LETTER

https://doi.org/10.1038/s41586-019-1347-4

Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer

Abhijit Parolia[1,2,3,12], Marcin Cieslik[1,2,4,12], Shih-Chun Chu[1,2], Lanbo Xiao[1,2], Takahiro Ouchi[1,2], Yuping Zhang[1,2], Xiaoju Wang[1,2], Pankaj Vats[1,2], Xuhong Cao[1,2,5], Sethuramasundaram Pitchiaya[1,2], Fengyun Su[1,2], Rui Wang[1,2], Felix Y. Feng[6,7,8,9], Yi-Mi Wu[1,2], Robert J. Lonigro[1,2], Dan R. Robinson[1,2] & Arul M. Chinnaiyan[1,2,5,10,11]*

Forkhead box A1 (FOXA1) is a pioneer transcription factor that is essential for the normal development of several endoderm-derived organs, including the prostate gland[1,2]. FOXA1 is frequently mutated in hormone-receptor-driven prostate, breast, bladder and salivary-gland tumours[3–8]. However, it is unclear how FOXA1 alterations affect the development of cancer, and FOXA1 has previously been ascribed both tumour-suppressive[9–11] and oncogenic[12–14] roles. Here we assemble an aggregate cohort of 1,546 prostate cancers and show that FOXA1 alterations fall into three structural classes that diverge in clinical incidence and genetic co-alteration profiles, with a collective prevalence of 35%. Class-1 activating mutations originate in early prostate cancer without alterations in ETS or SPOP, selectively recur within the wing-2 region of the DNA-binding forkhead domain, enable enhanced chromatin mobility and binding frequency, and strongly transactivate a luminal androgen-receptor program of prostate oncogenesis. By contrast, class-2 activating mutations are acquired in metastatic prostate cancers, truncate the C-terminal domain of FOXA1, enable dominant chromatin binding by increasing DNA affinity and—through TLE3 inactivation—promote metastasis driven by the WNT pathway. Finally, class-3 genomic rearrangements are enriched in metastatic prostate cancers, consist of duplications and translocations within the *FOXA1* locus, and structurally reposition a conserved regulatory element—herein denoted FOXA1 mastermind (FOXMIND)—to drive overexpression of *FOXA1* or other oncogenes. Our study reaffirms the central role of FOXA1 in mediating oncogenesis driven by the androgen receptor, and provides mechanistic insights into how the classes of FOXA1 alteration promote the initiation and/or metastatic progression of prostate cancer. These results have direct implications for understanding the pathobiology of other hormone-receptor-driven cancers and rationalize the co-targeting of FOXA1 activity in therapeutic strategies.

FOXA1 independently binds to and de-compacts condensed chromatin to reveal the binding sites of partnering nuclear hormone receptors[15,16]. In prostate luminal epithelial cells, FOXA1 delimits tissue-specific enhancers[17], and reprograms androgen receptor (AR) activity in prostate cancer[14]. Accordingly, FOXA1 and AR are co-expressed in prostate cancer cells, in which FOXA1 activity is indispensable for cell survival and proliferation[14] (Extended Data Fig. 1a–i). It is notable that, in AR-dependent prostate cancer, *FOXA1* is the third most-highly mutated gene[4,5] and—as shown here—is located at one of the most-highly rearranged genomic loci. Counterintuitively, recent studies have suggested these alterations are inactivating[18,19] and have described FOXA1 as a tumour suppressor in AR-driven metastatic prostate cancer[9–11]. However, FOXA1 alterations have not yet been fully characterized or experimentally investigated in cancer.

To study these alterations, we first curated an aggregate cohort of prostate cancer that comprised 888 localized and 658 metastatic samples[4,5,8,20], of which 498 and 357, respectively, had matched RNA-sequencing (RNA-seq) data. Here, *FOXA1* mutations recurred at a frequency of 8–9% in primary disease, which increased to 12–13% in metastatic castration-resistant prostate cancer (mCRPC) (Fig. 1a, Extended Data Fig. 1j). RNA-seq calls of structural variants revealed a high prevalence (Fig. 1b, Supplementary Table 1) and density (Extended Data Fig. 1k) of rearrangements within the *FOXA1* locus. The presence of structural variants was confirmed by whole-exome and whole-genome sequencing (Extended Data Fig. 1l, m, Supplementary Tables 2, 3). Overall, we estimated the recurrence of *FOXA1* locus rearrangements to be 20–30% in mCRPC (Extended Data Fig. 1n). All *FOXA1* mutations were heterozygous and *FOXA1* itself was copy-amplified in over 50% of cases with no biallelic deletions (Extended Data Fig. 2a, b). We also found a stagewise increase in FOXA1 expression in prostate cancer (Extended Data Fig. 2c, Supplementary Discussion).

When we mapped mutations onto the protein domains of FOXA1, we found two structural patterns: (1) missense and in-frame insertion and deletion (indel) mutations were clustered at the C-terminal end of the forkhead domain (FKHD); and (2) truncating frameshift mutations were restricted to the C-terminal half of the protein (Fig. 1c). *FOXA1* structural variants predominantly consisted of tandem duplications and translocations, which clustered in close proximity to the *FOXA1* gene without disrupting its coding sequence (Fig. 1d). Thus, we categorized FOXA1 alterations into three structural classes: class 1, which comprises all the mutations within the FKHD; class 2, which comprises mutations in the C-terminal end after the FKHD; and class 3, which comprises structural variants within the *FOXA1* locus (Fig. 1c, d, Extended Data Fig. 2d). We also found similar classes of FOXA1 alterations in breast cancer (Extended Data Fig. 2e, f).

We found that the majority of *FOXA1* mutations in primary prostate cancer belonged to class 1, which showed no enrichment in the metastatic disease (Fig. 1e). Conversely, class-2 mutations were significantly enriched in metastatic prostate cancer; in the rare primary cases with class-2 mutations, the mutant allele was detected at sub-clonal frequencies (Fig. 1e, f, Extended Data Fig. 2g, h). We found no cases that possessed both class-1 and class-2 mutations. Class-3 structural variants were also significantly enriched in mCRPC (odds ratio = 3.46) (Fig. 1g). Overall, we found the cumulative frequency of FOXA1 alterations to be over 34% in mCRPC (Fig. 1h). Assessment of concurrent alterations revealed that class-1 mutations are mutually exclusive with other primary events (for example, ETS fusions) (odds ratio = 0.078), whereas class-2-mutant mCRPC are enriched for *RB1* deletions (odds ratio = 4.17) (Extended Data Fig. 2i, j). Both mutational classes were further enriched for alterations in DNA repair, mismatch repair and

[1]Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, MI, USA. [2]Department of Pathology, University of Michigan, Ann Arbor, MI, USA. [3]Molecular and Cellular Pathology Program, University of Michigan, Ann Arbor, MI, USA. [4]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. [5]Howard Hughes Medical Institute, University of Michigan, Ann Arbor, MI, USA. [6]Helen Diller Family Comprehensive Cancer Center, University of California at San Francisco, San Francisco, CA, USA. [7]Department of Radiation Oncology, University of California at San Francisco, San Francisco, CA, USA. [8]Department of Urology, University of California at San Francisco, San Francisco, CA, USA. [9]Department of Medicine, University of California at San Francisco, San Francisco, CA, USA. [10]Department of Urology, University of Michigan, Ann Arbor, MI, USA. [11]Rogel Cancer Center, University of Michigan, Ann Arbor, MI, USA. [12]These authors contributed equally: Abhijit Parolia, Marcin Cieslik. *e-mail: arul@umich.edu

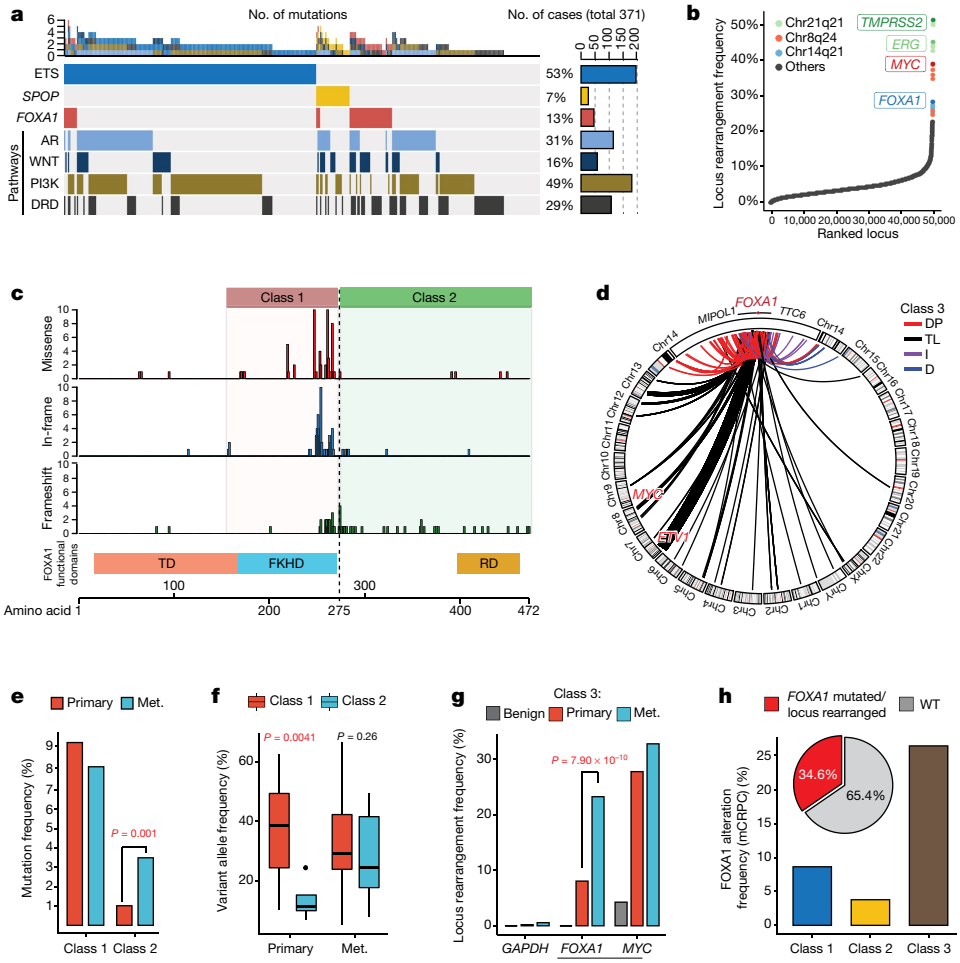


Fig. 1 | Structural classes of FOXA1 alterations. a, *FOXA1* mutations and key alterations in mCRPC. Alterations in ETS, AR, WNT, PI3K and DNA repair (DRD) were aggregated at the pathway or group level. **b**, Locus-level recurrence of RNA-seq structural variations. **c**, Structural classification of *FOXA1* mutations. TD, transactivation domain; RD, regulatory domain. **d**, Structural classification of *FOXA1* locus rearrangements. DP, tandem duplications; TL, translocations; I, inversions; D, deletions. **e**, Frequency of *FOXA1* mutational classes by prostate cancer stage ($n = 888$ primary, 658 metastatic (met.)) (two-sided Fisher's exact test). **f**, Variant allele frequency by stage and class (two-sided *t*-test). Box plot centre, median; box, quartiles 1–3, whiskers, quartiles 1–3 ± 1.5 × interquartile range (IQR). **g**, Locus-level recurrence of structural variants based on RNA-seq by prostate cancer stage (two-sided Fisher's exact test). **h**, Integrated (RNA-seq and whole-exome sequencing) recurrence of FOXA1-alteration classes in mCRPC (Stand Up 2 Cancer and Michigan Center for Translational Pathology (MCTP) cohort, $n = 370$).

WNT signalling pathways (Extended Data Fig. 2i, k), and had higher levels of expression of *FOXA1* mRNA relative to the wild-type cases (Extended Data Fig. 2l). Together, these data suggest that class-1 mutations emerge in localized prostate cancer, whereas class-2 and class-3 mutations are acquired or enriched, respectively, in the course of disease progression.

Class-1 mutations consist of missense and in-frame indels that cluster at the C-terminal edge of the winged-helix DNA-binding FKHD. The majority of the class-1 mutations were located either within the wing-2 region (residues 247–269) or a 3D hotspot that spatially protrudes towards wing 2[21] (Fig. 2a, b, Extended Data Fig. 3a, b). Notably, these mutations did not alter FKHD residues that make base-specific interactions with the DNA[22,23] (Fig. 2a, Extended Data Fig. 3c). In FOXA proteins, wing-2 residues make base-independent (that is, non-specific) contacts with the DNA backbone[23,24] that reportedly impede its nuclear movement[24]. Thus, we hypothesized that class-1 mutants with altered wing-2 regions would display faster nuclear mobility.

We cloned representative class-1 mutants of FOXA1: I176M (mutation of the 3D hotspot), R261G (missense) and R265–Q271del (in-frame deletion), all of which retained nuclear localization (Extended Data Fig. 3d). In fluorescence recovery after photobleaching (FRAP) assays, we found class-1 mutants had 5–6× faster nuclear mobility irrespective of the mutation type (Fig. 2c, d, Extended Data Fig. 3e, g). By contrast, class-2 mutants with intact wing-2 regions were sluggish in their nuclear movement (Fig. 2d, Extended Data Fig. 3f, g). Using single particle tracking, we verified that class-1 mutants have a higher overall rate of nuclear diffusion, with 3–4-fold fewer slow particles and shorter chromatin dwell times (Extended Data Fig. 3h, i). In chromatin immunoprecipitation with parallel DNA sequencing (ChIP–seq) assays, we found that ectopically expressed class-1 mutants in HEK293 cells bind DNA at the consensus FOXA1 motif (Extended Data Fig. 3j, k). In prostate cancer cells, the class-1 cistrome entirely overlapped with wild-type binding sites, with similar enrichment for FOXA1 and AR cofactor motifs, AR-binding sites and genomic distribution (Extended Data Fig. 3l–s). Furthermore, in growth rescue experiments using untranslated-region-specific small interfering (si) RNAs that targeted the endogenous *FOXA1* transcript, we found that exogenous class-1 mutants fully compensated for the wild-type protein (Extended Data Fig. 4a).

Next, we asked how class-1 mutations affect AR signalling. Similar to wild-type FOXA1, both class-1 and class-2 mutants interacted with the AR signalling complex (Extended Data Fig. 4b–d). In reporter assays, class-1 mutants induced 3–6-fold higher activation of AR signalling (Fig. 2e), which was evident even under stimulation with castrate levels of androgen or treatment with enzalutamide (Extended Data Fig. 4e, f). In parallel assays, class-2 mutants showed no differences relative to wild-type FOXA1 (Fig. 2e). Transcriptomic analyses of class-1 tumours

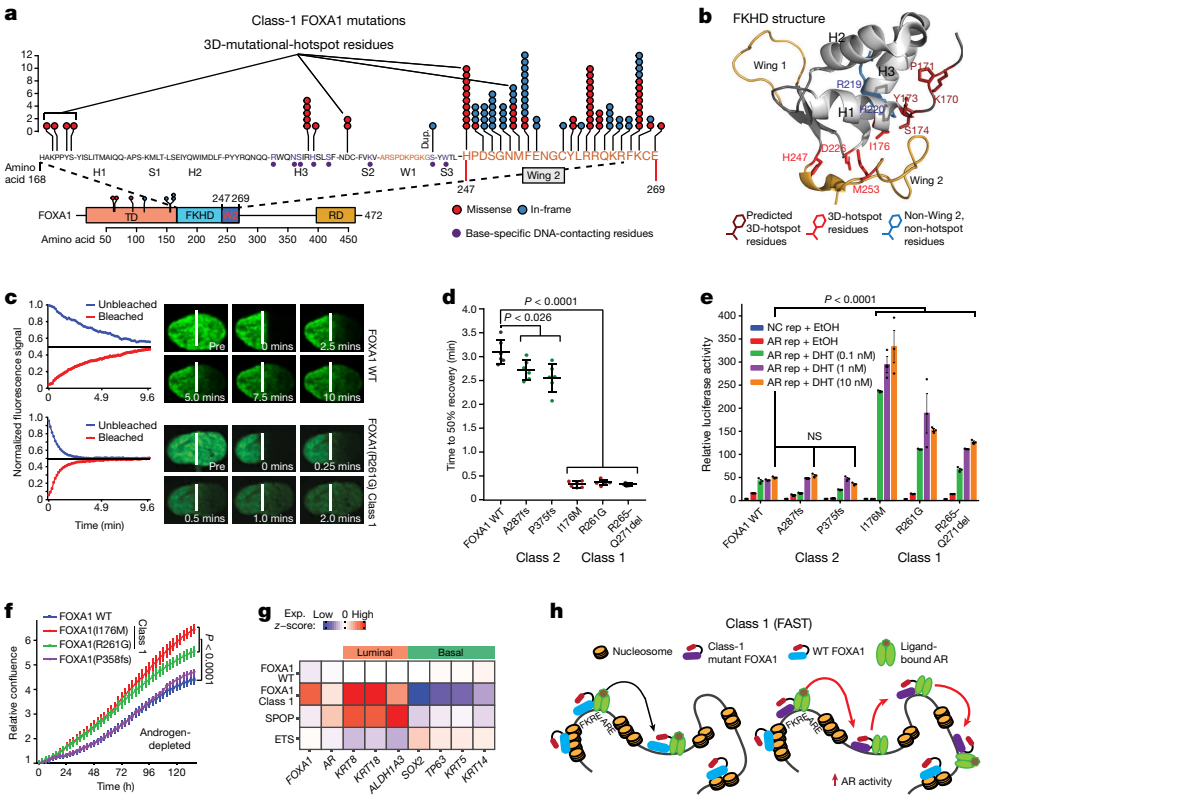


Fig. 2 | Functional characterization of class-1 mutations of *FOXA1*. **a**, Distribution of class-1 mutations on the protein map of FOXA1 functional domains and FKHD secondary structures. Dup., duplication. **b**, Crystal structure of the FKHD with visualization of non-wing-2 (that is, outside of amino acids 247–269) mutations. Mutations in the 3D hotspot are in red. **c**, FRAP kinetic plots (left) and representative time-lapse images from pre-bleaching to the equilibrated state (right; $n = 6$ biological replicates). Images are uniformly brightened for signal visualization. WT, wild type. **d**, FRAP durations until 50% recovery ($n = 6$ nuclei per variant). **e**, Negative control (NC) or AR reporter (rep) activity with overexpression of FOXA1 variants and dihydrotestosterone (DHT) stimulation ($n = 3$ biological replicates). **f**, Growth (IncuCyte) of 22RV1 cells that overexpress FOXA1 variants in androgen-depleted medium ($n = 5$ biological replicates). In **d**–**f**, mean ± s.e.m. is shown, and *P* values are from two-way analysis of variance (ANOVA) and Tukey's test. **g**, Relative expression of luminal and basal markers in class-1 ($n = 38$) tumours compared with wild-type ($n = 457$), SPOP ($n = 48$) and ETS ($n = 243$) primary prostate cancer tumours. **h**, Class-1 model. Wing-2-disrupted FOXA1 shows increased chromatin mobility and chromatin sampling frequency, which results in stronger transcriptional activation of oncogenic AR signalling. FKRE, forkhead-responsive element; ARE, androgen-responsive element.

from patients revealed the activation of hyperproliferative and pro-tumorigenesis pathways, and further enrichment of primary prostate cancer genes (Extended Data Fig. 4g–i). Notably, AR was predicted[25] to be the driver transcription factor for class-1 upregulated genes, which we experimentally confirmed for several targets (Extended Data Fig. 4j–l). Concordantly, overexpression of class-1 mutants in 22RV1 cells increased growth in androgen-depleted medium (Fig. 2f) but not in androgen-supplemented medium, and rescued proliferation upon treatment with enzalutamide (Extended Data Fig. 4m, n). For class-1 downregulated genes, the basal transcription factors TP63 and SOX2 were predicted to be transcriptional drivers (Extended Data Fig. 4j). Consistently, in class-1 specimens from patients, both of these transcription factors were significantly downregulated, with a concomitant downregulation of basal, and upregulation of luminal, markers (Fig. 2g, Extended Data Fig. 4o, p). In addition, class-1 tumours had a higher AR transcriptional signature, and a lower neuroendocrine transcriptional signature (Extended Data Fig. 4q). Together, these data suggest that class-1 mutations that alter the wing-2 region increase the nuclear speed and genome-scanning efficiency of FOXA1 without affecting its DNA sequence specificity (Supplementary Discussion), and drive a luminal AR program of prostate oncogenesis (Fig. 2h).

