AWARD NUMBER: W81XWH-15-1-0277

TITLE: ERF is a Potential ERK-Modulated Tumor Suppressor in Prostate Cancer

PRINCIPAL INVESTIGATOR: Rohit Bose

CONTRACTING ORGANIZATION: University of California, San Francisco, CA

REPORT DATE: April 2021

TYPE OF REPORT: Final Report

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
April 2021	FINAL	30Sep2015-31Dec2020

4. TITLE AND SUBTITLE

ERF Is a Potential ERK-Modulated Tumor Suppressor in Prostate Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-15-1-0277

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Rohit Bose

E-Mail: rohit.bose@ucsf.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

3333 California St Suite 315
San Francisco, California
94118
OSRInfo@ucsf.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Half of all prostate cancers contain an oncogenic gene fusion between the androgen-regulated upstream elements of the TMPRSS2-gene with the consequently upregulated ETS transcription factor ERG. Despite this high prevalence, detecting the presence of TMPRSS2-ERG in patients' tumors has little-to-no useful clinical utility, in part due to a lack of understanding of its mechanisms of oncogenesis. I have characterized a gene, ERF, which functions as a putative tumor suppressor. I had hypothesized that ERF is outcompeted by the TMPRSS2-ERG gene product. Currently, I have identified how ERF and ERG may compete with each other, and as a result, have opposing effects on cancer cell proliferation.

15. SUBJECT TERMS
prostate cancer, tumor suppressor

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	74	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

Page

1. **Introduction……04**
2. **Keywords……04**
3. **Accomplishments……05**
4. **Impact……29**
5. **Changes/Problems……29**
6. **Products……30**
7. **Participants & Other Collaborating Organizations……33**
8. **Special Reporting Requirements……33**
9. **Appendices……34**

ERF IS A POTENTIAL ERK-MODULATED TUMOR SUPPRESSOR IN PROSTATE CANCER

INTRODUCTION

Metastatic prostate cancer is the second most frequent cause of cancer death among American men and remains incurable. Half of such cancers contain a gene fusion, placing a slightly truncated ERG transcription factor downstream of the androgen-regulated promoter of TMPRSS2. This leads to significant upregulation of ERG, which drives tumorigenesis. ERG belongs to the ETS family of transcription factors, which contain the highly conserved ETS DNA-binding domain. There are differences in the biology of fusion-positive tumors versus those lacking the ERG fusion (fusion-negative). However, there is no targeted clinical strategy for treating either fusion-positive or fusion-negative cancers. This is partly due to our limited understanding of how exactly ERG promotes oncogenesis, as well as a lack of awareness of complimentary pathways in fusion-negative tumors. A recent multi-institutional effort sequencing metastatic prostate cancers yielded ERF as 1 of only 19 genes achieving the highest statistical criteria for containing cancer-relevant mutations, with several frameshift mutations as well as point mutations of key conserved residues noted in the critical ETS domain, presumably corresponding to loss-of-function mutations. ERF has known tumor suppressor ability and belongs to the same transcription factor family as ERG, contains a virtually identical DNA-binding ETS domain, but is a strong transcriptional repressor unlike the transcriptionally activating ERG. Most interestingly, all tumors with ERF mutations were fusion-negative. We hypothesize that in fusion-negative tumors, ERF functions as a tumor suppressor by occupying ETS binding sites and inhibiting gene expression, consistent with its repressor function. On the other hand, in fusion-positive tumors, loss of ERF function confers no discernible growth advantage, since ERF is already outcompeted at ETS binding sites by the upregulated ERG oncogene, which in turn activates gene transcription.

KEYWORDS

prostate cancer, tumor suppressor

ACCOMPLISHMENTS

Goals and Accomplishments

I. Specific Aim 1: To investigate how ERF and ERG may compete and have opposing effects on cancer cell proliferation

A. Major Task 1: Evaluate role of ERF in negatively regulating ERG-dependent genes

To investigate the possibility of competition between the two ETS transcription factors ERF and ERG, we used the VCaP cell line which expresses high levels of ERG (and contains wild-type ERF alleles) as a model. The expression of ERG or ERF was separately inhibited via shRNA (**Figure 1a-c**). The androgen transcriptome was analyzed by treatment with and without dihydrotestosterone (DHT). Consistent with earlier work showing that ERG expands the Androgen Receptor (AR) transcriptome in mouse prostatic tissue, inhibition of ERG expression ("ERG-low") resulted in a constricted androgen transcriptome compared to control VCaP cells ("ERG-high") (**Figure 1d-left**). Conversely, ERF inhibition increased the level of AR target gene expression and doubled the size of the androgen transcriptome (**Figure 1d-right**). ERG expression was unaffected by ERF knockdown and vice versa (**Figure 1c**); therefore, the opposing effects of ERF and ERG knockdown on AR output are unlikely due to one ETS factor directly affecting expression of the other.

There were two shortcomings to the approach, which we addressed next. Firstly, we wanted to ensure that the observations from VCaP cells extended to other ERG- dependent tumors. To this end, we interrogated another ERG-dependent model, which are prostate organoids derived from the Pten−/−; R26ERG/ERG genetically-engineered mouse model. When these cells are grafted back into mice, they form tumors rapidly in an ERG-dependent manner. We took these cells and infected them with a doxycycline-inducible FLAG-ERF construct (**Figure 2a**). We confirmed, as expected, that when doxycycline is added to cells, DHT- dependent upregulation of ERG-regulated gene expression is blocked (**Figure 2b**). Secondly, we wanted to show that the expanding effect of ERF knockdown on the DHT- transcriptome, can supersede the contracting effect of ERG knockdown, both described last report. To this end, we took the VCaP cells and simultaneously knocked down expression of both ERF and ERG via shRNA. We evaluated for the size of the DHT-dependent transcriptome. As before, the "ERG-low" cells demonstrated a constricted DHT-transcriptome, but a simultaneous inhibition of ERF expression resulted in a re-expansion of the DHT-transcriptome (**Figure 2c - left panel**), as assessed by RNA-seq. Furthermore, we observed that upregulated genes were now more upregulated with simultaneous ERF knockdown, and downregulated genes were more downregulated (**Figure 2c - right panel**). This is consistent with the model that ERF knockdown re-expands the constricted ERG-low DHT-transcriptome in both number of genes and magnitude of expression. In **Figure 2d**, we draw attention to two example DHT- regulated genes SCD and PLEKHD1, for which we will later show ChIP-seq tracks.

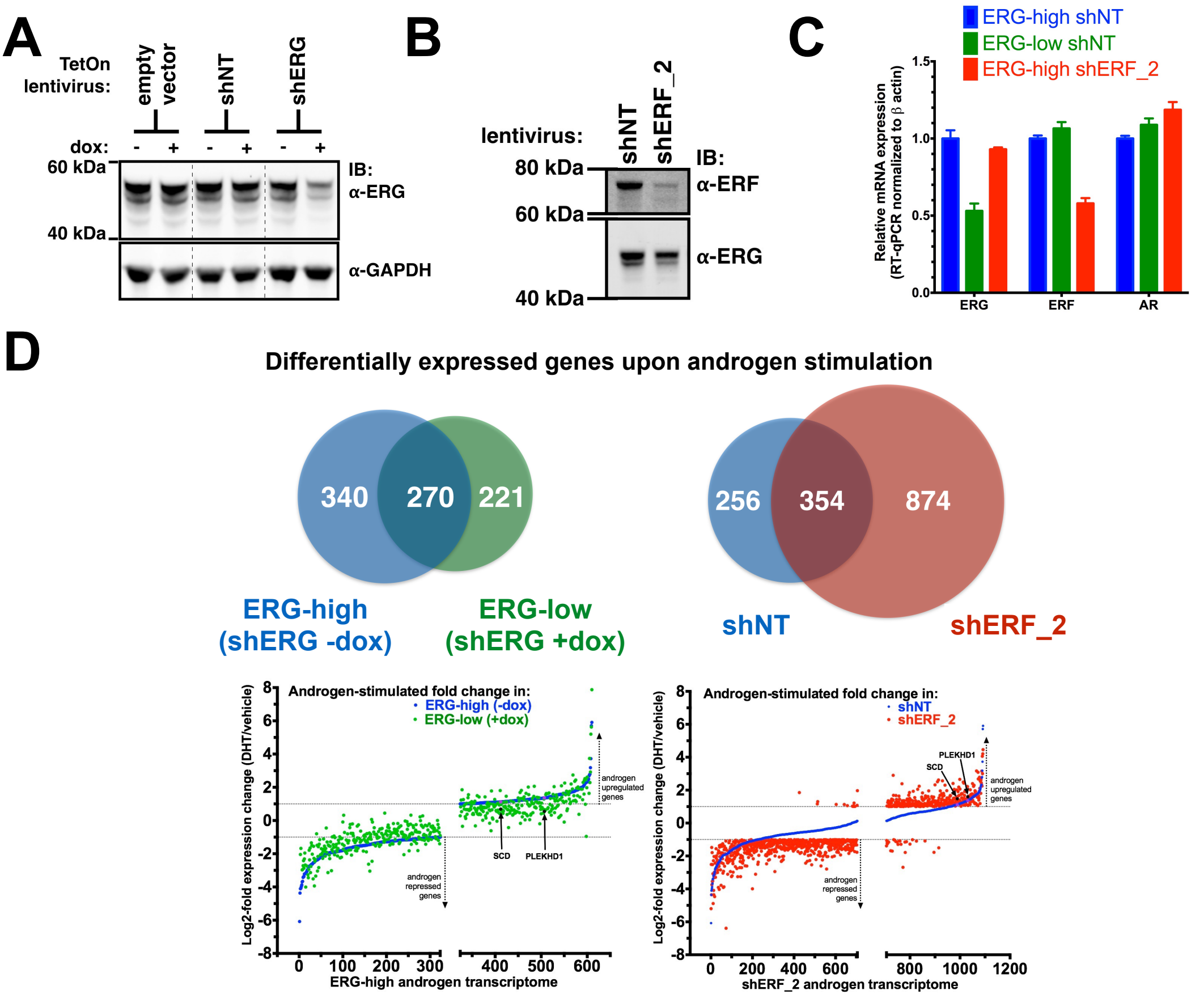


Figure 1 - A) VCaP cells were infected with a Tet-inducible non-targeting shRNA (shNT) or shRNA targeting ERG (shERG) and treated with or without 100ng/mL doxycycline (dox). From here on, cells with shERG without dox will be referred to as ERG-high and with dox as ERG-low. **B)** VCaP cells were infected with non-targeting shRNA (shNT) or shRNA targeting ERF (shERF_2). **C)** RT-qPCR (mean +/- s.e.m.) from VCaP cells infected with non-targeting shRNA (shNT) or shRNAs targeting ERF, as well as a doxycycline (dox)-inducible shRNA targeting ERG, +vehicle (ERG-high) or +dox (ERG-low). **D)** Top – Venn diagrams of androgen regulated genes (>2 fold, FDR<0.05 with 16 hours of 1nM DHT treatment). Bottom – Comparisons of androgen-stimulated fold change of gene expression between the indicated conditions.

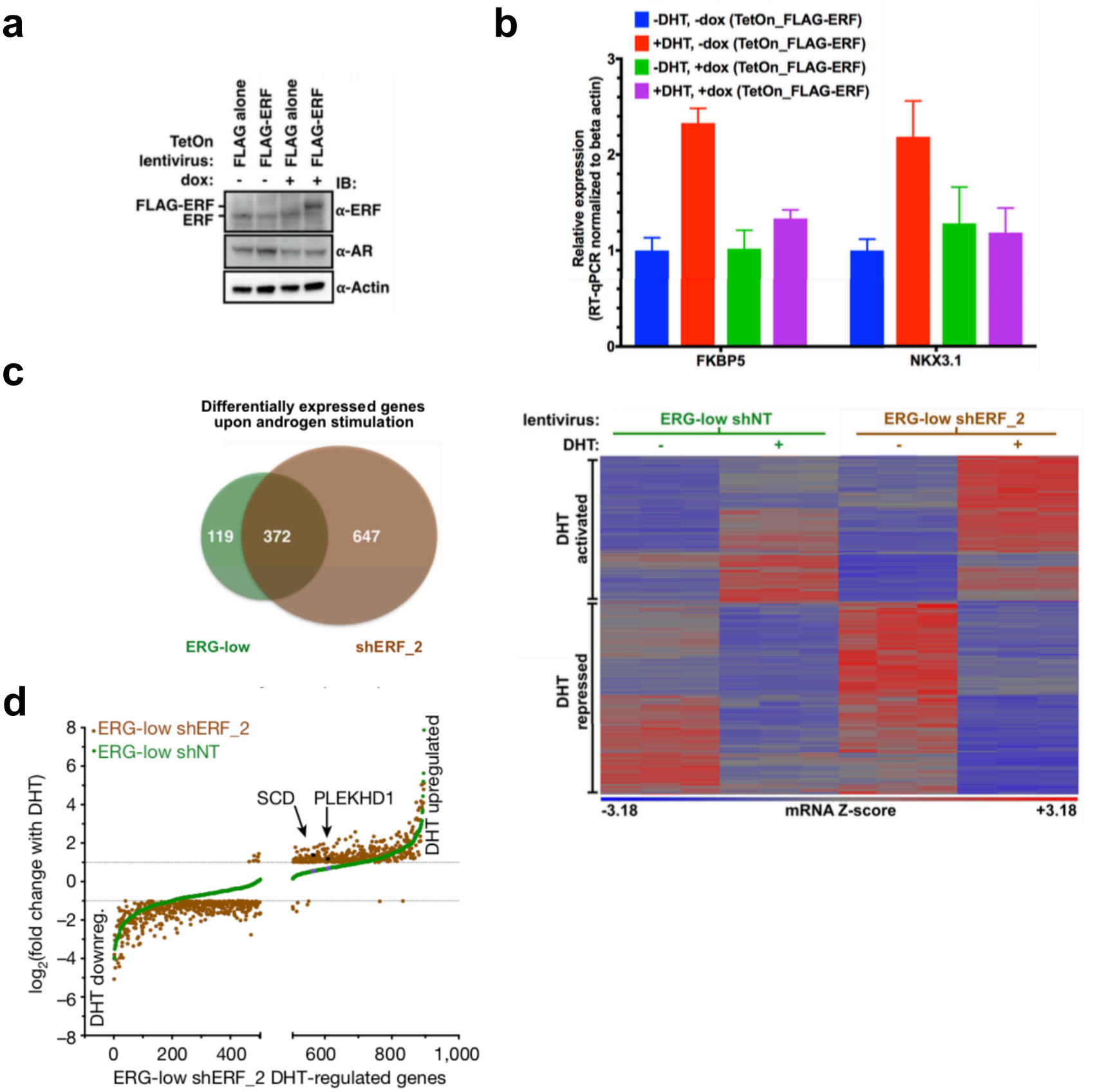


Figure 2 - a) Pten$^{-/-}$; R26$^{ERG/ERG}$ organoids infected with a doxycycline-inducible FLAG–ERF or empty vector (FLAG alone). **b)** RT–qPCR of mRNA isolated from the FLAG–ERF-infected organoids and treated with or without doxycycline, with or without dihydrotestosterone 1 nM for 16 h. Data are mean ± s.e.m.; n = 3 biological replicates. **c)** Androgen-regulated genes (at least a twofold change, FDR < 0.05 by RNA-seq with 1 nM DHT for 16 h) in VCaP cells infected with a doxycycline-inducible shRNA targeting ERG (with doxycycline, ERG-low) and either a constitutive shRNA targeting ERF (shERF_2) or a non-targeting shRNA (shNT). Data analysed by number (left, Venn diagram) and heat map (right); n = 3 biological replicates. **d)** Androgen-regulated genes (defined as in Fig. 1b) in ERG-low VCaPs and shERF_2 analysed by magnitude of change; n = 3 biological replicates. DHT downreg., DHT downregulated.

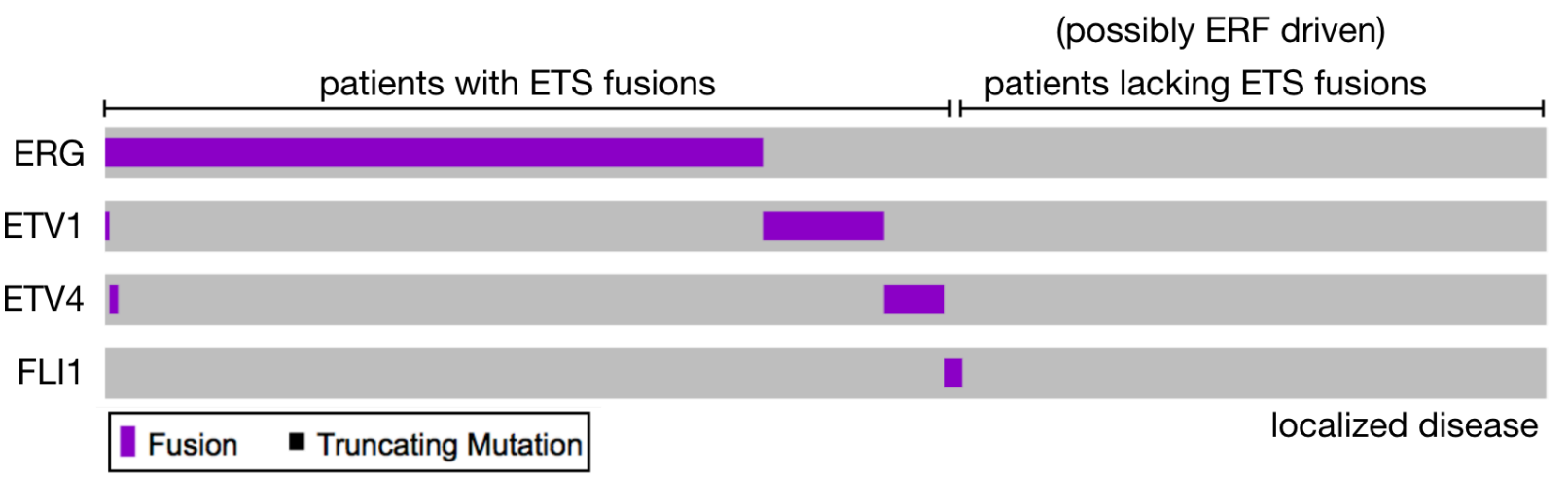


Figure 3 - cBio Oncoprint of patients with locoregional prostate cancer, estimating that up to 40% of tumors may be ERF driven by either genomic or non-genomic changes. Each column corresponds to a unique patient's tumor profile.

