Mutations in the transcription factor FOXA1 define a unique subset of prostate cancers but the functional consequences of these mutations and whether they confer gain or loss of function is unknown1–3. Here, by annotating the landscape of FOXA1 mutations from 3,086 human prostate cancers, we define two hotspots in the forkhead domain: Wing2 (around 50% of all mutations) and the highly conserved DNA-contact residue R219 (around 5% of all mutations). Wing2 mutations are detected in adenocarcinomas at all stages, whereas R219 mutations are enriched in metastatic tumours with neuroendocrine histology. Interrogation of the biological properties of wild-type FOXA1 and fourteen FOXA1 mutants reveals gain of function in mouse prostate organoid proliferation assays. Twelve of these mutants, as well as wild-type FOXA1, promoted an exaggerated pro-luminal differentiation program, and activated a mesenchymal and neuroendocrine transcriptional program. Assay for transposase-accessible chromatin using sequencing (ATAC-seq) of wild-type FOXA1 and representative Wing2 and R219 mutants revealed marked, mutant-specific changes in open chromatin at thousands of genomic loci and exposed sites of FOXA1 binding and associated increases in gene expression. Of note, ATAC-seq peaks in cells expressing R219 mutants lacked the canonical core FOXA1-binding motifs (GTAAAC/T) but were enriched for a related, non-canonical motif (GTAAAG/A), which was preferentially activated by R219-mutant FOXA1 in reporter assays. Thus, FOXA1 mutations alter its pioneering function and perturb normal luminal epithelial differentiation programs, providing further support for the role of lineage plasticity in cancer progression.

To investigate the role of mutant and wild-type FOXA1 in prostate cancer, we examined the landscape of FOXA1 mutations across a cohort of 3,086 patients with primary or metastatic disease. The overall frequency of FOXA1 mutation in these patients is around 11% (Fig. 1a, b), 3% of which are genomic amplifications and 8.4% are somatic point mutations, with less than 1% having both types of mutations (Fig. 1b). More than 50% of FOXA1 mutations map to a specific hotspot in the Wing2 region of the forkhead (FKHD) DNA-binding domain, often as missense mutations or indels in Wing2 (mainly between H247 and E254) and R261, which is conserved in all species examined. These results are shared by analyses in the Trento–Cornell–Broad12 dataset and MSK-IMPACT 1708. Data are expressed as percentage of the total number of samples with FOXA1 mutations at a given clinical stage. The prevalence of R219 mutations compared to all other point mutations found in FOXA1 in adenocarcinoma versus NEPC. Cases pooled from the Trento–Cornell–Broad12 dataset and MSK-IMPACT 1708. ***P = 0.0059, Fisher’s exact test, two-sided.
Distinct morphologies for different classes of alterations. a, FOXA1 luciferase reporter assay with results normalized to level of FOXA1(WT) activity. Colours indicate the position of the altered amino acid within the FKHD DNA-binding domain depicted in Fig. 1a. Grey indicates activity. Colours indicate the position of the altered amino acid within the luciferase reporter assay with results normalized to level of FOXA1(WT) (no EGF) and in restrictive medium conditions (dashed lines, no EGF), calculated using unpaired, two-tailed Student's t-test. b, Overexpression of wild-type (+WT) or various FOXA1 mutants promotes growth ten days after seeding in medium lacking EGF. c, Overexpression of wild-type (+WT) or various FOXA1 mutants promotes growth ten days after seeding in medium lacking EGF. d, Quantification of lumen-containing organoids for each line in the FOXA1 allelic series. All values of P are relative to empty vector control, calculated using unpaired, two-tailed Student's t-test. e, Histology and immunohistochemistry of organoid lines overexpressing variants of FOXA1 via the doxycycline-inducible pCW vector ten days after seeding.

Images from a single biological experiment. H&E, haematoxylin and eosin staining. Scale bars: top, 200 μm; other rows, 100 μm. f, Summary of GSEA comparing Foxa1 wild-type or mutant organoids to empty vector control for a basal low (luminal) gene set, the hallmark EMT gene set and a gene set of the top 100 genes induced following ERF knockdown (using short hairpin RNA directed against ERF, shERF) in organoids. Data are from RNA-sequencing (RNA-seq) analysis of three biological replicates for each organoid line. Only comparisons with an FDR of <0.25 are shown with the corresponding normalized enrichment score. Gene sets with a positive normalized enrichment score are enriched in organoids that have either Foxa1 wild-type or mutant alleles. Data in a–d are mean ± s.d. Numbers of biological replicates (indicated as dots) and specific P values are presented in the source data. *P < 0.05, **P < 0.01, NS, not significant. All P values are relative to overexpression of wild-type FOXA1 unless otherwise noted; unpaired, two-tailed Student's t-test.

F266), some of which are predicted to be sites of direct DNA contact10 (Fig. 1a, Extended Data Fig. 1). Another mutational hotspot is at R219, a DNA-contact site in α-helix 3, which is a highly conserved fold of the FKHD domain that sits in the major groove of target DNA (Extended Data Fig. 1). Finally, 20% of FOXA1 mutations encode truncations just downstream of the FKHD DNA-binding domain, resulting in the loss of the C-terminal transactivating domain. Annotation of all FOXA1 mutations in the MSK-IMPACT 504 cohort11 revealed that Wing2-hotspot mutations—the most common subclass—are found across all disease stages, but are more prevalent in primary locoregional cases (Fig. 1c). There are only four cases of R219 mutation in FOXA1 in this cohort but, notably, two of these had castration-resistant disease. We therefore expanded the analysis to 1,822 patients by including a larger cohort from MSK-IMPACT and a published cohort from Weill Cornell12, which is enriched for neuroendocrine prostate cancer (NEPC), and observed significant enrichment (P < 0.006) of R219 mutations versus other FOXA1 mutations in NEPC (3 out of 4) versus adenocarcinoma (8 out of 84) (Fig. 1d).

We next investigated whether FOXA1 mutation in patients is associated with clinical outcome. In the absence of appropriate longitudinal
FOXA1 expression constricts the AR cistrome and promotes AR-independent growth programs. a, Left, AR ChIP–seq in organoids overexpressing wild-type or mutant FOXA1 compared to control shows significant changes in the AR cistrome in response to FOXA1 expression. Right, FOXA1 ChIP–seq showing FOXA1 binding at same loci. ChIP–seq data are from two biological replicates. Statistical analysis of peaks is shown in Extended Data Fig. 6. b, Overexpression of FOXA1 promotes growth in prostate organoids in the setting of significantly reduced AR (CRISPR-mediated depletion in a bulk population), in both standard medium conditions (left) and in the absence of EGF (right). Two independent experiments result in the same growth trends for biological replicates 1 and 2.

data, we generated an RNA signature using mutant FOXA1 status of The Cancer Genome Atlas (TCGA) samples to interrogate the Decipher GRID cohort of 1,626 primary prostate cancer patients and found that tumours predicted to be FOXA1 mutant were significantly associated with higher Gleason Scores, shorter time to biochemical recurrence and more rapid progression to metastatic disease than unaltered cases (Extended Data Fig. 1b, c). Together with recent evidence, these data suggest that patients with FOXA1 mutations have a less favourable prognosis.

To characterize a large panel of the most recurrent alterations seen in prostate cancer, including truncating mutations, we generated a FOXA1 reporter construct (Extended Data Fig. 2), and found that all Wing2 mutations, D226N (a mutation in spatial proximity to Wing2 in the protein) and the truncation mutant G275X result in increased transcriptional activity (around twofold) compared to the wild type, whereas mutations at R219 (R219S and R219C) cause impaired activity (around 50% of wild-type activity) (Fig. 2a). To investigate the consequences of FOXA1 mutations on the growth of prostate cells, we used primary mouse prostate organoid culture (previously used to model tumour initiation) by introducing a series of wild-type or mutant mouse Foxa1 alleles using doxycycline-inducible lentiviral constructs (Extended Data Fig. 3a–c). Increased expression of wild-type (WT) FOXA1 resulted in a 2–3-fold increase in growth compared to vector control. This relative difference was substantially greater (around 50-fold) after removal of epidermal growth factor (EGF), a critical growth factor for normal organoid proliferation (Fig. 2b). In this setting, nearly all mutants tested showed an increase in growth relative to overexpression of FOXA1(WT), including the two α-helix 3 mutants (R219S and R219C) that had reduced reporter activity, as well as the truncation mutant G275X (Fig. 2c). All 14 mutants promoted growth relative to the empty vector control.

