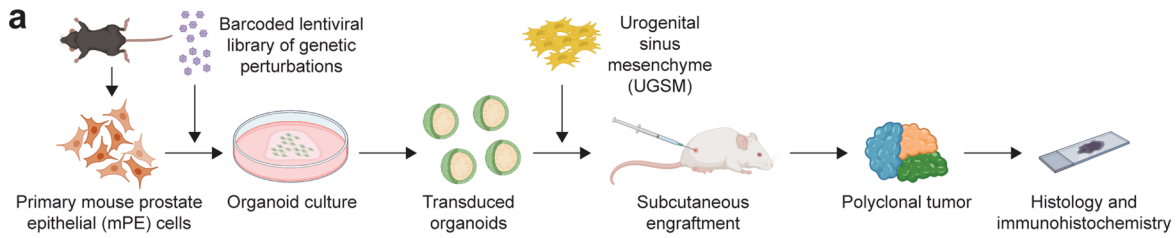
**Extended Data Fig. 6 | FHBT models demonstrate diverse cancer histologies.** (a–e) Low- and high-magnification images of H&E-stained sections and high-magnification images of IHC-stained sections for GFP and pan-cytokeratin (panCK) expression depicting characteristic histologies. Scale bars = 50  $\mu$ m.



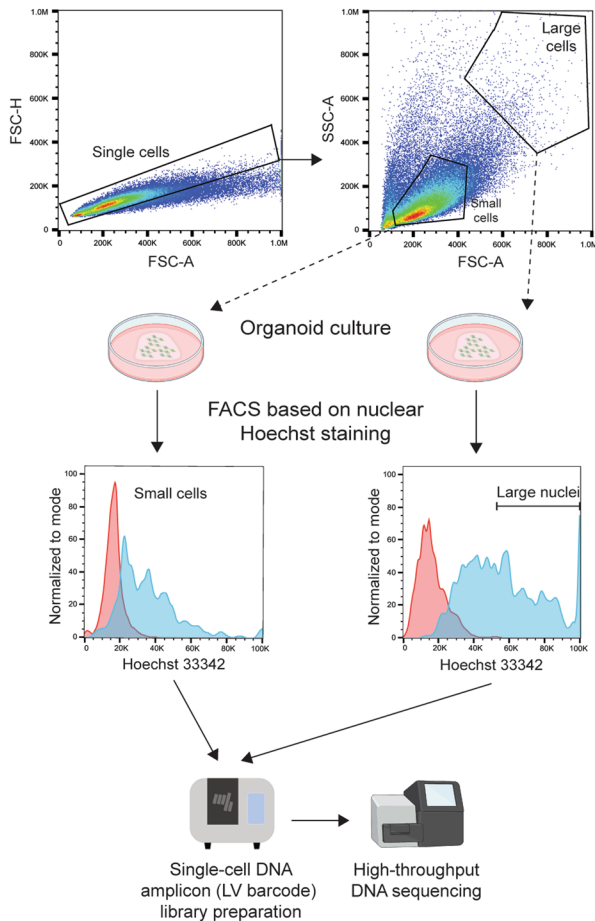


**Extended Data Fig. 7 | Phenotypic diversity and relevance of FHBT models.** (a) Heatmap showing the histologies of the Fred Hutch Bladder Tumor (FHBT) series relative to expressions of genes that constitute basal and luminal signatures for the BASE47 subtype predictor. (b) Principle component analysis

(PCA) projection plot of FHBT samples over BBN carcinogen-induced mouse bladder tumors color-coded based on histology or histology and Consensus Molecular Classification (UC = urothelial carcinoma, Sq = squamous, Src = sarcomatoid, NE = neuroendocrine) with 90% confidence ellipses shown.