Class-2 mutations consist of frameshifting alterations that truncate the C-terminal regulatory domain of FOXA1 (Fig. 3a). Thus, we characterized the class-2 cistrome by using N-terminal and C-terminal antibodies, with the C-terminal antibody binding exclusively to wild-type FOXA1 (Extended Data Fig. 5a, b). Notably, mCRPC-derived LAPC4 cells endogenously contained a *FOXA1* class-2 mutation (that is, a frameshift at amino acid P358 (P358fs)), and both wild-type and mutant variants interacted with the AR complex (Extended Data Fig. 5c–f). However, in ChIP–seq assays, only the N-terminal antibody detected FOXA1 binding to the DNA. By contrast, N-terminal and C-terminal FOXA1 cistromes substantially overlapped in wild-type prostate cancer cells (Fig. 3b, Extended Data Fig. 5g–i). Even with 13-fold overexpression of wild-type FOXA1 in LAPC4 cells, the endogenous class-2 mutant retained its binding dominance (Fig. 3b, Extended Data Fig. 5j, k). Conversely, overexpression of the FOXA1(P358fs) mutant in LNCaP cells markedly diminished the endogenous wild-type cistrome (Fig. 3b). In in vitro assays, class-2 mutants showed markedly stronger binding to the *KLK3* enhancer element (Fig. 3c, Extended Data Fig. 6a–d), and biolayer interferometry confirmed that the FOXA1(P358fs) mutant has an approximately fivefold-higher DNA-binding affinity (Extended Data Fig. 6e). In CRISPR-engineered class-2-mutant 22RV1 clones (Extended Data Fig. 6f, g), FOXA1 ChIP–seq data reaffirmed the cistromic dominance of class-2 mutants (Fig. 3d). Knockdown of either mutant FOXA1 or AR in 22RV1 or LNCaP class-2 CRISPR clones significantly attenuated proliferation (Fig. 3e, Extended Data Fig. 6h, i). Consistently, in rescue experiments, the FOXA1(P358fs) mutant fully compensated for the loss of wild-type FOXA1 (Extended Data Fig. 4a).

The class-2 cistrome was considerably larger than the wild-type cistrome (Extended Data Fig. 6j–l), and the acquired sites were enriched for the CTCF motif and distal regulatory regions (Extended Data Fig. 7a–e, Supplementary Discussion). In transcriptomic and motif analyses of the class-2 clones, LEF and TCF were predicted as

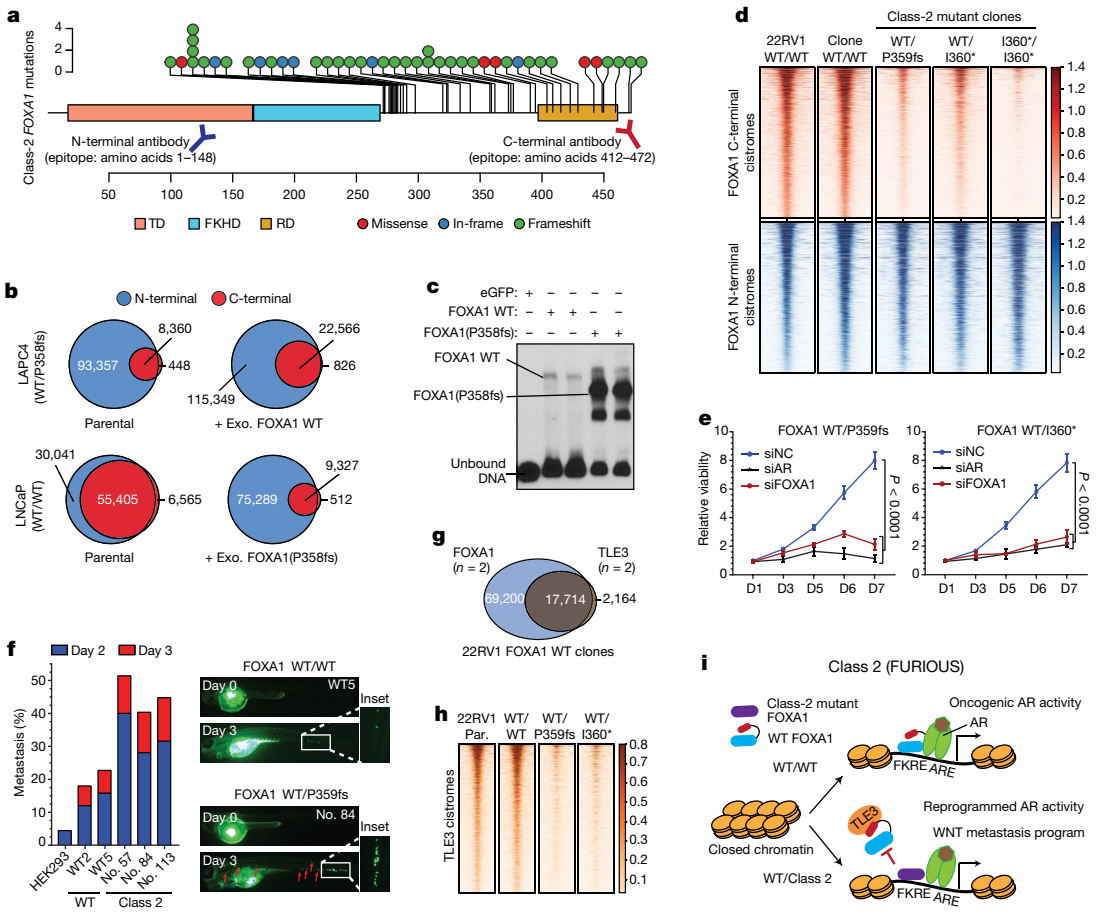


Fig. 3 | Functional characterization of class-2 mutations of *FOXA1*. **a**, Class-2 mutations and antibody epitopes on the protein map of FOXA1. **b**, N-terminal and C-terminal FOXA1 cistromes in FOXA1 wild-type (*FOXA1*$^{WT/WT}$ (WT/WT)) or mutant (*FOXA1*$^{WT/P358fs}$ (WT/P358fs)) prostate cancer cells that are untreated (left) or have exogenous (exo.) overexpression of FOXA1 variants (right). **c**, Electromobility shift of FOXA1 variants bound to the *KLK3* enhancer ($n = 3$ biological replicates). For gel source data, see Supplementary Fig. 1. **d**, FOXA1 ChIP–seq read-density heat maps in independent class-2-mutant 22RV1 CRISPR clones (*FOXA1*$^{WT/P359fs}$ (WT/P359fs), *FOXA1*$^{WT/I360*}$ (WT/I360*) and *FOXA1*^{I360*I360*} (I360*/I360*)). **e**, Growth of class-2-mutant 22RV1 clones treated with non-targeting (siNC), *AR*- or *FOXA1*-targeting siRNAs ($n = 5$ biological replicates; two-way ANOVA and Tukey's test). Mean ± s.e.m. is shown. D, day. **f**, Left, metastasis frequency in zebrafish embryos injected with HEK293 (negative control), wild-type 22RV1 clones or class-2-mutant 22RV1 clones ($n \geq 30$ embryos per group). Right, representative images of embryos, showing the disseminated prostate cancer cells. **g**, Overlap of wild-type FOXA1- and TLE3-binding sites in 22RV1 CRISPR clones ($n = 2$ biological replicates each). **h**, TLE3 ChIP–seq read-density heat maps in 22RV1 parental (par.) cells and distinct FOXA1 wild-type and class-2-mutant 22RV1 CRISPR clones. **i**, Class-2 model. Truncated FOXA1 shows dominant chromatin binding and displaces wild-type FOXA1 and TLE3 from the chromatin, which results in increased WNT signalling.

the top regulatory transcription factors for the upregulated genes (Extended Data Fig. 7g, h). The LEF–TCF complex is the primary nuclear effector of WNT signalling and remains inactive until it is bound by β-catenin[26]. Consistently, we found a marked accumulation of transcriptionally active β-catenin—that is, non-phosphorylated at S31, S37 and T41—in distinct mutant clones, as well as a concomitant increase in the expression of the WNT targets LEF1 and AXIN2 (Extended Data Fig. 7i, j). Class-2 clones showed 2–3-fold higher invasiveness in Boyden chamber assays (Extended Data Fig. 7k, l), and a higher rate and extent of metastatic dissemination in zebrafish embryos (Fig. 3f, Extended Data Fig. 7m). In these assays, class-1 mutant cells showed no differences relative to wild-type cells (Extended Data Fig. 7n). Furthermore, treatment with the WNT inhibitor XAV939 completely abrogated the class-2 invasive phenotype (Extended Data Fig. 7o). Investigating the mechanism that underlies this invasiveness, we found that FOXA1 transcriptionally activates and—through its C-terminal domain—recruits TLE3 (a bona fide WNT co-repressor[27]) to the chromatin (Extended Data Fig. 8a–e). Class-2 mutants had lost this interaction, which led to the untethering of TLE3 from chromatin and downstream activation of WNT signalling (Fig. 3g, h, Extended Data Fig. 8e–k, Supplementary Discussion). Together, these data suggest that class-2 mutations confer cistromic dominance and abolish TLE3-mediated repression of the WNT program of metastasis (Fig. 3i).

Class-3 rearrangements occur within the *PAX9* and *FOXA1* locus that is linearly conserved across the deuterostome superphylum[28] (Fig. 4a). Notably, almost all break ends were clustered within the *FOXA1* topologically associating domain (Extended Data Fig. 9a). We found that the genes located within the *FOXA1* topologically associating domain had the highest expression in the normal prostate, and the non-coding RP11-356O9.1 transcript had a prostate-specific expression (Extended Data Fig. 9b). Furthermore, in patient tumours, expression of RP11-356O9.1 was strongly correlated with *FOXA1* and *TTC6* expression (Extended Data Fig. 9c). Thus, to identify prostate-specific enhancers of the *FOXA1* topologically associating domain, we performed the assay for transposase-accessible chromatin using sequencing (ATAC-seq) and interrogated chromatin features in AR^{+} and AR^{-} prostate cells. Notably, a CTCF-bound intronic site in RP11-356O9.1 (hereafter denoted as *FOXMIND*) and a site within the 3′ untranslated region of *MIPOL1* were accessible and marked with active enhancer modifications only in AR^{+}FOXA1^{+} prostate cancer cells (Fig. 4b, Extended Data Fig. 9d). This strongly suggested that these conserved sites function as enhancer elements. Consistently, CRISPR knockout of these loci in VCaP cells led to a significant decrease in the expression of *FOXA1*

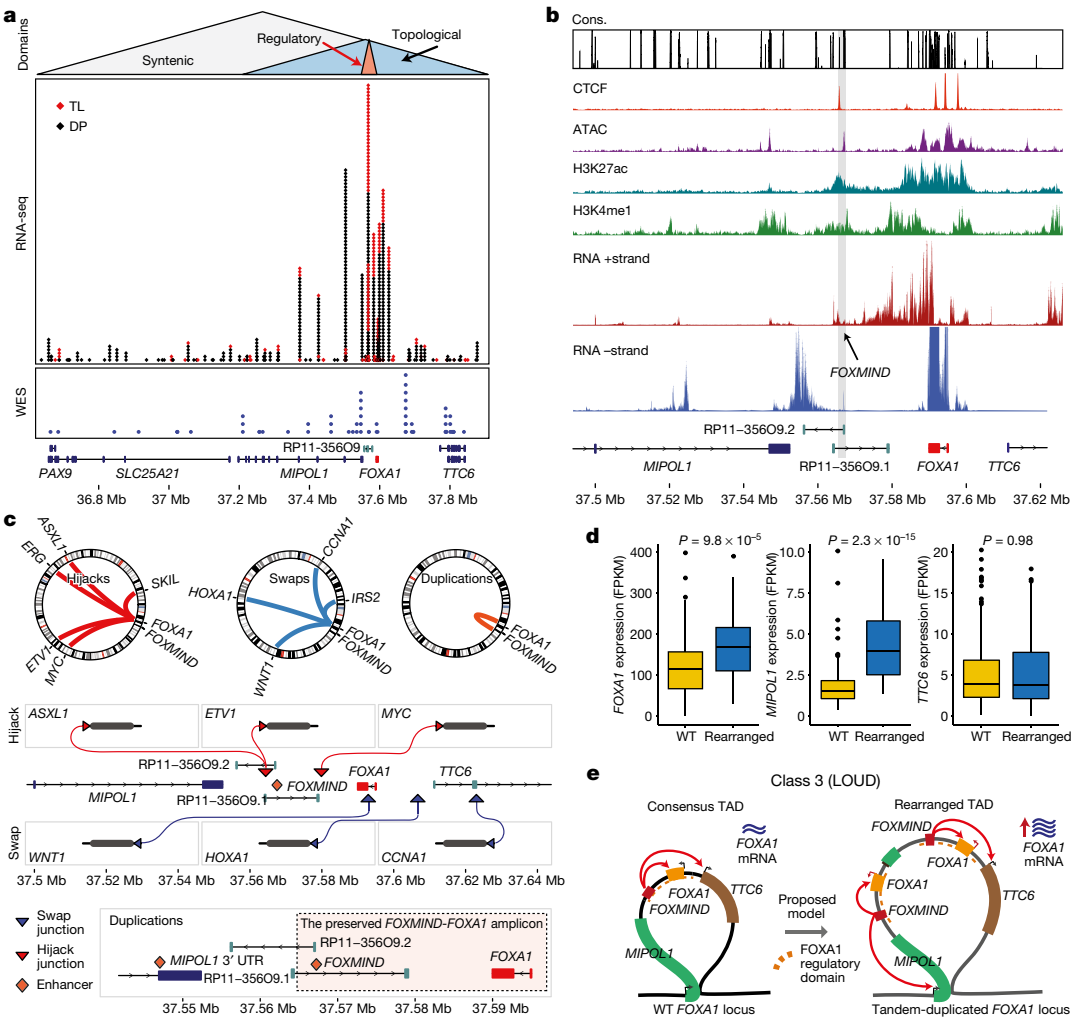


Fig. 4 | Genomic characterization of class-3 rearrangements of the *FOXA1* locus. a, Break ends in relation to the *FOXA1* syntenic, topological and regulatory domains. WES, whole-exome sequencing. **b**, Representative functional genomic tracks at the *FOXA1* locus. Base-level conservation (cons.), DNA accessibility (ATAC), enhancer-associated histone modifications (H3K27me1 and H3K27Ac), CTCF chromatin binding and stranded RNA-seq read densities are visualized. The *FOXMIND* enhancer is highlighted. **c**, Structural patterns of translocations and duplications. Hijacks occur between *FOXMIND* and *FOXA1*; swaps occur upstream of *FOXA1*. Duplications amplify the highlighted *FOXMIND-FOXA1* regulatory domain. **d**, Transcriptional changes in the *FOXA1*, *MIPOL1* and *TTC6* genes in wild-type ($n = 320$) and rearranged ($n = 50$) cases (two-sided *t*-test). Box plot centre, median; box, quartiles 1–3; whiskers, quartiles 1–3 ± 1.5 × IQR. FPKM, fragments per kilobase of transcript per million mapped reads. **e**, Class-3 model. Tandem duplications within the *FOXA1* topologically associating domain (TAD) amplify *FOXMIND* to drive overexpression of *FOXA1*.

and *TTC6*—but not of *MIPOL1*, which has its promoter outside of the *FOXA1* topologically associating domain (Extended Data Fig. 9d, e).

We found that translocations were largely within a 50-kb region between *FOXA1* and the 3′ untranslated region of *MIPOL1*, whereas break-end junctions from duplications mostly flanked the *FOXMIND-FOXA1* region (Fig. 4a, Extended Data Fig. 9f). For translocations, we delineated two patterns: (1) the hijacking of the *FOXMIND* enhancer; and (2) insertions upstream of the *FOXA1* promoter (Fig. 4c). The first pattern subsumes previously reported in-frame fusion genes that involve RP11-356O9.1, *ETV1*[29] and *SKIL*[30], as well as a newly reported *ASXL1* fusion (Supplementary Table 4). The second pattern inserts an oncogene (such as *CCNA1*) upstream of *FOXA1* (Fig. 4c). Notably, both mechanisms resulted in outlier expression of the translocated gene (Extended Data Fig. 9g). For duplications, which constitute 70% of all rearranged cases, we found *FOXMIND* and *FOXA1* to be co-amplified in 89% of the rearranged cases and never separated (Fig. 4c, bottom, Extended Data Fig. 9h), thus preserving the *FOXMIND-FOXA1* regulatory domain.

Next, while assessing the transcriptional effect of duplications, we found that levels of *FOXA1* mRNA were poorly correlated with copy number (Extended Data Fig. 10a), but highly sensitive to focal structural variants. Tandem duplications (ascertained at the RNA and DNA levels) significantly increased expression of *FOXA1* and *MIPOL1*, but not of *TTC6* (Fig. 4d). Translocations resulted in a modest decrease in expression levels of *FOXA1* (Extended Data Fig. 10b), despite a significant co-occurrence with tandem duplications (odds ratio = 3.89, Extended Data Fig. 10c). To investigate this further, we carried out haplotype-resolved, linked-read sequencing of MDA-PCA-2b cells, which contain a translocation of *FOXMIND* and *ETV1*. Here, *ETV1* translocation was accompanied by a focal tandem duplication in the non-translocated *FOXA1* allele (Extended Data Fig. 10d). The translocated *FOXA1* allele was inactivated, which resulted in monoallelic transcription (Extended Data Fig. 10e) without a net loss in *FOXA1* expression (266 fragments per kilobase of transcript per million mapped reads, 95th percentile in mCRPC). By contrast, RP11-356O9.1 retained biallelic expression (Extended Data Fig. 10f). In LNCaP cells, which also contain an *ETV1* translocation into the *FOXA1* locus, deletion of *FOXMIND* caused a significant reduction in *ETV1* expression (Extended Data Fig. 10g). Thus, translocations result in the loss of *FOXA1* expression from the allele *in cis*, which is rescued by tandem duplications of the allele *in trans*. Altogether, we propose a coalescent model in which class-3 structural variants duplicate or reposition

FOXMIND to drive overexpression of *FOXA1* or other oncogenes (Fig. 4e).

In summary, we identify three structural classes of FOXA1 alterations that differ in genetic associations and oncogenic mechanisms. We establish *FOXA1* as a principal oncogene in AR-dependent prostate cancer that is altered in 34.6% of mCRPC. Given the unique pathogenic features of the three classes, we have named them the 'FAST' (class-1), 'FURIOUS' (class-2) and 'LOUD' (class-3) alterations of FOXA1 (Figs. 2h, 3i, 4e, Supplementary Table 5, Supplementary Discussion). Structurally equivalent FOXA1 alterations are also found in other hormone-receptor-driven cancers, thus positioning FOXA1 as a promising target for therapeutic strategies in these malignancies.

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/s41586-019-1347-4.

Received: 22 May 2018; Accepted: 3 June 2019;
Published online 26 June 2019.

1. Gao, N. et al. Forkhead box A1 regulates prostate ductal morphogenesis and promotes epithelial cell maturation. *Development* **132**, 3431–3443 (2005).
2. Friedman, J. R. & Kaestner, K. H. The Foxa family of transcription factors in development and metabolism. *Cell. Mol. Life Sci.* **63**, 2317–2328 (2006).
3. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315–322 (2014).
4. Robinson, D. et al. Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015).
5. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* **163**, 1011–1025 (2015).
6. Ciriello, G. et al. Comprehensive molecular portraits of invasive lobular breast cancer. *Cell* **163**, 506–519 (2015).
7. Dalin, M. G. et al. Comprehensive molecular characterization of salivary duct carcinoma reveals actionable targets and similarity to apocrine breast cancer. *Clin. Cancer Res.* **22**, 4623–4633 (2016).
8. Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).
9. Jin, H.-J., Zhao, J. C., Ogden, I., Bergan, R. C. & Yu, J. Androgen receptor-independent function of FoxA1 in prostate cancer metastasis. *Cancer Res.* **73**, 3725–3736 (2013).
10. Jin, H.-J., Zhao, J. C., Wu, L., Kim, J. & Yu, J. Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. *Nat. Commun.* **5**, 3972 (2014).
11. Song, B. et al. Targeting FOXA1-mediated repression of TGF-β signaling suppresses castration-resistant prostate cancer progression. *J. Clin. Invest.* **129**, 156–582 (2019).
12. Robinson, J. L. L. et al. Androgen receptor driven transcription in molecular apocrine breast cancer is mediated by FoxA1. *EMBO J.* **30**, 3019–3027 (2011).
13. Robinson, J. L. L. et al. Elevated levels of FOXA1 facilitate androgen receptor chromatin binding resulting in a CRPC-like phenotype. *Oncogene* **33**, 5666–5674 (2014).
14. Pomerantz, M. M. et al. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat. Genet.* **47**, 1346–1351 (2015).
15. Cirillo, L. A. et al. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–289 (2002).
16. Iwafuchi-Doi, M. et al. The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol. Cell* **62**, 79–91 (2016).
17. Lupien, M. et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**, 958–970 (2008).
18. Barbieri, C. E. et al. Exome sequencing identifies recurrent *SPOP*, *FOXA1* and *MED12* mutations in prostate cancer. *Nat. Genet.* **44**, 685–689 (2012).
19. Yang, Y. A. & Yu, J. Current perspectives on FOXA1 regulation of androgen receptor signaling and prostate cancer. *Genes Dis.* **2**, 144–151 (2015).
20. Grasso, C. S. et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239–243 (2012).
21. Gao, J. et al. 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. *Genome Med.* **9**, 4 (2017).
22. Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364**, 412–420 (1993).
23. Li, J. et al. Structure of the forkhead domain of FOXA2 bound to a complete DNA consensus site. *Biochemistry* **56**, 3745–3753 (2017).
24. Sekiya, T., Muthurajan, U. M., Luger, K., Tulin, A. V. & Zaret, K. S. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. *Genes Dev.* **23**, 804–809 (2009).
25. Wang, Z. et al. BART: a transcription factor prediction tool with query gene sets or epigenomic profiles. *Bioinformatics* **34**, 2867–2869 (2018).
26. Behrens, J. et al. Functional interaction of β-catenin with the transcription factor LEF-1. *Nature* **382**, 638–642 (1996).
27. Daniels, D. L. & Weis, W. I. β-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nat. Struct. Mol. Biol.* **12**, 364–371 (2005).
28. Wang, W., Zhong, J., Su, B., Zhou, Y. & Wang, Y.-Q. Comparison of *Pax1/9* locus reveals 500-Myr-old syntenic block and evolutionary conserved noncoding regions. *Mol. Biol. Evol.* **24**, 784–791 (2007).
29. Tomlins, S. A. et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* **448**, 595–599 (2007).
30. Annala, M. et al. Recurrent SKIL-activating rearrangements in ETS-negative prostate cancer. *Oncotarget* **6**, 6235–6250 (2015).