The limitation of these above studies is that they were done *in vitro*. Thus, it is unclear how much validity they have with respect to *in vivo* patient tumors. Thus, we now shifted our focus to analyzing the relevant genomic data of patients. The available cohorts of prostate cancer patients are now considerably larger than they were at the onset of this project, enabling a more robust analysis of ERG-positive and similar ETS oncogenic tumors, as well as identifying the contrasting ETS fusion negative tumors, in which ERF would be expected to play a role. In our current analysis of our primary biopsies of locoregional prostate cancer, we have identified nearly 50% of patients possessing an ERG fusion, as well as an additional 10% of patients possessing fusions in the related oncogenic ETS factors: ETV1, ETV4 or FLI1 (**Figure 3**). The latter three genes are other oncogenic ETS transcription factors that form fusions with androgen-regulated genes, analogously to TMPRSS2 fusions with ERG. Like in that case, the most common fusion partner is still TMPRSS2, but differently to ERG, these fusions are interchromosomal rather than intrachromosomal, which might account for their lower frequency. The fact that these various oncogenic ETS fusions occur in a mutually exclusive pattern, for the most part, suggests that they have redundant functions. This is all critically relevant, as it leaves about 40% of prostate cancer patients that lack any discernible ETS fusion, and it is in these patients that we expect that the expression level and function of ERF will be most critical (**Figure 3**).

We believe that the mutual exclusivity between ERG fusions and patients with low ERF expression likely occurs because either is sufficient to drive prostate tumor initiation. However, as we showed in the previous report, even in ERG dependent prostate cancers, the relative level of ERF can significantly influence tumor formation. Thus, whereas exclusivity may be observed in early tumors, in advanced tumors in which the tumor mutation burden is higher, and have also undergone treatment with inhibitors of the AR axis, we think it is more likely that both ERG and ERF molecular alterations would be observed simultaneously. Further complicating this, is that more variable ETS expression is observed in advanced disease (owing to different levels of androgen signaling driving the fusion partner expression), so we will restrict our analysis to cells with above median expression of ERG, ETV1, ETV4 or FLI1, rather than using fusion detection. Our updated analysis of our metastatic biopsies of castration-resistant prostate cancers reveals that a similar fraction of patients possess high expression of an oncogenic ETS factor, to those with locoregional disease, consistent with prior reports (**Figure 4**). However, we expect that future analysis of such cohorts will reveal a greater percent of ERF molecular alterations (either at the genomic or transcriptomic levels), based on the synergy observed in cells between low ERF and high ERG.

B. Major Task 2: Evaluate role of ERF in competing for DNA binding with ERG

Given the similarity of the ETS domains between ERF and ERG, as well as the ability of ETS factors to displace one another from canonical GGA(A/T) motifs, we postulated that the opposing effects of ERF and ERG on androgen signaling could be explained by competition for

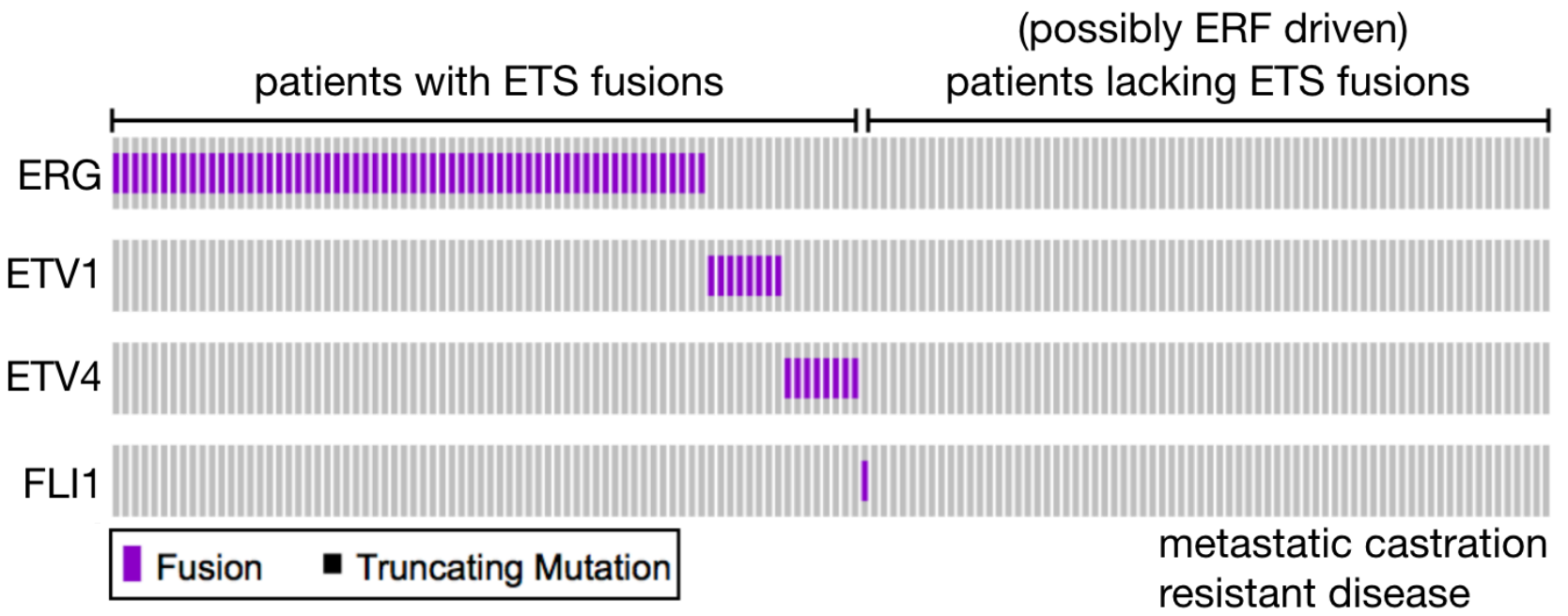


Figure 4 - cBio Oncoprint of patients with metastatic castration-resistant prostate cancer, estimating which tumors may be ERF driven by genomic or non-genomic changes. Each column corresponds to a unique patient's tumor profile.

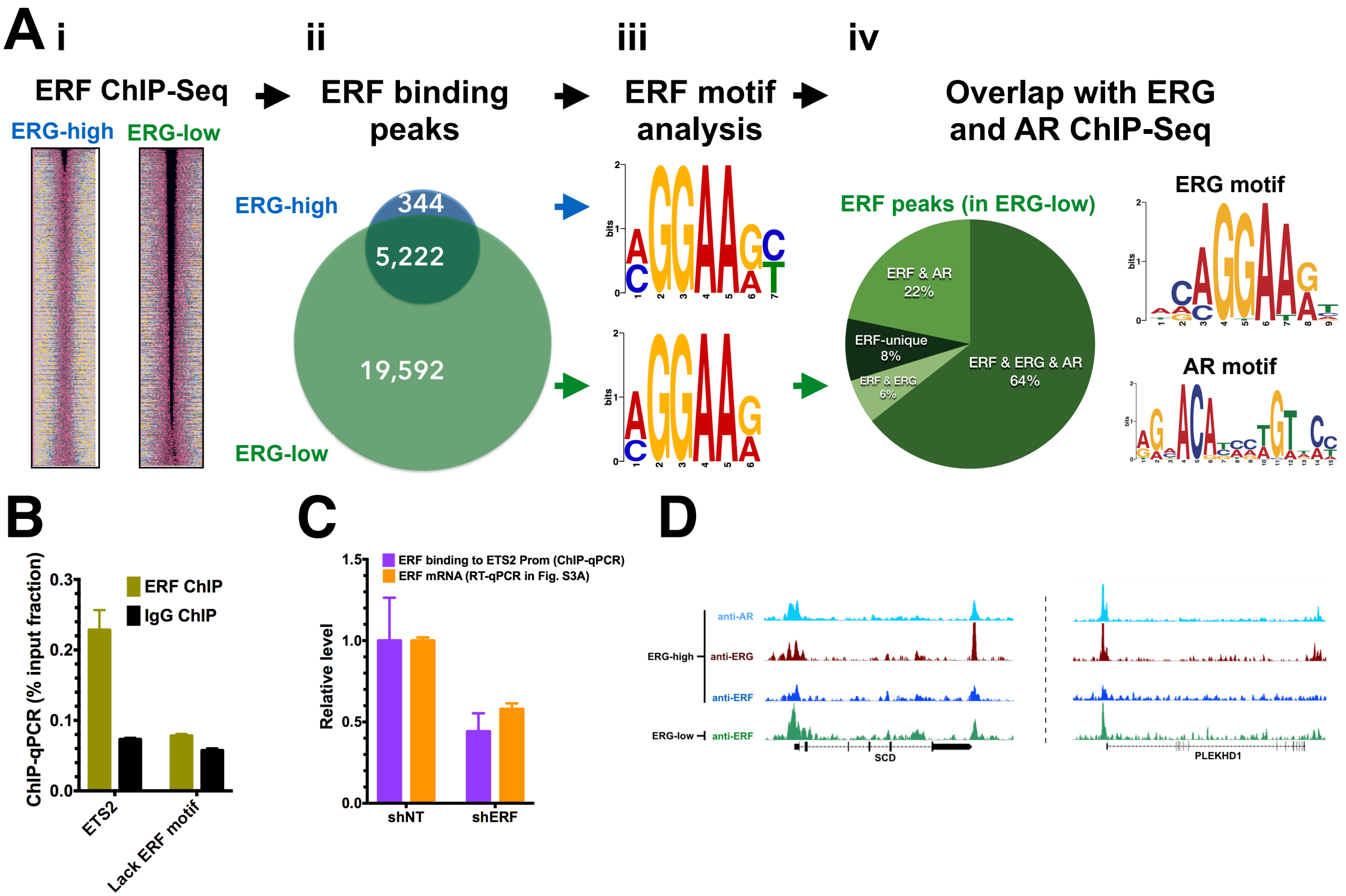


Figure 5- A) i – Heatmaps depicting ERF ChIP-seq read density in +/- 2kb regions of the ERF peak summits. The same number of ChIP-seq reads were loaded into the software seqMINER for creating the two maps, sorted by the read densities in the ERG-low condition. ii – Venn diagram showing the overlap of ERF ChIP peaks in the ERG-high and ERG-low states. iii – De novo motif analysis from the ERF peaks in the ERG-high (top) and ERG-low (bottom) states. iv – Pie chart illustrating the overlap between ERF ChIP-seq peaks with ERG and AR peaks, as well as the top motifs enriched in the ERG and AR ChIP-seq peaks. **B)** ChIP-qPCR (mean +/- s.e.m.) in VCaP cells for the ETS2 promoter region that contains a known ERF binding site as a positive control, and with an upstream element of PSA lacking the ERF binding motif noted in panel Aiii, as a negative control. **C)** Comparison of effect of ERF shRNA inhibition on ERF binding to ETS2 promoter by ChIP-qPCR versus ERF mRNA by RT-qPCR (mean +/- s.e.m.). **D)** ChIP-seq signals for the SCD and PLEKHD1 loci.

AR-associated ETS binding sites. To explore this possibility, we performed ERF ChIP-seq in the ERG-high and ERG-low states (**Figure 5A-D**). De novo motif analysis of the ChIP-seq peaks identified the canonical ETS motif as the primary ERF binding site, with 5,566 binding sites in the ERG-high condition (**Figure 5Ai-iii**). Remarkably, an additional 19,592 ERF binding sites were observed in the ERG-low state, nearly all of which were bound by ERG prior to ERG knockdown (**Figure 5Ai-iv**). Furthermore, the ChIP-seq signal intensity of the smaller number of ERF peaks observed in the ERG-high state was increased in the ERG-low state in almost all cases (**Figure 5Ai**). Finally, ERF peaks in the ERG-low state largely overlap with AR-binding peaks, as previously reported for ERG. PLEKHD1 and SCD are two AR-regulated genes that illustrate how ERG and ERF levels can modulate their expression. In the ERG-high state, both genes have prominent AR- and ERG-associated binding sites but only limited ERF binding (**Figure 5D**). However, ERF binding is substantially increased in the ERG-low state (**Figure 5D**) in parallel with a decrease in androgen-induced expression (**Figure 5D-lower left**). On the other hand, ERF inhibition yields only a small increase in androgen-induced expression for these two genes (**Figure 5D-lower right**), consistent with minimal ERF binding at baseline. These data suggest a model whereby the relative levels of ERF and ERG influence the size and output of the AR transcriptome and compete for ETS binding sites.

The above VCaP system was ideal for initial studies, since it has high levels of endogenous ERG. However, this left open the possibility that the competition that we observed was a cell-line specific phenomenon, and not generalizable to other prostate cancers or normal prostate tissue. To this end, and given the role that ERF also plays in benign prostate cells, I performed ChIP–seq to confirm that ERF binds also binds ETS sites in normal prostate organoids. It is important to note that ERG is not endogenously expressed in normal prostate, unlike half of prostate cancers (**Figure 6a**). De novo motif analysis of the ERF ChIP-seq peaks identified the canonical GGAA ETS DNA motif as the primary ERF binding site (**Figure 6e**). Remarkably, transient overexpression of ERG led to a significant decrease of ERF chromatin occupancy (**Figure 6a-e**), consistent with the competition for binding also seen previously in VCaP cancer cells. Furthermore, the ChIP-seq signal intensity of the smaller number of ERF peaks observed in the FLAG-ERG state was increased in the FLAG-alone state in almost all cases (**Figure 6b**). Sh3bp5 and Rph3aI are example genes demonstrating the ERG-mediated inhibition of ERF chromatin occupancy (**Figure 6d**). Given the similarity of the ETS domains between ERF and ERG, as well as the ability of ETS factors to displace one another from canonical GGAA motifs, we postulated that the opposing effects of ERF and ERG could be explained by competition for ETS binding sites. I explored the overlap between androgen receptor and ERF sites in normal prostate organoids by androgen receptor ChIP–seq and found 28% of the ERF sites overlap with androgen receptor binding sites (**Figure 6e**). This was lower than observed in the previously analyzed ERG-positive VCaP tumor cells, consistent with differences in androgen receptor chromatin occupancy between normal prostate and cancer. Collectively, these data suggest a model whereby the relative levels of ERF and ERG compete for ETS binding sites.