**c** FACS of dissociated mixed prostate adenocarcinoma and pleomorphic giant cell carcinoma tumor based on cell size



**d** Single-cell clonality analysis of small cells/nuclei

Genes encoded by LV	Clonal frequency (%)
Etv1	34.3
Spop F133C, Etv1	26.1
shZfx3	5.0
Spop F133C, shZfx3	2.9
Myc, shZfx3	2.6
Myc, Spop F133C, shZfx3	2.2
Etv1, shZfx3	1.9
Myc, Etv1	1.9
Spop F133C, Etv1, shZfx3	1.6
Trp53 R245Q, Etv1, shZfx3	1.5

Single-cell clonality analysis of large cells/nuclei

Genes encoded by LV	Clonal frequency (%)
shKmt2c, shTtn, Etv1	11.8
shKmt2c, shTtn, Spop F133C, Etv1	5.5
shKmt2c, shTtn, Erg, Etv1	5.1
Myc, Etv1	4.7
Spop F133C, Etv1	4.5
shKmt2c, Etv1	4.1
shKmt2c, shTtn, Spop F133C, Erg, Etv1	2.7
shKmt2c, Spop F133C, Etv1	2.7
shKmt2c, shTtn, Myc, Etv1	2.7
Myc, Spop F133C, Etv1	2.3
shKmt2c, Spop F133C, Erg, Etv1	2.1
shKmt2c, Myc, Etv1	1.9
shKmt2c, Pik3ca E545K, Spop F133C, Erg, Etv1	1.9
shKmt2c, Erg, Etv1	1.8
shKmt2c, Myc, Spop F133C, Etv1	1.7
shKmt2c, Pik3ca E545K, shTtn, Erg, Etv1	1.7
shKmt2c, Pik3ca E545K, shTtn, Etv1	1.5

Extended Data Fig. 8 | See next page for caption.

**Extended Data Fig. 8 | Association of adenocarcinoma with polymorphic giant cell carcinoma of the prostate with perturbation of *Kmt2c*.**

(a) Scheme of the mouse prostate epithelial (mPE) organoid transformation assay to uncover functional genotype-phenotype associations in prostate cancer. Created with [BioRender.com](https://www.biorender.com). (b) *Left*, Gross image of a tumor arising from mPE transformed with a prostate epithelial lentiviral pool (PE-LVp). *Right*, high-magnification images of H&E- and IHC-stained sections of regions with high-grade adenocarcinoma and pleomorphic giant cell carcinoma.

Scale bar = 50  $\mu$ m. (c) Overview of the experimental approach to enrich for prostate adenocarcinoma and pleomorphic giant cell carcinoma based on cell size and nuclear DNA content followed by single-cell lentiviral barcode enumeration. (d) Tables showing single-cell clonality analysis of: *Top*, tumor cells enriched for 'small cells/nuclei.' *Bottom*, tumor cells enriched for 'large cells/nuclei.' Highlighted in red is *shKmt2c* based on the enumeration of the associated lentiviral barcode. Created with [BioRender.com](https://www.biorender.com).

## Reporting Summary

Nature Portfolio 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 Portfolio 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

The single-cell or bulk DNA amplicon sequencing data collected for this manuscript was generated using a custom panel designed for the Mission Bio Tapestry to amplify segments of ten mouse genes at two exons each, the 5' and 3' lentiviral barcodes, and lentiviral GFP. Amplicon DNA libraries generated using Tapestry were sequenced on an Illumina MiSeq or HiSeq 2500 with 150 bp paired-end reads in the Fred Hutchinson Cancer Center Genomics Shared Resource. The bulk RNA sequencing libraries were prepared using a SMARTer Stranded Total RNA-Seq Kit v3 - Pico Input Mammalian (Takara Bio) and sequenced on an Illumina NovaSeq 6000 using a NovaSeq S4 flow cell with 100 bp paired end reads by MedGenome, Inc.