METHODS

Cell culture. Most cell lines were originally purchased from the American Type Culture Collection (ATCC) and were cultured as per standard ATCC protocols. LNCaP-AR and LAPC4 cells were gifts from the laboratory of C. Sawyers (Memorial Sloan Kettering Cancer Center). Unless otherwise stated, for all the experiments LNCaP, PNT2, LNCaP-AR, C42B, 22RV1, DU145 and PC3 cells were grown in the RPMI 1640 medium (Gibco) and VCaP cells in the DMEM with Glutamax (Gibco) medium supplemented with 10% full bovine serum (FBS; Invitrogen). LAPC4 cells were grown in IMEM (Gibco) supplemented with 15% FBS and 1 nM of R1881. For the immortalized normal prostate cells: RWPE1 cells were grown in keratinocyte medium with regular supplements (Lonza); PNT2 cells were grown in RPMI medium with 10% FBS. HEK293 cells were grown in DMEM (Gibco) medium with 10% FBS. All cells were grown in a humidified 5% CO_2 incubator at 37 °C. All cell lines were tested once a fortnight to be free of mycoplasma contamination and genotyped every month at the University of Michigan Sequencing Core using Profiler Plus (Applied Biosystems) and compared with corresponding short tandem repeat profiles in the ATCC database to authenticate their identity in culture between passages and experiments.

Antibodies. For immunoblotting, the following antibodies were used: FOXA1 N-terminal (Cell Signaling Technologies: 58613S; Sigma-Aldrich: SAB2100835); FOXA1 C-terminal (Thermo Fisher Scientific: PA5-27157; Abcam: ab23738); AR (Millipore: 06-680); LSD1 (Cell Signaling Technologies: 2139S); vinculin (Sigma Aldrich: V9131); H3 (Cell Signaling Technologies: 3638S); GAPDH (Cell Signaling Technologies: 3683); β-actin (Sigma Aldrich: A5316); β-catenin (Cell Signaling Technologies: 8480S); vimentin (Cell Signaling Technologies: 5741S); phospho(S33/S37/T41)-β-catenin (Cell Signaling Technologies: 8814S); LEF1 (Cell Signaling Technologies: 2230S); AXIN2 (Abcam: ab32197); and TLE3 (Proteintech: 11372-1-AP).

For co-immunoprecipitation and ChIP–seq experiments, the following antibodies were used: FOXA1 N-terminal (Cell Signaling Technologies: 58613S); FOXA1 C-terminal (Thermo Fisher Scientific: PA5-27157); AR (Millipore: 06-680); V5 tag (R960-25); and TLE3 (Proteintech: 11372-1-AP).

Immunoblotting and nuclear co-immunoprecipitation. Cell lysates were prepared using the RIPA lysis buffer (Thermo Fisher Scientific; cat. no. 89900) and denatured in the complete NuPage 1× LDS/reducing agent buffer (Invitrogen) with 10 min heating at 70 °C. Between 10 and 25 μg of total protein was loaded per well, separated on 4–12% SDS polyacrylamide gels (Novex) and transferred onto 0.45-μm nitrocellulose membrane (Thermo Fisher Scientific; cat. no. 88018) using a semi-dry transfer system (Trans-blot Turbo System; BioRad) at 25 V for 1 h. The membrane was incubated for 1 h in blocking buffer (Tris-buffered saline, 0.1% Tween (TBS-T), 5% non-fat dry milk) and incubated overnight at 4 °C with primary antibodies. When samples were run on multiple gels for an experiment, multiple loading control proteins (GAPDH, β-actin, total H3 and vinculin) were probed on each membrane separately. Host-species-matched secondary antibodies conjugated to horseradish peroxidase (HRP; BioRad) were used at 1/20,000 dilution to detect primary antibodies and blots were developed using enhanced chemiluminescence (ECL Prime, Thermo Fisher Scientific) following the manufacturer's protocol.

For nuclear co-immunoprecipitation assays, 8–10 million cells ectopically overexpressing different V5-tagged FOXA1 variants and wild-type AR (or TLE3) were fractionated to isolate intact nuclei using the NE-PER kit reagents (Thermo Fisher Scientific; cat. no. 78835) and lysed in the complete IP lysis buffer (Thermo Fisher Scientific; cat. no. 87788). Nuclear lysates were incubated for 2 h at 4 °C with 30 μl of magnetic protein-G Dynabeads (Thermo Fisher Scientific; cat. no. 10004D) for pre-clearing. A fraction of the pre-cleared lysate was saved as input and the remainder was incubated overnight (12–16 h) with 10 μg of target protein antibody at 4 °C with gentle mixing. Next day, 50 μl of Dynabeads protein-G beads were added to the lysate–antibody mixture and incubated for 2 h at 4 °C. Beads were washed three times with IP buffer (150 nM NaCl; Thermo Fisher Scientific) and directly boiled in 1× NuPage LDS/reducing agent buffer (ThermoFisher Scientific; cat. no. NP0007 and NP0009) to elute and denature the precipitated proteins. These samples were then immunoblotted as described above with the exception of using protein A-HRP secondary (GE HealthCare; cat. no. NA9120-1ML) antibody for detection.

RNA extraction and quantitative polymerase chain reaction. Total RNA was extracted using the the miRNeasy Mini Kit (Qaigen), with the inclusion of the on-column genomic DNA digestion step using the RNase-free DNase Kit (Qaigen), following the standard protocols. RNA was quantified using the NanoDrop 2000 Spectrophotometer (ThermoFisher Scientific) and 1 μg of total RNA was used for complementary DNA (cDNA) synthesis using the SuperScript III Reverse Transcriptase enzyme (Thermo Fisher Scientific) following the manufacturer's instructions. Twenty nanograms of cDNA was input per polymerase chain reaction (PCR) using the FAST SYBR Green Universal Master Mix (Thermo Fisher Scientific) and every sample was quantified in triplicate. Gene expression was calculated relative to *GAPDH* and *HPRT1* (loading control) using the $\Delta\Delta C_t$ method and normalized to the control group for graphing. Quantitative PCR (qPCR) primers were designed using the Primer3Plus tool (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and synthesized by Integrated DNA Technologies.

Primer used in this study are listed below: *GAPDH*: forward (F), TGCACCACCA ACTGCTTAGC and reverse (R), GGCATGGACTGTGGTCATGAG; *HPRT1*: F, AGGCGAACCTCTCGGCTTTC and R, CTAATCACGACGCCAGGGCT; *ACTB*: F, AGGATGCAGAAGGAGATCACTG and R, AGTACTTGCGCTCAGGAGGAG; *AR*: F, CAGTGGATGGGCTGAAAAAT and R, GGAGCTTGGTGAGCTGGTAG; *FOXA1*-3′: F, GAAGACTCCAGCCTCCTCAACTG and R, TGCCTTGAAGTCCA GCTTATGC; *FOXA1*-5′: F, CTACTACGCAGACACGCAGG and R, CCGCTCGTAGTCATGGTGTT; *TLE3*: F, AAGGACAGCTTGAGCCGATA and R, TTTGGTCTTGGAGGAAGGTG; *TTC6*: F, CGAACAGAGCCAGGAGGT AG and R, GTTCTCCCTGGGCTCCTAAC; *MIPOL1*: F, GCAAACGGTTAGAGC AGGAG and R, GGGTCTGGATTTCCTCTTCC; *ETV1*: F, TACCCCATGGACC ACAGATT and R, CACTGGGTCGTGGTACTCCT; *TUBB*: F,CTGGACCGCATC TCTGTGTACT and R,GCCAAAAGGACCTGAGCGAACA.

siRNA-mediated gene knockdown. Cells were seeded in a 6-well plate at the density of 100,000–250,000 cells per well. After 12 h, cells were transfected with 25 nM of gene-targeting ON-TARGETplus SMARTpool siRNAs or non-targeting pool siRNAs as negative control (Dharmacon) using the RNAiMAX reagent (Life Technologies; cat. no. 13778075) on two consecutive days, following the manufacturer's instructions. Both total RNA and protein were extracted on day 3 (total 72 h) to confirm efficient (>80%) knockdown of the target genes. For crystal-violet staining, at day 9 growth medium was aspirated and cells were first fixed with 4% formaldehyde solution, followed by a 30-min incubation in 0.5% crystal-violet solution in 20% methanol, and then scanned. Catalogue numbers and guide sequences (5′ to 3′) of siRNA SMARTpools (Dharmacon) used are: non-targeting control (cat. no. D-001810-10-05; UGGUUUACAUGUCGACUAA, UGGUUUACAUGUUGUGUGA, UGGUUUACAUGUUUUCUGA, UGGUUUA CAUGUUUUCCUA); *AR* (cat. no. L-003400-00-0005; GAGCGUGGACUUUCCG GAA, UCAAGGAACUCGAUCGUAU, CGAGAGAGCUGCAUCAGUU, CAGAAAUGAUUGCACUAUU); *FOXA1* (cat. no. L-010319-00-0005; GCACUGCAAUACUCGCCUU, CCUCGGAGCAGCAGCAUAA, GAACAGCU ACUACGCAGAC, CCUAAACACUUCCUAGCUC); *TLE3* (cat. no. L-019929-00-0005; GCCAUUAUGUGAUGUACUA, GCAUGGACCCGAUAGGUAU, GAACCACCAUGAACUCGAU, UCAGGUCGAUGCCGGGUAA).

The *FOXA1* SMARTpool consists of siRNAs targeting 5′ as well as 3′ ends of the *FOXA1* transcript. Thus, both wild-type and class-2 mutant transcripts are degraded using the SMARTpool siRNAs. This was experimentally confirmed in LAPC4 cells that endogenously contain a *FOXA1* class-2 mutation (Extended Data Fig. 1d, e).

CRISPR–Cas9-mediated gene or enhancer knockout. Cells were seeded in a 6-well plate at the density of 200,000–300,000 cells per well and infected with viral particles with lentiCRISPR-V2 plasmids coding either non-targeting (sgNC) or single guide RNAs (sgRNAs) targeting the exon 1 or the FKHD of *FOXA1* (both resulting in FOXA1 inactivation). This was followed by three days of puromycin selection, after which proliferation assays were carried out as described below. The lentiCRISPR-V2 vector was a gift from the laboratory of F. Zhang (Addgene plasmid no. 52961).

sgRNA sequences used are as follows: sgNC no. 1: 5′-GTAGCGAACGTGTCC GGCGT-3′; sgNC no. 2: 5′-GACCGGAACGATCTCGCGTA-3′; sg*FOXA1* exon 1: 5′-GTAGTAGCTGTTCCAGTCGC-3′; sg*FOXA1* FKHD: 5′-GCCGTTCTCGAACATGTTGC-3′.

Alternatively, for functional interrogation of the *FOXA1* topologically associating domain (TAD) enhancer elements, VCaP or LNCaP cells were transfected with pairs of sgRNAs targeting the *MIPOL1* untranslated region (UTR) or *FOXMIND* or a control locus within the *FOXA1* TAD. Transfected cells were then selected with puromycin (1.0 μg/ml) for 48 h, followed by incubation for an additional 72 h. Total RNA was extracted and qPCR was performed as described above.

Pairwise sgRNA sequences are as follows (5′ to 3′): control sgRNA (sgCtrl): CA CCGATTAGCCTCAACTATACCA and CACCGTGCAATATCTGAATCACACG; sgMIPOL1 UTR: CACCGTGAAAAAAAACGACAGTCTG and CACCGAACTC AAGTCAGCAGCAAAG; sgFOXMIND 1: CACCGCTTTAATAAAGCTATTTGC and CACCGATAGAGTGACTAATGCCCTG; sgFOXMIND 2: CACCGTAACAGT TGACCTACTAAC and CACCGATTTAGATAAGGGGATAGAA; sgFOX-MIND 3: CACCGCTTTAATAAAGCTATTTGC and CACCGATTTAG ATAAGGGGATAGAA.

CRISPR knockout screen. For the genome-wide CRISPR knockout screen, a two-vector system was used. First, LNCaP cells were engineered to stably overexpress the enzymatically active Cas9 protein. These cells were then treated with the human GeCKO knockout sgRNA library (GecKO V2) that was a gift from the Zhang laboratory (Addgene; cat. no. 1000000049). This was followed by puromycin

selection for 48 h, after which a fraction of these cells was processed to isolated genomic DNA as the input sample. The remaining cells were then cultured for 30 days, and genomic DNA was extracted at this time point. sgRNA sequences were amplified using common adaptor primers and sequenced on the Illumina HiSeq 2500 (125-nucleotide read length). Sequencing data were analysed as described[31] and depletion or enrichment of individual sgRNAs at 30 days was calculated relative to the input sample. Note that only a subset of genes—including essential controls, epigenetic regulators and transcription factors from the GeCKO-V2 screen—was plotted in Extended Data Fig. 1i.

Proliferation assays. For siRNA growth assays, cells were directly plated in a 96-well plate at the density of 2,500–8,000 cells per well and transfected with gene-specific or non-targeting siRNAs, as described above, on day 0 and day 1. Every treatment was carried out in six independent replicate wells. CellTiter-Glo reagent (Promega) was used to assess cell viability at multiple time points after tranfection, following the manufacturer's protocol. Data were normalized to readings from siNC treatment on day 1, and plotted as relative cell viability to generate growth curves.

Alternatively, for CRISPR sgRNA growth assays, cells were treated as described above for target-gene inactivation and seeded into a 24-well plate at 20,000 cells per well, with 2 replicates per group. After 12 h, plates were placed into the IncuCyte live-cell imaging machine (IncuCyte) set at the phase-contrast option to record cell confluence every 3 h for between 7 and 9 days. Similarly, for class-1 growth assays (Fig. 2f), stable doxycycline-inducible 22RV1 cells were grown in 10% charcoal-stripped-serum (CSS)-supplemented medium for 48 h. Androgen-starved cells were then seeded into a 96-well plate at 5,000 cells per well in 10% CSS medium with or without addition of doxycycline (1 μg/ml) to induce control or mutant protein expression (6 replicates per group). Once adherent, treated cells were placed in the IncuCyte live-cell imaging machine set at phase contrast to record cell confluence every 3 h for between 7 and 9 days. In all IncuCyte assays, confluence measurements from all time points were normalized to the matched measurement at 0 h and plotted as relative confluence to generate growth curves.

Cloning of representative FOXA1 mutants. Wild-type *FOXA1* coding sequence was purchased from Origene (cat. no. SC108256) and cloned into the pLenti6/V5 lentiviral vector (Thermo Fisher Scientific; cat. no. K4955-10) using the standard TOPO cloning protocol. Class-1 missense mutations (I176M, H247Q and R261G) were engineered from the wild-type FOXA1 vector using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Tech) as per the manufacturer's instructions. All point mutations were confirmed using Sanger sequencing through the University of Michigan Sequencing Core Facility. Engineered mutant plasmids were further transfected in HEK293 cells to confirm expression of the mutant protein. For truncated class-2 variants, the wild-type coding sequence up to the amino acid before the intended mutation was cloned. All FOXA1 variants had the V5 tag fused on the C terminus. Selected mutants were cloned into a doxycycline-inducible vector (Addgene: pCW57.1; cat. no. 41393) to generate stable lines. For FRAP and single particle tracking assays, the pCW57.1 vector was edited to incorporate an in-frame GFP or Halo coding sequences at the C-terminal end, respectively.

FRAP assay and data quantification. PNT2 cells were seeded in a 6-well plate at 200,000 cells per well, and transfected with 2 μg of doxycycline-inducible vectors that encoded different variants of FOXA1 fused to GFP on the C-terminal end. After 24 h, cells were plated in glass-bottom microwell dishes (MatTek; #P35G-1.5-14-C) in phenol-free growth medium supplemented with doxycycline (1 μg/ml). Cells were then incubated for 48 h to allow for robust expression of the exogenous GFP-tagged protein and strong adherence to the glass surface. Microwell dishes were placed in humidity-controlled chamber set at 37 °C (Tokai-Hit) and mounted on the SP5 Inverted 2-Photon FLIM Confocal microscope (Leica). FRAP Wizard from the Leica Microsystems software suite was used to conduct and analyse FRAP experiments. Fluorescence signals were automatically computed in regions of interest using in-built tools in the FRAP Wizard. Roughly half of the nucleus was photobleached using the argon laser at 488 nm and 100% intensity for 20–30 iterative frames at 1.2-s intervals. Laser intensity was reduced to 1% for imaging post bleaching. Immediately after photobleaching, 2 consecutive images were collected at 1.2-s intervals followed by images taken at 10-s intervals for 60 frames (that is, 10 min).

For data analyses, recovery of signal in the bleached half and loss of signal in the unbleached half were measured as average fluorescence intensities in at least 80% of the respective areas, excluding the immediate regions flanking the separating border. All intensity curves were generated from background-subtracted images. The fluorescence signal measured in a region of interest was normalized to the signal before bleaching using the following formula[32]: $R = (I_t - I_{bg})/(I_o - I_{bg})$, in which I_o is the average intensity in the region of interest before bleaching, I_t is the average intensity in the region of interest at any time-point after bleaching and I_{bg} is the background fluorescence signal in a region outside of the cell nucleus. Raw recovery kinetic data from above were fitted with best hyperbolic curves using the GraphPad Prism software and the time until 50% recovery was calculated from the resulting best-fit equations. For representative time-lapse nuclei images shown in the FRAP figures, the fluorescence signal was uniformly brightened for ease of visualization.

Single particle tracking experiment and data quantification. PNT2 cells were transiently transfected with doxycycline-inducible vectors encoding C-terminal Halo-tagged wild-type or class-1 mutant variants of FOXA1. Transfected cells were seeded in glass-bottomed DeltaT culture dishes (Bioptechs; cat. no. 04200417C) and incubated for 24 h with 0.01 μg/ml of doxycycline. Cells were then treated with phenol-red-free medium containing 2% FBS and 5 nM cell permeable JF549 Halo ligand dye[33] for 30 min at 37 °C. Cells were subsequently washed twice, 10 min per wash at 37 °C, with phenol-red-free medium containing 2% FBS. Before imaging, cells were washed once with the 1× HBSS buffer and were imaged in the buffer.

Single particle tracking was performed on an Olympus IX81 microscope via HILO illumination, as previously described[34], at a spatial accuracy of 30 nm and temporal resolution of 33 ms. Image analysis was performed as previously described[35]. In brief, tracking was done in Imaris (bitplane) and particles that were at least visible for four continuous frames were used for further analysis. Diffusion constants were calculated as previously described[36], assuming a Brownian diffusion model under steady-state conditions. Dwell time histograms were fit to a double-exponential function to extract fast and slow dwell times of 'bound' particles that displayed a frame-to-frame displacement of <300 nm. All particles that were visible for less than 4 consecutive frames, or those that moved >300 nm between frames, were counted as 'unbound' particles. At least 5 cells were imaged for each transcription factor variant and >500 particles were tracked to extract diffusion constants and dwell time.

Dual luciferase AR reporter assay. HEK293 cells stably overexpressing the wild-type AR protein (that is, HEK293-AR) were used for the AR reporter assays. HEK293-AR cells were seeded in a 12-well plate at 300,000 cells per well and transfected with 2 μg of the pLenti6/V5 vector encoding different variants of FOXA1, or GFP (control). After 8 h, medium was replaced with 10% CSS-supplemented phenol-free medium (androgen-depleted) and cells were transfected with the AR reporter Firefly luciferase or negative-control constructs from the Cignal AR-Reporter(luc) kit (Qiagen; cat. no. CCS-1019L) as per the manufacturer's instructions. Both constructs were premixed with constitutive *Renilla* luciferase vector as control. After 12 h, cells were treated with different dosages of DHT or enzalutamide (at 10 μM dosage); and additional 24 h later dual luciferase activity was recorded for every sample using the Dual-Glo Luciferase assay (Promega; E2980) and luminescence plate reader (Promega-GLOMAX-Multi Detection System). Each treatment condition had four independent replicates. Firefly luciferase signals were normalized with the matched *Renilla* luciferase signals to control for variable cell number and/or transfection efficiencies, and normalized signals were plotted relative to the negative control reporter constructs.