Next, we performed a set of large scale ERF ChIP-seq in both the ERG-high and ERG-low states with multiple biologic replicates. As predicted, due to putative competition between ERF and

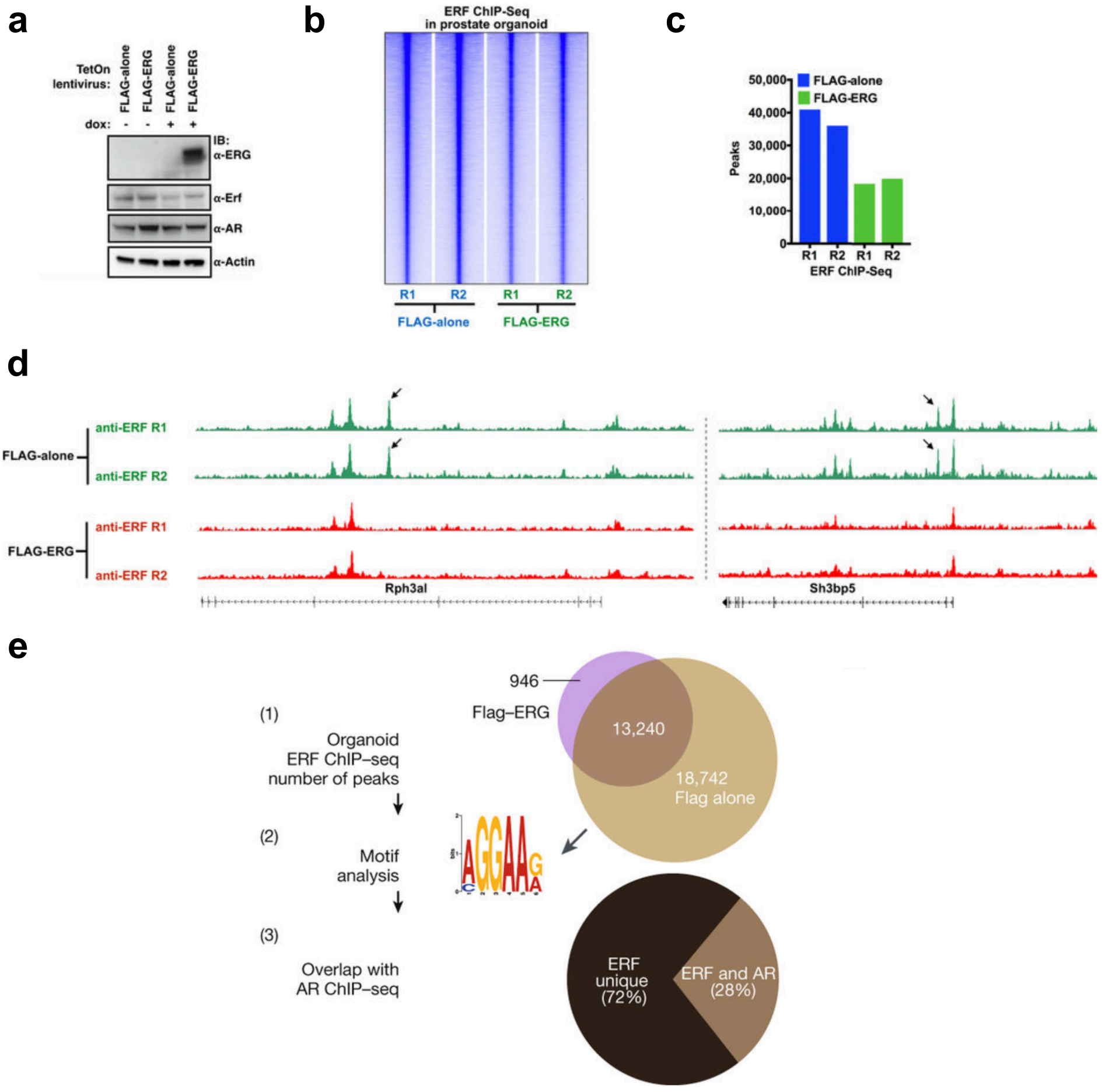


Figure 6 - a) Mouse normal prostate organoids were infected with a TetOn doxycycline-inducible Flag–ERG or empty vector (Flag-alone) and treated with or without doxycycline. **b)** ERF ChIP–seq in normal prostate organoids infected with either Flag-alone or Flag–ERG lentivirus, both treated with doxycycline (n = 2 biological replicates: R1, R2), and analysed with heat maps. **c)** Comparison of ERF ChIP–seq peak numbers, n = 2 biological replicates (R1, R2). **d)** ERF ChIP–seq signals for example genes Rph3a1 and Sh3bp5 loci, normalized by read depths. **e)** ERF ChIP–seq in normal prostate organoids infected with either a Flag–ERG lentivirus or Flag alone, both treated with doxycycline (n = 2 biological replicates: R1, R2) analysed by (1) peak overlap, (2) motif analysis, and (3) overlap with ERG and androgen receptor ChIP–seq.

ERG, ERF binding to chromatin was globally increased in the ERG-low state (**Figure 7a**). To appreciate this more clearly, we have chosen a representative 10 Mb region, where clearly both biologic replicates of ERF-ChIP demonstrated markedly increased binding in the ERG-low state (**Figure 7b**). Moreover, by performing simultaneous AR and ERG ChIP, it can be noted that the newly acquired binding sites by ERF in the ERG-low state correspond to previously occupied ERG binding sites at AR loci. Finally, to tie this data together with **Figure 2d**, we have demonstrated the ChIP-seq tracks of two representative genes transcriptionally regulated by DHT, PLEKHD1 and SCD. In both cases, you can clearly see examples of increased ERF-chromatin occupancy in the ERG-low state, at sites previously occupied by ERG, at AR loci (**Figure 7c**). Collectively, these data suggest a model whereby the relative levels of ERF and ERG compete for ETS binding sites.

Thus, at the cistromic level, when ERG levels are reduced, ERF is able to occupy DNA binding sites previously bound by ERG. This is consistent with a competition model, but not proof of one. Thinking about this model with a greater degree of granularity, there are several levels as to how this competition occurs. One is that from sheer stoichometry and relative levels of ERG and ERF, there is differential occupancy of GGA(A/T) binding sites. Another is that ERG and ERF may have different affinities for those GGA(A/T) sites. Thirdly, the subcellular localization of these ETS factors may be differentially regulated. Fourthly, not all GGA(A/T) sites are necessarily equal; while clearly some can be occupied by either factor, there are individual sites that may be more ERF or ERG specific. We are just beginning to untangle these possibilities. Again, focusing on our efforts on analysis of our human tumors, we can begin to address the first level, whereby we can see about a 30 fold upregulation of ERG expression in the presence of a TMPRSS2-ERG fusion, compared to tumors lacking an ERG fusion, or normal prostate (**Figure 8**). Incidentally, the expression of ERF in normal tissue (**Figure 9**) is highest in reproductive tissues, including prostate tissue. While not proof, this is consistent our model of high ERF being critical in these tissues to maintain homeostasis.

C. Major Task 3: Evaluate for in vitro role of shERF

The near complete lack of ERF mutations in tumors with TMPRSS2–ERG fusions led us to investigate whether ERF loss recapitulates the phenotype of ERG gain, keeping in mind that ERG is not expressed in benign prostate epithelium. To the phenotype of ERF loss in vitro, we infected normal murine prostate organoids with shRNA targeting murine Erf (shErf) (**Figure 10A**). The shErf organoids acquired morphologic characteristics of ERG overexpression: they formed single-cell luminal structures lacking basal cells (**Figure 10B**) and RNA-seq profiling revealed significant enrichment for genes whose expression is upregulated by androgen in human prostate cancer cells (**Figure 10C**).

However, ERG only displays potent oncogenicity in a mouse Pten-/- background. To this end, and to determine whether ERF loss yields a phenotype similar to ERG gain, I now infected

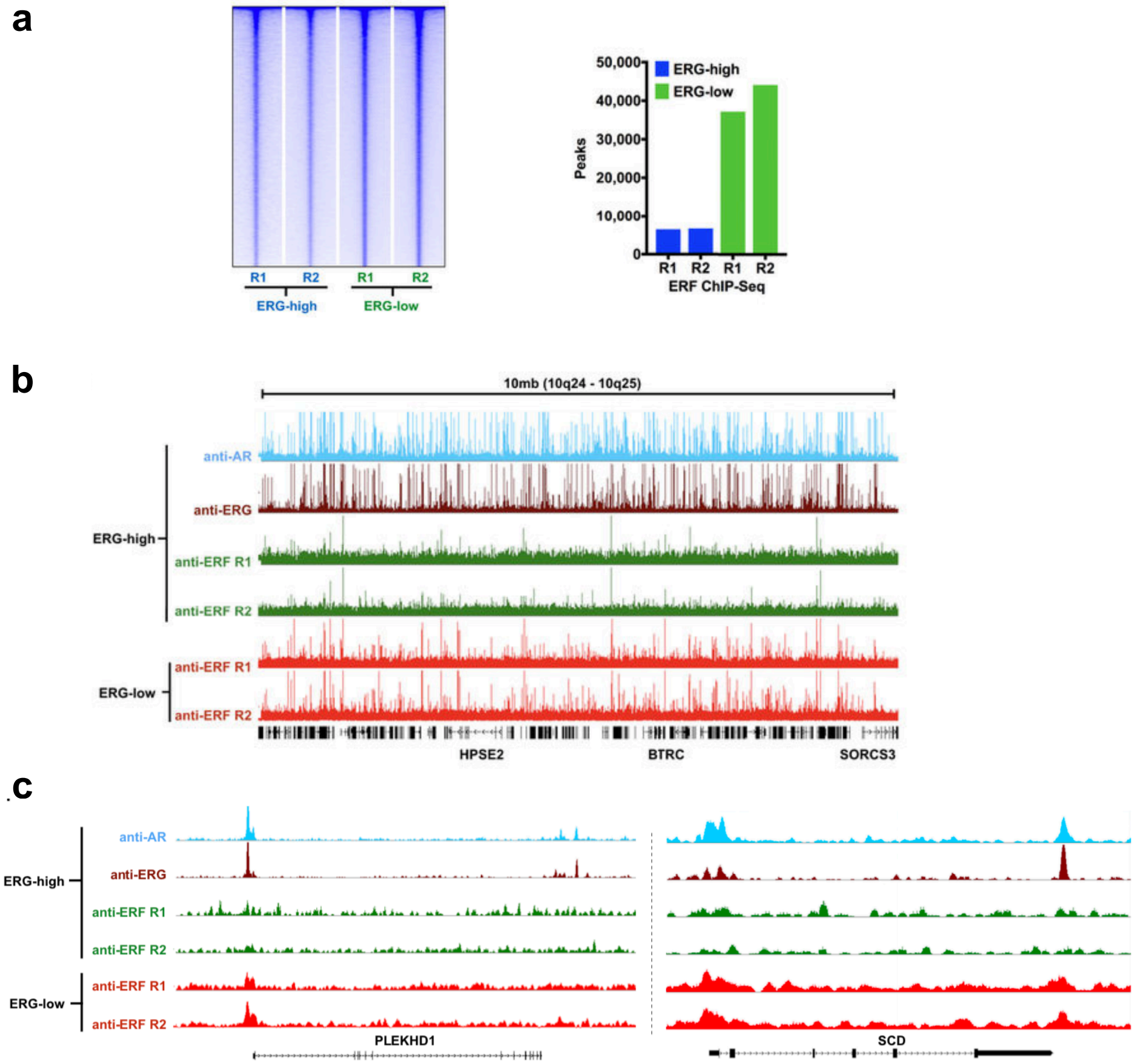


Figure 7 - a) ERF ChIP–seq in VCaPs infected with a doxycycline-inducible shRNA targeting ERG in the absence (ERG-high) or presence of 100ng/mL doxycycline (ERG-low); (n = 2 biological replicates: R1, R2) analysed by heat maps (LEFT panel) or peak number (RIGHT panel) **b)** A 10-Mb region illustrating ChIP–seq of ERF binding in ERG-high condition compared with ERG-low. **c)** ChIP–seq signals for the SCD and PLEKHD1 loci (examples of androgen-regulated genes). In both b) and c), ChIP–seq signals at the y axis were normalized by read depths.

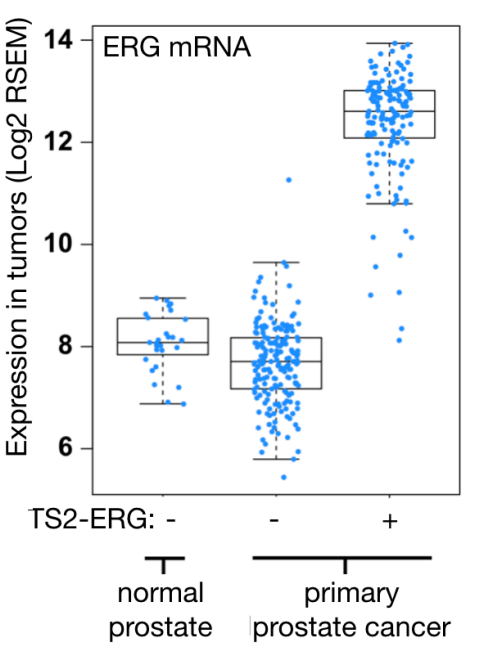


Figure 8 - ERG mRNA levels in patient samples from normal prostate, fusion-negative primary prostate cancer, and fusion-positive primary prostate cancer

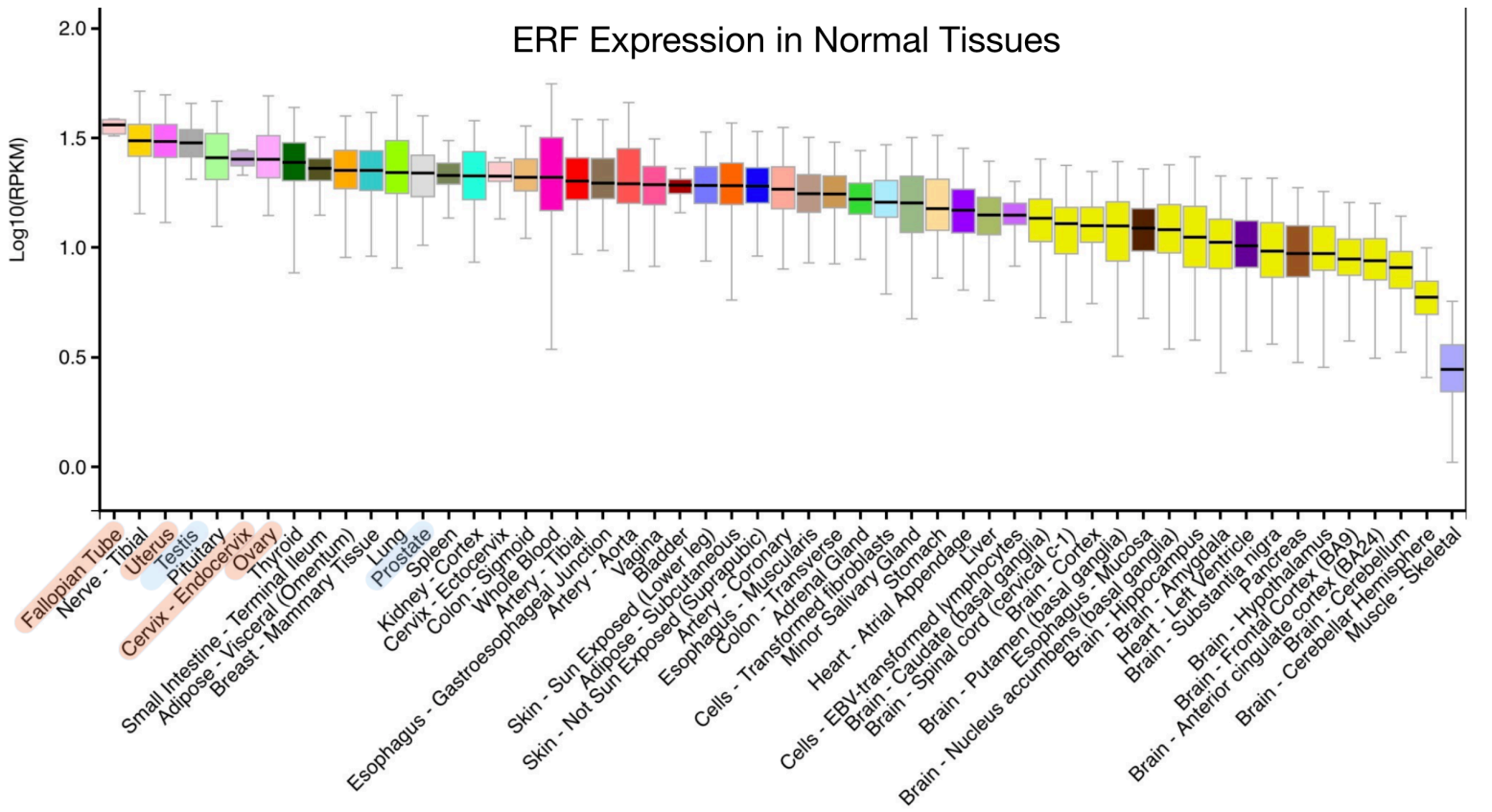


Figure 9 - Log10 RPKM mRNA levels demonstrate higher expression of ERF in reproductive tissues compared to others