#### Data analysis

Data and statistical analysis were performed using Microsoft Excel Office v16.65 and Graph Pad Prism v9.4.1. Standard statistical tests were used to analyze biological data including Student's  $t$  test, Wilcoxon rank-sum test, Fisher's exact test, two-way ANOVA with post hoc Tukey's or Sidak's multiple comparisons. Results for immunohistochemical (IHC) analysis were plotted using QuPath 0.2.3. Results from flow cytometric analysis were acquired using BD FACSCanto and Sony SH800 Cell Sorter instruments. Flow data were analyzed using FlowJo 10.8.0.

Bulk RNA-seq and single-cell or bulk DNA amplicon sequencing data were analyzed on a Linux workstation using R v4.1.0.

The sequencing data reported in the manuscript was processed using the following tools or packages: RStudio v4.1.0, biomaRt package v2.24.1, the ConsensusMIBC package v1.1, pheatmap package v1.0.12, prcomp package v3.6.2, factoextra package v1.0.7, ggpubr package v0.6.0, DESeq2 package v1.38.3, Gene Set Enrichment Analysis 4.3.1, samtools v1.11, BWA-mem v0.7.17-r1188, Bowtie2 v2.4.4, Cutadapt v4.1, UMI-tools v1.0.0, python 3.7, toil-rnaseq v4.1.2, sva (ComBat-seq) v3.36.0



For functional experiments, each was repeated at least three times independently and results were expressed as mean  $\pm$  SD or mean  $\pm$  SEM.

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 Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Sequencing data pertaining to this study is available from Gene Expression Omnibus (GEO) as SuperSeries GSE229783. RNA-seq data from FHBT models is available from accession number GSE229780. Bulk DNA amplicon sequencing data from lentiviral library representation studies and from FHBT models are available from accession numbers GSE231542 and GSE229781, respectively. Single-cell DNA amplicon sequencing data related to determining the unique proviral copies per cell after lentiviral transduction across a range of MOIs is available from accession number GSE231543. Single-cell DNA amplicon sequencing data from FHBT models and enriched cells from the tumor model with prostate adenocarcinoma and focal pleomorphic giant cell carcinoma is available from accession number GSE229782.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	NA
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

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

## 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](https://www.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	In vitro experiments were performed using three independent replicates and each experiment was repeated at least three times. Sample sizes for in vitro experiments were either based on ensuring sufficient statistical power or based on the standard in the field. For in vivo studies, transduced mouse bladder or prostate cells were subcutaneously injected in 5-6 mice and those which formed tumors were collected and screened by histology. Sample size was based on prior experience with dissociated-cell tissue recombination/transplantation assays and inherent variability due to technical complexity and pilot studies of the frequency of transformation of prostate (30-40%) and bladder (70-80%) epithelial cell grafts using the methodology (see Extended Data Table 1). 5-6 grafts from each transformation study ensured the generation of tumors from at least one graft for each experiment.
Data exclusions	No data was excluded from the analyses.
Replication	Experiments have been repeated multiple times using different independent biological samples with similar experimental conditions or otherwise mentioned in the figure legends, main text or methods.
Randomization	Randomization was not applicable as there was no pre-specified comparison of interventions.
Blinding	Blinding was not applied to the study as a therapeutic intervention was not investigated.

# 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
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

## Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

### Antibodies used

Antibodies used in this study for FACS:

Human/mouse/bovine integrin alpha 6/CD49f PE-conjugated antibody (FAB13501P, R&D Systems, 1:40);  
PE/Cyanine 7 anti-mouse CD325 (Ep-CAM) antibody (118216, Biolegend, 1:40);  
CD31 (PECAM-1) monoclonal antibody (390), FITC (11-0311-82, eBioscience, 1:100);  
CD45 monoclonal antibody (30-F11), FITC (11-0451-85, eBioscience, 1:100);  
TER-119 monoclonal antibody (TER-119), FITC (11-5921-82, eBioscience, 1:100).