Electrophoretic mobility shift assay. HEK293 cells were plated in 10-cm dishes at 1 million per plate and transfected with 10 μg of the pLenti6/V5 vector coding GFP (control) or different variants of FOXA1. After 48 h, cells were trypsinized and nuclear lysates were prepared using the NE-PER kit reagents (Thermo Fisher Scientific). Immunoblots were run to confirm comparable expression of recombinant FOXA1 variants in 2 μl (that is, equal volume) of final nuclear lysates. Next, FOXA1 and AR ChIP–seq data were used to identify the *KLK3* enhancer element. Sixty base pairs of the *KLK3* enhancer, centred at the *FOXA1* consensus motif 5′-GTAAACAA-3′, were synthesized as single-stranded oligonucleotides (IDT) and biotin-labelled using the Biotin 3′-End DNA labelling kit (Thermo Fisher Scientific), and then annealed to generate a labelled double-stranded DNA duplex.

Binding reactions were carried out in 20-μl volumes containing 2 μl of the nuclear lysates, 50 ng/μl poly(dI.dC), 1.25% glycerol, 0.025% Nonidet P-40 and 5 mM $MgCl_2$. Biotin-labelled *KLK3* enhancer probe (10 fmol) was added at the very end with gentle mixing. Reactions were incubated for 1 h at room temperature, size-separated on a 6% DNA retardation gel (100 V for 1 h; Invitrogen) in 0.5× TBE buffer, and transferred on the Biodyne Nylon membrane (0.45 μm; Thermo Fisher Scientific) using a semi-dry system (BioRad). Transferred DNA was crosslinked to the membrane using the UV light at 120 mJ/cm^2 for 1 min. Biotin-labelled free and protein-bound DNA was detected using HRP-conjugated streptavidin (Thermo Fisher Scientific) and developed using chemiluminescence according to the manufacturer's protocol.

Protein synthesis and purification. First, wild-type FOXA1 and FOXA1(P358fs) proteins were purified using the *Escherichia coli* bacterial expression system and nickel-affinity chromatography. In brief, wild-type FOXA1 or FOXA1(P358fs) coding sequences were cloned into the pFC7A (HQ) Flexi vector (Promega; cat. no. C8531) with a C-terminal HQ tag, following the manufacturer's protocol. These expression constructs were used to transform Single Step (KRX) Competent *E. coli* cells (Promega; cat. no. L3002), which have been modified for synthesis of mammalian proteins. A starter broth of 2 ml was inoculated with a single colony of transformed bacterial cells and incubated at 37 °C with constant shaking at

250 rpm until an optical density at 600 nm (OD_{600}) of 0.4–0.5 was reached. The starter brother was then used to inoculate 1,000 ml of LB broth containing ampicillin, and protein synthesis was induced using 0.1% v/v of rhamanose. Induced culture was incubated at 20 °C for 16 h with constant shaking at 250 rpm. Bacterial cells were then pelleted by centrifugation at 4,000 rpm for 30 min and mechanically lysed through sonication in 50 mM Tris (pH 7.4), 150 mM NaCl, 1 mM $MgCl_2$, 0.5 mM EDTA, 1 mM DTT and 1% glycerol in the presence of protease inhibitors (Roche). HisLink Purification Resin (Promega; cat. no. V8821) was used to purify untagged recombinant proteins from the crude bacterial lysates as per the manufacturer's protocol (this also includes removal of the His tag). Purified protein fractions were then tested for purity by Coomaisse staining relative to the crude input lysates, and purified protein concentrations were estimated using protein standards of known concentrations (Thermo Fisher Scientific; cat. no. 23208). The identities of purified proteins were confirmed via immunoblotting using an N-terminal FOXA1 antibody (Cell Signaling Technology; cat. no. 58613S).

Biolayer interferometry assay. Biolayer interferometry (BLI) assays were carried out using the Octet-RED96 system (PALL ForteBio) and in-built analysis software. In brief, a biotin-labelled, 60-bp *KLK3* enhancer element centred at the FOXA1 consensus motif was immobilized on the Super Streptavidin Biosensors (PALL ForteBio, part no. 18-5057) with the loading step carried out for 1,000 s with shaking at 500 rpm. This was followed by baseline measurements for 120 s and association for 900 s using varying concentrations of purified FOXA1 proteins (3.125–100 nM; two replicate biosensors per concentration). A control DNA element with no FOXA1 motif was used in the negative-control reaction with 100 nM of the protein. The association step was followed by the dissociation step for 3,000 s. Signal from all the biosensors was adjusted for the background signal from the control sensors and normalized data of DNA binding kinetics were analysed using the Octet-RED96 (PALL ForteBio) analysis software, as previously described[37].

Generation of CRISPR clones and stable lines. 22RV1 or LNCaP cells were seeded in a 6-well plate at 200,000 cells per well and transiently transfected with 2.5 μg of lentiCRISPR-V2 (Addgene; 52961) vector using the Lipofectamine 3000 reagent (cat. no. L3000008), encoding the Cas9 protein and sgRNA that cuts either at amino acid 271 (5′-GTCAAGTGCGAGAAGCAGCCG-3′) or 359 (5′-GCCGGGCCCGGAGCTTATGGG-3′) in exon 2 of *FOXA1*. Cells were treated with non-targeting control sgRNA (5′-GACCGGAACGATCTCGCGTA-3′) vector to generate isogenic wild-type clones. Transfected cells were selected with puromycin (Gibco) for 3–4 days and sorted by fluorescence-activated cell sorting as single cells into 96-well plates. Cells were maintained in 96-well plates for 4–6 weeks, with replacement of the growth medium every 7 days to allow for the expansion of clonal lines. Clones that successfully seeded were further expanded and genotyped for *FOXA1* using Sanger sequencing, and immunoblotting with the N-terminal FOXA1 antibody. Sequence- and expression-validated 22RV1 and LNCaP clones with distinct class-2 mutations were used for growth, invasion and metastasis assays as described.

To generate stable cells, doxycycline-inducible vectors coding different variants of FOXA1 or GFP (control) were packaged into viral particles at the University of Michigan Vector Core. Prostate cancer cells were seeded in a 6-well plate at 100,000–250,000 cells per well and infected with 0.5 ml of 10× viral titres packaged at the University of Michigan Vector Core. This was followed by 3–4 days of puromycin (Gibco) selection to generate stable lines.

Rescue growth and functional compensation experiments. Stable 22RV1 cells with doxycycline-inducible expression of empty vector (control), FOXA1 wild type, or distinct FOXA1 mutants were seeded in a 6-well plate in the completed growth medium supplemented with 1.0 μg/ml of doxycycline. Notably, the exogenous genes only contain the coding sequence of *FOXA1* without its intron and UTRs. After 24 h, cells were transfected with 30 nM of either distinct 3′ UTR-specific *FOXA1*-targeting siRNAs or a non-targeting control siRNA using the RNAiMAX (Life Technologies; cat. no. 13778075) reagent. *FOXA1* UTR-specific siRNAs were purchased from Thermo Fisher Scientific (cat. no. siNC, 4390844 (sequence is proprietary); siRNA no. 3, s6687 (sense sequence: 5′-GCAAUACUCUUAACCAUAA-3′); siRNA no. 4, 5278 (sense sequence: 5′-AACACATAAAATTAGTTTC-3′); and siRNA no. 5 – 107428 (sense sequence: 5′-AAGTTATAGGGAGCTGGAT-3′)). On the following day, cells were counted and seeded in a 96-well plate at a density of 5,000 cells per well with 6 replicates for each treatment condition. Cell growth was then assessed using the IncuCyte assay, as described above.

Testing the GFP-tagged wild-type FOXA1 variant. 22RV1 cells were seeded in 10-cm dishes and transfected with 8 μg of mammalian expression plasmids encoding either FOXA1(WT) or FOXA1(WT)–GFP (the exact construct used in the FRAP assay) using the Lipofectamine 3000 (Life Technologies; cat. no. L3000008) reagent, as per the manufacturer's protocol. Transgene expression was induced using 1.0 μg/ml of doxycycline and cells were cultured for 96 h with doxycycline replenishment every 48 h. Total RNA was extracted and RNA-seq was performed as described. A portion of these cells was used for the rescue growth experiments using UTR-specific *FOXA1* siRNAs as described above.

Matrigel invasion assay. 22RV1 CRISPR clones were grown in 10% CSS-supplemented medium for 48 h for androgen starvation. A matrigel-coated invasion chamber was used, which was additionally coated with a light-tight polyethylene terephthalate membrane to allow for fluorescent quantification of the invaded cells (Biocoat: 24-well format, no. 354166). Fifty thousand starved cells were resuspended in serum-free medium and were added to each invasion chamber. Twenty per cent FBS-supplemented medium was added to the bottom wells to serve as a chemoattractant. After 12 h, medium from the bottom well was aspirated and replaced with 2 μg/ml Calcein-green AM dye (Thermo Fisher Scientific; C3100MP) in 1× HBSS (Gibco) and incubated for 30 min at 37 °C. Invasion chambers were then placed in a fluorescent plate reader (Tecan-Infinite M1000 PRO) and fluorescent signals from the invaded cells at the bottom were averaged across 16 distinct regions per chamber to determine the extent of invasion.

ChIP–seq. ChIP experiments were carried out using the HighCell# ChIP-Protein G kit (Diagenode) as per the manufacturer's protocol. Chromatin from five million cells was used per ChIP reaction with 6.5 μg of the target protein antibody. In brief, cells were trypsinized and washed twice with 1× PBS, followed by crosslinking for 8 min in 1% formaldehyde solution. Crosslinking was terminated by the addition of 1/10 volume 1.25 M glycine for 5 min at room temperature followed by cell lysis and sonication (Bioruptor, Diagenode), resulting in an average chromatin fragment size of 200 bp. Fragmented chromatin was then used for immunoprecipitation using various antibodies, with overnight incubation at 4 °C. ChIP DNA was de-crosslinked and purified using the iPure Kit V2 (Diagenode) using the standard protocol. Purified DNA was then prepared for sequencing as per the manufacturer's instructions (Illumina). ChIP samples (1–10 ng) were converted to blunt-ended fragments using T4 DNA polymerase, *E. coli* DNA polymerase I large fragment (Klenow polymerase) and T4 polynucleotide kinase (New England BioLabs (NEB)). A single A base was added to fragment ends by Klenow fragment (3′ to 5′ exo minus; NEB) followed by ligation of Illumina adaptors (Quick ligase, NEB). The adaptor-ligated DNA fragments were enriched by PCR using the Illumina Barcode primers and Phusion DNA polymerase (NEB). PCR products were size-selected using 3% NuSieve agarose gels (Lonza) followed by gel extraction using QIAEX II reagents (Qiagen). Libraries were quantified and quality checked using the Bioanalyzer 2100 (Agilent) and sequenced on the Illumina HiSeq 2500 Sequencer (125-nucleotide read length).

Zebrafish embryo metastasis experiment. Wild-type AB^{TL} zebrafish were maintained in aquaria according to standard protocols. Embryos were generated by natural pairwise mating and raised at 28.5 °C on a 14 h light/10 h dark cycle in a 100-mm Petri dish containing aquarium water with methylene blue to prevent fungal growth. All experiments were performed with 2–7-day-old embryos post-fertilization, and were done in approved University of Michigan fish facilities using protocols approved from the University of Michigan Institutional Animal Care and Use Committee (UM-IACUC). Cell injections were carried out as previously described[38]. In brief, GFP-expressing normal (control) or cancer cells were resuspended in PBS at the concentration of 1×10^7 cells/ml. Forty-eight hours after fertilization, wild-type embryos were dechorionated and anaesthetized with 0.04 mg/ml tricaine. Approximately 10 nl (approximately 100 cancer cells) were microinjected into the perivitelline space using a borosilliac micropipette tip with filament. Embryos were returned to aquarium water and washed twice to remove tricaine, then moved to a 96-well plate with one embryo per well and kept at 35 °C for the duration of the experiment. All embryos were imaged at 24-h intervals to follow metastatic dissemination of injection cells. Water was changed daily to fresh aquarium water. More than 30 fish were injected for each condition (wild-type no. 2, $n = 30$; wild-type no. 5, $n = 50$; no. 57, $n = 35$; no. 84, $n = 57$; no. 113, $n = 38$) and metastasis was visually assessed daily up to 5 days after injection (that is, for a total of 7 days post-fertilization) by counting the total number of distinct cellular foci in the body of the embryos. All of the metastasis studies were terminated at seven days post-fertilization in accordance with the approved embryo protocols. Embryos were either imaged directly in the 96-well plates or placed onto a concave glass slide to capture representative images using a fluorescent microscope (Olympus-IX71). For quantification, evidently distinct cell foci in the embryo body were counted 72 h after the injections.

For all these experiments, relevant ethical regulations were carefully followed. No statistical methods were used to predetermine sample size for any of the cohort analyses or experiments. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment unless otherwise stated.

ATAC-seq and data analysis. ATAC-seq was performed as previously described[39]. In brief, 25,000 normal prostate or prostate cancer cells were washed in cold PBS and resuspended in cytoplasmic lysis buffer (CER-I from the NE-PER kit, Invitrogen, cat. no. 78833). This single-cell suspension was incubated on ice for 10 min with gentle mixing by pipetting at every 2 min. The lysate was centrifuged at 1,300*g* for 5 min at 4 °C. Nuclei were resuspended in 2× TD buffer, then incubated with Tn5 enzyme for 30 min at 37 °C (Nextera DNA Library Preparation Kit;

cat. no. FC-121-1031). Samples were immediately purified by Qiagen minElute column and PCR-amplified with the NEBNext High-Fidelity 2X PCR Master Mix (NEB; cat. no. M0541L). qPCR was used to determine the optimal PCR cycles to prevent over-amplification. The amplified library was further purified by Qiagen minElute column and SPRI beads (Beckman Coulter; cat. no. A63881). ATAC-seq libraries were sequenced on the Illumina HiSeq 2500 (125-nucleotide read length).

Paired-end .fastq files were uniquely aligned to the hg38 human genome assembly using Novoalign (Novocraft) (with the parameters -r None -k -q 13 -k -t 60 -o sam –a CTGTCTCTTATACACATCT), and converted to .bam files using SAMtools (version 1.3.1). Reads mapped to mitochondrial or duplicated reads were removed by SAMtools and PICARD MarkDuplicates (version 2.9.0), respectively. Filtered .bam files from replicates were merged for downstream analysis. MACS2 (2.1.1.20160309) was used to call ATAC-seq peaks. The coverage tracks were generated using the program bam2wig (http://search.cpan.org/dist/Bio-ToolBox/) with the following parameters:–pe–rpm–span–bw. Bigwig files were then visualized using the IGV (Broad Institute) open source genome browser.

ChIP–seq data analysis. Paired-end 125-bp reads were trimmed and aligned to the GRCh38 human reference using the STAR (version 2.4.0g1) aligner with splicing disabled; the resulting reads were filtered using samtools 'samtools view -@ 8 -S -1 -F 384'. The resulting .bam file was sorted and duplicate-marked using Novosort, and converted into a bigwig file for visualization using 'bedtools genomecov -bg -split -ibam' and 'bedGraphToBigWig'. The coverage signal was normalized to total sequencing depth/1 $\times$ 10^6 reads. Peak calling was performed using MACS2 with the following settings: 'macs2 callpeak–call-summits–verbose 3 -g hs -f BAM -n OUT–qvalue 0.05'. ChIP peak profile plots and read-density heat maps were generated using deepTools2[40], and cistrome overlap analyses were carried out using the ChIPpeakAnno[41] package in R. It is important to note that, given the cistromic dominance of class-2 mutants, in heterozygous class-2 mutant clones part of the FOXA1 antibody binds to the wild-type protein that does not interact with, or immunoprecipitate, the DNA. This confounds all analyses involving peak-read density comparisons between the wild-type and class-2-mutant FOXA1 ChIP–seq data; we therefore largely avoided this strategy in our study. For the same reason, the read densities from only the heterozygous clones were factored by 1.5 for heat map generation in Fig. 3d.

De novo and known motif enrichment analysis. All de novo and known motif enrichment analyses were performed using the HOMER (v.4.10) suite of algorithms[42]. Peaks were called by the findPeaks function (-style factor -o auto) at 0.1% false discovery rate; de novo motif discovery and enrichment analysis of known motifs were performed with findMotifsGenome.pl (-size 200 -mask). For motif analysis of common wild-type- and mutant-specific chromatin binding sites, the top 5,000 peaks ranked by score were used as input. A common set of background sequences was generated by di-nucleotide shuffling of the input sequences using the fasta-shuffle-letters function from MEME[43]. Alternatively, we ranked peaks by the relative signal fold change between mutant and wild type, and selected the top and bottom 5,000 peaks (keeping the requirement that mutant-specific peaks are not called in the wild-type cistrome, and vice versa) for motif discovery. For class-2 mutants, only heterozygous 22RV1 clones were used, which more accurately recapitulate the clinical presentation of *FOXA1* mutations. Also, for both mutational classes, cistromes from biological replicates were merged to define a union cistrome that was compared to the union wild-type cistrome generated from matched FOXA1 wild-type cells. For the supervised motif analyses, we identified all instances of the FOXA canonical motif (5′-T[G/A]TT[T/G]AC-3′) within cistromes (ChIP–seq peaks) of class-1 and wild-type FOXA1 proteins using motifmatchR, and calculated nucleotide frequencies in the flanking positions.

Cohorts, datasets and resources. This study uses previously published public or restricted patient genetic data. Genetic calls for primary prostate cancer and breast cancer were obtained from the Genomic Data Commons (GDC)[44] for the prostate cancer PRAD[5] and breast cancer BRCA[6,45] cohorts, respectively. Raw RNA-seq data (paired-end reads from unstranded polyA libraries) for the samples were downloaded from the GDC and processed with our standard clinical RNA-seq pipeline CRISPR/CODAC (see below). For The Cancer Genome Atlas (TCGA) PRAD and BRCA cohorts, we downloaded mutational calls from multiple sources (GDC, cBio Portal and UCSC Xena) and additionally used the BAM-slicing tool to download sequence alignments from whole-exome sequencing libraries to the *FOXA1* locus. We then used our internal pipeline (see below) to call single-nucleotide variants and indels within *FOXA1*. We also used the downloaded aligned data for manual review of *FOXA1* mutation calls. Mutation calls for advanced primary and metastatic cases were obtained from the MSK-IMPACT cohort (downloaded from the cBio portal[46]). The main MCTP mCRPC cohort includes 360 previously reported cases (the location of all raw .bam files is provided in ref. [47]), the 10 additional mCRPC cases included here (but not in ref. [47]) will be included in the Database of Genotypes and Phenotypes (dbGaP) under accession code phs000673.v3.p1, and belong to a continuous sequencing program with the same IRB-approved protocol (MI-Oncoseq program, University of Michigan Clinical Sequencing Exploratory Research). The genetic sequencing data (WXS) for rapid autopsy cases are available from dbGaP with accession codes hs000554.v1.p1and phs000567.v1.p1. De-identified somatic mutation calls, RNA-seq fusion calls, processed and segmented copy-number data, and RNA-seq expression matrices across the full 370 cases of the MCTP mCRPC cohort are available on request from the authors.

Preparation of whole-exome sequencing and RNA-seq libraries. Integrative clinical sequencing (comprising exome sequencing and polyA and/or capture RNA-seq) was performed using standard protocols in our Clinical Laboratory Improvement Amendments-compliant sequencing laboratory. In brief, tumour genomic DNA and total RNA were purified from the same sample using the AllPrep DNA/RNA/miRNA kit (Qiagen). Matched normal genomic DNA from blood, buccal swab or saliva was isolated using the DNeasy Blood & Tissue Kit (Qiagen). RNA-seq was performed using the exome-capture transcriptome platform[48]. Exome libraries of matched pairs of tumour and normal DNA were prepared as previously described[49], using the Agilent SureSelect Human All Exon v4 platform (Agilent). All the samples were sequenced on an Illumina HiSeq 2000 or HiSeq 2500 (Illumina) in paired-end mode. The primary base call files were converted into FASTQ sequence files using the bcl2fastq converter tool bcl2fastq-1.8.4 in the CASAVA 1.8 pipeline.