We next examined the histological features of the resulting organoids. We observed that increased expression of FOXA1(WT), FOXA1(D226N) or the Wing2-hotspot mutations all promote exaggerated lumen formation and size (Fig. 2e, Extended Data Fig. 3d). By contrast, organoids expressing FOXA1(R219S), and to a lesser extent those expressing FOXA1(R219C), were unable to form measurable lumens, and the bilayer orientation of basal (p63+) and luminal (androgen receptor-positive (AR+)) cell layers appeared disrupted (Fig. 2e, Extended Data Fig. 3e). This phenotype resembles that of FOXA1-deficient organoids generated using CRISPR–Cas9 (Extended Data Fig. 4a–c), consistent with mouse models. We also repeated the overexpression studies in endogenous-Foxa1-deleted organoids using CRISPR-resistant cDNAs encoding two pro-luminal FOXA1 mutants (ΔF254/E255 and D226N) and found that the pro-luminal phenotype was unchanged (Extended Data Fig. 4d–g). Findings from RNA sequencing were consistent with these histologies. Mutants conferring a pro-luminal phenotype showed similarity to ETS-mutant luminal organoids by gene-set enrichment analysis (GSEA), with the notable exception of FOXA1(R219S), which instead showed enrichment of an epithelial–mesenchymal-transition (EMT) program and a repression of the ETS-mutant gene set (Fig. 2f), consistent with
**FOXA1 mutations cause marked shifts in the chromatin landscape.**

**a.** Number of significant peaks open or closed (log2(fold change) > 2 for open peaks, log2(fold change) < -2 for closed peaks) after doxycycline treatment of organoids transfected with pCW-FOXA1 for expression of wild-type or mutant FOXA1 relative to empty vector. Right, includes counts for organoids following CRISPR-mediated deletion of FOXA1. Data are from three biological replicates, mean ± s.d. Unpaired, two-tailed Student’s t-test. No significant difference between activity of wild type and R219S on the GTAAAR reporter (P = 0.2314). FOXA1(ΔF254/E255) has significantly less activity on GTAAAR than either FOXA1(WT) (P = 0.0059) or FOXA1(R219S) (P = 0.0033).

As FOXA1 is a cofactor for AR and FOXA1 mutant cases in the TCGA cohort have higher AR scores than either normal samples or other subtypes,

its distinct morphology. We also examined the activity of FOXA1 in an in vivo setting, and observed increased proliferation across all lines, an increase in subcutaneous tumour size in two of four lines (FOXA1(WT) and FOXA1(G275X)), and an increased prevalence of invasive, intraductal basal disease (defined by the loss of AR expression) in tumours derived from FOXA1(R219S) organoids transduced with single-guide RNA (sgRNA) directed against PTEN—consistent with FOXA1(R219S) histology in vitro (Extended Data Fig. 4h–j).
This suggests that FOXA1 may replace AR function at these sites, supported by retention of the increased growth advantage conferred by FOXA1 despite CRISPR-mediated deletion of Ar (Fig. 3b, Extended Data Fig. 5b). To reconcile the high AR scores seen in TCGA with this AR-independent growth program, we examined expression levels of the mouse orthologues of the human AR gene signature20 and found that the majority are induced by FOXA1 (Extended Data Fig. 5c). Thus, while the number of AR-binding sites is substantially reduced, a core set of AR-target genes are maintained in the setting of increased FOXA1 activity. We also investigated whether transcriptomic changes observed in the Foxa1 mutant mouse organoids were similar to those observed in FOXA1 mutant human tumours. Remarkably, the human orthologues of differentially expressed genes in FOXA1(ΔF254/E255) mouse organoids were sufficient to cluster FOXA1 mutant tumours within the TCGA cohort (P = 2.1 × 10⁻⁸, Extended Data Fig. 5d).

Given the role of FOXA1 as a pioneer transcription factor, we conducted a genome-wide analysis of changes in open and closed chromatin using ATAC-seq. Expression of FOXA1(WT) led to an increase in open chromatin after five days (more than 1,000 open peaks with significant change in accessibility, false discovery rate (FDR) < 0.05, log fold change of 2 in peak read coverage compared to control) whereas deletion of FoxA1 led to the opposite, with the closing of around 1,000 peaks. Organoids expressing FOXA1(ΔF254/E255) and FOXA1(R219S) also had increased peak numbers, but these changes occurred substantially faster (in one day) and involved many more peaks (Fig. 4a), consistent with altered pioneering activity.

Unsupervised clustering analysis identified distinct sets of peaks for FOXA1(ΔF254/E255) and FOXA1(R219S) (Fig. 4b). Cluster 0 is largely defined by marked peak changes observed with both FOXA1(WT) and FOXA1(ΔF254/E255), demonstrating that overexpression of wild-type FOXA1 opens new regions of chromatin compared with controls; this effect is amplified in cells expressing FOXA1(ΔF254/E255). By contrast, organoids expressing FOXA1(R219S) gain thousands of distinct peaks (defined by clusters 3 and 5) without changes in cluster 0. ChiP-seq reveals that FOXA1 protein binds at these same ATAC-seq loci (Fig. 4c, Extended Data Fig. 6a–d) and cumulative distributive function plots confirm that there are mutation-specific changes in expression of the genes that map to these newly open chromatin peaks (Extended Data Fig. 6e–h).

Motif analysis revealed enrichment of FOXA1-binding motifs in clusters 0 and 1 (FOXA1(WT) and FOXA1(ΔF254/E255)) (Extended Data Fig. 7a) but not in clusters 3 and 5 (FOXA1(R219S)), despite evidence of FOXA1(R219S)–DNA binding and associated gene-expression changes. However, de novo motif analysis of cluster 3 peaks identified a motif with similarities to the core GATAA(A/G)(T) FOXA1 binding motif but with substitution of (G/A) for (C/T) at position 6 (Extended Data Fig. 7b). This impression was confirmed by selective enrichment of the (G/A) motif in clusters 3 and 5 versus the (C/T) motif in clusters 0 and 1 (Fig. 4d). To determine whether this motif is functional, we repeated the reporter assays described in Fig. 2a with FOXA1(R219S) and found that FOXA1(R219S) preferentially activates a DNA template modified to reflect the (G/A) bias at position 6, whereas FOXA1(WT) and FOXA1(ΔF254/E255) exhibit substantially higher activity on the canonical (C/T) sequence (Fig. 4e, Extended Data Fig. 7c–e), suggesting a mechanism by which FOXA1(R219S) selectively targets novel genomic loci. Finally, two motifs recently associated with FOXA1 dimers—termed convergent and divergent21—were relatively enriched in cluster 0 versus cluster 1, potentially explaining the novel pioneering activity of FOXA1(ΔF254/E255) (Fig. 4d).

Collectively, our analysis of mutant FOXA1 alleles in prostate cancer reveals unanticipated and diverse consequences for the pioneering function of FOXA1. Wing2 mutants display a gain in pioneering activity that is substantially greater than that observed by overexpression of comparable levels of FOXA1(WT), but both alterations affect nearly identical regions of the genome (cluster 0) that are distinguishable from endogenous FOXA1 sites (cluster 1) on the basis of enrichment of FOXA1 dimer motifs. We postulate that the changes in gene expression associated with these novel open regions contribute to oncogenesis. By contrast, FOXA1 R219 mutants display pioneering function over distinct regions of the genome (clusters 3 and 5) enriched for a variant FOXA1-binding motif that, on the basis of reporter assays, is permissive for the binding of FOXA1 R219 mutants despite mutation of the α-helix 3 consensus DNA-binding residue. Further investigation of the relative DNA-binding affinities of these mutants for the different motifs and the potential role of the Wing2 domain in this retained DNA binding (based on known DNA contacts in the minor groove) is warranted. In both classes of mutations, the biological consequence is lineage plasticity for pro-luminal versus anti-luminal programs.

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/s41586-019-1318-9.

Received: 10 June 2018; Accepted: 22 May 2019;
Published online 26 June 2019.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019
METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized. For all assays except lumen area quantification and histological assessment of xenograft tumors, the investigators were not blinded to allocation during experiments and outcome assessment.

Pan-prostate cancer mutation analysis. The 12 cohorts used for analysis (total of 3,086 samples) included published datasets as well as unpublished data from the MSK-IMPACT 1708 cohort (frozen 25 May 2018), across all stages of prostate cancer (see Supplementary Table 1). Samples were compiled and duplicate samples were pruned to generate a master list of 3,086 prostate cancer cases, which were then stratified on the basis of their FOXA1 alteration status and the class of mutation in the samples. The Wing2 hotspot includes cases with mutations or indels between H247 and E269. Truncations after the FKHD domain were defined as any frameshift alteration distal to residue E269. Any mutation that did not specifically fall into one of the distinct classes was called 'other'. Sample analysis was performed in part using the CBioPortal for Cancer Genomics22,23.