Antibodies used for immunohistochemistry:

Rabbit polyclonal panCK (ab9377, Abcam, 1:100);  
Rabbit monoclonal GFP antibody (clone D5.1, Cell Signaling, 1:100);  
Rabbit polyclonal p63 antibody (12143-1-AP, Proteintech, 1:200);  
Mouse monoclonal p53 antibody (clone 1C12, Cell Signaling, 1:500);  
Rabbit monoclonal HOXB13 antibody (clone D7N8O, Cell Signaling, 1:50);  
Rabbit polyclonal AR antibody (06-680, Millipore, 1:2,000);  
Rabbit monoclonal GATA3 antibody (clone D13C9, Cell Signaling, 1:200);  
Rabbit monoclonal CD44 antibody (clone E7K2Y, Cell Signaling, 1:100).

### Validation

For each antibody, the validation statement has been taken from the manufacturer's website or data sheet and detailed as follows:

Human/mouse/bovine integrin alpha 6/CD49f PE-conjugated antibody (FAB13501P, R&D Systems) - the antibody is validated to detect human, mouse, and bovine Integrin alpha 6/CD49f. Recognizes an epitope in the extracellular domain of the Integrin alpha 6 subunit.

PE/Cyanine 7 anti-mouse CD325 (Ep-CAM) antibody (clone- G8.8, 118216, BioLegend)- validated to be used by flow and is cited in multiple publication can be seen on this website: <https://www.biolegend.com/fr-lu/products/pe-cyanine7-anti-mouse-cd326-ep-cam-antibody-5303>.

CD31 (PECAM-1) monoclonal antibody (390), FITC (11-0311-82, eBioscience) - validated by staining in more than 40 publications found on this website: <https://www.thermofisher.com/antibody/product/CD31-PECAM-1-Antibody-clone-390-Monoclonal/11-0311-82>.

CD45 monoclonal antibody (30-F11), FITC (11-0451-85, eBioscience) - the 30-F11 antibody has been tested by flow cytometric analysis of mouse bone marrow cells. <https://www.thermofisher.com/antibody/product/CD45-Antibody-clone-30-F11-Monoclonal/11-0451-82>

TER-119 monoclonal antibody (TER-119), FITC (11-5921-82, eBioscience) - has been tested by flow cytometric analysis of mouse bone marrow cells. <https://www.thermofisher.com/antibody/product/TER-119-Antibody-clone-TER-119-Monoclonal/17-5921-82>

Rabbit polyclonal panCK (ab9377, Abcam, 1:100) - suitable for: IHC-P, ICC, ICC/IF, Flow Cyt, WB, IHC-Fr. <https://www.abcam.com/products/primary-antibodies/wide-spectrum-cytokeratin-antibody-ab9377.html>

Rabbit monoclonal GFP antibody (clone D5.1, Cell Signaling, 1:100) - validated for WB and IHC. <https://www.cellsignal.com/products/primary-antibodies/gfp-d5-1-rabbit-mab/2956>

Rabbit polyclonal p63 antibody (12143-1-AP, Proteintech, 1:200) - validated for WB, IHC, IP and IF. <https://www.ptglab.com/products/TP63-Antibody-12143-1-AP.htm>

Mouse monoclonal p53 antibody (clone 1C12, Cell Signaling, 1:500) - validated for WB, IHC, flow, and ChIP. <https://www.cellsignal.com/products/primary-antibodies/p53-1c12-mouse-mab/2524>

Rabbit monoclonal HOXB13 anti-body (clone D7N8O, Cell Signaling, 1:50) - Validated for WB, IP, and IHC. <https://www.cellsignal.com/products/primary-antibodies/hoxb13-d7n8o-rabbit-mab/90944>

Rabbit polyclonal AR antibody (06-680, Millipore, 1:2,000) - validated for WB and IHC. [https://www.emdmillipore.com/US/en/product/Anti-Androgen-Receptor-Antibody,MM\\_NF-06-680](https://www.emdmillipore.com/US/en/product/Anti-Androgen-Receptor-Antibody,MM_NF-06-680)

Rabbit monoclonal GATA3 antibody (clone D13C9, Cell Signaling, 1:200) - validated for WB, IHC, ChIP, IF and flow. <https://www.cellsignal.com/products/primary-antibodies/gata-3-d13c9-xp-rabbit-mab/5852>