Analysis of whole-exome sequencing data. The .fastq sequence files from whole-exome libraries were processed through an in-house pipeline constructed for analysis of paired tumour and normal data. The sequencing reads were aligned to the GRCh37 reference genome using Novoalign (version 3.02.08) (Novocraft) and converted into .bam files using SAMtools (version 0.1.19). Sorting, indexing, and duplicate marking of .bam files used Novosort (version 1.03.02). Mutation analysis was performed using freebayes (version 1.0.1) and pindel (version 0.2.5b9). Variants were annotated to RefSeq (via the UCSC genome browser, retrieved on 22 August 2016), as well as COSMIC v.79, dbSNP v.146, ExAC v.0.3 and 1000 Genomes phase 3 databases using snpEff and snpSift (v.4.1g). Single nucleotide variants and indels were called as somatic if they were present with at least 6 variant reads and 5% allelic fraction in the tumour sample, and present at no more than 2% allelic fraction in the normal sample with at least 20$\times$ coverage. Additionally, the ratio of variant allelic fractions between tumour and normal samples was required to be at least six to avoid sequencing and alignment artefacts at low allelic fractions. Minimum thresholds were increased for indels observed to be recurrent across a pool of hundreds of platform- and protocol-matched normal samples. Specifically, for each such indel, a logistic regression model was used to model variant and total read counts across the normal pool using PCR duplication rate as a covariate, and the results of this model were used to estimate a predicted number of variant reads (and therefore allelic fraction) for this indel in the sample of interest, treating the total observed coverage at this genomic position as fixed. The variant read count and allelic fraction thresholds were increased by these respective predicted values. This filter eliminates most recurrent indel artefacts without affecting our ability to detect variants in homopolymer regions from tumours exhibiting microsatellite instability. Germline variants were called using 10 variant reads and 20% allelic fraction as minimum thresholds, and were classified as rare if they had less than 1% observed population frequency in both the 1000 Genomes and ExAC databases. Exome data were analysed for copy-number aberrations and loss of heterozygosity by jointly segmenting B-allele frequencies and $\log_2$-transformed tumour/normal coverage ratios across targeted regions using the DNAcopy (version 1.48.0) implementation of the Circular Binary Segmentation algorithm. The expectation–maximization algorithm was used to jointly estimate tumour purity and classify regions by copy-number status. Additive adjustments were made to the $\log_2$-transformed coverage ratios to allow for the possibility of non-diploid tumour genomes; the adjustment resulting in the best fit to the data using minimum mean-squared error was chosen automatically and manually overridden if necessary.

Detection of copy-number break ends from whole-exome sequencing. The output of our clinical whole-exome sequencing pipeline includes segmented copy-number data, inferred absolute copy numbers and predicted parent-specific genotypes (for example, AAB), detection of loss of heterozygosity, and detection of copy-neutral loss of heterozygosity (uniparental disomy). Together, these data enable the detection of joint discontinuities in the copy-number profile (log-ratio and B-allele frequencies) at exon-level resolution. A subset of genomic rearrangements results in changes in copy number or allelic shifts, and the presence of such discontinuities in paired tumour-normal whole-exome sequencing data are therefore strongly indicative of a somatic breakpoint. For example, one copy gain will result in a segment with an increased log-ratio, and a corresponding zygosity deviation (see above). This segment will be discontinuous with adjacent segments, which will result in the call of a whole-exome sequencing break end (discontinuity) on either side of the copy gain. The size of the break end depends on the density of covered exons and in general the resolution is better in genic versus intergenic regions. We assessed the presence of such breakpoints within the gene-dense and exon-dense *FOXA1* locus; all copy-number break ends met statistical thresholds of the circular binary segmentation (CBS) algorithm (see above) at either the log-ratio or B-allele level.

Genetic characterization of mCRPC tumour samples at the pathway level. The co-occurrence or mutual exclusivity of *FOXA1* alterations with other previously described genetic events in prostate cancer has been carried out at the pathway level, but grouping putative functionally equivalent (and largely genetically mutually exclusive) events. All known types of ETS fusion (*ERG*, *ETV1*, *FLI1*, *ETV4* and *ETV5*) were considered as ETS-positive tumours, PI3K alterations included *PTEN* homozygous loss, *PIK3CA* activating mutations and *PIK3R1* inactivating mutations, AR pathway alterations included *AR*, *NCOR1*, *NCOR2* and *ZBTB16* mutations or deletions, but excluded *AR* amplifications and copy gains. The KMT category included mutations in all recurrently mutated lysine methyltransferases. The WNT category included inactivating alterations in *APC* and activating mutations in *CTNNB1*. DRD included cases with mutations in *BRCA1*, *BRCA2*, *PALB2* and *ATM* (all common mismatch repair genes), and *CDK12*.

Assessment of two-hit biallelic alterations. To assess the frequency of genetic inactivations of both alleles we integrated mutational, copy-number and RNA-seq (fusion) data. A gene was considered as having both alleles inactivated for any combination (pair) of the following events: copy loss, mutation, truncating fusion and copy-number breakpoint, in addition to homozygous deletion of both copies and two independent mutations. Ambiguous cases were manually reviewed to increase the accuracy and ascertain whether both events, for example, copy-number breakpoint and gene fusion, are probably independent events.

Unified mutation calling and variant classification of FOXA1. Mutation calls for *FOXA1* obtained or downloaded from the GDC and TCGA flagship manuscripts[5,6] as well as our internal pipelines were lifted over to GRCh38 (using the Bioconductor package rtracklayer) and annotated with respect to the canonical RefSeq *FOXA1* isoform. For TCGA samples or cases, multiple call sets were available and we manually reviewed all discrepancies in *FOXA1* mutation calls, resulting in a unified call set with improved sensitivity and specificity. Mutational effect (consequence) was simplified into three categories: missense, in-frame indel and frameshift (the last category included stop-gain, stop-loss and splice-site mutations). The resulting mutations were dichotomized into class 1 and class 2 based on their position relative to amino acid residue 275. Variant allele frequencies were only available for TCGA and the in-house mCRPC cohorts.

Analysis of whole-genome sequencing data. The bcbio-nextgen pipeline version 1.0.3 was used for the initial steps of tumour whole-genome data analysis. Paired-end reads were aligned to the GRCh38 reference using BWA (bcbio default settings), and structural variant calling was done using LUMPY[50] (bcbio default settings), with the following post-filtering criteria: "(SR> = 1 & PE> = 1 & SU> = 7) & (abs(SVLEN)>5e4) & DP <1000 & FILTER == "PASS". The following settings were chosen to minimize the number of expected germline variants: false discovery rate (FDR) < 0.05 for germline status for both deletions and duplications. Additionally, common structural germline variants were filtered.

Analysis of 10X genomics long-read sequencing data. High-molecular mass DNA from MDA-PCA-2b and LNCaP cell lines was isolated and processed into linked-read next-generation sequencing libraries per the manufacturer's instructions (10X WGS v2 kit). The resulting paired-end sequencing data were sequenced on an Illumina Hi-Seq 2500 instrument and analysed (demultiplexing, alignment, phasing and structural variant calls) using the longranger 2.2.1 pipeline with all default settings. The resulting libraries met all 10X-recommended quality control parameters including molecule size, average phasing length, and sequencing coverage (~50×). Here, we focused on structural variant calls within the *FOXA1* TAD and confirmed the presence of the previously reported *FOXMIND-ETV1* fusions; that is, translocation for MDA-PCA-2b, and balanced insertional translocation for LNCaP. Both cell lines were confirmed to contain three copies of *FOXA1* (that is, one translocated allele and two duplicated alleles).

RNA-seq data pre-processing and primary analysis. RNA-seq data processing—including quality control, read trimming, alignment, and expression quantification by read counting—was carried out as previously described[49], using our standard clinical RNA-seq pipeline CRISP (available at https://github.com/mcieslik-mctp/bootstrap-rnascape). The pipeline was run with default settings for paired-end RNA-seq data of at least 75 bp. The only changes were made for unstranded transcriptome libraries sequenced at the Broad Institute and the TCGA and CCLE cohorts, for which quantification using featureCounts[51] was used in unstranded mode '-s0'. The resulting counts were transformed into fragments per kilobase of transcript per million mapped reads using upper-quartile normalizations as implemented in EdgeR[52]. For mCRPC samples *FOXA1* expression estimates were adjusted by tumour content estimated from whole-exome sequencing (see above) given the highly prostate-specific *FOXA1* expression profile. For the quantification of *FOXMIND* expression levels, a custom approach was necessary given the poor annotation and unspliced nature of this transcript. First, we delineated regions of sense and antisense transcription from the *FOXMIND* ultra-conserved regulatory elements, chr14:37564150-37591250:+ and chr14:37547900-37567150:-, respectively. Next, to make the expression estimates reliable in unstranded libraries, we identified regions of substantial overlap between the sense and antisense RP11-356O9.1 transcripts, and *FOXA1* and *MIPOL1*. These overlaps have been excluded from quantification, resulting in the following trimmed target regions: chr14:37564150-37589500, and chr14:37553500-37567150. Within these regions, the average base-level coverage normalized to sequencing depth was computed as an expression estimate.

Differential expression analyses. All differential expression analyses were done using limma R-package[53], with the default settings for the voom[54], lmFit, eBayes and topTable functions. The contrasts were designed as follows to identify transcriptional signatures of class-1 mutants. Given the mutual exclusivity of the genotypes in primary and metastatic tumours, the overall MCTP mCRPC cohort of 371 cases was partitioned into 4 groups: (1) ETS-fused or *SPOP*-mutant tumours, (2) class-1 mutant tumours, (3) class-2 mutant tumours, and (4) tumours that were wild type for ETS, *SPOP* and *FOXA1*. To avoid confounding effects, the class-2 and ETS and SPOP groups were excluded from class-1 transcriptional analyses. Next, the class-1 samples were contrasted with the wild-type samples with additional independent regressors for assay type (capture vs polyA, as previously described[49], and mutational status (see above) for the following genes and pathways: PI3K, WNT, DRD, *RB1* and *TP53*. In other words, we constructed a design matrix with coefficients for class-1 mutational status, in addition to coefficients for confounding variables and recurrent genetic heterogeneity. This allowed us to estimate the fold changes (expressed logarithmically) and adjusted *P* values associated with *FOXA1* mutations and other genotypes (for example, PI3K status). An analogous procedure was carried out for the primary class-1 samples (TCGA) and for class-2 mutations in mCRPC (MCTP), but given the lack of mutual-exclusivity between class-2 mutations and ETS and SPOP group, only class-1 mutations were excluded.

Pathway and signature enrichment analyses. The Molecular Signatures Database (MSigDB)[55] was used as a source of gene sets comprising cancer hallmarks, molecular pathways, oncogenic signatures and transcription factor targets. The enrichment of signatures was assessed using the parametric random-set method[56], and visualized using the gene-set enrichment analysis (GSEA) enrichment statistic[57] and barcode plots. All *P* values have been adjusted for multiple-hypothesis testing using a false discovery rate correction. To identify putative transcription factors regulating differentially expressed genes, we used the transcription factor prediction tool BART[25]. BART was run with all default settings, and the provided transcription factor databases. We used voom- and limma-based gene-level fold-changes as input to the algorithm.

Detection of structural variants from RNA-seq. The detection of chimeric RNAs (gene fusions, structural variants, circular RNAs and read-through events) was carried out using our previously published[49] in-house toolkit for the comprehensive detection of chimeric RNAs, CODAC (available at https://github.com/mctp/codac). In brief, three separate alignment passes (STAR 2.4.0g1) against the GRCh38 (hg38) reference with known splice junctions provided by Gencode v.27 (ref. [58]) are made for the purposes of expression quantification and fusion discovery. The first pass is a standard paired-end alignment followed by gene-expression quantification. The second and third pass are for the purpose of gene fusion discovery and to enable the chimeric alignment mode of STAR (chimSegmentMin: 10, chimJunctionOverhangMin: 1, alignIntronMax: 150000, chimScoreMin: 1). Fusion detection was carried out using CODAC with default parameters to balance sensitivity and specificity (annotation preset:balanced). CODAC uses MOTR v.2, a custom reference transcriptome based on a subset of Gencode 27 (available with CODAC). Prediction of topology (inversion, duplication, deletion and translocation), and distance (adjacent, breakpoints in two directly adjacent loci; cytoband, breakpoints within the same cytoband based on UCSC genome browser; arm, breakpoints within the same chromosome arm). The high specificity of our pipeline has been assessed through Sanger sequencing[49]. To create fusion circos plots, we have colour-coded the CODAC variants on the basis of the inferred topology of the breakpoints. Unbiased discovery of recurrently rearranged loci has been carried out by breaking the genome into 1.5-Mb windows with a step of 0.5 Mb. For each window, the percentage of patients with at least one RNA break end has been calculated. The resulting genomic windows were ranked and clustered by proximity for visualization. CODAC has the ability to make fusion calls independent of known transcriptome references or annotations and is therefore capable of detecting fusions involving intergenic or poorly annotated regions.

Classification of *FOXA1* locus genomic rearrangements. Structural variants within the *FOXA1* locus have been partitioned into two broad topological patterns: (1) translocations (including inversions and deletions involving distal loci on the same chromosome) and (2) focal duplications. The translocations have been further subdivided into hijacking and swapping events on the basis of their position relative to *FOXMIND* (GRCh38: chr14:37564150-37591250) and *FOXA1*. Hijacking translocations position a translocation partner within the *FOXMIND-FOXA1* regulatory domain (defined as GRCh38: chr14:37547501-37592000, based on manual review of chromatin conformation Hi-C, CTCF, H3K4me1, H3K27ac, evolutionary conservation and synteny data). Swapping

translocations preserve the *FOXMIND-FOXA1* regulatory domain but insert the translocation partner upstream of the *FOXA1* promoter, frequently 'swapping-out' the *TTC6* gene. Notably, one isoform of *TTC6* gene can be transcribed from the bi-directional *FOXA1* promoter. Focal duplications within the *FOXA1* locus have been derived from the CODAC structural-variant output file. In brief, for each case independently, all RNA-seq fusion junctions annotated by CODAC as tandem duplications and overlapping the *FOXA1* topologically associating domain (GRCh38: chr14:37210001-37907919) have been collated and used to infer the minimal duplicated region. Because RNA-seq chimeric junctions generally coincide with splice junctions (limited resolution) and generally cannot be phased (ambiguous haplotype), the inference of minimal duplicated regions makes the necessary and parsimonious assumption that overlapping tandem duplications are due to a single somatic genetic event, and not multiple independent events.

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

Data availability

All raw data for the graphs, immunoblot and gel electrophoresis figures are included in the Source Data or Supplementary Information. All materials are available from the authors upon reasonable request. All the raw next-generation sequencing, ChIP and RNA-seq data generated in this study have been deposited in the Gene Expression Omnibus (GEO) repository at NCBI (accession code GSE123625).

Code availability

All custom data analysis software and bioinformatics algorithms used in this study are publically available on Github: https://github.com/mcieslik-mctp/ and https://github.com/mctp/.

31. Shalem, O. et al. Genome-scale CRISPR–Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
32. Phair, R. D. et al. Global nature of dynamic protein–chromatin interactions in vivo: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. *Mol. Cell. Biol.* **24**, 6393–6402 (2004).
33. Grimm, J. B. et al. A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat. Methods* **12**, 244–250 (2015).
34. Pitchiaya, S. et al. Dynamic recruitment of single RNAs to processing bodies depends on RNA functionality. *Mol. Cell* **74**, 521–533 (2019).
35. Swinstead, E. E. et al. Steroid receptors reprogram FoxA1 occupancy through dynamic chromatin transitions. *Cell* **165**, 593–605 (2016).
36. Pitchiaya, S., Androsavich, J. R. & Walter, N. G. Intracellular single molecule microscopy reveals two kinetically distinct pathways for microRNA assembly. *EMBO Rep.* **13**, 709–715 (2012).
37. Shah, N. B. & Duncan, T. M. Bio-layer interferometry for measuring kinetics of protein–protein interactions and allosteric ligand effects. *J. Vis. Exp.* **84**, e51383 (2014).
38. Teng, Y. et al. Evaluating human cancer cell metastasis in zebrafish. *BMC Cancer* **13**, 453 (2013).
39. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
40. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
41. Zhu, L. J. et al. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010).
42. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
43. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
44. Wilson, S. et al. Developing cancer informatics applications and tools using the NCI genomic data commons API. *Cancer Res.* **77**, e15–e18 (2017).
45. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
46. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
47. Wu, Y.-M. et al. Inactivation of *CDK12* delineates a distinct immunogenic class of advanced prostate cancer. *Cell* **173**, 1770–1782 (2018).
48. Cieslik, M. et al. The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res.* **25**, 1372–1381 (2015).
49. Robinson, D. R. et al. Integrative clinical genomics of metastatic cancer. *Nature* **548**, 297–303 (2017).
50. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
51. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2013).
52. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
53. Smyth, G. K., McCarthy, D. J. & Smyth, G. K. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (eds Dudoit, S. & Carey, V. J.) 397–420 (Springer, New York, 2005).
54. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
55. Liberzon, A. et al. The molecular signatures database hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
56. Newton, M. A., Quintana, F. A., Boon, J. A. D., Sengupta, S. & Ahlquist, P. Random-set methods identify distinct aspects of the enrichment signal in gene-set analysis. *Ann. Appl. Stat.* **1**, 85–106 (2007).
57. Subramanian, A., Kuehn, H., Gould, J., Tamayo, P. & Mesirov, J. P. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics* **23**, 3251–3253 (2007).
58. Searle, S. et al. The GENCODE human gene set. *Genome Biol.* **11**, P36 (2010).

Acknowledgements We thank D. Macha, L. Wang, S. Zelenka-Wang, I. Apel, M. Tan, Y. Qiao, A. Delekta, K. Juckette and J. Tien for technical assistance, and S. Gao for assistance with the manuscript. This work was supported by the Prostate Cancer Foundation (PCF), Early Detection Research Network (UO1 CA214170), NCI Prostate SPORE (P50 CA186786) and Stand Up 2 Cancer-PCF Dream Team (SU2C-AACR-DT0712) grants to A.M.C. A.M.C. is an NCI Outstanding Investigator, Howard Hughes Medical Institute Investigator, A. Alfred Taubman Scholar and American Cancer Society Professor. A.P. is supported by a Predoctoral Department of Defense (DoD) - Early Investigator Research Award (W81XWH-17-1-0130). M.C. is supported by a DoD - Idea Development Award (W81XWH-17-1-0224) and a PCF Young Investigator Award.

Author contributions A.P., M.C. and A.M.C. conceived and designed the study; A.P. performed all the experiments with assistance from L.X., T.O., X.W. and S.P. M.C. carried out bioinformatics analyses with assistance from A.P., Y.Z., R.J.L. and P.V. S.-C.C. and A.P. performed zebrafish in vivo experiments. A.P. is responsible for the following experimental figures: Figs. 2b–f, h, 3b–i, 4e, as well as Extended Data Figs. 1a–i, 3b–n, 4a–f, k–n, 5a–k, 6a–l, 7i–o, 8a–h, j, 9a, d, e, 10g. M.C. is responsible for the following computational figures: Figs. 1a–h, 2a, g, 3a, 4a–d, as well as Extended Data Figs. 1j–n, 2a–l, 3a, p, q, 4g–j, o–q, 7a–c, g, h, 9b, c, f–h, 10a–f. Y.Z. is responsible for the following computational figures: Extended Data Figs. 3o, r, s, 7d–f, 8i, k. F.S. and R.W. generated ChIP–seq and RNA-seq libraries. X.C. performed sequencing. F.Y.F. provided genomic validation data. Y.-M.W. and D.R.R. coordinated clinical sequencing. A.P., M.C. and A.M.C. wrote the manuscript and organized the figures.

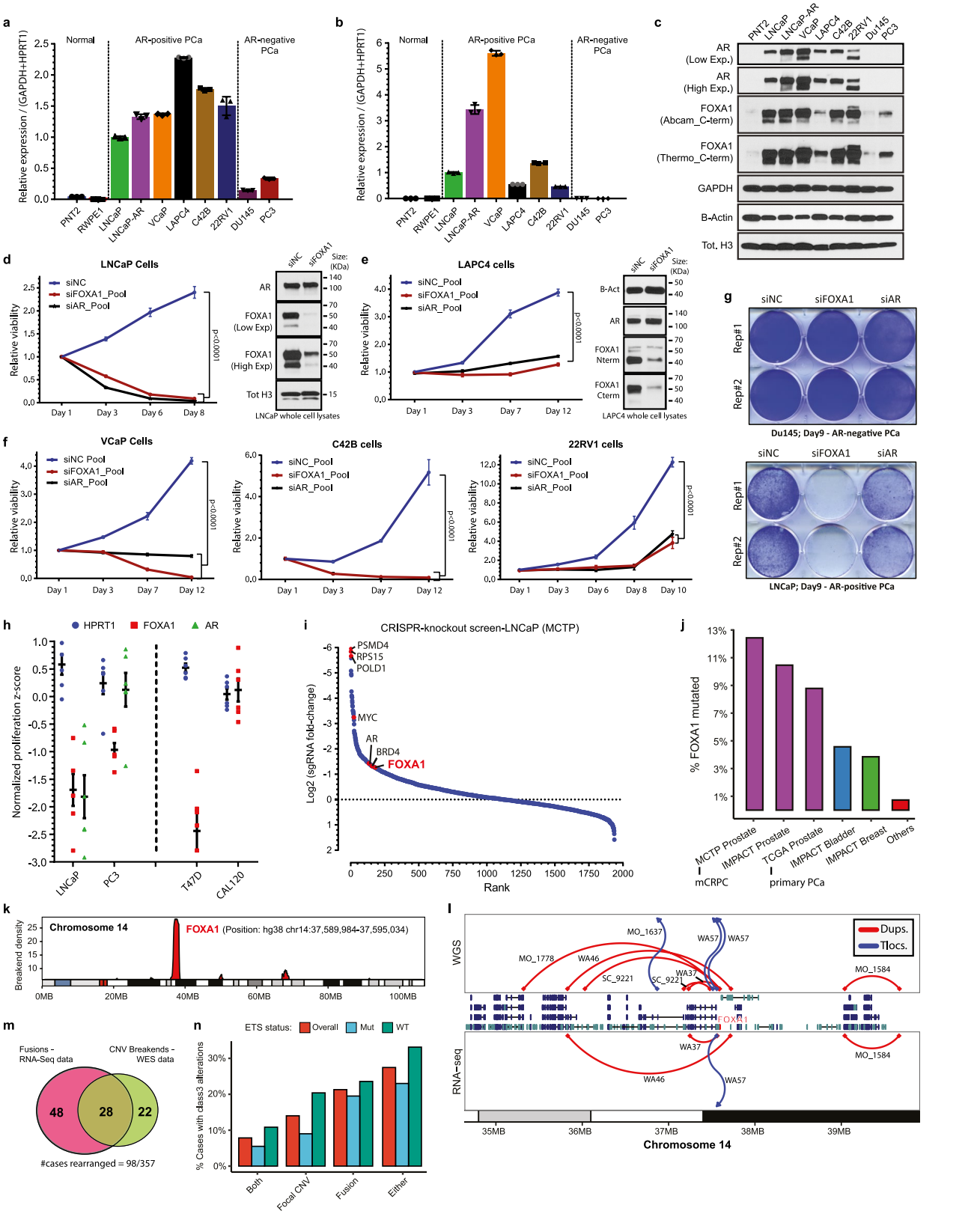
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1347-4.