3D modelling. Three-dimensional representation of the FKHD domain of FOXA1 complexed with DNA was generated using PyMOL (PDB: 1VTN).

Constructs. To create pCW-Flag-2A-dsRED (pCW-EV), sequences for p2A and DsRED were cloned in the pCW-Cas9 plasmid (Addgene Plasmid #50661) using in-fusion cloning (Takara Bio). To generate pCW-Flag-Foxa1-1-2a-dsRED (pCW-Foxa1), mouse Foxa1 cDNA was cloned into pCW-Flag-2A-dsRED using in-fusion cloning (Takara Bio). All primers and sequences are listed in Supplementary Table 2. To generate the sgRNA vector CRISPR-Zeo, GFP from plPKp sgRNA. EFS.GFP (a gift from B. Ebert, Addgene plasmid no. 57822) was excised with BamHI and MluI. The neomycin-resistance gene was excised from lentil-sgRNA (MS2), zeo backbone (a gift from E. Zhang, Addgene plasmid no. 61427) using BsrGI and EcoRI. ZeoR was ligated into the plPKp5G-sgRNA.EFS backbone in a four-way ligation using BamHI–BsrGI and EcoRI–MluI adaptors. To create the LVX-Ubc-EGFP-Luc2_Hygro construct for visualization of injected cells by live imaging or immunohistochemistry, we first generated the plasmid LVX-Ubc-EGFP-Luc2_puro as follows: 0.72 kb EGPFl cDNA from pQCXIP-EGFP was cloned into the BamHI and NotI sites of pLVX-TRE3G-IREs (Clontech, 631362) via a EcoRI–NotI cloning adaptor to make pLVX-TRE3G-EGFP-IREs. The TRE3G promoter was then removed with an XhoI and BamHI digestion, and replaced with the 1.26-kb Ubc promoter obtained from Duet011 (Addgene) with a PacI and BamHI digest and using a XhoI–PacI cloning adaptor to make pLVX-Ubc-EGFP-IREs. pLVX-Ubc-EGFP-Luc2 was then constructed by cloning the 1.7-kb Luc2 cDNA derived from pGL4.10(luc2) (Promega) with a HindIII and XbaI digest into the MluI and EcoRI sites of pLVX-Ubc-EGFP-IREs via MluI–HindIII and XbaI–EcoRI cloning adaptors. The pyromycin cassette was replaced with the hygromycin cassette to generate LVX-Ubc-EGFP-Luc2_Hygro.

Generation of FOXA1 mutant cDNA. Site directed mutagenesis was carried out on pCW-FOXA1 via CRISPR–Cas9 to induce point mutations in the cDNA using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent), according to the manufacturer’s protocol. Primers were designed using Agilent’s QuikChange Primer Design tool (https://www.genomics.agilent.com/primerDesignProgram.jsp). To prevent CRISPR–Cas9 targeting by sgFOXA1_1 sgRNA mutagenesis was carried out on the 3′ end using the Agilent’s QuikChange II XL Site-Directed Mutagenesis Kit (Agilent), according to the manufacturer’s protocol. Primers were designed using Agilent’s QuikChange Primer Design tool (https://www.genomics.agilent.com/primerDesignProgram.jsp). To prevent CRISPR–Cas9 targeting by sgFOXA1_1 sgRNA mutagenesis was carried out on the 3′ end using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent), according to the manufacturer’s protocol.

Organoids were transfected with varying levels of FOXA1 introduced into the system, ratios of pCMV6-mFOXA1.mycDDK and pCW-EV constructs were altered, keeping the total amount of DNA transfected into each well constant. In evaluating the relative response ratios between FOXA1(WT) and various mutants, one concentration of cDNA (170 ng per well) was used and relative response ratios reflect activity of given variant on the reporter. Luminescence measurements were taken 24 h after transfection. All results are means and standard deviations from experiments performed in biological triplicates (as a minimum, n ranging from 3 to 7) (see source data for Fig 2 for details), and Firefly luciferase activity of individual wells was normalized against Renilla luciferase activity.

Organoid lines. All parental organoid lines were established in our laboratory as previously described24. The blue red organoids (BRO) line was established from mice harbouring red fluorescent protein (RFP) driven by a composite human kera- tin 18 promoter and cerulean fluorescent protein (CFP) driven by a bovine keratin 5 promoter25. BROs were transduced with lentilCrispVR2 carrying either sgNT or sgFoxa1_1 and selected using puromycin. BRO lines were maintained in standard mouse organoid media conditions21. K14-1 organoids were derived from mice harbouring an actin–GFP fusion protein driven by a human keratin 14 promoter26. K14-1 organoids were transduced with the allelic series of pCW-Foxa1 wild-type or mutant constructs, as well as pCW-EV as a control. Bulk cells were selected using puromycin. K14-1 organoids were maintained in standard mouse organoid media conditions21, with 2.5 ng/ml EGFP supplementation. For rescue experiments of either Foxa1 or Ar deletion, K14-1 organoids carrying pCW-Foxa1 constructs were subsequently transduced with lentiCas9-Blast, bulk selected with blastcidin, and then transduced with either CRISPR-Zeo-sgFoxa1_1 or sgGFp, or sgAr and bulk selected with puromycin. Organoids were derived from a homozygous Rosa26 Lox-stop-Lox Cas9 mouse (C67BL/J]6 background, Jackson Laboratory 026175) and transduced with adenoCre-GFP in vitro for expression of Cas9. These cells were then transduced with lentilGuide-Puro-sgPten and bulk selected with puromycin, transduced with LVX-Ubc-EGFP-Luc2_Hygro and bulk selected with hygromycin, transduced with the allelic series of pCW-Foxa1 wild-type or mutant constructs or pCW-EV, as well as pCW-EV as a control, and sorted for dsRED expression to enrich for transduced cells.

Organoid culture. Mouse organoids were sorted, cultured in 3D and transduced with lentiviruses as described previously24. Organoids infected with pCW-EV, pCW-Foxa1, or LentivCrispV2 constructs were selected with 2 μg/ml puromycin for 5 days, 3–4 days after transduction, while those infected with CRISPR-Zeo were selected with 30 μg/ml for 7 days, 3–4 days after transduction. Transduction with Lenti-Cas9-Blast was followed by five days of selection in 10 μg/ml blastcidin. Preparation of 3D organoids for histology was carried out as previously described24. Haematoxylin and eosin staining and immunohistochemistry was carried out by the MSKCC Molecular Cytology Core. Cells were confirmed to be free of mycoplasma using the Lonza MycoAlert Mycoplasma Detection Kit (LT07-318). pCW-FOXA1, or LentivCrispV2 constructs were selected with 2 mg/ml puromycin for 5 days, 3–4 days after transduction, while those infected with CRISPR-Zeo were selected with 30 μg/ml for 7 days, 3–4 days after transduction. Transduction with Lenti-Cas9-Blast was followed by five days of selection in 10 μg/ml blastcidin. Preparations of 3D organoids for histology was carried out as previously described24. Haematoxylin and eosin staining and immunohistochemistry was carried out by the MSKCC Molecular Cytology Core. Cells were confirmed to be free of mycoplasma using the Lonza MycoAlert Mycoplasma Detection Kit (LT07-318). pCW-FOXA1, or LentivCrispV2 constructs were selected with 2 μg/ml puromycin for 5 days, 3–4 days after transduction, while those infected with CRISPR-Zeo were selected with 30 μg/ml for 7 days, 3–4 days after transduction. Transduction with Lenti-Cas9-Blast was followed by five days of selection in 10 μg/ml blastcidin. Preparation of 3D organoids for histology was carried out as previously described24. Haematoxylin and eosin staining and immunohistochemistry was carried out by the MSKCC Molecular Cytology Core. Cells were confirmed to be free of mycoplasma using the Lonza MycoAlert Mycoplasma Detection Kit (LT07-318).
supplemented for the first feeding at 10 μM. After 10 days, the area of each visible lumen was measured using light microscopy and Nikon NIS Elements software. In a typical experiment, ~30–50 organoids were measured by an observer blinded to organoid genotypes.

**Western blot.** Membranes were probed with antibodies directed against AR (1:1,000, ER179(2), Abcam), FOXA1 (1:1,000, Ab2, Sigma), cyclinophilin B (1:1,000, EPR12703(B), Abcam), Flag (1:1,000, M2, Sigma) or PTEN (1:1,000, D4.3, Cell Signaling). Membranes were processed with secondary HRP-conjugated antibodies and chemiluminescent detection.

**Immunohistochemistry.** Organoids and tumours were processed and stained as described previously.17 The following antibodies were used for staining on mouse organoids and organoid-derived xenografts: HNF-3-α/FOXA1 antibody (5 μg/ml, 3B3NB, Novus Biologicals), AR (1:1,000, N-20, Santa Cruz), p63 (1:800, 4A4, Ventana). Staining was visualized with BrightVision (Dako), Ki67 (Abcam ab15580 at 1 μg/ml).