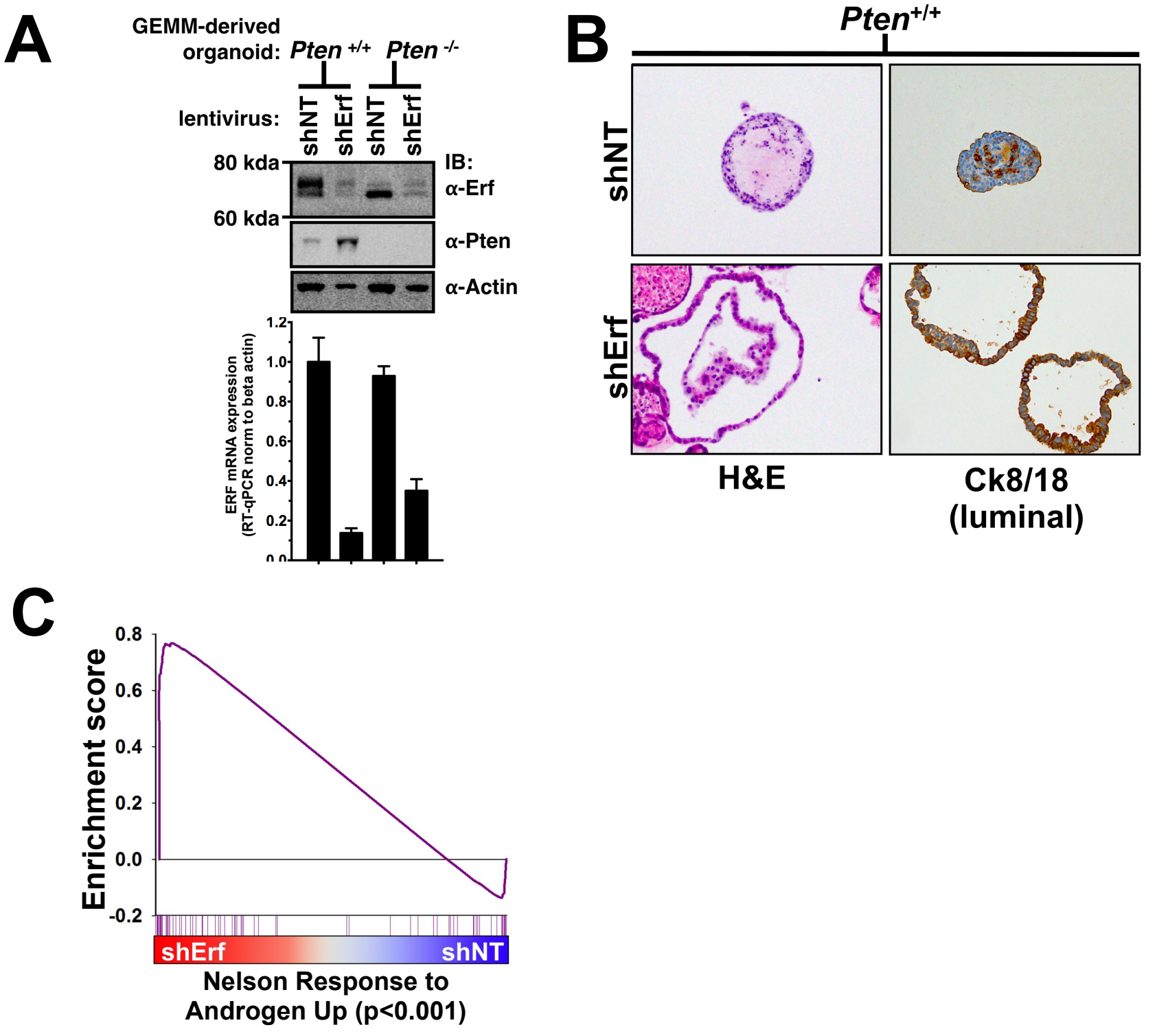


Figure 10 - A) Organoids derived from Pten+/+ and Pten-/- GEMM prostates infected with non-targeting shRNA (shNT) or mouse Erf (shErf). Western blot (top) and RT-qPCR (bottom). Error bars represent s.e.m. from 3 biological replicates **B)** Histology and IHC analysis of the murine prostate organoids. **C)** Expression profiles of shErf and shNT in normal prostate compared by GSEA for an androgen expression signature

prostate organoids derived from Pten−/− mice with short hairpin RNA (shRNA) targeting mouse Erf (shErf_m). The *Pten*$^{-/-}$ shErf_m organoids acquired morphological characteristics of ERG overexpression: they formed luminal structures lacking basal cells (**Figure 11a**), and profiling by RNA sequencing (RNA-seq) and qPCR demonstrated profound loss of expression of most basal signature genes (**Figure 11b–e**). Consistent with these changes, RNA-seq profiling revealed significant enrichment for genes whose expression is upregulated by androgen in human prostate cancer cells (**Figure 11c - right panel**), which will be explored further in Specific Aim 3 in the future. Collectively, this supports the hypothesis that ERF loss leads to a pro-luminal shift, consistent with the changes observed as normal prostate tissue transitions to adenocarcinoma.

We have now performed an unpublished analysis of our genome-wide screen in VCaP cells, whereby we reduced ERG expression and thereby reduced cell proliferation. We then identified which shRNAs specifically rescue this phenotype. As can be clearly seen from the Z-score normalized the data, shERF hairpins are some of the most depleted specifically in the ERG-knockdown cells, as compared to the high ERG cells (**Figure 12**). Moreover, this was observed in all 6 shRNAs targeting ERF. This supports in an unbiased way, that ERF is a critical negative regulator of ERGs biologic function.

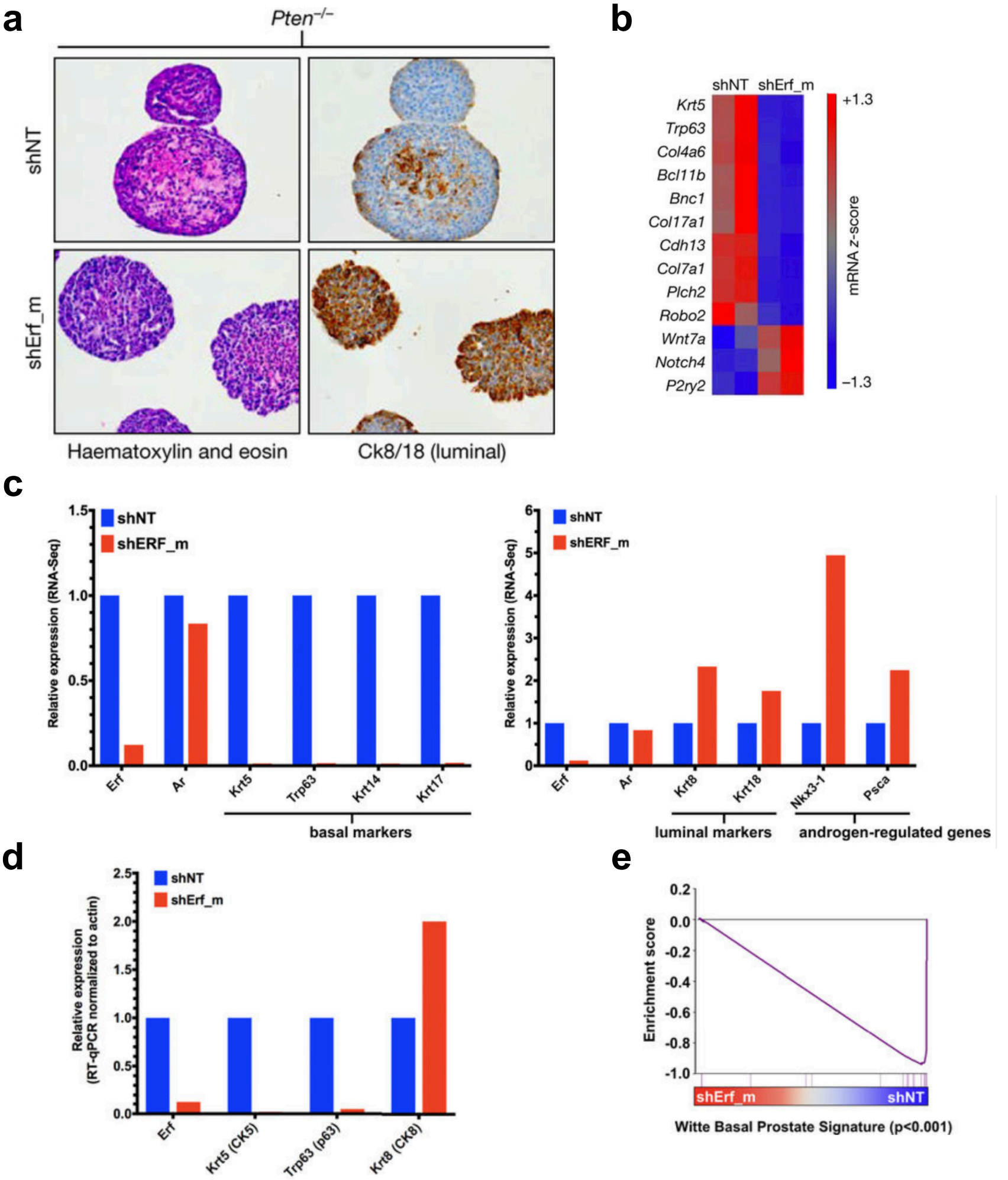


Figure 11 - a) Immunohistochemistry of *Pten*$^{-/-}$ mouse prostate organoids infected with non-targeting shRNA (shNT) or targeting ERF (shErf_m). **b)** Basal signature applied to *Pten*$^{+/+}$ organoid RNA-seq (n = 2 biological replicates). **c)** RNA-seq analysis of organoids derived from *Pten*$^{+/+}$ and *Pten*$^{-/-}$ mouse prostates infected with non-targeting shNT or shErf_m; *n* = 2 biological replicates. **d)** RT–qPCR analysis of the *Pten*$^{+/+}$ organoids; n = 2 biological replicates. **e)** *Pten*$^{+/+}$ organoid RNA-seq (n = 2 biological replicates) interrogated by GSEA for expression signature of Witte basal prostate cancer

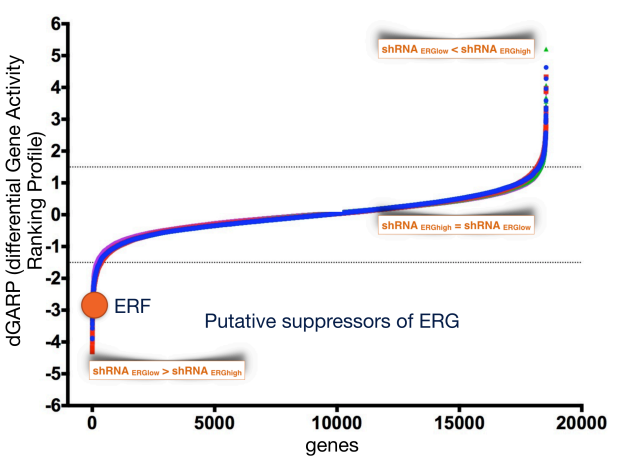


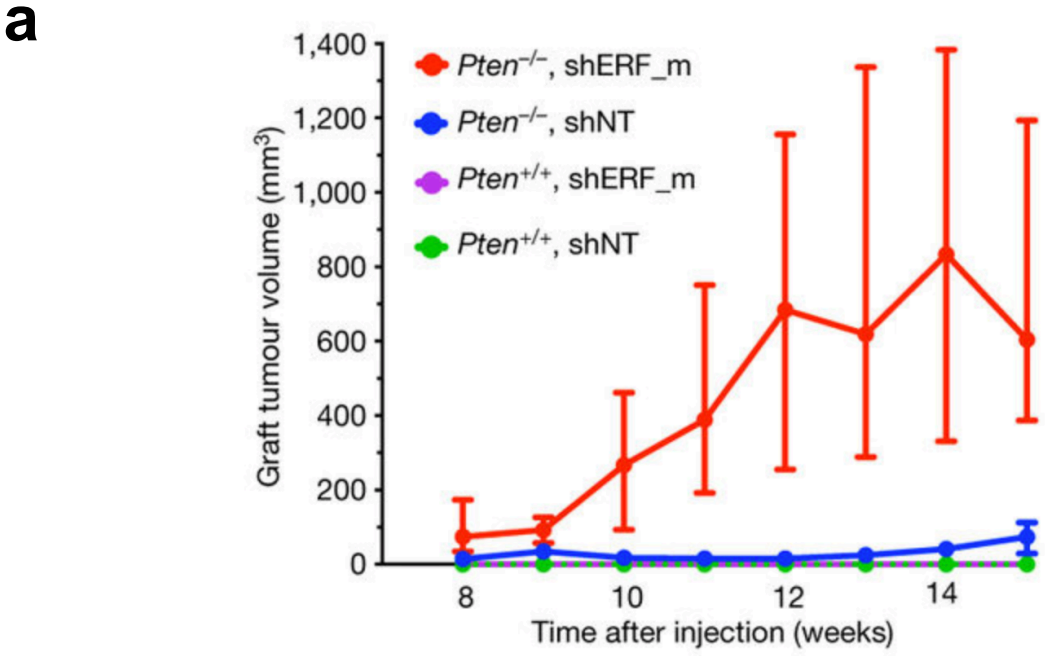
Figure 12 - Z-score normalized data from genome-wide screen. The genes have been rank ordered along the X-axis. Genes on the extreme left are targeted by shRNAs that are most enriched in the ERG-low state, and genes on the extreme right are targeted by shRNAs that are most depleted in the ERG-low state.

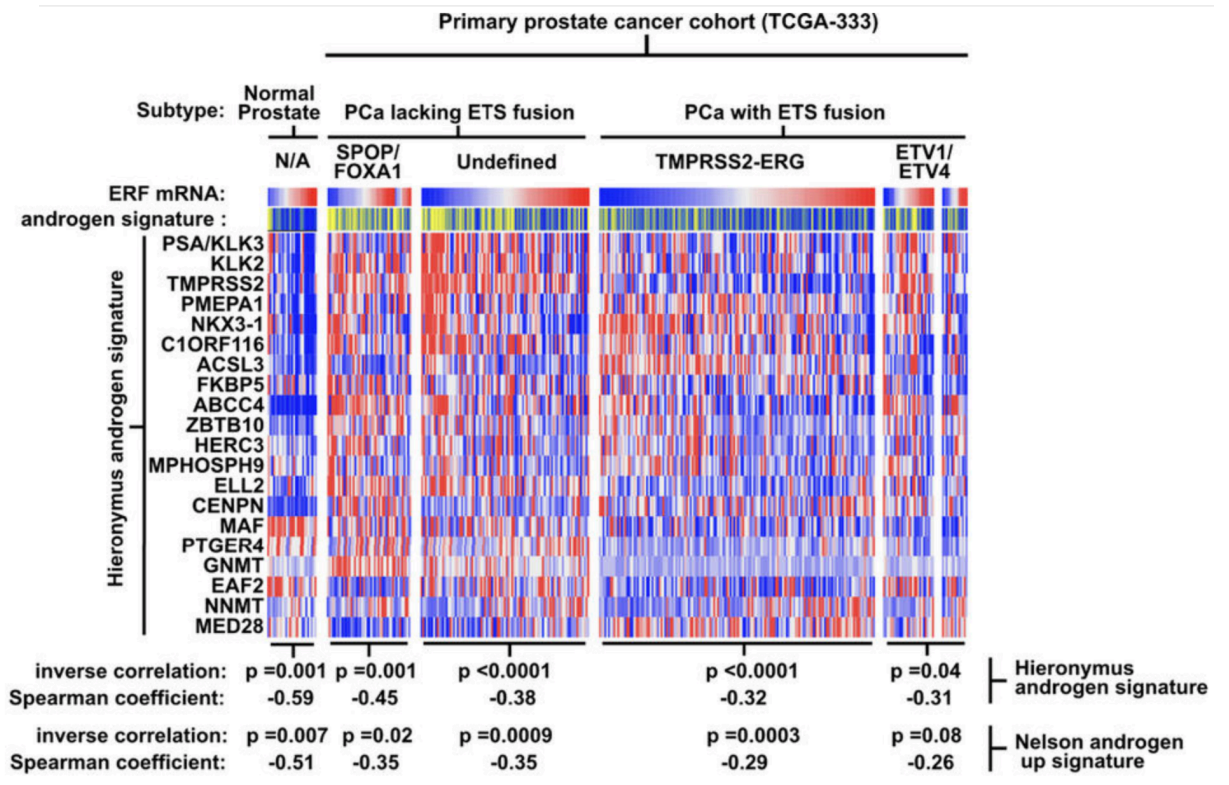
Specific Aim 2: To evaluate for in vivo tumor suppressor function of ERF

A. Major Task 1: Evaluate role of ERF as in vivo tumor suppressor

The Pten−/− organoids infected with shErf_m were indeed able to form tumors when grafted back into mice, recapitulating the phenotype of ERG overexpression and confirming that ERF is a bona fide tumor suppressor (**Figure 13a**). Next, to assess the impact of ERF in human prostate cancers, we interrogated the mRNA expression profiles of the primary and metastatic human prostate cohorts TCGA-333 and SU2C-150 respectively. In agreement with the prior functional studies, ERF mRNA levels are inversely correlated with two luminal transcriptional activity signatures both in normal human prostate and in all primary tumor subtypes (**Figure 13b**). This reciprocal association is also observed in metastatic cancers if the analysis is limited to tumors without amplification of or mutations in the AR gene, which encodes the androgen receptor (**Figure 13c**). The fact that reduced ERF expression enhances luminal transcriptional output even in the absence of mutation or deletion, raises the possibility that ERF may have a broader role in prostate oncogenesis.