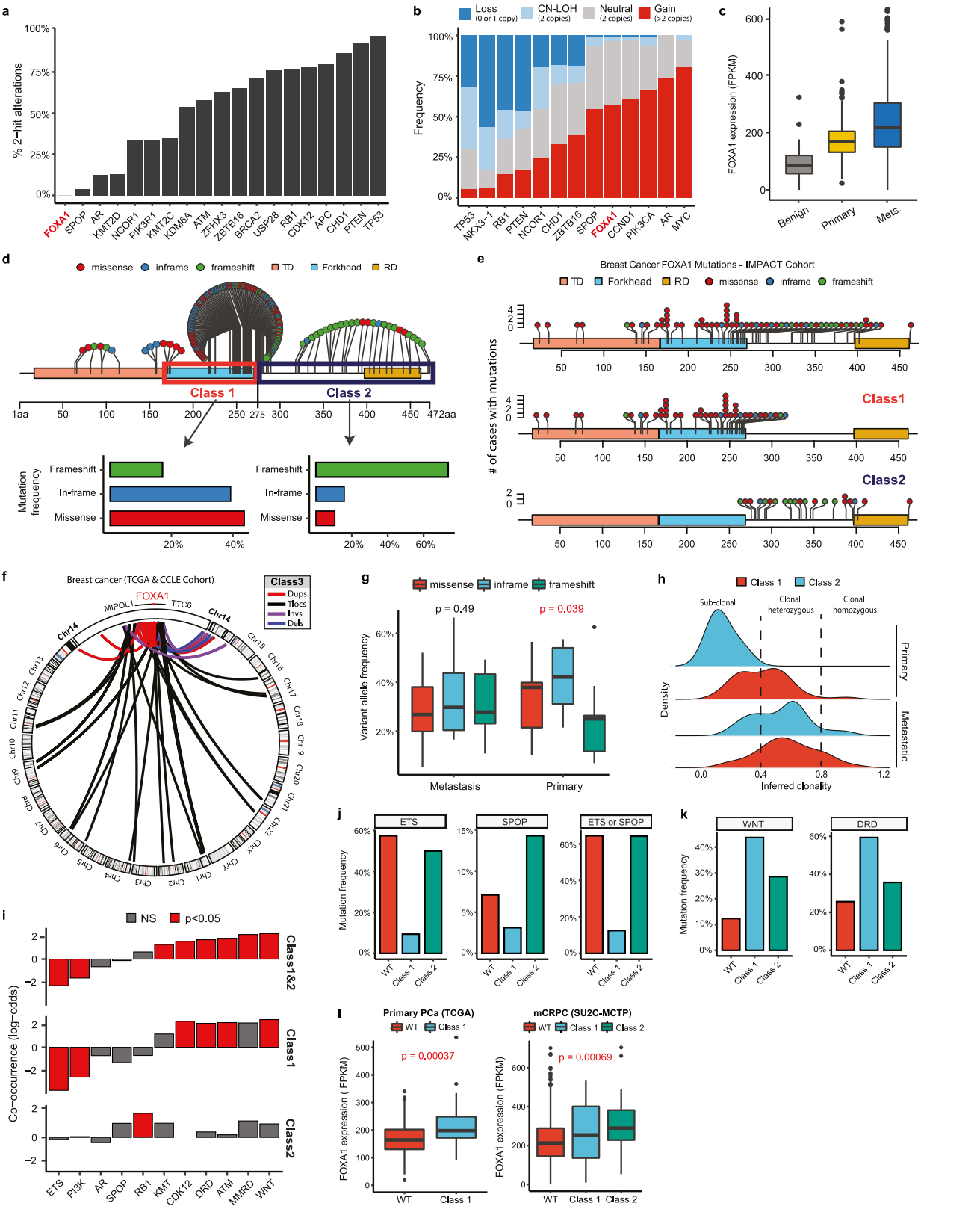
Correspondence and requests for materials should be addressed to A.M.C.

Peer review information *Nature* thanks Myles Brown, William Nelson, Mark A. Rubin and the other anonymous reviewer(s) for their contribution to the peer review of this work.



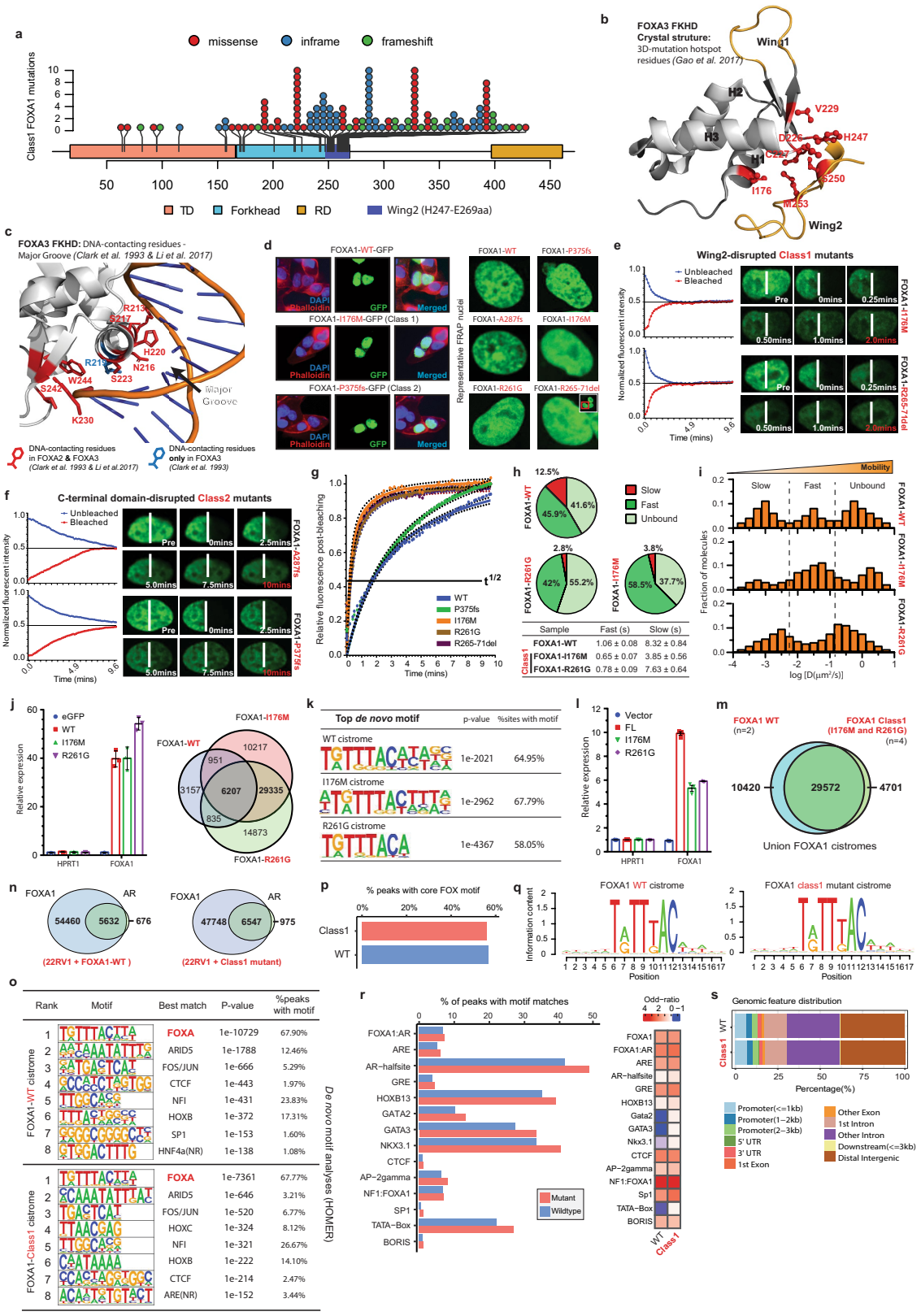
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Functional essentiality and recurrent alterations of FOXA1 in AR$^+$ prostate cancer. a–**c**, AR (**a**) and FOXA1 (**b**) mRNA (qPCR) and (**c**) protein expression in a panel of prostate cancer cells ($n = 3$ technical replicates). Mean ± s.e.m. is shown and dots are individual data points. **d**–**f**, Growth curves of AR$^+$ prostate cancer cells treated with non-targeting control (siNC), *AR*- or *FOXA1*-targeting siRNAs (25 nM at day 0 and 1; $n = 6$ biological replicates). Immunoblots confirm knockdown of FOXA1 protein in LNCaP and LAPC4 72 h after siRNA treatment. For all gel source data, see Supplementary Fig. 1. **g**, Crystal-violet stain of AR$^-$ DU145 prostate cancer and LNCaP (control) cells treated with siNC, *AR*- or *FOXA1*-targeting siRNAs. Results represent 3 independent experiments ($n = 2$ biological replicates). **h**, Averaged proliferation *z*-scores for 6 independent *FOXA1*-targeting sgRNAs extracted from publically available CRISPR Project Achilles data (BROAD Institute) in prostate and breast cancer cells. *HPRT1* and *AR* data serve as negative and positive controls, respectively. Mean ± s.e.m. is shown; dots are proliferative *z*-scores for independent sgRNAs. **i**, Ranked depletion or enrichment of sgRNA read counts from GeCKO-V2 CRISPR knockout screen in LNCaP cells (at day 30) relative to the input sample. Only a subset of genes—including essential controls, chromatin modifiers and transcription factors—is visualized. **j**, Recurrence of *FOXA1* mutations across TCGA, MSK-IMPACT and SU2C cohorts. **k**, Density of break ends (RNA-seq chimeric junctions) within overlapping 1.5-Mb windows along chr14 in mCRPC tumours. **l**, Whole-genome sequencing (WGS) of seven mCRPC index cases with distinct patterns of *FOXA1* translocations (Tlocs) and duplications (Dups), nominated by RNA-seq (WA46, WA37, WA57 and MO_1584) or whole-exome sequencing (MO_1778, SC_9221 and MO_1637). **m**, Concordance of RNA-seq (chimeric junctions) and whole-exome-sequencing-based *FOXA1* locus rearrangements calls (mCRPC cohort). CNV, copy-number variation. **n**, Frequency of *FOXA1* locus rearrangements in mCRPC based on RNA-seq and whole-exome sequencing.



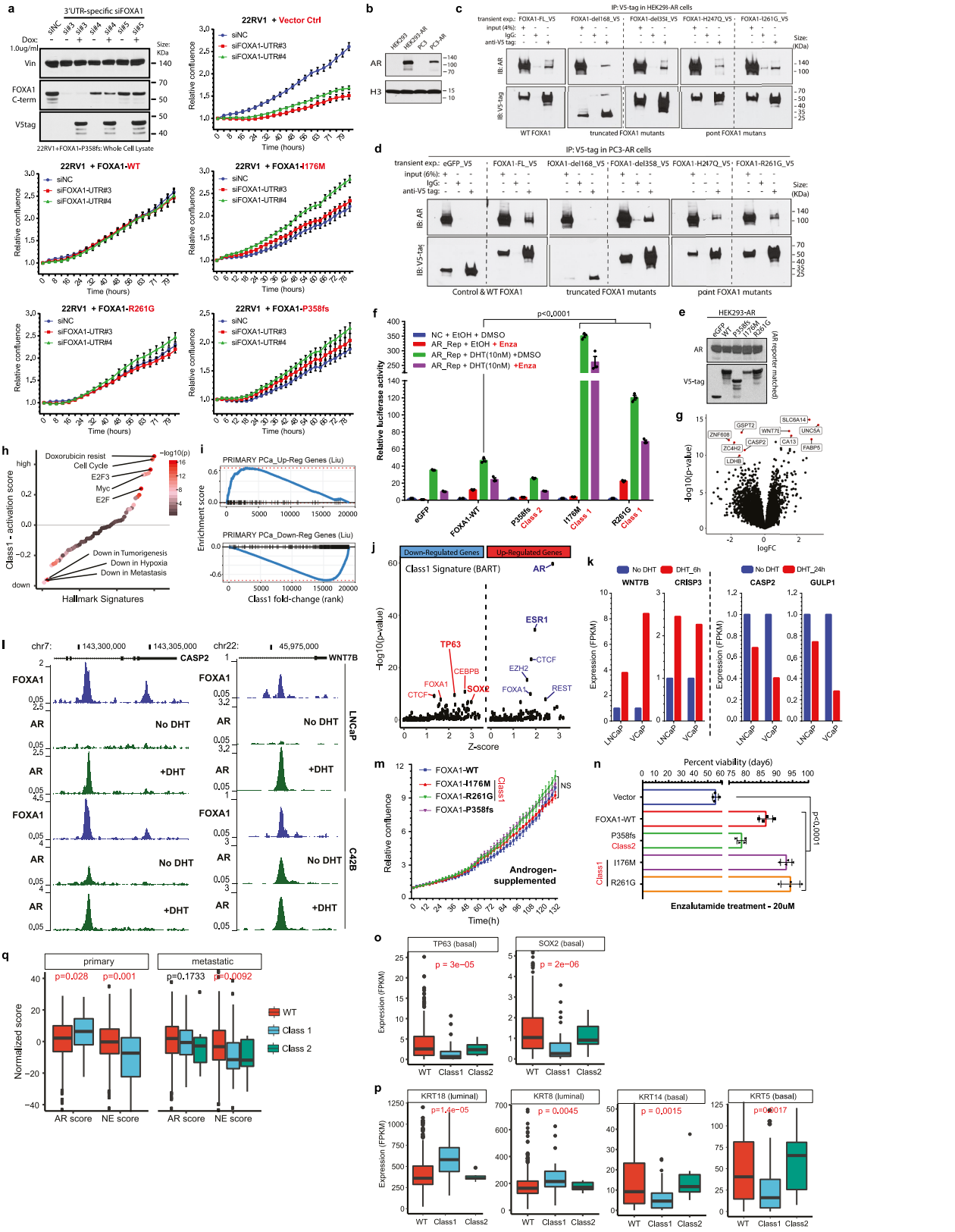
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Genomic characteristics of the three classes of FOXA1 alterations in prostate and breast cancer. a, **b**, Bi-allelic inactivation (**a**) and copy-number variations (**b**) of *FOXA1* across mCRPC ($n = 371$). CN-LOH, copy-neutral loss of heterozygosity. **c**, FOXA1 expression (RNA-seq) in benign ($n = 51$), primary ($n = 501$) and metastatic ($n = 535$) prostate cancer. **d**, Distribution and functional categorization of *FOXA1* mutations (all cases in the aggregate cohort) on the protein map of FOXA1. **e**, Aggregate and class-specific distribution of *FOXA1* mutations in advanced breast cancer (MSK-IMPACT cohort). **f**, Structural classification of *FOXA1* locus rearrangements in breast cancer (TCGA and CCLE cell lines). **g**, **h**, Variant allele frequency of *FOXA1* mutations by tumour stage (**g**) and clonality estimates of class-1 and class-2 mutations (**h**) in tumour-content-corrected primary prostate cancer ($n = 500$) and mCRPC ($n = 370$) specimens. **i**, Mutual exclusivity or co-occurrence of *FOXA1* mutations (two-sided Fisher's exact test). Mutations in AR, WNT, and PI3K were aggregated at the pathway level. ETS, ETS gene fusions; DRD, DNA repair defects and included alterations in *BRCA1*, *BRCA2*, *ATM* and *CDK12*; MMRD, mismatch repair deficiency (total $n = 371$). **j**, Mutual exclusivity of ETS and/or SPOP ($n = 26$) alterations with FOXA1 ($n = 46$) alterations distinguished by class in mCRPC ($n = 371$). **k**, Co-occurrence of WNT ($n = 58$) and DRD ($n = 107$) pathway alterations with FOXA1 alteration classes in mCRPC ($n = 371$). **l**, Stage- and class-specific increase in FOXA1 expression levels in primary ($n = 500$) and metastatic prostate cancer ($n = 357$). Left, two-sided *t*-test. Right, two-way ANOVA. For all box plots, centre shows median, box marks quartiles 1–3 and whiskers span quartiles 1–3 $\pm$ 1.5 $\times$ IQR.



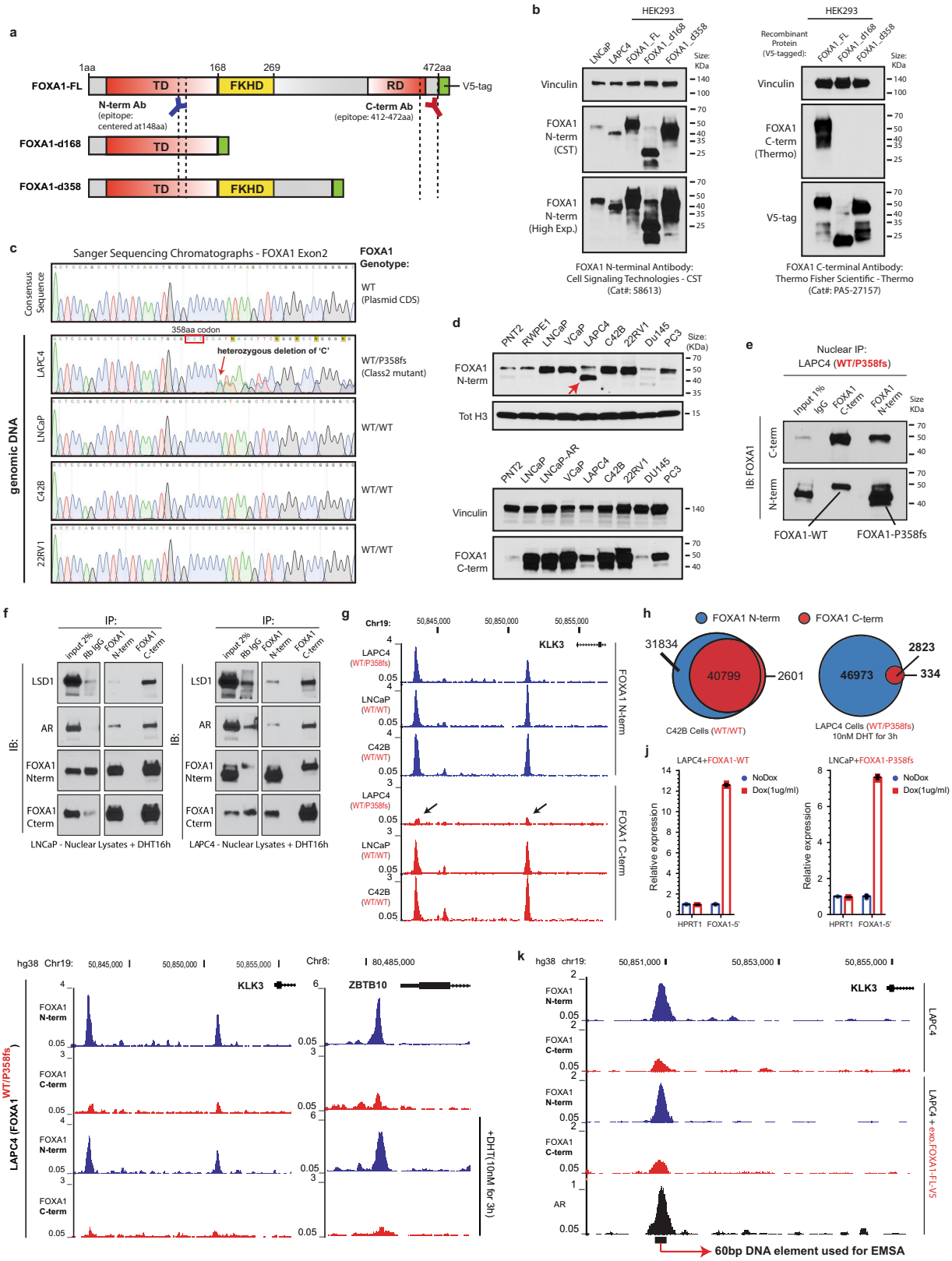
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Biophysical and cistromic characteristics of the class-1 FOXA1 mutants. a, Distribution of class-1 mutations on the protein map of FOXA1. **b**, Three-dimensional structure of FKHD (FOXA3) with visualization of all mutated residues collectively identified as the 3D-mutational hotspot in FOXA1 across cancers. **c**, DNA-bound 3D structure of FKHD with visualization of all residues shown through crystallography to make direct base-specific contacts with the DNA in FOXA2 and FOXA3 proteins. **d**, Representative fluorescent images of nuclei expressing different variants of FOXA1 fused to GFP at the C termini. **e**, **f**, FRAP kinetic plots (left) and representative time-lapse images (right) from pre-bleaching (pre) to 100% recovery (red timestamps) for wing-2-altered class-1 mutants (**e**) and truncated class-2 mutants (that is, A287fs and P375fs) (**f**) ($n = 6$ nuclei per variant; quantified in Fig. 2d). White lines indicate the border between bleached and unbleached areas. **g**, Representative FRAP kinetics in the bleached area for indicated FOXA1 variants. $t_{1/2}$ line indicates the time to 50% recovery. Coloured dots show raw data; superimposed solid curves show a hyperbolic fit with 95% confidence intervals. **h**, Single particle tracking quantification of chromatin-bound (slow and fast) and unbound (freely diffusing) particles of wild-type and class-1 FOXA1 variants, and average chromatin dwell times (mean ± s.d.) for the bound fractions ($n \geq 500$ particles per variant). **i**, Diffusion constant histograms of single particles of wild-type or distinct class-1 FOXA1 mutants. Particles were categorized into chromatin-bound (slow and fast) or unbound fractions using cut-offs marked by dashed lines ($n \geq 500$ particles per variant imaged in 3–5 distinct nuclei). **j**, Left, mRNA expression (qPCR) of labelled FOXA1 variants in stable, isogenic HEK293 cells ($n = 3$ technical replicates). Right, overlaps between FOXA1 wild-type and class-1 mutant cistromes from these cells ($n = 2$ biological replicates). **k**, Top de novo motifs identified from the three FOXA1 cistromes from HEK293 cells (HOMER, hypergeometric test). **l**, mRNA expression (qPCR) of labelled FOXA1 variants in stable, isogenic 22RV1 cells ($n = 3$ technical replicates). For **j** and **l**, centres show mean values and lines mark s.e.m. **m**, Overlap between wild-type ($n = 2$ biological replicates) and class-1 ($n = 4$ biological replicates) cistromes from stable 22RV1 overexpression models. **n**, Overlap between the FOXA1 wild-type and AR union cistromes generated from 22RV1 cells overexpressing wild-type ($n = 2$ biological replicates) or class-1 mutant (I176M or R216G; $n = 2$ biological replicates each) FOXA1 variants. **o**, De novo motif results for the wild-type or class-1 mutant FOXA1-binding sites from prostate cancer cells (HOMER, hypergeometric test). **p**, **q**, Per cent of wild-type or class-1 binding sites with perfect match to the core FOXA1 motif (5′-T[G/A]TT[T/G]AC-3′) (**p**) and the consensus FOXA1 motifs identified from these sites (**q**). **r**, Left, per cent of wild-type or class-1 binding sites containing known motifs of the labelled FOXA1 or AR cofactors. Right, enrichment of the cofactor motifs in the two cistromes relative to the background ($n =$ top 5,000 peaks by score for each variant, see Methods). **s**, Genomic distribution of wild-type and class-1 binding sites in prostate cancer cells.