**In vivo experiments.** All in vivo xenograft experiments were performed by subcutaneous injection of 2 × 10⁴ dissociated organoid cells (Rosa26-Cas9-sgPTEN-luc2-pCw-FOXA1 or ERG) resuspended in 100 μl of 50% matrigel (BD Biosciences) and 50% growth medium into the flanks of five 8- to 12-week-old male NOD. Cg–PrkdcscidIl2rgtm1Wjl/Sg mice (005557, The Jackson Laboratory) to yield ten tumours per group. As soon as palpable, tumour volume was measured weekly using the tumour-measuring system Peira TM900 (Peira). Tumours were then collected at given time points for histology using 4% paraformaldehyde. Histological assessment was carried out by an observer blinded to tumour genotypes. All animal experiments were performed in compliance with the guidelines of the Research Animal and Biosafety Committee of the Memorial Sloan Kettering Cancer Center. In accordance with our IACUC and our approved protocol, none of the mice exceeded the maximum tumour burden allowed (total for both sides) of 2,000 mm³.

**RNA isolation and sequencing.** RNA was extracted from organoids using an RNeasy Kit (Qiagen). Freshly sorted dsRED+ cells were seeded in triplicate per infected construct at the start of the assay, and moving forward, replicates were processed independently, collected at the appropriate time points. Library preparation and sequencing were performed by the New York Genome Center, where RNA-seq libraries were prepared using the TruSeq Stranded mRNA Library Preparation Kit (Illumina) in accordance with the manufacturer’s instructions. In brief, 100 ng of total RNA was used for purification and fragmentation of mRNA. Purified mRNA underwent first and second strand cDNA synthesis. cDNA was then adenylated, ligated to Illumina sequencing adapters, and amplified by PCR (using 10 cycles). Final libraries were evaluated using fluorescent-based assays including PicoGreen (Life Technologies) or Qubit Fluorometer (Invitrogen) and Fragment Analyzer (Advanced Analytics) or BioAnalyzer (Agilent 2100), and were sequenced on an Illumina HiSeq2500 sequencer (v4 chemistry, v2 chemistry for Rapid Run) using 2 × 50-bp reads. Reads were aligned to the mm10 mouse reference genome using STARaligner.30 Read counts for infected organoids were calculated using featureCounts (v.1.4.3) and transcripts were quantified using DESeq2 (http://bioconductor.org/packages/release/bioc/html/DESeq2.html) based on the gene read count data. Multiple-hypothesis testing was considered using Benjamin–Hochberg correction. The statistical significance of the overlap between two groups of genes was tested using Fisher’s exact test. GSEA was performed using the JAVA program (http://www.broadinstitute.org/gsea) and run in pre-ranked mode to identify enriched signatures. The GSEA plot, normalized enrichment score and FDR and q values were derived from GSEA output. The following gene sets were used: Hallmark gene sets, Neuroendocrine high, Basal low, and shERF up.

**Prostate cancer tumour samples and microarray data.** A total of 1,959 radical prostatectomy tumour tissues, amplified and hybridized to Human Exon 1.0 ST microarrays (Thermo Fisher).

**FOX1A1 mutant transcriptional signature.** By following the similar strategy as previously reported for SPOP mutants,17 we developed the FOXA1A1 mutant transcriptional signature that includes 67 genes differentially expressed between FOXA1A1 mutant and wild-type samples from TCGA prostate cancer RNA-seq data. The low-expressed genes (mean RNA-seq expression value) among mutant samples and high-expressed genes among wild-type samples (scATAC – I 10 X 20), and paired reads that mapped to organoid genotypes were filtered before the analysis. Specifically, we identified significantly differentially expressed genes by comparing FOXA1A1 mutants within FKHD DNA-binding domain and wild-type cases as determined from DNA mutational analyses among TCGA samples lacking ET5 family gene fusions (ET5, ETV1, ETV4 and FLII), with Wilcoxon rank-sum test and controlled for false discovery using Benjamini–Hochberg adjustment (FDR < 0.05).

**SCaPT development based on FOXA1A1 mutant transcriptional signature and SVM model.** To predict tumours in the FOXA1A1 mutant subclass in the absence of DNA sequencing data (that is, microarray datasets), we developed the subclass predictor based on transcriptional data (SCaPT model) based on the support vector machine (SVM) model. Given a set of training data marked with two categories, SVM builds a model that assigns testing data into one category or the other, making it a non-probabilistic binary linear classifier. In our SCaPT model, the training data were defined as the transcriptional scores of FOXA1A1 mutant signature from TCGA cohort. The testing data would be the transcriptional z scores from RNA-seq or microarray expression data of FOXA1A1 mutant signature.

**Prostate cancer molecular subclass prediction by decision tree.** In each individual study of retrospective and prospective GRID cohorts, the FOXA1A1 mutant subclass was first predicted using the SCaPT model. Next, using a decision tree and previously developed microarray-based classifiers for the ERG+ and ET5+ subtypes, we classified the remaining cases in each cohort. Some cases with both predicted FOXA1A1 mutant and ERG+ ET5+ status were classified as conflict subclass, and the rest without FOXA1A1 mutant calling and outlier expression were considered as ‘other’ subclass.

**Statistical analysis of human data.** Statistical analyses were performed in R v.3.4.0 (R Foundation). All statistical tests were two-sided with a significance level of P < 0.05. Univariate logistic regression analyses were performed on the combined cohort to test the statistical association between FOXA1A1 mutant status and clinical variables, including age, race, pre-operative prostate-specific antigen (PSA), Gleason score, lymph node invasion (LNI), surgical margin status (SMS), extracapsular extension (ECE) and seminal vesicle invasion (SVI). We evaluated the associations between FOXA1A1 mutant status and patient outcomes including biochemical recurrence (BCR), metastasis (MET) and prostate cancer specific mortality (PCSM) on the basis of Kaplan–Meier analysis.

**ATAC-seq.** Freshly sorted cells carrying pCw constructs (dsRED+) were seeded in triplicate per infected construct at the start of the assay, and moving forward, replicates were processed independently, collected at the appropriate time points. CRISPR cells were transfected with either the control guide (sgNT), guide 14 for FOXA1 (sgFOXA1_1) or guide 15 for FOXA1 (sgFOXA1_2). At time of collection, cells were trypsinized, and 50,000 cells (counted by using trypsin blue exclusion) were processed for ATAC-seq as follows. After a wash step in cold cell wash buffer (CWB; 10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂), outer membranes were disrupted in lysis buffer (CWB + 0.1% NP40) for 2 min on ice. The lysis reaction was stopped with the addition of 1 ml CWB buffer (CWB + 0.1% NP40) for 2 min on ice. After a centrifugation step at 1,500g for 10 min, pelleted nuclei were kept for the next step. In a 90-μl final volume, fragmentation was performed for 30 min at 37°C, using the Nextera DNA library prep kit (illumina FC-121-1030). After addition of SDS to 0.2% final concentration, DNA was purified on AMPure XP beads (Beckman Coulter A63881) using a 2:1 ratio (v/v) of beadstranded DNA. Freshly eluted DNA was barcoded and amplified in 110-μl PCR volume (NEB Next Q5 Hot Start HiFi PCR, M0543L) to generate a library with the following PCR program: 65°C, 5 min; 98°C, 30 s; 11 cycles of (98°C, 10 s ramping to 65°C, 30 s), 4°C hold. Quality control of the libraries was performed with a Bioanalyzer 2200 (Agilent Technologies, D1000 screntapes and reagents, 5067-5582) to assess size range and fragment size of the DNA libraries (that is, microarray datasets), we developed the subclass predictor based on transcriptional data (SCaPT model) based on the support vector machine (SVM) model. Given a set of training data marked with two categories, SVM builds a model that assigns testing data into one category or the other, making it a non-probabilistic binary linear classifier. In our SCaPT model, the training data were defined as the transcriptional scores of FOXA1A1 mutant signature from TCGA cohort. The testing data would be the transcriptional z scores from RNA-seq or microarray expression data of FOXA1A1 mutant signature.

**Prostate cancer molecular subclass prediction by decision tree.** In each individual study of retrospective and prospective GRID cohorts, the FOXA1A1 mutant subclass was first predicted using the SCaPT model. Next, using a decision tree and previously developed microarray-based classifiers for the ERG+ and ET5+ subtypes, we classified the remaining cases in each cohort. Some cases with both predicted FOXA1A1 mutant and ERG+ ET5+ status were classified as conflict subclass, and the rest without FOXA1A1 mutant calling and outlier expression were considered as ‘other’ subclass.