This above models the approximate 5% of patients that have genomic mutation or focal deletion of the ERF locus, and lack TMPRSS2-ERG gene fusions. However, that data did not confirm an in vivo role for ERF in ERG-high tumors, which represent about 30-50% of American prostate cancer patients. To this end, we interrogated the previously-described prostate organoids derived from the Pten−/ −; R26ERG/ERG genetically-engineered mouse model, which form tumors rapidly in an ERG- dependent manner when grafted back into mice. We infected these organoids with a lentiviral shRNA targeting ERF (shErf), and confirmed knockdown among pools of cells (**Figure 14a**). We did not select for single clones, hence the knockdown was incomplete. We injected these organoids back into mouse flanks and assessed for tumor formation. As predicted from our model, CRISPR-mediated knockdown of Erf accelerated ERG-high tumor formation (**Figure 14b**). We also took the same cells infected instead with a doxycycline-inducible FLAG-ERF (described in **Figure 2a**) and injected these into mice, which were subjected to either normal feed, or doxycycline- containing feed. As predicted, the doxycycline-fed mice had a complete block of tumor formation, consistent with our in vivo competition model between ERG and ERF (**Figure 14c**). The above experiments confirm that in vivo ERF and ERG function in opposing directions, but does not confirm that they compete against each other in the same pathway (i.e. bind to the same chromatin sites) as would be predicted from the ChIP-seq data in **Figure 5**. To this end, we exploited the property of VCaP cells, which like the organoids above, are completely dependent on ERG for proliferation; knockout of ERG will halt proliferation. We, however, first inhibited the expression of ERF via shRNA, prior to CRISPR-mediated knockout of ERG (**Figure 14d**), and then assessed the cells for proliferation. As before, a simple CRISPR of ERG halts proliferation of the cells. However, when ERF expression is knocked down first, the cells remarkably lose their dependence on ERG, and continue to proliferate (**Figure 14e**). In other words, when the tumor suppressor ERF is inhibited first, the cells appear to no longer require the ERG oncogene for proliferation. This implies that a major oncogenic activity of ERG is to functionally inactivate ERF and these findings are likely applicable to the 30-50% of





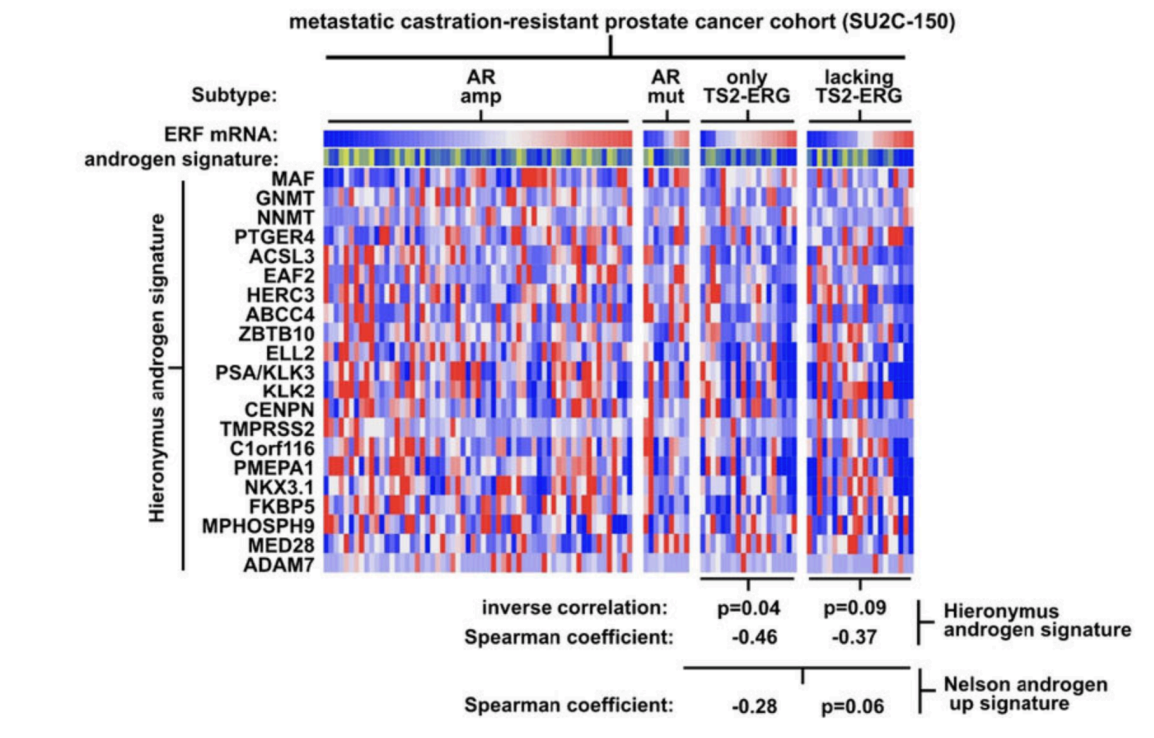


Figure 13 - a) Graft tumour volumes of GEMM-derived prostate organoids. Data are median ± interquartile range; n = 10 tumours per condition. **b)** Expression profiles of the TCGA-333 primary prostate cancer cohort (n = 333 patients) were interrogated for correlation between the ERF mRNA level and two androgen transcriptional activity signatures. P values were calculated by the Spearman correlation test. **c)** The same analysis as **b** was applied to the SU2C-150 (*n* = 150 patients) metastatic castration-resistant prostate cancer cohort[3] (mCRPC).

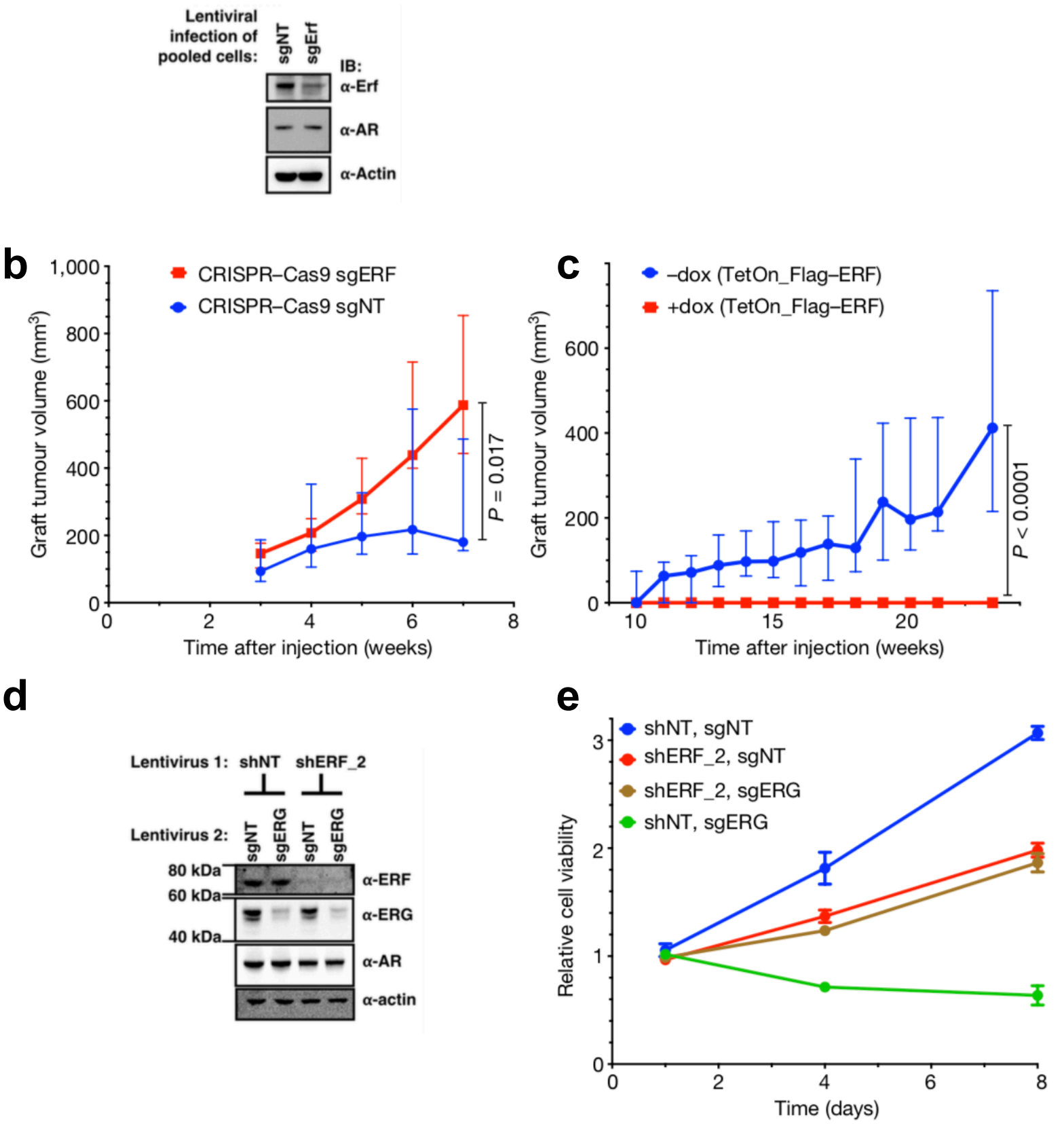


Figure 14 - a) Pooled mouse Pten$^{-/-}$; R26$^{ERG/ERG}$ organoids infected with CRISPR–Cas9 targeting AAVS1 (sgNT) or ERF (sgErf). **b)** Tumor volumes of grafts derived from Pten$^{-/-}$;R26$^{ERG/ERG}$ organoids infected with CRISPR–Cas9 targeting ERF (sgERF). Data are median ± interquartile range; n = 10 tumors per condition, P value via Mann–Whitney exact two-tailed test. **c)** Similar to **b** but infected instead with doxycycline-inducible Flag–ERF and mouse fed with control or doxycycline-containing feed. **d)** Pooled VCaP cells were first infected with non-targeting shRNA (shNT) or ERF (shERF_2), followed by sgNT or sgERF CRISPR–Cas9. **e)** Cell viability assay in pooled VCaP cells infected first with shRNA targeting ERF (shERF_2) followed by sgERF. Data are mean ± s.e.m.; n = 3 biological replicates.

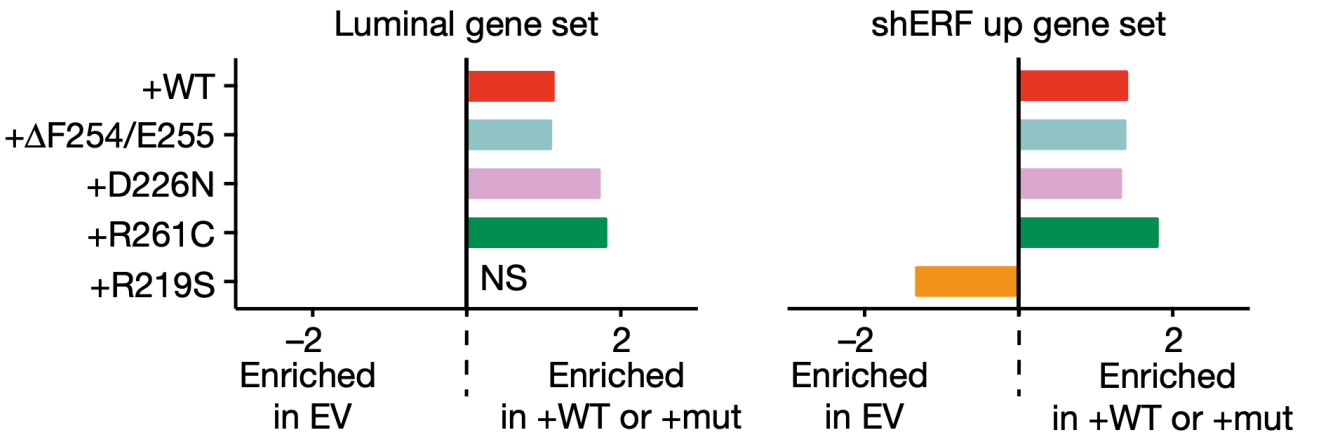


Figure 15 - Summary of GSEA comparing Foxa1 wild-type or mutant organoid lines to empty vector control for a basal low (luminal) gene set, a gene set of the top 100 genes induced following ERF knockdown (using short hairpin RNA directed against ERF, shERF) in organoids. Data are from RNA-sequencing (RNA-seq) analysis of three biological replicates for each organoid line. Only comparisons with an FDR of <0.25 are shown with the corresponding normalized enrichment score. Gene sets with a positive normalized enrichment score are enriched in organoids that have either Foxa1 wild-type or mutant alleles.

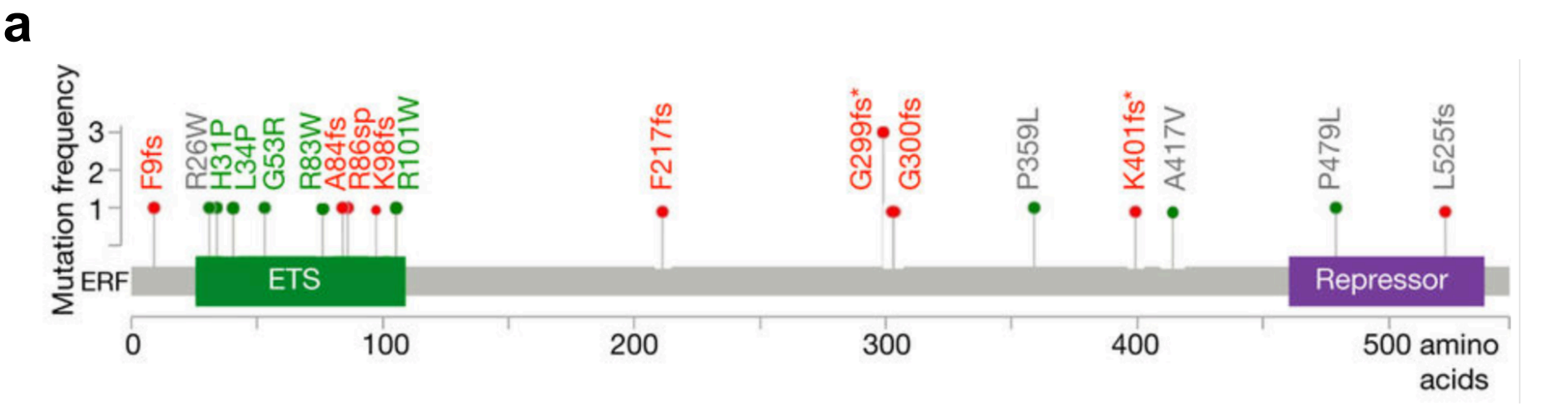
prostate cancers that contain high levels of ERG. This is a major finding in the field of human prostate oncology.

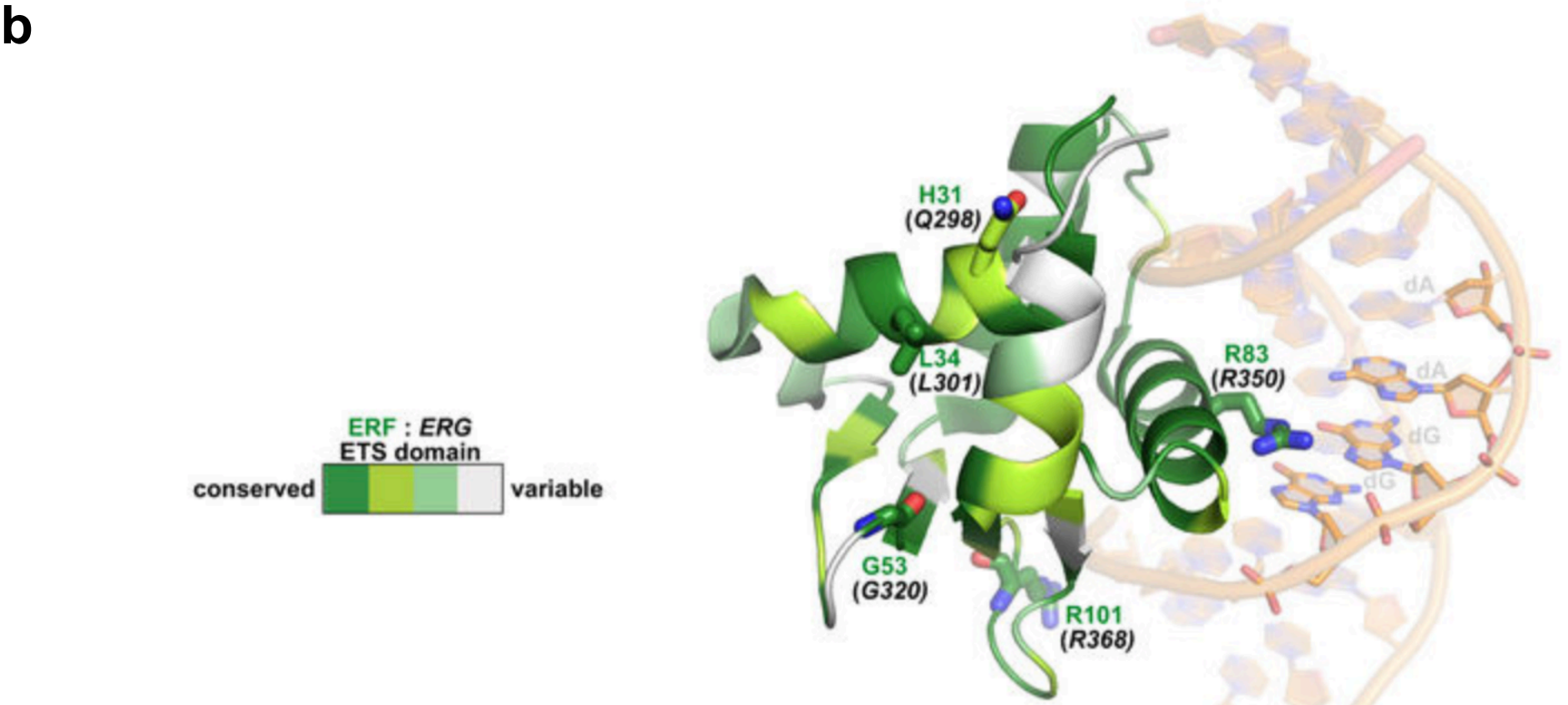
There is a class of mutations in another transcription factor, FOXA1 identified in prostate cancer, which tend to be mutual exclusive from ERF and ERG mutations. By modeling expression of wildtype FOXA1 in normal prostate organoids (which have no other mutations), as well as FOXA1 tumor point mutations (such as D226N, R261C etc.), they were found to activate a pro-luminal epithelial gene expression program, very similar to that observed with ERF knockdown (**Figure 15**). Moreover, when we directly compared the gene expression program of most FOXA1 tumor mutations, as well as overexpression of wildtype FOXA1, they activated a very similar gene expression to ERF (**Figure 15**). This is a critical finding re: ERF for several reasons. Firstly, FOXA1 and ERFs DNA binding motifs are completely distinct, so it is compelling that two different classes of transcription factors cause a similar gene program. Secondly, since FOXA1 and ERF mutations are mutually exclusive, it suggests that different oncogenic transcription programs converge on a functionally similar pathway with similar consequences. By analyzing the downstream genes that are common to either ERF-driven or FOXA1-driven tumors, we can begin to understand how ERF functions as tumor suppressor. Please see further details in the attached *Nature* paper in the Appendix.