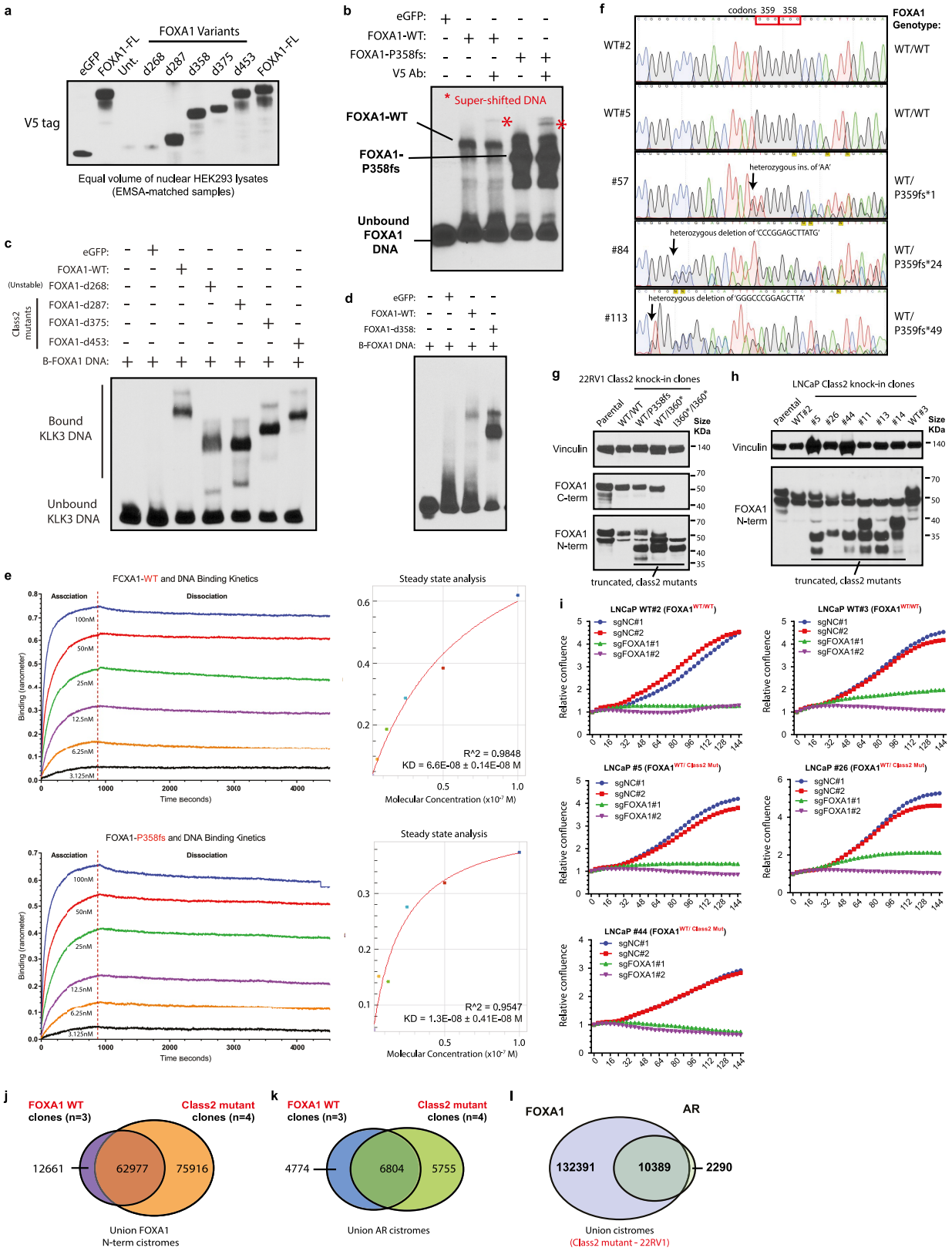
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Functional effect of *FOXA1* mutations on oncogenic AR signalling. a, Immunoblot showing expression of endogenous and V5-tagged exogenous FOXA1 proteins in doxycycline (dox)-inducible 22RV1 cells transfected with distinct UTR-specific *FOXA1*-targeting siRNAs (no. 3–5) or a non-targeting control siRNA (siNC). These results represent two independent experiments. IncuCyte growth curves of 22RV1 cells overexpressing empty vector (control), wild-type or mutant FOXA1 variants upon treatment with UTR-specific *FOXA1*-targeting siRNAs ($n = 5$ biological replicates). Mean ± s.e.m. is shown. **b**, Immunoblots confirming stable overexpression of the wild-type AR protein in HEK293 and PC3 cells. **c**, **d**, Co-immunoprecipitation assay of indicated recombinant FOXA1 variants using a V5-tag antibody in HEK293 (**c**) and PC3 (**d**) cells stably overexpressing the AR protein (referred to as HEK293-AR and PC3-AR cells). eGFP is a negative control. FOXA1-FL, full-length wild-type FOXA1. del168 and del358 are truncated FOXA1 variants with only the first 168 amino acids (that is, before the FKHD) or 358 amino acids of the FOXA1 protein. H247Q and R261G are missense class-1 mutant variants. **e**, Immunoblots confirming comparable expression of AR and recombinant FOXA1 variants in AR reporter assay-matched HEK293 lysates. Immunoblots show representative results from 2 or 3 independent experiments and class-1 and class-2 mutants serve as biological replicates. For all gel source data (**a**, **b**–**e**), see Supplementary Fig. 1. **f**, AR dual-luciferase reporter assays with transient overexpression of indicated FOXA1 variants in HEK293-AR cells with or without DHT stimulation and enzalutamide treatment ($n = 3$ biological replicates per group). Mean ± s.e.m. is shown (two-way ANOVA and Tukey's test). **g**, Genes differentially expressed in class-1 tumours from patients ($n = 38$) compared to FOXA1 wild-type tumours (see Methods). The most significant genes are shown in red and labelled (limma two-sided test). **h**, Differential expression of cancer-hallmark signature genes in class-1 mutant prostate-cancer tumours (GSEA statistical test). **i**, Localized, primary prostate cancer gene signature showing concordance between class-1 tumour and primary prostate cancer genes. **j**, BART prediction of specific transcription factors mediating observed transcriptional changes. The significant and strong (z-score) mediators of transcriptional responses in class-1 tumours are labelled (BART, Wilcoxon rank-sum test). **k**, mRNA expression (RNA-seq) of class-1 signature genes in LNCaP and VCaP cells either starved for androgen (no DHT) or stimulated with DHT (10 nM). RNA-seq from two distinct prostate cancer cell lines is shown. **l**, Representative FOXA1 and AR ChIP–seq normalized signal tracks at the *WNT7B* or *CASP2* gene loci in LNCaP and C42B cells. ChIP–seq assays were carried out in two distinct prostate cancer cell lines with similar results. **m**, Growth curves (IncuCyte) of 22RV1 cells overexpressing distinct FOXA1 variants in complete, androgen-supplemented growth medium ($n = 2$ biological replicates). Mean ± s.e.m. is shown. **n**, Per cent viable 22RV1 stable cells, overexpressing either empty vector, wild-type or mutant FOXA1 variants upon treatment with enzalutamide (20 μM for 6 days; $n = 4$ biological replicates). Mean ± s.e.m. is shown. P values in **m** and **n** were calculated using two-way ANOVA and Tukey's test. **o**, **p**, mRNA expression (RNA-seq) of labelled basal and luminal transcription factors or canonical markers in FOXA1 wild-type, class-1 or class-2 mutant tumours in primary prostate cancer (total $n = 500$; two-way ANOVA). **q**, Extent of AR and neuroendocrine (NE) pathway activation in FOXA1 wild-type, class-1 or class-2 mutant cases from both primary ($n = 500$) and metastatic ($n = 370$) prostate cancer. Both AR and NE scores were calculated using established gene signatures (see Methods). Left, two-sided t-test; right, two-way ANOVA. For all box plots, centre shows median, box marks quartiles 1–3 and whiskers span quartiles 1–3 ± 1.5 × IQR.



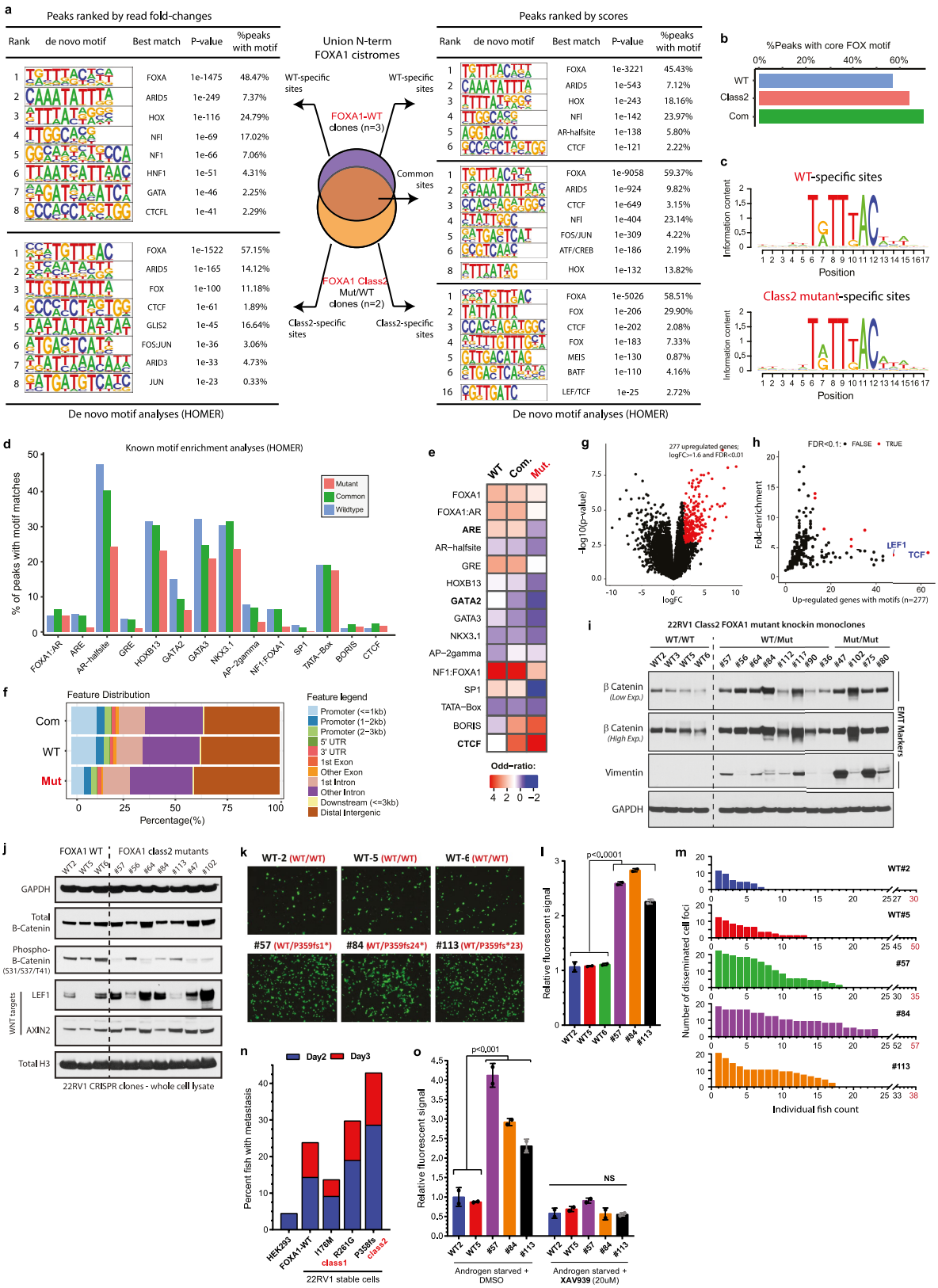
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | DNA-binding dominance of the class-2 FOXA1 mutants. a, FOXA1 protein maps showing the recombinant proteins used to validate the N-terminal (N-term) and C-terminal (C-term) FOXA1 antibodies. **b**, Immunoblots depicting detection of all variants by the N-terminal antibody (left), and of only the full-length wild-type FOXA1 protein by the C-terminal antibody (right). These results were reproducible in two independent experiments. Antibody details are included in the Methods. **c**, Sanger sequencing chromatograms showing the heterozygous class-2 mutation in LAPC4 cells after the P358 codon in exon 2 ($n = 2$ technical replicates). All other tested prostate cancer cell lines were wild type for FOXA1. **d**, Immunoblots confirming the expression of the truncated FOXA1 variant in LAPC4 at the expected approximately 40-kDa size (top, red arrow). The short band is detectable only with the N-terminal (top) FOXA1 antibody and not the C-terminal (bottom) antibody. These results were reproducible in two independent experiments. **e**, Co-immunoprecipitation and immunoblotting of FOXA1 using N-terminal and C-terminal antibodies from LAPC4 nuclei with species-matched IgG used as control. **f**, Nuclear co-immunoprecipitation of FOXA1 from LAPC4 or LNCaP cells stimulated with DHT (10 nM for 16 h) using N-terminal and C-terminal antibodies. Species-matched IgG are controls. Immunoprecipitations and immunoblots in **d**–**f** were reproducible in two and three independent experiments, respectively. For gel source data (**b**, **d**, **e**, **f**), see Supplementary Fig. 1. **g**, FOXA1 N-terminal and C-terminal ChIP–seq normalized signal tracks from FOXA1 wild-type or class-2 mutant prostate cancer cells at canonical AR target *KLK3*. **h**, Left, overlap between global N-terminal and C-terminal FOXA1 cistromes in untreated C42B cells. Right, overlap between global N-terminal and C-terminal FOXA1 cistromes in LAPC4 cells treated with DHT (10 nM for 3 h). **i**, FOXA1 ChIP–seq normalized signal tracks from N-terminal and C-terminal antibodies in LAPC4 cells with or without DHT stimulation (10 nM for 3 h) at *KLK3* and *ZBTB10* loci. ChIP–seq assays in **g** and **i** were carried out in two distinct FOXA1 wild-type prostate cancer cells. For LAPC4 ChIP–seq experiments, results were reproducible in two independent experiments. **j**, mRNA (qPCR) expression of *FOXA1* in LAPC4 cells with exogenous overexpression of wild-type FOXA1 (left), and in LNCaP cells with exogenous overexpression of the P358fs mutant (right) ($n = 3$ technical replicates). Mean ± s.e.m. is shown and dots are individual data values. **k**, FOXA1 ChIP–seq normalized signal tracks from N-terminal and C-terminal antibodies in parental LAPC4 cells and LAPC4 cells overexpressing wild-type FOXA1 at the *KLK3* locus. This experiment was independently repeated twice with similar results. The 60-bp AR- and FOXA1-bound *KLK3* enhancer element used for electrophoretic mobility shift assay (EMSA) is shown.