**Statistical analysis of human data.** Statistical analyses were performed in R v.3.4.0 (R Foundation). All statistical tests were two-sided with a significance level of P < 0.05. Univariate logistic regression analyses were performed on the combined cohort to test the statistical association between FOXA1A1 mutant status and clinical variables, including age, race, pre-operative prostate-specific antigen (PSA), Gleason score, lymph node invasion (LNI), surgical margin status (SMS), extracapsular extension (ECE) and seminal vesicle invasion (SVI). We evaluated the associations between FOXA1A1 mutant status and patient outcomes including biochemical recurrence (BCR), metastasis (MET) and prostate cancer specific mortality (PCSM) on the basis of Kaplan–Meier analysis.
Clusters were defined by cutting the hierarchical clustering at the first 8 bifurcations transformation to the full peak atlas, selecting the differentially accessible peaks, annotations (Extended Data Fig. 9). These analyses produced a set of ~20,500 differentially accessible peaks (described above). Each ATAC-seq peak in the atlas was scanned for 718 transcription factor motifs in the *Mus musculus* CIS-DB database using FIMO 4.5 MEME suite, using the default P value cut-off of 1 × 10⁻⁴. The background sequence distribution for motif detection was based on nucleotide frequencies in the full set of 20,523 differentially accessible peaks (A = T = 0.2711, C = G = 0.2289). Of the 718 motifs in the database, 713 had a match within at least one peak among the differentially accessible peaks.

**FIMO motif search.** We restricted the analysis to 298 transcription factors that had a median ATAC-seq expression across biological replicates of above 5 reads per kilo base of transcription per million mapped reads in at least one organoid line or time point. In addition, CTCF and CTCFL, DNA-binding proteins associated with 3D chromatin structure, were excluded. To rank the level of enrichment of transcription factor motifs in each cluster relative to the background, the number of peaks containing each motif was calculated for each cluster and for the full set of differentially accessible peaks. Enrichment–depletion scores for each motif in a cluster were reported as binomial Z-scores relative to the background of motif occurrences in the set of differential ATAC-seq peaks. Namely, if \( P \) represents the probability that a peak in the background set contains an occurrence of the motif, then the binomial Z-score for a cluster of size \( N \) with \( C \) peaks containing the motif is

\[
Z = \frac{C - N \cdot P}{\sqrt{N \cdot P \cdot (1 - P)}}
\]

While these Z-scores do not incorporate a correction for multiple hypotheses, in the practice the top-ranked motifs have such strong enrichments that they would still be highly significant after correction.

**Non-canonical FOXA1 motif analysis.** To examine enrichment–depletion of non-canonical FOXA1 motifs, we considered four additional motifs. First, we examined previously reported convergent and divergent FOXA1 dimer motifs. For each of these motifs we computed the permutation \( Z \) score of the core GTAAAC/T pattern with either and equal probability of C/T (similar to canonical) or an equal probability of A/G (non-canonical). We used FIMO to search for hits of these motifs across differential ATAC-seq peaks and reported enrichment–depletion within clusters as binomial Z-scores as before.

**ChiP-seq.** Freshly sorted cells carrying pCW constructs (dsRED⁺) were seeded in duplicate per infected construct at the start of the assay, and moving forward, replicates were processed independently, collected following five days of doxycycline treatment. At time of collection, cells were trypsinized and 70,000 cells (counted by using trypan blue exclusion) were processed for ChiP-seq as follows. Cells were fixed with formaldehyde (1%) and the reaction was quenched with glycine 1.25 M and Tris 1 M pH 8. Fixed cells were lysed with SDS lysis solution containing protease inhibitors. Unpaired, solid-phase read samples were sonicated, precipitated with antibodies (HHN-3 alpha/FoxA1 antibody (3B3NB) (Novus Biologicals), AR (ER179(2), Abcam) and protein A/G bead complex. The chromatin and immune complex were sequentially washed with a low-salt solution, high-salt solution, LiCl solution and Tris-NaCl solution. Chromatin was eluted from the complex with a solution containing 1% of SDS and 0.1 M NaHCO₃. Cross-linking between DNA and protein was reversed by adding NaCl solution and incubating at 65 °C overnight. Libraries were made using NEBNext Ultra II DNA library prep kit for Illumina (NEB E7645L). Quality control was performed with Bioanalyzer 2200 (Agilent Technologies, D10000 sniffmap and reagents, 5067-5582) to assess size range of amplified DNA fragments, and with Quant-IT PicoGreen dsDNA Assay Kit (Thermo Fisher P11496) to quantify the DNA fragments generated. ChiP libraries were then pooled at equimolar concentration and were sequenced multiplexed on the Illumina HiSeq with 50-bp paired-end sequencing.

**Biinformatic analysis of ChiP-seq.** Raw reads were first trimmed with Trimmomatic v0.35 (bcl2fastq v.2.2.6, options:--local mm--no-mixed--no-discardant) using mm10 genome. After alignment, PCR duplicates were removed with Picard.
tools (http://broadinstitute.github.io/picard/) (MarkDuplicates v2.9.0) and peaks were called individually for each replicate with MACS2\(^{24}\) (v.2.1.0.20151222, --options: keep-dup 1 -g mm -p 0.05). These called peaks between replicates were then used with IDR\(^{40}\) (v.2.0.2) framework to identify reproducible peaks. Depthool\(^{23}\) was used for visualization and HOMER (v.4.10.3) was used for discovering de novo motifs.

**ChiP-seq normalization and analysis.** To analyse ChiP-seq signal for AR and FOXA1 in each organoid line relative to ATAC-seq clusters, we normalized ChiP-seq data across experiments based on background signal, namely by defining flanking regions of reproducible peaks and using DESeq scaling factors relative to these regions for library size normalization. To compare AR or FOXA1 binding between a pair of organoid lines with respect to an ATAC-seq cluster, we compared the corresponding distributions of normalized ChiP-seq signal over peaks in the cluster by a one-sided Wilcoxon rank-sum test.

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

**Data availability** The described RNA-seq, ATAC-seq and ChiP-seq data have been deposited in the Gene Expression Omnibus under the following accession numbers: GSE128667 (all data), GSE128421 (ATAC-seq sub-series), GSE128666 (RNA-seq sub-series) and GSE128867 (ChiP-seq sub-series). Patient predicted FOXA1 mutant status and outcome data from Decipher GRID are available from the authors upon reasonable request.

22. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the CBioPortal. Sci. Signal. 6, pl1 (2013).

23. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2, 401–404 (2012).

24. Watson, P. A. et al. Constitutively active androgen receptor splice variants expressed in castration-resistant prostate cancer require full-length androgen receptor. Proc. Natl Acad. Sci. USA 107, 16759–16765 (2010).

25. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. Science 343, 80–84 (2014).

26. Mollembaal, M. et al. Differential binding and co-binding pattern of FOXA1 and FOXA3 and their relation to H3K4me3 in HepG2 cells revealed by ChiP-seq. Genome Biol. 10, R129 (2009).

27. Peng, W., Bao, Y. & Sawicki, J. A. Epithelial cell-targeted transgene expression enables isolation of cyan fluorescent protein (CFP)-expressing prostate stem/progenitor cells. Transgenic Res. 20, 1073–1086 (2011).

28. Veizi, A., Bauer, C, Vasioukhin, V. & Fuchs, E. Actin cable dynamics and Rho/ROCK orchestrate a polarized cytoskeletal architecture in the early steps of lumen area quantification, processed organoids for immunohistochemistry, performed immunoblots, in vitro cell growth assays, lumen formation assays, and FOXA3 and their relation to H3K4me3 in HepG2 cells revealed by ChiP-seq. Genome Biol. 10, R129 (2009).

29. Smith, B. A. et al. A basal stem cell signature identifies aggressive prostate cancer phenotypes. Proc. Natl Acad. Sci. USA 112, E6544–E6552 (2015).

30. Klein, E. A. et al. A genomic classifier improves prediction of metastatic disease within 5 years after surgery in node-negative high-risk prostate cancer patients managed by radical prostatectomy without adjuvant therapy. Eur. Urol. 67, 778–786 (2015).

31. Boomsma, J. L. et al. Identification of TDRD5 as a direct target gene of ERG in primary prostate cancer. Int. J. Cancer 133, 335–343 (2013).

32. Ross, A. E. et al. Tissue-based genomics augments post-prostatectomy risk stratification in a natural history cohort of intermediate- and high-risk men. Eur. Urol. 69, 157–165 (2016).