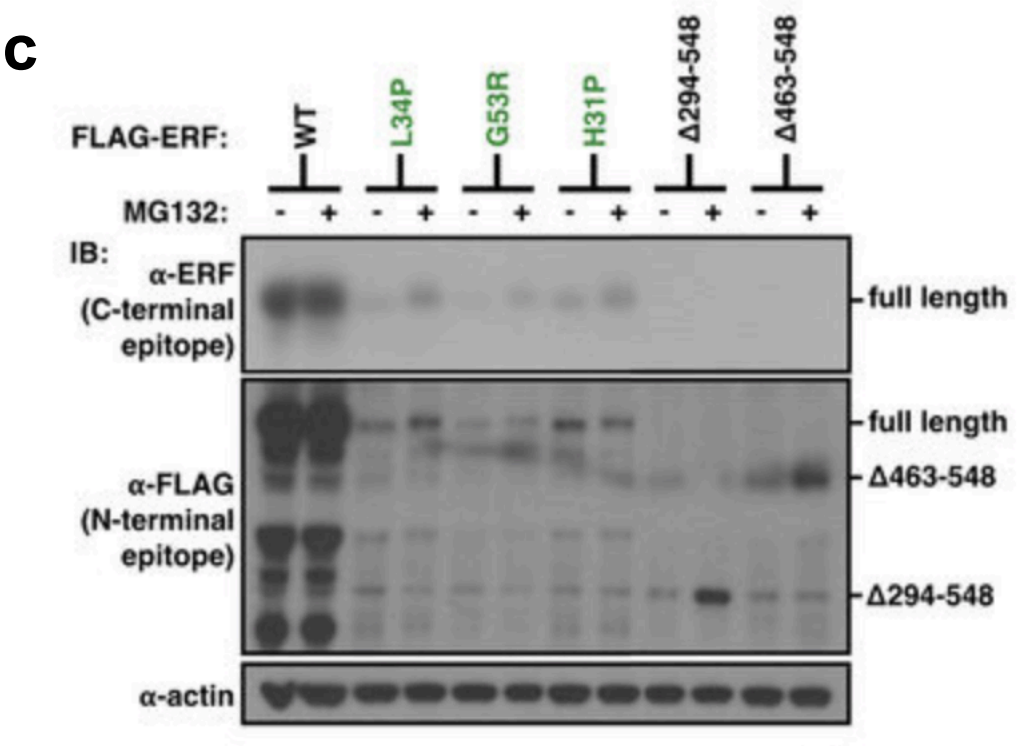
B. Major Task 2: Evaluation of role of point mutants in ERF tumor

We queried all published prostate cancer genome cohorts (n = 930 patients) and complied a list of all ERF mutations (**Figure 16a**), present in 1–3% of patients. The mutations include the specific K401fs and G299fs loss-of-function truncations (initially defined in craniosynostosis families), as well as similar missense mutations in the ETS domain. Mapping of these ETS missense mutations onto the known crystal structure of ERG revealed that the altered residues are located within conserved helices (**Figure 16b**), with four of the five conserved mutations predicted to be destabilizing by in silico analysis. Expression of complementary DNAs (cDNAs) containing the ETS missense or truncating mutations in LNCaP prostate cancer cells led to reduced ERF steady- state levels relative to wild-type ERF despite robust messenger RNA (mRNA) expression (**Figure 16c, d**), in agreement with prior craniosynostosis studies. Collectively, this pointed to ERF point mutations in prostate cancer as being destabilizing.

Although we suspected poor protein folding in the missense mutants due to the high conservation of the ETS domain residues that were consistently mutated, other explanations, such as increased degradation by mammalian ubiquitin ligases, were possible. To this end, we examined two other systems. Firstly, there is a patient-derived organoid line called MSK-PCa3 which contains a truncating point mutation in one allele of ERF at K401. When blotting total cell lysates, or ERF immunoprecipitation via an antibody with an amino- terminal epitope, we were able to identify the full length ERF, but no band corresponding to where we would expect to see a truncated ERF 1-401 mutant (**Figure 17a**). Secondly, we expressed the ETS domain of ERF in bacteria, and were able to purify this easily. However, when we introduced point mutations found in prostate







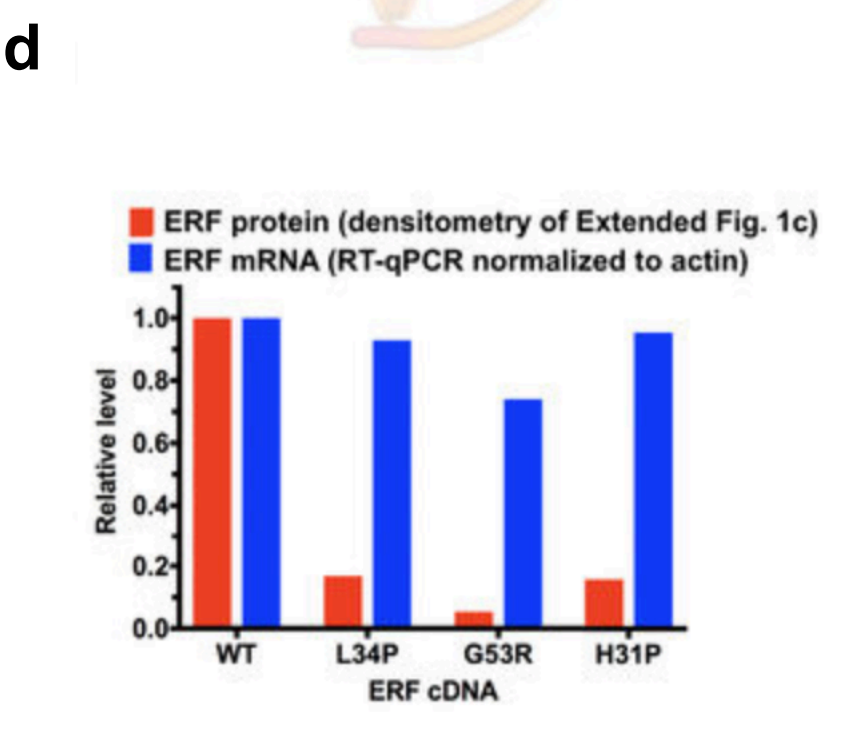


Figure 16 - a) ERF point mutations identified in all prostate cancer cohorts. *Mutations shared with patients having craniosynostosis. Green: ETS missense mutations of residues conserved with ERG. Red: frameshift (fs) / splice-site (sp) mutations upstream of repressor domain. **b)** ERG ETS domain (Protein Data Bank accession number 4IRI) illustrating sequence conservation with ERF. ERF missense mutations indicated by sticks. **c)** Expression of ERF mutant proteins in LNCaP cells. **d)** ERF mutant protein densitometry from immunoblot in **c** compared with mRNA RT–qPCR.

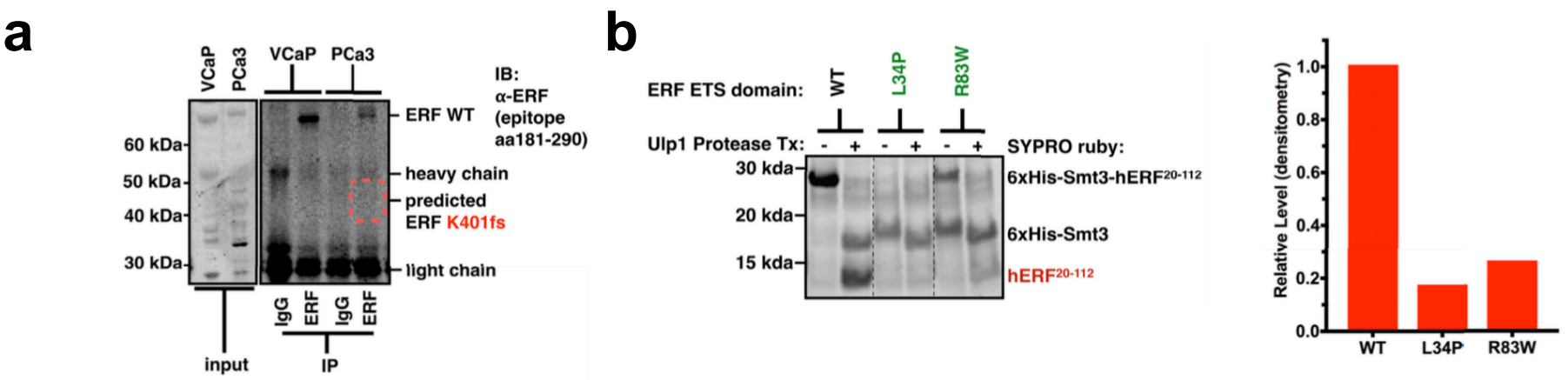


Figure 17 - a) ERF immunoprecipitation in VCaP and MSK-PCa3 cells. **b)** Left, bacterially expressed ETS domain (hERF20–112) with ERF tumour mutations before (−) or after (+) Ulp1 protease cleavage. Right, densitometry of in vitro ETS domains.

Cohorts	Source of specimen	Clinical State	Cohort size	# of patients with ERF point mutations	ERF Point Mutation Rate	Reference	Notes
TCGA-333	Primary site (prostate)	Localized, treatment naive	333	1	0.3%	Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. Cell 163, 1011–1025 (2015).	
TCGA-498	Primary site	Localized, treatment naive	498	4	0.8%	Broad Institute FireBrowse portal: (http://firebrowse.org/?cohort=PRAD)	Includes TCGA-333
CRUK	Primary site	Mixed	108	1	0.9%	ICGC Portal: https://icgc.org/icgc/cgp/70/508/71331	
SU2C-150	Metastatic site	Castration-resistant, metastatic prostate cancer	150	4	2.7%	Robinson, D. et al. Integrative clinical genomics of advanced prostate cancer. Cell 161, 1215– 1228 (2015).	
SU2C-294	Metastatic site	Castration-resistant, metastatic prostate cancer	294	8	2.7%	Pending publication from International SU2C/PCF Consortium.	Includes SU2C-150
FHCRC	Primary or metastatic site	Castration-resistant, metastatic prostate cancer	54	2	3.7%	Kumar, A. et al. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. Nature Medicine 1–13 (2016).	
Trento/Cornell/ Broad 2016	Metastatic site	Castration-resistant, metastatic prostate cancer	77	1	1.3%	Beltran, H. et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. Nature Medicine 22, 298–305 (2016).	
MSK Organoids	Local recurrence or metastatic site or circulating tumor cell	Castration-resistant, metastatic prostate cancer	7	1	14.3%	Gao, D. et al. Organoid Cultures Derived from Patients with Advanced Prostate Cancer. Cell (2014).	

Table 1 - Summary of published prostate cohorts with the ERF point mutation rate calculated, demonstrating a higher rate of ERF mutations in advanced disease.

cancer patients, we were no longer able to purify those ETS domains any longer (**Figure 17b**). Collectively, this pointed to ERF point mutations in prostate cancer as being destabilizing.

Above, we described the relevance of ERF missense mutations in two different in vitro systems: MSK-PCa3 organoids, and expression of the ETS domain of ERF in bacteria. Now, we shift our analyses to in vivo human data, to better understand the relevance of ERF mutations in prostate cancer. As such, we have collated a list of all ERF point mutations in published cohorts (**Table 1**). This table brings to light a few points. Firstly, the point mutation rate of ERF in most primary cohorts is <1% of patients, with the caveat that most published cohorts are biased towards men of European-American origin. On the other hand, the mutation rate ranges between 3-14% of metastatic castration-resistant prostate cancer (mCRPC). This is consistent with our earlier findings that in early disease, other mutations are sufficient to initiate tumorigenesis, but in later disease, ERF loss can cooperate with other mutations in order to drive AR transcription.

B. Major Task 3: Evaluation of ability of MEK/ERK inhibitor to augment tumor suppressor activity of ERF

Nothing to report.

Specific Aim 3: To interrogate and map how ERF may regulate androgen-mediate transcription via its novel interaction with AR

A. Major Task 1: Determination of role of ERF in AR signaling

The inverse correlation between ERF mRNA level and the androgen receptor signature is also observed in the half of prostate cancers that possess the TMPRSS2–ERG fusion, suggesting that ERF may have an androgen receptor repressive function in this subtype as well. To investigate this possibility, the expression of ERF or ERG was separately inhibited via shRNA in the ERG-fusion-positive VCaP cell line and the androgen transcriptome analyzed, looking for genes with a two-fold change in expression by RNA-seq upon 1nM dihydrotestosterone (DHT) addition, FDR<0.05. Consistent with prior literature, inhibition of ERG expression ('ERG-low') resulted in a contracted androgen transcriptome compared with the wild-type state ('ERG-high'). Conversely, ERF inhibition ('shERF_2') increased the change in expression of androgen receptor target genes and doubled the size of the androgen transcriptome. DHT treatment had no effect on ERF mRNA levels (**Figure 18b - Mid panel**) and doxycycline alone had no effect on gene expression (**Figure 18b - Right panel**). ERF knockdown had no effect on ERG protein level, on androgen receptor protein levels, or on androgen receptor subcellular localization (**Figure 18a**). ERF knockdown in VCaP cells led to enrichment of androgen-signaling genes, both noted by GSEA interrogation of the above RNA-seq experiment (**Figure 18c**) or by validation RT-qPCR experiments (**Figure 18d**). Collectively, ERF serves as an endogenous repressor of AR signaling, both in normal prostate and in prostate cancer cells.

The above, however, does not explain how ERF functions as a tumor suppressor in prostate cancers lacking ERG expression, and whether such activity is related to AR signaling. To this end, we interrogated CWR22Pc human prostate cancer cells, which lack high expression of any known oncogenic ETS factor (e.g. ERG, ETV1, ETV4) and lack any known ETS fusion. We infected these cells with a lentiviral shRNA targeting ERF (shERF) and confirmed knockdown via Western blot (**Figure 19a**). Next, we treated these cells with DHT and assessed the AR transcriptome. Consistent with our broad model, ERF knockdown increased the number of AR-regulated genes (**Figure 19b - left panel**), as well as the magnitude of those expression changes (**Figure 19b - center and right panels**). As before, upregulated genes were more upregulated and downregulated genes were more downregulated. Finally, we confirmed that such changes were not due simply to ERF regulating AR expression; indeed, we demonstrated that AR mRNA levels were constant despite shERF manipulation (**Figure 19c**).