Rabbit monoclonal CD44 antibody (clone E7K2Y, Cell Signaling, 1:100) - validated for WB and IHC. <https://www.cellsignal.com/products/primary-antibodies/cd44-e7k2y-xp-rabbit-mab/37259>

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	HEK293T (CRL-3216) were obtained from the American Type Culture Collection and were cultured in DMEM medium supplemented with 10% FBS, 100 U/mL penicillin and 100 µg/mL streptomycin, and 4 mmol/L GlutaMAX (Thermo Fisher).
Authentication	Cell line authentication was done via short tandem repeat (STR) profiling at the IDEXX BioAnalytics, 4011 Discovery Drive, Columbia, MO 65201.
Mycoplasma contamination	All cell lines routinely tested negative for Mycoplasma contamination.
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	No misidentified cell lines were used in this study.

## Animals and other research organisms

Policy information about [studies involving animals; ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	For studies using immunocompromised mice, six- to eight-week-old male NSG (NOD-SCID-IL2Ry-null) mice were obtained from The Jackson Laboratory and were 2-4 months old when used for the studies. Eight- to twelve-week-old male wild-type C57Bl/6 (C57Bl/6J) mice were obtained from The Jackson Laboratory and were 2-4 months old when used for the studies.
Wild animals	The study did not involve wild animals.
Reporting on sex	Male mice were used for all the experiments.
Field-collected samples	The study did not include field-collected samples.
Ethics oversight	All animal care and studies were performed in accordance with an approved Fred Hutchinson Cancer Center Institutional Animal Care and Use Committee protocol (PROTO000051048) and Comparative Medicine regulations.

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

## Flow Cytometry

### Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

Sample preparation	Bladder and prostates from eight- to twelve-week-old male C57BL/6 mice were dissociated into single cells and were stained with antibodies for fluorescence-activated cell sorting on a Sony SH800 Cell Sorter. Bladder urothelial and prostate epithelial cells were sorted and collected based on a Lin(-) CD49f(high) EpCAM(high) immunophenotype. For prostate polymorphic giant cells were analyzed based on forward and side scatter and further staining for nuclear DNA content with Hoechst 33342 dye.
Instrument	Flow cytometric analysis or sorting were performed using BD FACSCanto and Sony SH800 Cell Sorter instruments.
Software	FlowJo 10.8.0 software.

## Cell population abundance

Bladder and prostates from eight- to twelve-week-old male C57BL/6 mice were dissociated into single cells and were stained with antibodies for fluorescence-activated cell sorting on a Sony SH800 Cell Sorter. Bladder urothelial and prostate epithelial cells were sorted and collected based on a Lin(-) CD49f(high) EpCAM(high) immunophenotype. For prostate polymorphic giant cells were analyzed based on forward and side scatter and further staining for nuclear DNA content with Hoechst 33342 dye. Cell population abundance is shown in associated flow plots showing the gating strategy.

## Gating strategy

Stained prostate and bladder epithelial cells were plotted based on side scatter-area (SSC-A) and FITC (lineage markers) and the lineage-negative population was selected for gating by CD49f-PE and EpCAM-APC. The CD49f(high) EpCAM(high) population was sorted for experimental use.

Dissociated cells from mixed prostate adenocarcinoma an pleomorphic giant cell carcinoma tumors were plotted based on forward scatter-height (FSC-H) and forward scatter-area (FSC-A) to select single cells. Single cells were then gated based on high SSC-A and high FSC-A to isolate larger cells. After passage in organoid culture, dissociated and stained tumor cells were plotted based on SSC-A and Hoechst 33342 staining. Tumor cells were sorted based on high and low Hoechst 33342 staining.

The gating strategies for isolating prostate and bladder epithelial cells and the polymorphic giant cells isolation are provided in the Extended Data Figs. 1a, and 8c, respectively.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.