33. Taylor, B. S. et al. Integrative genome profiling of human prostate cancer. Cancer Cell 18, 11–22 (2010).

34. Erho, N. et al. Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. PLoS ONE 8, e66653 (2013).
Extended Data Fig. 1 | See next page for caption.
Patients with predicted FOXA1 mutant status have worse outcomes. a, Co-crystal structure of the FKHD domain of FOXA3 in complex with DNA resembling the FKHD consensus sequence (PDB 1VTN), with residues and folds of interest indicated, including α-helix 3 (orange), which sits in the major groove of DNA, and Wing2 (cyan), which undergoes frequent mutation in prostate cancer. b, Kaplan–Meier plot showing significantly different clinical outcomes of time to biochemical recurrence (BCR, top) or progression to metastatic disease (MET, bottom) for predicted FOXA1 mutant cases vs wild type in the GRID cohort. The difference of MET and BCR survival curves was tested with the R survdiff function, using the G-rho family of tests, without adjustments for multiple comparisons. RP, radical prostatectomy. c, Associations between predicted FOXA1 mutation status and clinical variables using univariate analysis of the GRID cohort, with FOXA1 wild type as reference. The GRID cohort included 1,626 radical prostatectomy tumour samples. The centre values represent the median odds ratio via univariate analysis. The error bars represent first and third quartiles of odds ratio. The lines represent minimum and maximum odds ratio. Univariate logistic regression analyses were performed on the GRID cohort to test the statistical association between FOXA1 mutant status and clinical variables via generalized linear test, without adjustments for multiple comparisons. The test was two-sided with the significance level of $P < 0.05$ as the cut-off. ADT, androgen deprivation therapy; PSA, prostate-specific antigen; RT, radiotherapy.
Extended Data Fig. 2 | Details of FOXA1 luciferase reporter assay.

a, Schematic of FOXA1 luciferase reporter, depicting the modified response elements (at wobble positions within the canonical FOXA1 motif) cloned in tandem upstream of a minimal promoter driving luciferase expression. b, Dose–response curve of FOXA1 luciferase reporter activity in response to increased amounts of Foxa1WT cDNA introduced into the system, expressed as a relative response ratio with 100% Foxa1WT cDNA set to 1 and 0% Foxa1WT cDNA (100% ‘stuffer’ DNA) set to 0. Data are from three biological replicates, central line and error bars represent mean ± standard deviation. c, Western blot of allelic series of FOXA1 mutants in HEK293T cells 24 h after transfection with equal amounts of cDNA as used in FOXA1 luciferase reporter assay. The ΔF254/E255 and ΔM253–N256 mutants are shown as F254_E255del and M253_N256del, respectively, in the Extended Data. CYCLO B, loading control cyclophilin B. Representative blot, experiment repeated three independent times with similar results. For source gel data, see Supplementary Fig. 1.
Extended Data Fig. 3  See next page for caption.
Extended Data Fig. 3 | Inducible overexpression of FOXA1 variants influences organoid lumen size and morphology. a, Schematic of doxycycline-inducible pCW-FOXA1 constructs used in the study. b, Western blot analysis of lysates from pCW-FOXA1(WT) organoids following acute doxycycline treatment. Representative blot, experiment repeated two independent times with similar results. For source gel data, see Supplementary Fig. 1. c, Western blot analysis of lysates from organoids following long-term doxycycline treatment. Sizes of endogenous and Flag-tagged FOXA1 are noted, as well as the smaller truncated form from G275X at the expected size ~38 kDa. Representative blot, experiment repeated three independent times with similar results. For source gel data, see Supplementary Fig. 1. d, Quantification of lumen areas measured at ten days after seeding. Solid black bar represents geometric mean. Values for sample size (indicated as dots) and $P$ values are as follows: EV (292), +WT (284, $P < 0.0001$ over EV), +R219S (60, $<0.0001$), +F254_E255del (119, $<0.0001$), +D226N (120, $<0.0001$), +R261C (114, $<0.0001$), +R219C (333, 0.2915), +G275X (75, $<0.0001$), +M253_N256del (150, 0.2006), +M253K (63, 0.2343), +Y259C (32, 0.0082), +F266L (107, 0.1219), +H247Q (63, 0.8343), +H247R (180, $<0.0001$), +H247Y (71, 0.9104). All $P$ values are relative to WT unless noted, calculated using unpaired, two-tailed Student's $t$-test. Colours represent location of mutation within FOXA1. e, Histology and immunohistochemistry of organoid lines overexpressing additional alleles of FOXA1 via the doxycycline-inducible pCW vector 10 days after seeding. Images from a single biological experiment.
Extended Data Fig. 4 | See next page for caption.
Extended Data Fig. 4 | Analysis of FOXA1 alterations in FOXA1-deleted or PTEN-deleted contexts. 