Next, we leveraged the genome-wide screen mentioned in Aim 1, Task 3, to explore the interplay between the dose response of different concentrations of the androgen dihydrotestoterone (DHT) in concert with different levels of ERG and ERF. We have graphed the abundance of 3 sample shRNAs targeting ERF over the duration of the screen. If the shRNA relative abundance decreases over time, this implies that the targeted gene is required for proliferation (a potential oncogene or housekeeping gene). If the shRNA relative abundance increases over time, this implies that the targeted gene inhibits proliferation (a potential tumor suppressor). As you can

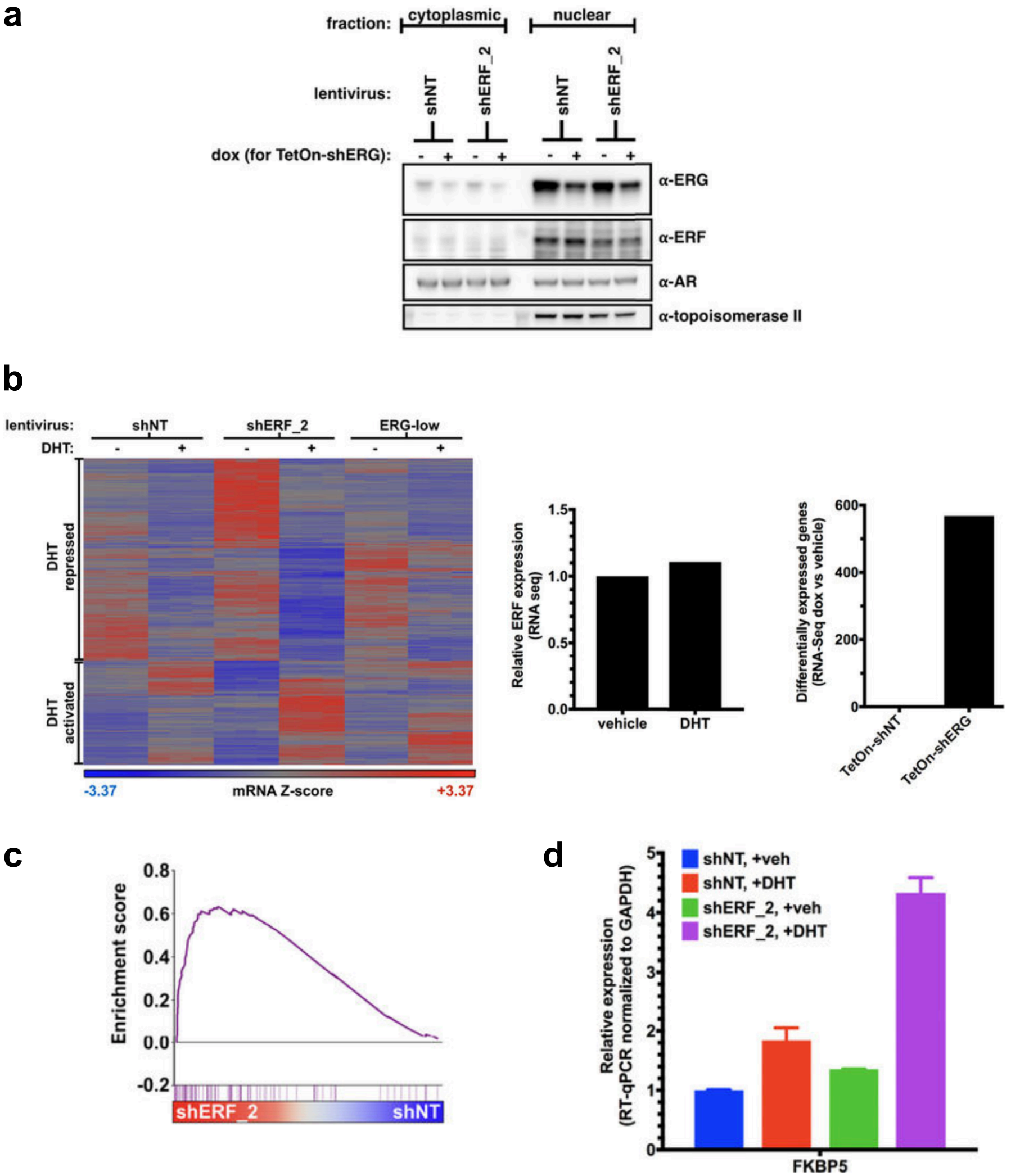


Figure 18 - a) Nuclear/cytoplasmic fractionation of VCaP cells infected with a non-targeting shRNA (shNT) or shRNA targeting ERF (shERF_2). **b)** Left - Androgen-regulated genes (at least a two-fold change, FDR < 0.05 by RNA-seq with 1 nM DHT for 16h) in VCaP cells infected with a doxycycline-inducible shRNA targeting ERG (ERG-low) or targeting ERF (shERF_2) and analysed by heat map; n = 3 biological replicates. Center - Analysis of same RNA-seq data evaluating the effect of dihydrotestosterone (DHT) on ERF expression. Right - Analysis of RNA-seq data to evaluate the effect of doxycycline addition alone. **c)** Interrogation of RNA-seq in shERF_2 VCaP cells (n = 3 biological replicates) by GSEA for Nelson androgen up expression signature. **d)** RT–qPCR of shERF_2-infected VCaP cells treated ± DHT. Data are mean ± s.e.m.; n = 3 biological replicates.

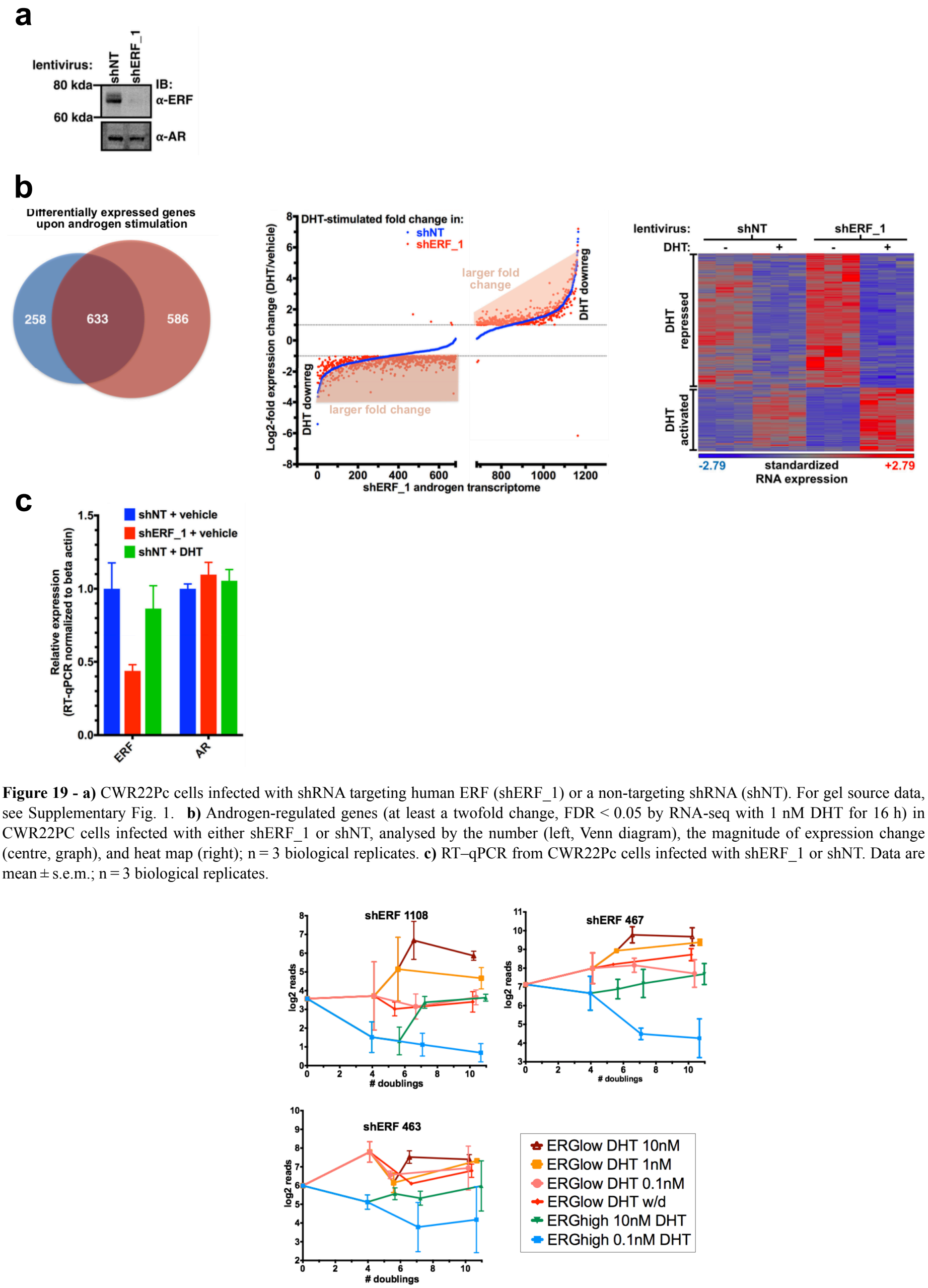


Figure 19 - a) CWR22Pc cells infected with shRNA targeting human ERF (shERF_1) or a non-targeting shRNA (shNT). For gel source data, see Supplementary Fig. 1. **b)** Androgen-regulated genes (at least a twofold change, FDR < 0.05 by RNA-seq with 1 nM DHT for 16 h) in CWR22PC cells infected with either shERF_1 or shNT, analysed by the number (left, Venn diagram), the magnitude of expression change (centre, graph), and heat map (right); n = 3 biological replicates. **c)** RT–qPCR from CWR22Pc cells infected with shERF_1 or shNT. Data are mean ± s.e.m.; n = 3 biological replicates.

Figure 20 - Tracking of the relative abundance of shRNAs targeting ERF in the presence of different levels of ERG and differing concentrations of dihydrotestosterone (DHT). Over time, in the presence of low ERG and high DHT concentrations, all ERF-targeted shRNAs are enriched over time. This is supports the possibility that the tumor suppressor function of ERF in this system is DHT- and ERG-dependent.

see in **Figure 20**, all ERF shRNAs (including the 3 not pictured) show that the degree to which ERF serves as a tumor suppressor increases with increasing DHT concentration. In other words, this suggests that the mechanism by which ERF serves as a tumor suppressor is AR-related. This dose dependence is a completely novel finding and may have implications for patients under AR-based therapy.

B. Major Task 2: Confirmation of AR-ERF interaction

Nothing to report

C. Major Task 3: Significance of AR-ERF interaction

Nothing to report

Training Opportunities and Professional Development

Since moving to UCSF and starting my independent research program here, I attend oncology grand rounds weekly. I also attend our weekly Genitourinary Medical Oncology Program meetings, our monthly Benioff Initiative for Prostate Cancer fundamental science meetings, and our biannual Prostate Cancer Program retreats. At these meetings, I interact with several other CDMRP Prostate Cancer funded investigators including Dr. Eric Small and Dr. Felix Feng. I continue to communicate and collaborate with my mentor Dr. Charles Sawyers. Regarding conferences, I attended the Prostate Cancer Foundation meeting, ASCO, and GU-ASCO, all of which have major prostate cancer research components.

Dissemination of Results

The data presented was published (please see 'Products' section) in two high impact papers.

IMPACT

On principal discipline

The findings over the reporting period have furthered our understanding of tumor genetics responsible for prostate cancer progression. Specifically, this project has investigated the poorly understood nature of the ERG fusion-negative tumors and further clarified the nature of the ERG fusion-positive tumors as well. The identification of an ETS factor, ERF, that serves as a tumor suppressor in fusion-negative tumors, but is outcompeted in ERG fusion-positive prostate cancer, is a significant contribution to our understanding of the tumor genetics of prostate cancer.

The results from this proposed research project could also have a major impact on prostate cancer care. Existing treatments for advanced prostate cancer increase survival only on the order of months, and can be associated with significant morbidities. Half of all such cancers lack ERG gene fusions, and at present, we lack a unifying understanding of the key regulators in such tumors. If the project hypothesis is correct, then we may be able to identify a novel treatment strategy for patients possessing fusion-negative tumors. This approach could lead to increased survival and decreased morbidity for men suffering from this disease. In addition to the clinical applications above, this project could shed light on how tumors with and without ERG mutations may actually be similar. This, in turn, would contribute to our understanding of the 50% who do have the ERG mutation, and assist in developing a targeted strategy for them as well.

On other disciplines

Nothing to report

On technology transfer

Nothing to report

Beyond science and technology

Nothing to report

CHANGES/PROBLEMS

Nothing to report

PRODUCTS

Publications, conference papers, and presentations

Peer Reviewed Publications:
Bose R, Karthaus, WR, Armenia J, Abida W, Iaquinta PJ, Zhang Z, Wongvipat J, Wasmuth EV, Shah N, Sullivan PS, Doran MG, Wang P, Patruno A, Zhao Y, International SU2C/PCF Prostate Cancer Dream Team, Zheng D, Schultz N, Sawyers CL. ERF mutations reveal a balance of ETS factors controlling prostate oncogenesis. Nature. 2017 June 29;546(7660):671-5. Published. Federal support was acknowledged.

Adams EJ, Karthaus WR, Hoover E, Liu D, Gruet A, Zhang Z, Cho H, DiLoreto R, Chhangawala S, Liu Y, Watson PA, Davicioni E, Sboner A, Barbieri CE, Bose R, Leslie CS, Sawyers CL. FOXA1 mutations alter pioneering activity, differentiation and prostate cancer phenotypes. Nature. 2019 Jul;571(7765):408-412. PubMed PMID: 31243370; NIHMSID: NIHMS1530156; PubMed Central PMCID: PMC6661172. Federal support was acknowledged.

Conference Presentations of this work:

Bose R, Abida W, Karthaus W, Iaquinta P, Armenia J, Wongvipat J, Doran M, Hieronymus H, Watson P, Liang Y, Schultz N, Sawyers CL. Genomic Mutations of an ETS Tumor Suppressor Expand the ETS-related Subset of Prostate Cancer. Multi-institutional Prostate Cancer Program Retreat – Fort Lauderdale (2016)

Bose R, Karthaus W, Armenia J, Abida W, Iaquinta P, Wongvipat J, Wasmuth E, Shah N, Doran M, Wang P, Sullivan P, Patruno A, International SU2C/PCF Consortium, Zheng D, Schultz N, Sawyers CS. Loss of function mutations in ETS2 repressor factor, ERF, reveal a balance between positive and negative ETS factors controlling prostate oncogenesis. Prostate Cancer Foundation Annual Scientific Retreat – Carlsbad (2016)

Bose R, Abida W, Iaquinta P, Karthaus W, Armenia J, Wongvipat J, Doran M, Hieronymus H, Watson P, Liang Y, Schultz N, Sawyers CL. Investigation of a frequently mutated transcriptional repressor in prostate cancer, in particular its role in modulating androgen signaling and its effects on TMPRSS2-ERG dependent tumor maintenance. GU ASCO – San Franscisco (2016)

Bose R, Abida W, Karthaus W, Iaquinta P, Armenia J, Wongvipat J, Doran M, Hieronymus H, Watson P, Liang Y, Schultz N, Sawyers CL. Genomic Mutations of an ETS Tumor Suppressor Expand the ETS-related Subset of Prostate Cancer. Multi-institutional Prostate Cancer Program Retreat – Fort Lauderdale (2016)

Bose R, Abida W, Karthaus W, Armenia J, Iaquinta P, Wongvipat J, Doran M, Hieronymus H, Watson P, Sullivan P, Liang Y, Schultz N, Sawyers CS. Loss of Function Mutations of an ETS

Repressor Expand the ETS Positive Subset of Prostate Cancer. Nature Conference: Cancer as an Evolving and Systemic Disease – New York (2016)

Bose R, Abida W, Karthaus W, Iaquinta P, Armenia J, Wongvipat J, Doran M, Hieronymus H, Watson P, Sullivan P, Liang Y, Schultz N, Sawyers CL. Genomic Mutations of a Novel ETS Tumor Suppressor Expand the ETS-related Subset of Prostate Cancer. Geoffrey Beene Cancer Research Retreat – Crystal Springs, NJ (2016)

Bose R, Abida W, Karthaus W, Armenia J, Iaquinta P, Wongvipat J, Doran M, Hieronymus H, Watson P, Sullivan P, Liang Y, Schultz N, Sawyers CS. Loss of function mutations of an ETS Repressor Expand the ETS Positive Subset of Prostate Cancer. AACR – New Orleans (2016)

Bose R, Abida W, Karthaus W, Armenia J, Iaquinta P, Wongvipat J, Sullivan P, Doran M, Hieronymus H, Watson P, Liang Y, Schultz N, Sawyers CS. Genomic mutations of an ETS tumor suppressor expand the ETS-related subset of prostate cancer. Targeting Cancer – CSHL (2016)

Bose R, Karthaus W, Armenia J, Abida W, Iaquinta P, Wongvipat J, Wasmuth E, Shah N, Doran M, Wang P, Sullivan P, Patruno A, International SU2C/PCF Consortium, Zheng D, Schultz N, Sawyers CS. Loss of function mutations in ETS2 repressor factor, ERF, reveal a balance between positive and negative ETS factors controlling prostate oncogenesis. Dept of Defense PCRP IMPACT – Baltimore (2016)

Bose R, Karthaus W, Armenia J, Abida W, Iaquinta P, Wongvipat J, Wasmuth E, Shah N, Doran M, Wang P, Sullivan P, Patruno A, International SU2C/PCF Consortium, Zheng D, Schultz N, Sawyers CS. Loss of function mutations in ETS2 repressor factor, ERF, reveal a balance between positive and negative ETS factors controlling prostate oncogenesis. Prostate Cancer Foundation Annual Scientific Retreat – Carlsbad (2016)

Bose R, Abida W, Karthaus W, Iaquinta P, Armenia J, Wongvipat J, Doran M, Hieronymus H, Watson P, Liang Y, Schultz N, Sawyers CL. Genomic Mutations of an ETS Tumor Suppressor Expand the ETS-related Subset of Prostate Cancer. Multi-institutional Prostate Cancer Program Retreat – Fort Lauderdale (2016)

Bose R, Sullivan P, Liang Y, Sawyers CL. Functional genomic interrogation of androgen signaling co-dependencies in advanced mCRPC – Fort Lauderdale (2017)

Ryan M, Jin P, Nadkarni T, Bose R. Functional genomic interrogation of androgen signaling co-dependencies in advanced mCRPC. Los Angeles (2020).