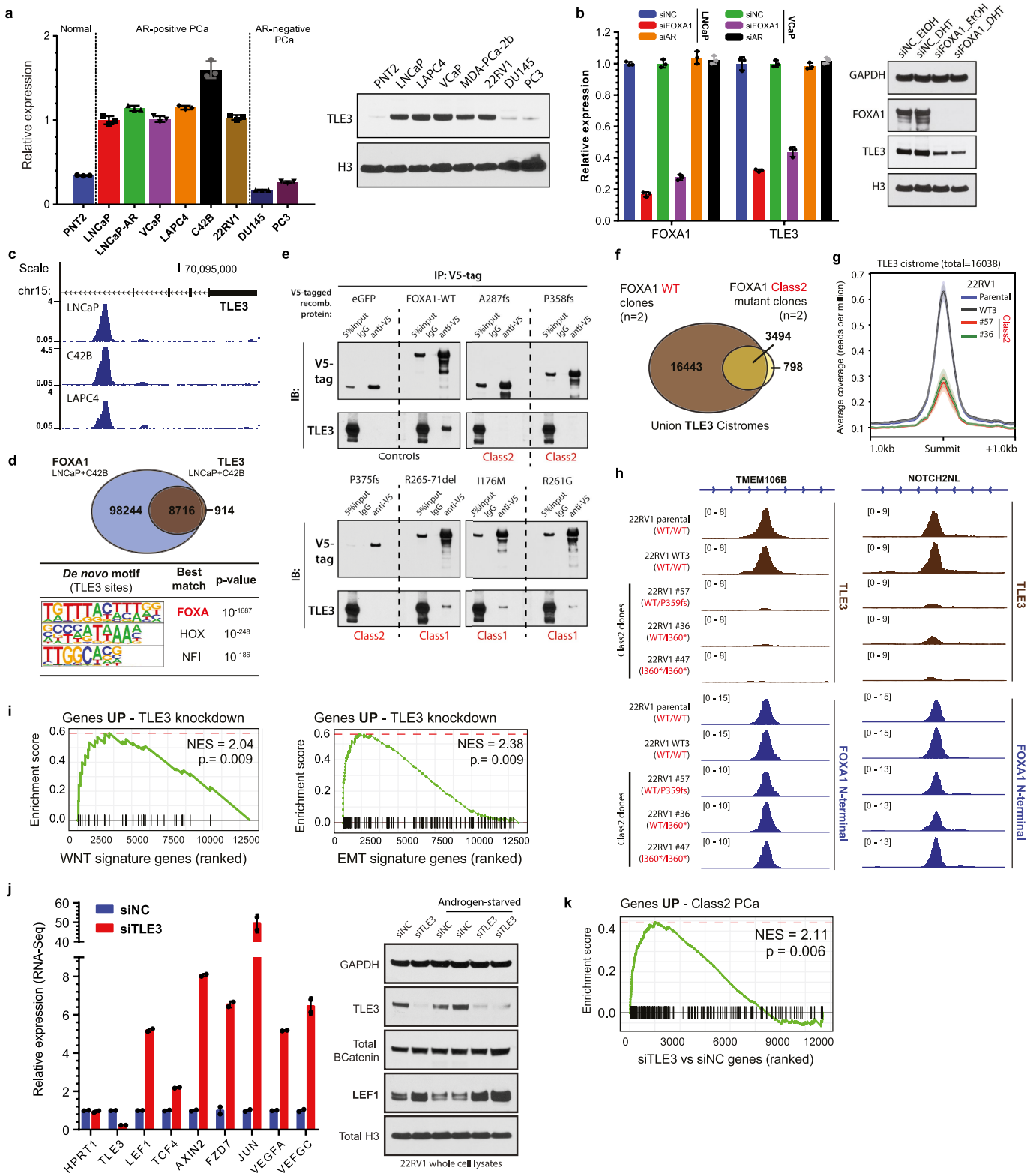
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | DNA-binding affinity and functional essentiality of the class-2 FOXA1 mutants. a, Immunoblot showing comparable expression of recombinant FOXA1 variants in equal volume of nuclear HEK293 lysates used to perform EMSAs. **b**, Higher exposure of EMSA with recombinant wild-type or P358fs mutant and *KLK3* enhancer element, showing the super-shifted band with addition of the V5 antibody (red asterisks; matched to Fig. 3f). **c**, **d**, EMSA with recombinant wild-type or different class-2 mutants (truncated at 268, 287, 358, 375 and 453 amino acids) and *KLK3* enhancer element. Class-2 mutants display higher affinity than wild-type FOXA1. Each class-2 mutant serves as a biological replicate and these results were reproducible in two independent experiments. **e**, DNA association and dissociation kinetics at varying concentrations of purified wild-type or P358fs class-2 FOXA1 mutants from the biolayer-interferometry assay performed using OctetRED system. Overall binding curves and equilibrium dissociation constants (mean ± s.d.) are shown. These results were reproducible in two independent experiments. **f**, Sanger sequencing chromatograms from a set of 22RV1 CRISPR clones confirming the introduction of distinct indels in the endogenous *FOXA1* allele, resulting in a premature stop codon ($n = 2$ technical replicates). Protein mutations are identified on the right. **g**, Immunoblots showing the expression of endogenous wild-type or class-2 mutant FOXA1 variants in parental and distinct CRISPR-engineered 22RV1 clones. **h**, Immunoblots showing expression of FOXA1 (N-terminal antibody) in parental and CRISPR-engineered LNCaP clones expressing distinct class-2 mutants with truncations closer to the FKHD domain. For gel source data (**a**–**d**, **g**, **h**), see Supplementary Fig. 1. **i**, Growth curves of wild-type or mutant clones upon treatment with the non-targeting or *FOXA1*-targeting sgRNAs and CRISPR–Cas9 protein (see Methods). For **i**, distinct class-2 clones and distinct sgRNAs serve as biological replicates. **j**, **k**, Overlap between union FOXA1 (**j**) and AR (**k**) cistromes from wild-type ($n = 3$ biological replicates) and class-2-mutant ($n = 4$ biological replicates) 22RV1 clones. **l**, Overlap between union FOXA1 and AR cistromes from class-2 mutant 22RV1 cells.



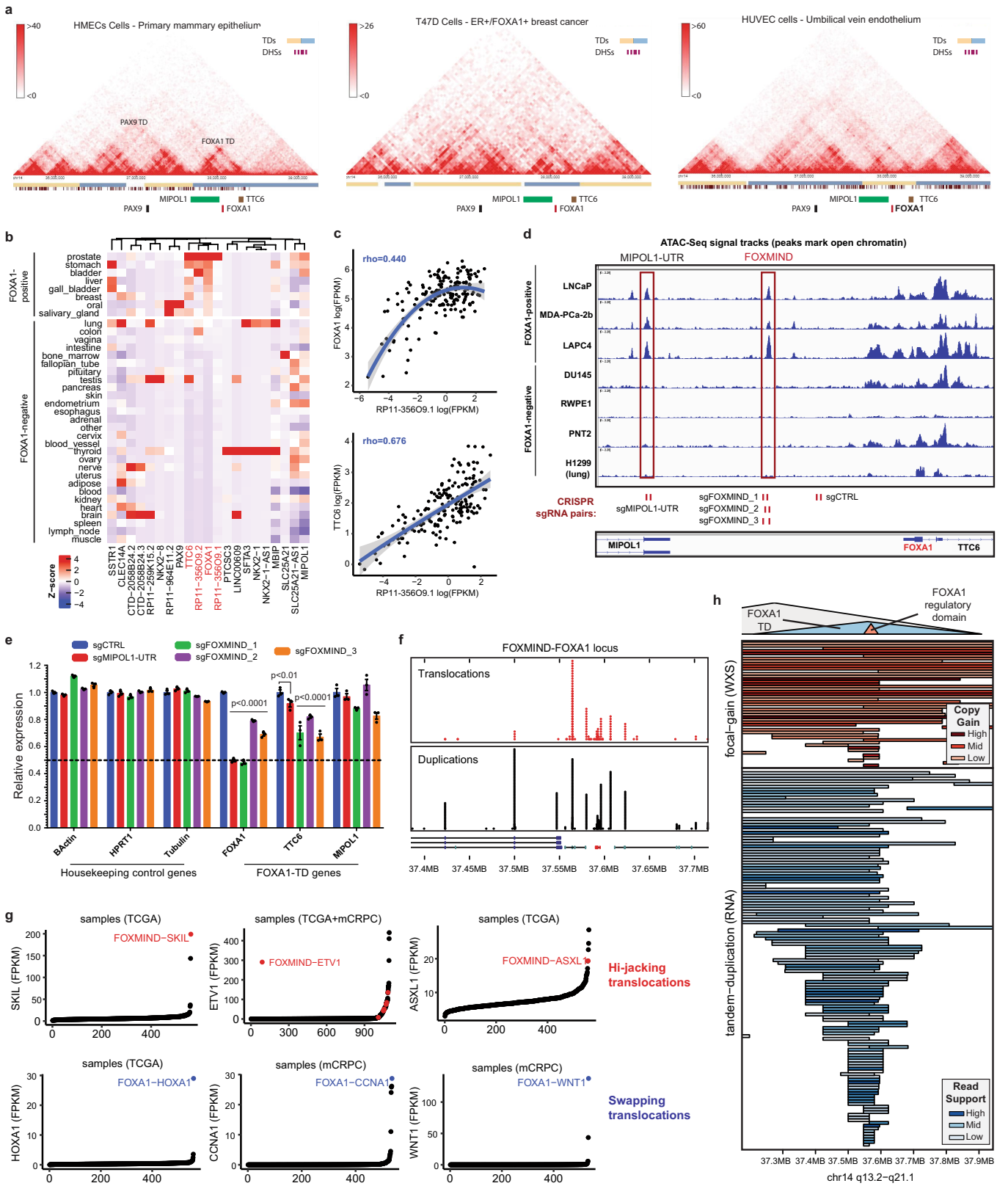
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Cistromic and WNT-driven phenotypic characteristics of the class-2 FOXA1 mutants. a, De novo motif analyses of the wild-type-specific, common and class-2-specific FOXA1-binding site subsets defined from either sequencing-read fold changes (left) or peak-calling scores (right) of ChIP–seq data. Wild-type and class-2 cistromes were generated from $n = 3$ and $n = 2$ independent biological replicates, respectively. Only the top 5,000 or 10,000 peaks from each subset were used as inputs for motif discovery (see Methods) (HOMER, hypergeometric test). **b**, **c**, Per cent of wild-type or class-2 binding sites with perfect match to the core FOXA1 motif (5′-T[G/A]TT[T/G]AC-3′) (**b**) and the consensus FOXA1 motifs identified from these sites (**c**). **d**, **e**, Per cent of binding sites in the three FOXA1-binding-site subsets containing known motifs of the labelled FOXA1 or AR cofactors (**d**), and enrichment of the cofactor motifs in the three binding site subsets relative to the background (**e**). **f**, Genomic distribution of wild-type-specific, common and class-2-specific binding sites in prostate cancer cells. **g**, Differential expression of genes in FOXA1 class-2 mutant CRISPR clones relative to FOXA1 wild-type clones ($n = 2$ biological replicates (limma two-sided test)). **h**, Distinct transcription factor motifs within the promoter (2-kb upstream) of differentially expressed genes. Transcription factors with the highest enrichment (fold change, per cent of upregulated genes with the motif and significance) are highlighted and labelled (two-tailed Fisher's exact test). **i**, Immunoblots showing the expression of β-catenin and vimentin in a panel of wild-type and heterozygous or homozygous class-2 mutant 22RV1 CRISPR clones. **j**, Immunoblots showing the phosphorylation status of β-catenin and expression of direct WNT target genes in select class-2 mutant 22RV1 clones. Immunoblots in **i** and **j** are representative of two independent experiments; every individual clone serves as a biological replicate. For gel source data, see Supplementary Fig. 1. **k**, Representative images of Boyden chambers showing invaded cells stained with calcein AM dye. **l**, Quantified fluorescence signal from invaded cells ($n = 2$ biological replicates per group; two-way ANOVA and Tukey's test). Mean ± s.e.m. is shown and dots are individual data points. **m**, Absolute counts of disseminated cell foci in individual zebrafish embryos as a measure of metastatic burden. **n**, Per cent metastasis at day 2 and day 3 in zebrafish embryos injected with either the normal HEK293 cells (negative controls) or 22RV1 prostate cancer cells virally overexpressing wild-type, class-1 or class-2 mutant FOXA1 variants ($n > 20$ for each group). **o**, Fluorescent signal from the invaded wild-type or class-2-mutant 22RV1 cells after androgen starvation (5% charcoal-stripped serum medium for 72 h) or treatment with the WNT inhibitor XAV939 (20 μM for 24 h; $n = 2$ biological replicates per group; two-way ANOVA and Tukey's test). Mean ± s.e.m. and individual data points are shown.



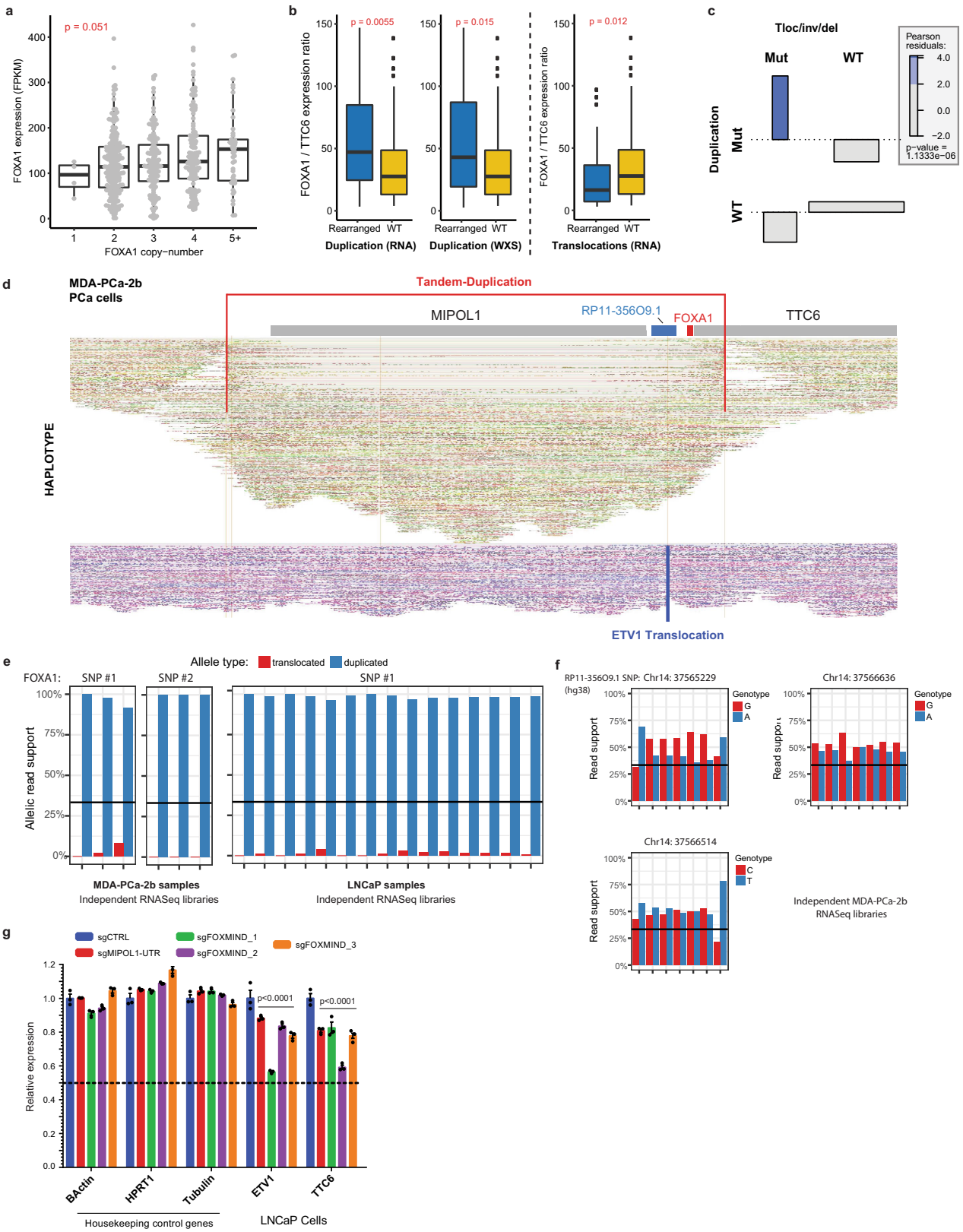
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Functional association of FOXA1 and TLE3 in prostate cancer. a, mRNA (qPCR) and protein (immunoblot) expression of TLE3 in a panel of prostate cancer cells. Mean ± s.e.m. and individual data points are shown. **b**, Left, mRNA expression of *FOXA1* and *TLE3* in LNCaP and VCaP cells treated with siRNAs targeting either *FOXA1* or *AR* ($n = 3$ technical replicates). Two FOXA1 wild-type prostate cancer cells serve as biological replicates. Mean ± s.e.m. and individual data points are shown. Right, protein expression of FOXA1 and TLE3 in matched LNCaP lysates. **c**, FOXA1 N-terminal ChIP–seq normalized signal tracks from LNCaP, C42B and LAPC4 prostate cancer cells at the *TLE3* locus. Each cell line serves as a biological replicate. **d**, Overlap of the union wild-type FOXA1- and TLE3-binding sites from LNCaP and C42B prostate cancer cells ($n = 1$ for each), and top de novo motifs discovered (HOMER, hypergeometric test) in the TLE3 cistrome. **e**, Co-immunoprecipitation assays of labelled recombinant FOXA1 wild-type, class-1 or class-2 variants using a V5-tag antibody in HEK293 cells overexpressing the TLE3 protein. V5-tagged GFP protein was used as a negative control. These results were reproducible in two independent experiments and distinct class-1 and class-2 mutants serve as biological replicates. **f**, Overlap of union TLE3 cistromes from isogenic wild-type ($n = 2$ biological replicates) or heterozygous class-2-mutant ($n = 2$ biological replicates) 22RV1 CRISPR clones. **g**, ChIP peak profile plots from TLE3 ChIP–seq in isogenic FOXA1 wild-type or class-2-mutant 22RV1 clones ($n = 2$ biological replicates each). **h**, Representative TLE3 and FOXA1 ChIP–seq read signal tracks from independent 22RV1 CRISPR clones with or without endogenous *FOXA1* class-2 mutation ($n = 2$ biological replicates each). **i**, GSEA showing significant enrichment of WNT (left) and EMT (right) pathway genes in 22RV1 cells treated with *TLE3*-targeting siRNAs ($n = 2$ biological replicates for each treatment; GSEA enrichment test). **j**, Left, mRNA (RNA-seq) expression of direct WNT target genes in 22RV1 upon siRNA-mediated knockdown of *TLE3* ($n = 2$ biological replicates). Right, Immunoblot showing LEF1 upregulation upon TLE3 knockdown in 22RV1 prostate cancer cells with and without androgen starvation (representative of two independent experiments). For gel source data (**a**, **b**, **e**, **j**), see Supplementary Fig. 1. **k**, Gene enrichment plots showing significant enrichment of class-2 upregulated genes upon *TLE3* knockdown in 22RV1 cells ($n = 2$ biological replicates for each treatment; GSEA enrichment test).



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Topological, physical and transcriptional characteristics of the *FOXA1* locus in normal tissues and prostate cancer. a, HI-C data (from: http://promoter.bx.psu.edu/hi-c/view.php) depicting conserved topological domains within the *PAX9* and *FOXA1* syntenic block in normal and FOXA1$^+$ cancer cell lines. DHSs, DNase I hypersensitive sites. **b**, Highly tissue-specific patterns of gene expression within the *PAX9* and *FOXA1* syntenic block. Tissues were dichotomized into FOXA1$^+$ and FOXA1$^-$ on the basis of FOXA1 expression levels; genes were subject to unsupervised clustering. *z*-score normalization was performed for each gene across all tissues. **c**, Correlation of RP11-356O9.1 (Methods) and *FOXA1* or *TTC6* expression levels across metastatic tissues ($n = 370$; Spearman's rank correlation coefficient). The 95% confidence interval is shown. **d**, Representative ATAC-seq ($n = 1$) read signal tracks from normal basal epithelial prostate (RWPE1 and PNT2 cells) or prostate cancer cells. Cells are grouped on the basis of expression of FOXA1, and differentially pioneered loci are marked with red boxes. CRISPR sgRNA pairs used for genomic deletion of the labelled elements are shown at the bottom. Distinct FOXA1$^+$ and FOXA1$^-$ cell lines serve as biological replicates for ATAC-seq. **e**, mRNA (qPCR) expression of housekeeping control genes, genes located within the *FOXA1* topologically associated domain, and *MIPOL1* in VCaP cells treated with CRISPR sgRNA pairs targeting a control site (sgCTRL), *FOXMIND* or the *MIPOL1* UTR regulatory element (see Extended Data Fig. 2c for sgRNA binding sites). Distinct sgRNA pairs cutting at *FOXMIND* serve as biological replicates. Mean ± s.e.m. is shown ($n = 3$ technical replicates; two-way ANOVA and Tukey's test). **f**, Distribution of tandem duplication and translocation break ends (chimeric junctions or copy-number segment boundaries) focused at the *FOXMIND-FOXA1* regulatory domain. **g**, Outlier expression of genes involved in translocations with the *FOXA1* locus. Translocations positioning a gene between *FOXMIND* and *FOXA1* (hijacking) are shown on top (red). Translocations positioning a gene upstream of the *FOXA1* promoter (swapping) are shown on the bottom (blue). **h**, Inferred duplications within the *FOXA1* locus on the basis of RNA-seq (tandem break ends) and whole-exome sequencing (copy-gains), zoomed-in at the *FOXA1* topologically associating domain.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Transcriptional and genomic characteristics of class-3 FOXA1 rearrangements in prostate cancer. a, Dosage sensitivity of the *FOXA1* gene. Expression of *FOXA1* (RNA-seq) across mCRPC tumours ($n = 370$) as a function of gene ploidy (as determined by absolute copy number at the *FOXA1* locus (two-way ANOVA)). **b**, Relative expression of *FOXA1* (within the minimally amplified region) to *TTC6* (outside the amplified region) in rearranged ($n = 50$) (duplication or translocation) versus wild-type ($n = 320$) *FOXA1* loci (two-sided *t*-test). For all box plots, centre shows median, box marks quartiles 1–3 and whiskers span quartiles 1–3 ± 1.5 × IQR. **c**, Association plot visualizing the relative enrichment of cases with both translocation and duplications within the *FOXA1* locus ($n = 370$). Overabundance of cases with both events is quantified using Pearson residuals. Significance of this association is based on the χ^2 test without continuity correction. Inv, inversion; del, deletion. **d**, *FOXA1* locus visualization of linked-read (10X platform) whole-genome sequencing of the MDA-PCA-2b cell line. Alignments on the haplotype-resolved genome are shown in green and purple. Translocation and tandem-duplication calls are indicated in blue and red, respectively. **e**, Monoallelic expression of *FOXA1* cell lines with *FOXMIND-ETV1* translocations in MDA-PCA-2b ($n = 6$ biological replicates) and LNCaP ($n = 15$ biological replicates). Phasing of *FOXA1* SNPs to structural variants is based on linked-read sequencing (Methods). **f**, Biallelic expression of the RP11-356O9.1 transcript assessed using three distinct SNPs in MDA-PCA-2b cells that contain *ETV1* translocation into the *FOXA1* locus ($n = 7$ biological replicates). **g**, mRNA (qPCR) expression of *ETV1* and *TTC6* upon sgRNA-mediated disruption of the *FOXMIND* or the *MIPOL1* UTR enhancer in LNCaP cells, which also contain *ETV1* translocation into the *FOXA1* locus (see Extended Data Fig. 9d for sgRNA binding sites). Distinct sgRNA pairs cutting at *FOXMIND* serve as biological replicates. Mean ± s.e.m. are shown ($n = 3$ technical replicates; two-way ANOVA and Tukey's test).