a, CRISPR–Cas9-mediated knockdown of Foxa1 results in a markedly altered morphology. Organoids lacking Foxa1 (sgFoxa1) have a reduced capacity to form lumens while maintaining expression of AR and the basal marker p63. sgNT (guide RNA targeting human gene AAVS1) serves as a negative control. 
b, Western blot analysis of lysates from organoids carrying control guide RNA (sgNT) or guide RNA targeting Foxa1. Representative blot, experiment repeated three times with similar results. For source gel data, see Supplementary Fig. 1. 
c, Quantification of organoids containing lumens, seven days after trypsinization in normal organoid media. Data are from three biological replicates, bars represent mean ± standard deviation, *P* value calculated using unpaired, two-tailed Student’s *t*-test. 
d, Sequence indicating the location of three silent point mutations introduced upstream of the PAM sequence for Foxa1-targeting RNA sgFoxa1_1. 
e, Western blot analysis of lysates from organoids carrying either CRISPR-Zeo-sgGFP (CZsgGFP) or CRISPR-Zeo-sgFoxa1_1 (CZsgFoxa1) in addition to the pCW construct indicated, either EV or with a Foxa1 allele present, plus or minus doxycycline treatment for ten days. Representative blot, experiment repeated twice with similar results. For source gel data, see Supplementary Fig. 1. 
f, Images of organoid lines carrying various combinations of guide RNA and cDNAs, ten days after doxycycline treatment. 
g, Quantification of lumen-containing organoids in lines with endogenous Foxa1 deleted via CRISPR–Cas9 (sgFoxa1, sgGFP as control guide) and overexpression of CRISPR-resistant Foxa1 WT or mutant cDNA ten days after seeding. Data are from two biological replicates, bars represent mean. 
h, Western blot analysis of lysates from PTEN-deficient organoids grafted into mice, with doxycycline-induced overexpression of appropriate FOXA1 variants. Representative blot, experiment repeated twice with similar results. For source gel data, see Supplementary Fig. 1. 
i, Overexpression of FOXA1(WT) or FOXA1(G275X) in sgPTEN organoids promotes tumour growth in mice at six weeks after engraftment into the flank of NOD scid gamma mice. Data are from the following number of tumours: EV = 8, +WT = 8, +R219S = 10, +F254_E255del = 10, +G275X = 9, +ERG = 10. Data are mean ± s.d., *P* values calculated using unpaired, two-tailed Student’s *t*-test vs EV. Colours represent location of mutation within FOXA1. 
j, Representative histology and immunohistochemistry (IHC) of a single tumour for given PTEN-deficient, FOXA1-expressing lines. Histology and immunohistochemistry were performed on 5–9 tumours per line, from a single in vivo experiment, with similar results.
Extended Data Fig. 5 | See next page for caption.
Extended Data Fig. 5 | Analysis of the interplay between AR and FOXA1 in mouse organoids expressing FOXA1 variants. a, Box plot representations of normalized counts from AR (left) and FOXA1 (right) ChIP–seq shown in Fig. 3a to quantify the reduction in AR binding following FOXA1 wild-type or mutant overexpression, and the increase in FOXA1 wild-type binding at those sites where AR is lost. Box: 25th to 75th percentile; band: median; top whisker: 75th percentile plus 1.5 times interquartile range; bottom whisker: 25th percentile minus 1.5 times interquartile range. Sample size = 2,914 peaks. P values calculated using an unpaired, one-sided Wilcoxon test. b, Western blot analysis of lysates from AR-deficient organoids generated using CRISPR–Cas9 carrying representative Foxa1 alleles. Levels are significantly reduced but AR is not completely absent (as seen on the long exposure); this is a bulk population rather than single cell clones and thus a small number of cells escaped CRISPR–Cas9-mediated Ar deletion. Cells were treated with doxycycline for at least ten days. Representative blot, experiment repeated twice with similar results. For source gel data, see Supplementary Fig. 1. c, Expression of mouse orthologues of AR-target genes found in the AR signature used in TCGA cohort analysis based on mouse organoid RNA-seq. Genes depicted are those that have a mouse orthologue of the human gene found in the signature, and a significant expression change (DESeq2 adjusted $P < 0.05$) compared to EV control at 11 days of doxycycline treatment, as well as Pec, an AR-target gene expressed in mouse organoids. Data are from RNA-seq of three biological replicates. FE, F254_E255del. d, FOXA1(F254_E255del) signature can predict mutant tumours in TCGA. Hierarchical clustering and heat map of significantly differentially expressed genes between mouse FOXA1(F254_E255del) organoids and EV control (FDR $\leq 1 \times 10^{-10}$). Human homologues of differentially expressed genes (DEGs) from this analysis were used to cluster FOXA1 mutant tumours ($n=14$) and can detect nearly all FOXA1 mutant human tumours ($P = 2.1 \times 10^{-8}$) out of the 333 TCGA samples, 199 of which are ETS+. Two-sided Fisher’s exact test was used to test the enrichment of FOXA1 mutant samples within each sub-cluster, without adjustments for multiple comparisons.
Extended Data Fig. 6 | See next page for caption.
Extended Data Fig. 6 | Integrated analysis of ChIP–seq, ATAC-seq and RNA-seq data in FOXA1 mutant organoid lines. a, Cluster 0 peaks have higher FOXA1 ChIP–seq signal in F254_E255del mutant organoid than empty vector control. Box plots show normalized day five AR ChIP–seq signal and FOXA1 ChIP–seq signal across different organoid lines at peaks from cluster 0, where normalization is based on background ChIP signal. FOXA1 ChIP signal is significantly higher in F254_E255del (FE) and in WT compared to EV control (P values are listed in Supplementary Table 11). Sample size = 5,260 peaks. b, Cluster 1 peaks have higher FOXA1 ChIP–seq signal and lower AR ChIP–seq signal in FOXA1(WT)-overexpressing organoids than in EV control. Box plots show normalized day five AR ChIP–seq signal and FOXA1 ChIP–seq signal across different organoid lines at peaks from cluster 1, where normalization is based on background ChIP signal. FOXA1 ChIP signal is significantly higher, and AR ChIP signal significantly lower, in WT compared to EV control. Sample size = 1,493 peaks. c, Cluster 3 peaks have higher FOXA1 ChIP–seq signal in R219S organoid than EV control. Box plots show normalized day five AR ChIP–seq signal and FOXA1 ChIP–seq signal across different organoid lines at peaks from cluster 3, where normalization is based on background ChIP signal. FOXA1 ChIP signal is significantly higher in R219S compared to EV control. Sample size = 6,641 peaks. d, Cluster 5 peaks have higher FOXA1 ChIP–seq signal and lower AR ChIP–seq signal in R219S organoid than EV control. Box plots show normalized day five AR ChIP–seq signal and FOXA1 ChIP–seq signal across different organoid lines at peaks from cluster 5, where normalization is based on background ChIP signal. FOXA1 ChIP signal is significantly higher, and AR ChIP signal significantly lower, in R219S compared to EV control. Sample size = 1,983 peaks. a–d, box: 25th to 75th percentile; band: median; top whisker: 75th percentile plus 1.5 times interquartile range; bottom whisker: 25th percentile minus 1.5 times interquartile range. P values calculated using an unpaired, one-sided Wilcoxon test. e, Genes associated with cluster 0 are significantly induced in F254_E255del mutant organoids. Top, plots show empirical cumulative distribution of log2 expression changes at 24 h vs day 0 in WT (left), F254_E255del mutant (middle) and R219S mutant (right) organoids for all expressed genes (black), genes associated with at least one ATAC-seq peak in cluster 0 (cluster 0-associated genes, red), and the top quartile of these genes based on number of assigned cluster 0 peaks (strong cluster 0-associated genes, yellow). Cluster 0-associated genes show strong expression induction compared to all genes in F254_E255del as well as in WT but not in R219. Bottom, As a control, similar cumulative log2 expression changes for cluster 1-associated genes (red) or strong cluster 1-associated genes (yellow) do not show significant induction in F254_E255del. All P values are listed in Supplementary Table 12 and are one-sided Wilcoxon rank-sum tests. f, Genes associated with cluster 0 are significantly induced in F254–E255del mutant organoids. Top, plots show empirical cumulative distribution of log2 expression changes at 11 days vs day 0 in WT (left), F254_E255del mutant (middle) and R219S mutant (right) organoids for all expressed genes (black), genes associated with at least one ATAC-seq peak in cluster 0 (cluster 0-associated genes, red), and the top quartile of these genes based on number of assigned cluster 0 peaks (strong cluster 0-associated genes, yellow). Cluster 0-associated genes show strong expression induction compared to all genes in F254_E255del as well as in WT but not in R219. Bottom, As a control, similar cumulative log2 expression changes for cluster 3-associated genes (red) or strong cluster 3-associated genes (yellow) do not show significant induction in F254_E255del as well as in WT but not in R219. All P values are listed in Supplementary Table 12 and are one-sided Wilcoxon rank-sum tests. g, Genes associated with clusters 3 and 5 are significantly induced in R219S mutant organoid. Top, plots show empirical cumulative distribution of log2 expression changes at 24 h vs day 0 in WT (left), F254_E255del mutant (middle) and R219S mutant (right) organoids for all expressed genes (black), genes associated with at least one ATAC-seq peak in cluster 3 (cluster 3-associated genes, red), and the top quartile of these genes based on number of assigned cluster 0 peaks (strong cluster 3-associated genes, yellow). Cluster 3-associated genes show strong expression induction compared to all genes in R219S but not in WT or F254_E255del. Bottom, similar analysis for cumulative log2 expression changes for cluster 5-associated genes (red) and strong cluster 5-associated genes (yellow). These genes are significantly induced in R219S and repressed in F254_E255del in WT for this time point. All P values are listed in Supplementary Table 12 and are one-sided Wilcoxon rank-sum tests. h, Genes associated with clusters 3 and 5 are significantly induced in R219S mutant organoids. Top, Plots show empirical cumulative distribution of log2 expression changes at day 11 vs day 0 in WT (left), F254_E255del mutant (middle) and R219S mutant (right) organoids for all expressed genes (black), genes associated with at least one ATAC-seq peak in cluster 3 (cluster 3-associated genes, red), and the top quartile of these genes based on number of assigned cluster 0 peaks (strong cluster 3-associated genes, yellow). Cluster 3-associated genes show strong expression induction compared to all genes in R219S but not in WT or F254_E255del. Bottom, similar analysis for cumulative log2 expression changes for cluster 5-associated genes (red) and strong cluster 5-associated genes (yellow). These genes are significantly induced in R219S and repressed in F254_E255del. All P values are listed in Supplementary Table 12 and are one-sided Wilcoxon rank-sum tests.
Extended Data Fig. 7 | Motif analysis of ATAC-seq data and FOXA1 reporter assay for evaluation of non-canonical FOXA1 motif. a, FIMO motif analysis of ATAC-seq clusters. Summary of motif enrichments and depletion results for each cluster relative to the background of all differentially accessible peaks, as reported by binomial Z-score. The top 15 enriched database motifs for expressed transcription factors are shown for each cluster. In addition, enrichment–depletion results for four additional FOXA1-related motifs are shown: convergent and divergent dimer motifs, and altered FOXA1 core binding motifs with either G/A or C/T at position 6. Transcription factors in parentheses represent motifs inferred from other species. Complete lists can be found in Supplementary Tables 3–10. b, Top motif identified de novo using HOMER on ATAC-seq cluster 3 (R219S-specific) with motif core indicated, and variation from canonical FOXA1 motif depicted. P values derived from one-sided binomial test. c, Schematic of reporter design. The canonical response element reporter is the same reporter used in Fig. 2, with various iterations of the canonical FOXA1 motif in tandem. The non-canonical motif has substitutions at position 6, indicated in pink, to reflect the newly identified motif enriched in cluster 3 of ATAC-seq. Note that the orientation of the upper motif cartoon and the sequence in the reporter schematic are the reverse complement of the motif identified by HOMER (GTAAAG). Modified base is noted in position 6. d, Dose–response curve for activity of both FOXA1 luciferase reporters in response to increased amounts of Foxa1 WT cDNA introduced into the system. Data shown are one representative biological replicate of three carried out, all showing same trends, but absolute luciferase/Renilla ratios vary from experiment to experiment. e, Results of reporter assays expressed as a relative response ratio, normalized to level of FOXA1(WT) activity for a given reporter. Data are from three biological replicates, mean ± s.d. Unpaired, two-tailed Student’s t-test.
Extended Data Fig. 8 | Insert size distributions for ATAC-seq experiments, and track figures demonstrating peak reproducibility across ATAC-seq replicates. **a**, Representative insert size distributions computed from individual ATAC-seq experiments based on aligned read pairs, showing modes corresponding to nucleosome-free regions, mono-nucleosomal fragments, and di-nucleosomal fragments. **b**, Signal tracks for individual replicate ATAC-seq experiments at the Runx2, Plekha5 and Mbnl1 loci show reproducibility of accessibility events. DEseq scaling factors estimated from the atlas of IDR-reproducible peaks were used for library size normalization.
Extended Data Fig. 9 | ATAC-seq peak annotation distributions.
Fraction of peaks annotated as promoter, intergenic, intronic and exonic for full atlas of reproducible peaks, differentially accessible peaks, and by ATAC-seq cluster. See Supplementary Table 15 for full annotation counts.
Extended Data Fig. 10 | MA plots for differential accessibility analysis. 

a, MA plots for differential accessibility analysis relative to EV controls. Representative MA plots (log (fold change) vs mean read counts) for differential peak accessibility analysis of FOXA1 mutant- and WT-expressing organoid lines vs empty vector controls at day 0, day 1, and day 5. Peaks that are significantly differential at FDR-corrected $P < 0.05$ are shown in colour. Dotted lines at log (fold change) = 2 and log (fold change) = −2 show cut-offs used for requiring robust accessibility changes in pairwise comparisons. 

b, MA plots for differential accessibility analysis at different time points relative to day 0. Representative MA plots (log (fold change) vs mean read counts) for differential peak accessibility analysis in each organoid line at day 1 vs day 0 and day 5 vs day 0. In a, b, all sample sizes are $n = 183,093$ (number of peaks in the atlas). Peaks that are significantly differential at FDR-corrected $P < 0.05$ are shown in colour, using two-sided Wald test with Benjamini–Hochberg correction.
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 Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of 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 statistics 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
- Clearly defined error bars
- State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

See methods "RNA isolation and sequencing", "Assay for Transposase Accessible chromatin (ATAC) coupled with Next Generation Sequencing (NGS)", and "Chromatin Immuno-Precipitation (ChIP) coupled with Next Generation Sequencing (NGS)". Specific software includes Nikon NIS elements software.

Data analysis

See methods, "ATAC data and preprocessing", "ATAC-seq atlas creation", "Assignment of ATAC-seq peaks to genes", "ScAPT development based on FOXA1 mutant transcriptional signature and SVM model", and "Prostate cancer molecular subclass prediction by decision tree" and "Bio-informatics analysis ChIP-seq". Specific software includes: STARaligner (v2.4.2a), featureCounts (v1.4.3), Picard (v1.83), R v3.4.0, Trim Galore v0.4.5, CutAdapt v1.16, FastQC v0.11.7, Bowtie2 v2.3.4.1, DESeq2 v1.18.1, deeptools 3.0.2, Homer v4.10, Trimmomatic v0.35, IDR (v2.0.2).

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/reviewers upon request. 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

All data will be made available from the authors upon reasonable request. The described RNA-seq and ChIP-seq and ATAC-seq data will be deposited in the Gene Expression Omnibus in the future.

Field-specific reporting
Please select 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/authors/policies/ReportingSummary-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 done, for most assays, 3 biological replicates were carried out. For organoid lines, there are often more than 3 because lines were often included in a replicate experiment before results of previous experiment was analyzed, resulting in a greater number of replicates than what was likely needed. |
| Data exclusions | Entire growth assays were excluded if the control line did not grow as expected (i.e. didn't recover well from sorting so there was no growth seen). In that situation, data from entire experiment (all lines) were excluded. Mouse tumor measurements were excluded from data if it was noted at the time of injection that there was a technical issue. Both of these circumstances were pre-established and were held true for all lines. |
| Replication | All findings were reproducible, as more than 1 biological replicate was generally carried out, or the sample sizes were large enough to ensure good representation of the population behavior. |
| Randomization | No randomization was done for our experiments because there were no drug treatment groups, and all work was done in vitro. |
| Blinding | For lumen size assays, actual measurements were carried out by a member of the lab who did not have any knowledge of which alleles were associated with growth/lumen formation phenotypes. All other experiments were not blinded and it was not necessary to be as they were much more quantitative and less subjective. |

Reporting for specific materials, systems and methods

| Materials & experimental systems | Methods |
|--------------------------------|---------|
| n/a | n/a |
| □ | □ |
| Investigators involved in the study | Involved in the study |
| □ | □ |
| Unique biological materials | ChIP-seq |
| □ | |
| Antibodies | Flow cytometry |
| □ | |
| Eukaryotic cell lines | MRI-based neuroimaging |
| □ | |
| Palaeontology | |
| □ | |
| Animals and other organisms | |
| □ | |
| Human research participants | |

Unique biological materials
Policy information about availability of materials

Obtaining unique materials | All unique materials are readily available upon request.
Antibodies

| Antibodies used | Validation |
|-----------------|------------|
| AR (1:1,000, N-20, Santa Cruz), FOXA1 (1:100, Ab2, Sigma), Cyclophilin B (1:1000, EPR12703(B), Abcam), or FLAG (1:1000, M2, Sigma), HNF-3 alpha/FoxA1 Antibody (3B3NB) 5ug/mL (Novus Biologicals), p63 (1:800, 4A4, Ventana), PTEN (1:1000, D4.3, Cell Signaling), Ki67 (Abcam #ab15580 at 1ug/ml). AR (1:1,000, ER179(2), Abcam) |
| AR (1:1,000, N-20, Santa Cruz, recommended for detection of AR of mouse, rat and human origin by WB, IP, IF, IHC(P) and ELISA) FOXA1 (1:1000, Ab2, Sigma, species reactivity: rat, human, mouse, canine; application: western blot, suitable) Cyclophilin B (1:1000, EPR12703(B), Abcam, Tested applications: Suitable for: WB, Unsuitable for: Flow Cyt, ICC/IF or IHC-P Species reactivity: Reacts with: Mouse, Rat, Human) FLAG (1:1000, M2, Sigma, For highly sensitive and specific detection of FLAG fusion proteins by immunoblotting, immunoprecipitation, immunohistochemistry, immunofluorescence and immunocytochemistry. Optimized for single banded detection of FLAG fusion proteins in mammalian, plant, and bacterial expression systems. HNF-3 alpha/FoxA1 Antibody (3B3NB) 5ug/ml (Novus Biologicals, Reactivity Hu, Mu, Rt; Applications WB, Flow, ICC/IF, IHC, IHC-P) p63 (1:800, 4A4, Ventana, Anti-p40 (BC28) Mouse Monoclonal Primary Antibody intended for laboratory use in the detection of the p40 protein in formalin-fixed, paraffin-embedded tissue. PTEN (1:1000, D4.3, Cell Signaling) Rabbit Monoclonal antibody with reactivity to mouse, human, rat, dog, monkey. Applications: WB, IP, IHC, IF, F, ChIP. Ki67 (Abcam #ab15580 at 1ug/ml, suitable for IHC - Wholemount, IHC-P, IHC-FrFl, Flow Cyt, ICC/IF, ICC) |

Eukaryotic cell lines

| Policy information about | cell lines |
|--------------------------|------------|
| Cell line source(s)      | Primary mouse organoid lines were established in our laboratory for this study |
| Authentication           | since these were developed in our lab, authentication was not necessary |
| Mycoplasma contamination  | All organoid lines used were tested for mycoplasma contamination and were found to be negative |
| Commonly misidentified lines (See | none of the cell lines used are in this register |
| FLAT register)           | |

Animals and other organisms

| Policy information about | studies involving animals, ARRIVE guidelines recommended for reporting animal research |
|--------------------------|---------------------------------|
| Laboratory animals       | Mus Musculus strain NOD.Cg-Prkdcsid Il2rgtm1.Wjj/SzJ, males between 8-12 weeks at beginning of study (study duration ~12 weeks). |
| Wild animals             | Study did not involve wild animals |
| Field-collected samples  | Study did not involve field collected samples |

ChIP-seq

| Data deposition |
|-----------------|
| Confirm that both raw and final processed data have been deposited in a public database such as GEO. |
| Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks. |

| Data access links |
|-------------------|
| May remain private before publication. For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data. |

| Files in database submission |
|-----------------------------|
| Provide a list of all files available in the database submission. |

| Genome browser session (e.g. UCSC) |
|-----------------------------------|
| Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents. |

Methodology

| Replicates |
|------------|
| Describe the experimental replicates, specifying number, type and replicate agreement. |

| Sequencing depth |
|------------------|
| Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end. |

| Antibodies |
|-----------|
| HNF-3 alpha/FoxA1 Antibody (3B3NB) (Novus Biologicals), AR (ER179(2), Abcam) |
| Section                      | Description                                                                                          |
|------------------------------|------------------------------------------------------------------------------------------------------|
| Peak calling parameters      | Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used. |
| Data quality                 | Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment. |
| Software                     | Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details. |