Website(s) or other Internet site(s)

Nothing to report

Technologies or techniques

Nothing to report

Inventions, patent applications, and/or licenses

Nothing to report

Other Products

Nothing to report

PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

Individuals

Name:	Rohit Bose
Project Role:	PI
ORCID ID:	0000-0002-6785-0697
Nearest person month worked:	8.0 months (average over 5 years)
Contribution to Project:	PI, Designed, performed and analyzed experiments.

Name:	Peng Jin
Project Role:	Postdoctoral fellow
Nearest person month worked:	8.9 months (average over 2 years)
Contribution to Project:	Performed experiments.

Name:	Tejasveeta Nadkarni
Project Role:	Technician / Lab Specialist
Nearest person month worked:	7.3 months (average over 2 years)
Contribution to Project:	Performed experiments.

Name:	Hui Yu
Project Role:	Technician / Lab Specialist
Nearest person month worked:	0.8 months (average over 2 years)
Contribution to Project:	Performed experiments.

Change in Other Active Support

None

Other Organizations

Nothing to report

SPECIAL REPORTING REQUIREMENTS

None

LETTER

doi:10.1038/nature22820

ERF mutations reveal a balance of ETS factors controlling prostate oncogenesis

Rohit Bose[1,2], Wouter R. Karthaus[1], Joshua Armenia[1,3], Wassim Abida[2], Phillip J. Iaquinta[1], Zeda Zhang[1,4], John Wongvipat[1], Elizabeth V. Wasmuth[1], Neel Shah[1,4], Patrick S. Sullivan[1], Michael G. Doran[1], Ping Wang[5], Anna Patruno[2], Yilin Zhao[5], International SU2C/PCF Prostate Cancer Dream Team*, Deyou Zheng[5,6,7], Nikolaus Schultz[1,3],& Charles L. Sawyers[1,8]

Half of all prostate cancers are caused by the *TMPRSS2–ERG* gene-fusion, which enables androgens to drive expression of the normally silent E26 transformation-specific (ETS) transcription factor *ERG* in prostate cells[1,2]. Recent genomic landscape studies of such cancers[3–8] have reported recurrent point mutations and focal deletions of another ETS member, the ETS2 repressor factor *ERF*[9]. Here we show these *ERF* mutations cause decreased protein stability and mostly occur in tumours without *ERG* upregulation. *ERF* loss recapitulates the morphological and phenotypic features of *ERG* gain in normal mouse prostate cells, including expansion of the androgen receptor transcriptional repertoire, and *ERF* has tumour suppressor activity in the same genetic background of *Pten* loss that yields oncogenic activity by *ERG*. In the more common scenario of *ERG* upregulation, chromatin immunoprecipitation followed by sequencing indicates that ERG inhibits the ability of ERF to bind DNA at consensus ETS sites both in normal and in cancerous prostate cells. Consistent with a competition model, ERF overexpression blocks ERG-dependent tumour growth, and ERF loss rescues *TMPRSS2–ERG*-positive prostate cancer cells from ERG dependency. Collectively, these data provide evidence that the oncogenicity of ERG is mediated, in part, by competition with ERF and they raise the larger question of whether other gain-of-function oncogenic transcription factors might also inactivate endogenous tumour suppressors.

Recent exome sequencing revealed that 3% of patients in the SU2C-294 metastatic prostate cancer cohort[3] (Extended Data Fig. 1a) have somatic point mutations, but not gene fusions, involving the ETS member *ERF*. *ERF* was one of a small number of genes whose mutation frequency and predicted functional impact reached significance by the MutSig algorithm[10]. Loss-of-function *ERF* germline mutations and lower *ERF* expression have been previously implicated in the disease complex craniosynostosis[11]. Notably, the DNA-binding ETS domain of ERF is most similar to the ERG subfamily[12]. However, unlike oncogenic ETS factors[12], ERF possesses a transferable carboxy-terminal domain that mediates transcriptional repression[9].

We queried additional prostate cancer genome cohorts ($n = 930$ patients)[4–8] (Extended Data Fig. 1a) and found further evidence of *ERF* mutations (Fig. 1a) in 1–3% of patients. The mutations include the specific *K401fs* and *G299fs* loss-of-function truncations also found in craniosynostosis families[11], as well as similar missense mutations in the ETS domain. Mapping of these ETS missense mutations onto the known crystal structure of ERG[13] revealed that the altered residues are located within conserved helices (Extended Data Fig. 1b), with four of the five conserved mutations predicted to be destabilizing[14]. Expression of complementary DNAs (cDNAs) containing the ETS missense or truncating mutations in LNCaP prostate cancer cells led to reduced ERF steady-state levels relative to wild-type ERF despite robust messenger RNA (mRNA) expression (Extended Data Fig. 1c, d), in agreement with the craniosynostosis studies[11]. In addition, only full-length ERF was detected following immunoprecipitation in prostate cancer cells possessing an endogenous heterozygous *K401fs* allele[5] (Extended Data Fig. 1e). Consistent with destabilization, we were unable to isolate mutant ERF ETS domains through recombinant expression in bacteria, whereas appreciable amounts of wild-type ERF ETS were obtained (Extended Data Fig. 1f).

In addition to destabilizing mutations, we observed remarkably narrow deletions of the *ERF* locus within the TCGA-333 primary tumour cohort[4] (Fig. 1b). Median *ERF* expression in those tumours containing focal hemizygous deletions is lower than in normal prostate and diploid *ERF* tumours (Extended Data Fig. 2a, $P = 0.019$). Intriguingly, tumours with either *ERF* mutations or focal deletions are mostly exclusive to TMPRSS2–ERG-negative tumours (Fig. 1c, $P = 0.022$). Metastatic tumours in the SU2C-294 cohort[3] containing *ERF* point mutations are also mostly exclusive to those without upregulated ERG, but this distinction does not reach statistical significance (Extended Data Fig. 2b, $P = 0.066$).

The near complete lack of *ERF* mutations in tumours with *TMPRSS2–ERG* fusions led us to investigate whether ERF loss recapitulates the phenotype of ERG gain. *ERG* is not expressed in benign prostate epithelium and displays potent oncogenicity in a mouse $Pten^{-/-}$ background[2,15]. On the other hand, *ERF* is endogenously expressed in normal prostate (Extended Data Fig. 2a). To determine whether its loss yields a phenotype similar to ERG gain, we infected prostate organoids[16] derived from $Pten^{+/+}$ and $Pten^{-/-}$ mice[15] with short hairpin RNA (shRNA) targeting mouse *Erf* (shErf_m) (Extended Data Fig. 3a). The $Pten^{+/+}$ shErf_m organoids acquired morphological characteristics of ERG overexpression[17]: they formed single-cell luminal structures lacking basal cells (Fig. 2a), and profiling by RNA sequencing (RNA-seq) demonstrated profound loss of expression of most basal signature genes[18] (Fig. 2b and Extended Data Fig. 2b–d). Consistent with these changes[17,18], RNA-seq profiling revealed significant enrichment for genes whose expression is upregulated by androgen in human prostate cancer cells (Extended Data Fig. 2c, e). The $Pten^{-/-}$ organoids[16] infected with shErf_m also demonstrated a marked luminal shift (Fig. 2a) and were able to form tumours when grafted back into mice, recapitulating the phenotype of ERG overexpression[15] (Fig. 2c).

To assess the impact of ERF loss in the half of human prostate cancers that lack ERG expression similarly to normal prostate, we infected a human cancer-derived cell line lacking the *TMPRSS2–ERG* gene fusion, CWR22Pc, with shRNA targeting *ERF* (Extended Data Fig. 4a) and analysed its androgen transcriptome. We observed both an increase in

[1]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. [2]Department of Medicine, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. [3]Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. [4]Louis V. Gerstner, Jr. Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [5]Department of Genetics, Albert Einstein College of Medicine, Bronx, New York 10461, USA. [6]Department of Neurology, Albert Einstein College of Medicine, Bronx, New York 10461, USA. [7]Department of Neuroscience, Albert Einstein College of Medicine, Bronx, New York 10461, USA. [8]Howard Hughes Medical Institute, Chevy Chase, Maryland 20185, USA.
*Lists of participants and their affiliations appear at the end of the paper.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

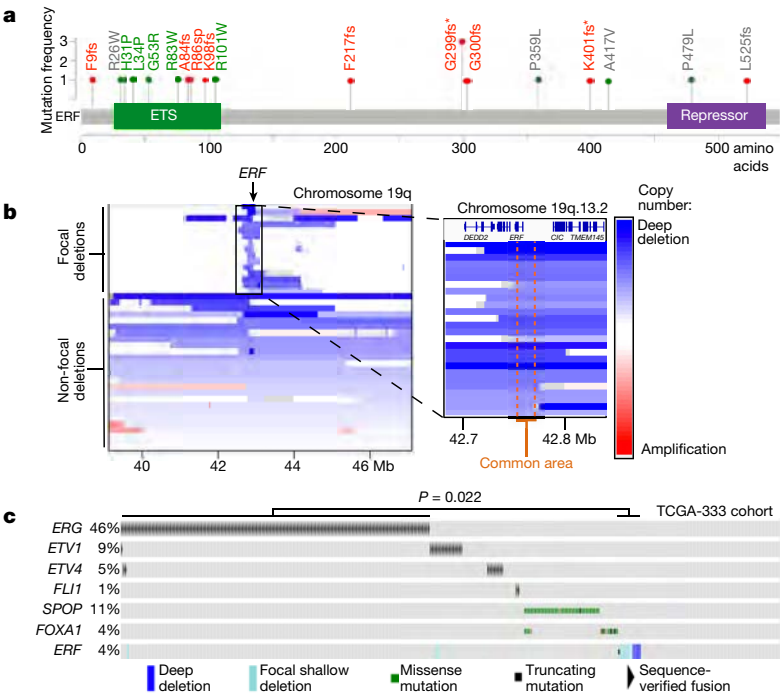


Figure 1 | Recurrent ERF loss-of-function mutations and focal deletions are found in prostate cancer and are mostly exclusive to tumours without TMPRSS2–ERG. a, ERF point mutations identified in all prostate cancer cohorts[3–8]. *Mutations shared with patients having craniosynostosis. Green: ETS missense mutations of residues conserved with ERG. Red: frameshift (fs)/splice-site (sp) mutations upstream of repressor domain. **b**, ERF copy number deletions in TCGA-333 human primary prostate cancer cohort[4]. **c**, cBio Oncoprint of patients with primary prostate cancer ($n = 333$ patients)[4]. Each column corresponds to a unique patient's tumour profile. *P* value calculated by Fisher's exact two-tailed test.

the number of differentially expressed androgen receptor target genes and in the magnitude of the expression changes (Extended Data Fig. 4b), despite no alteration of androgen receptor mRNA or protein levels (Extended Data Fig. 4a, c). Next, we interrogated the mRNA expression profiles of the primary and metastatic human prostate cohorts TCGA-333 and SU2C-150 (refs 3, 4), respectively. In agreement with our functional studies, *ERF* mRNA levels are inversely correlated with two androgen transcriptional activity signatures[19,20], both in normal human prostate and in all primary tumour subtypes (Fig. 2d and Extended Data Fig. 5a). This reciprocal association is also observed in metastatic cancers if the analysis is limited to tumours without amplification of or mutations in the *AR* gene, which encodes the androgen receptor (Extended Data Fig. 5b). The fact that reduced *ERF* expression enhances androgen receptor transcriptional output even in the absence of mutation or deletion, raises the possibility that *ERF* may have a broader role in prostate oncogenesis.

The inverse correlation between *ERF* mRNA level and the androgen receptor signature is also observed in the half of prostate cancers

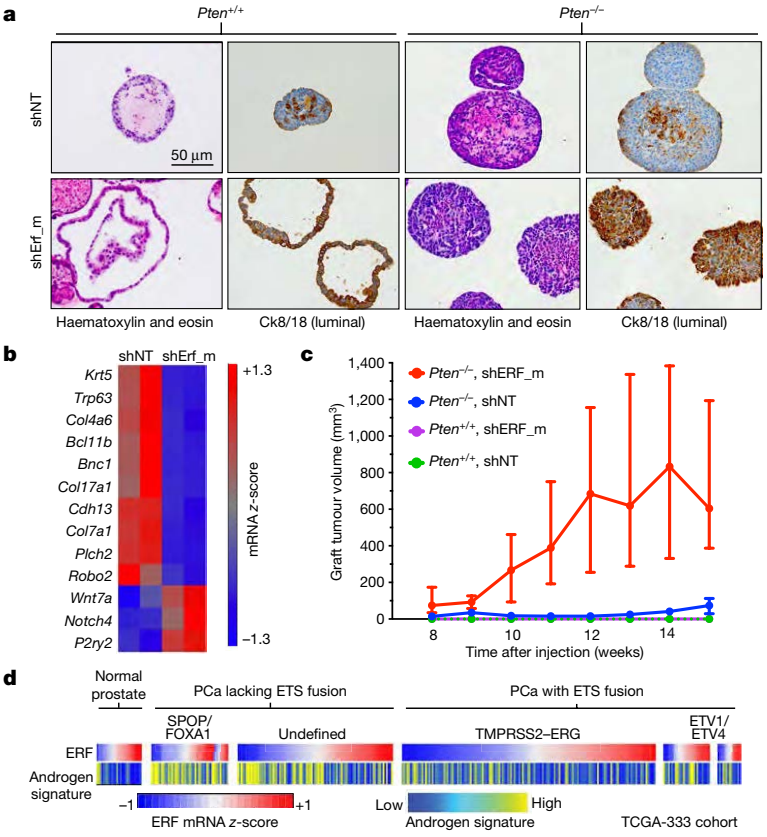


Figure 2 | ERF is a tumour suppressor, a negative regulator of androgen signalling, and its loss phenocopies TMPRSS2–ERG gain. a, Immunohistochemistry of mouse prostate organoids infected with non-targeting shRNA (shNT) or targeting ERF (shErf_m). **b**, Basal signature[18] applied to $Pten^{+/+}$ organoid RNA-seq ($n = 2$ biological replicates). **c**, Tumour volumes of organoid grafts. Data are median ± interquartile range; $n = 10$ tumours per condition. **d**, Human TCGA primary prostate cancer (PCa) cohort[4] ($n = 333$ patients) interrogated for ERF mRNA level and the Hieronymus androgen transcriptional activity signature[4,19,20].

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

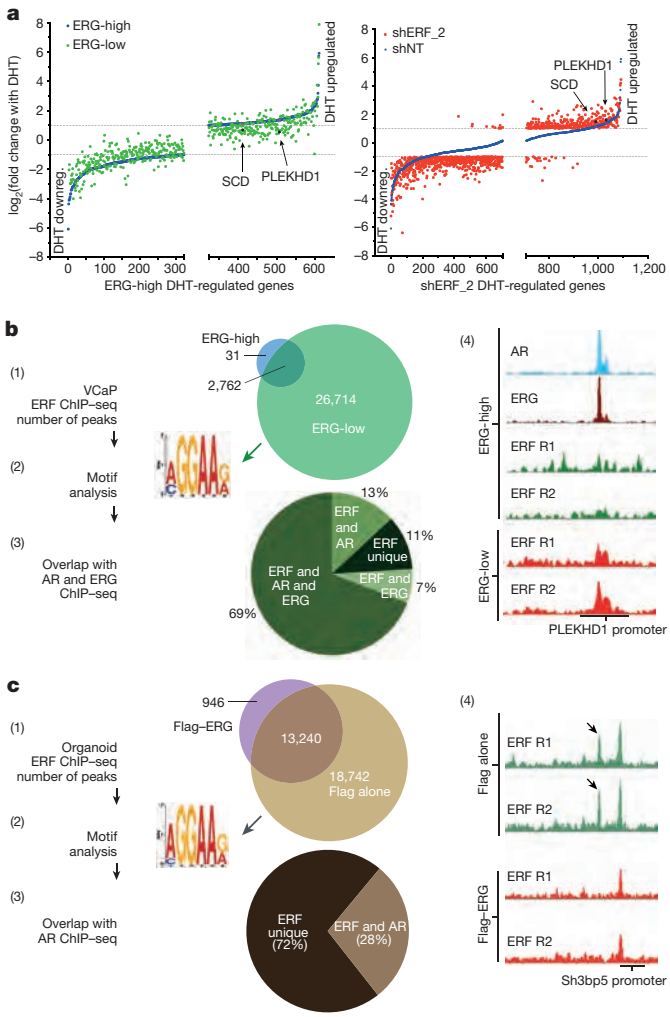


Figure 3 | ERF and TMPRSS2–ERG have opposing effects on the androgen receptor transcriptome and cistrome. a, Androgen-regulated genes (as in Extended Data Fig. 4b) analysed by magnitude of change in VCaP cells infected with a doxycycline-inducible shRNA targeting ERG (with doxycycline, ERG-low; without doxycycline, ERG-high) or a constitutive shRNA targeting ERF (shERF_2); $n = 3$ biological replicates. DHT downreg., DHT downregulated. **b**, ERF ChIP–seq in VCaP cells ($n = 2$ biological replicates: R1, R2) analysed by (1) peak overlap, (2) motif analysis, (3) overlap with ERG and androgen receptor ChIP–seq, and (4) example ChIP–seq signals. **c**, ERF ChIP–seq in normal prostate organoids infected with either a Flag–ERG lentivirus or Flag alone, both treated with doxycycline ($n = 2$ biological replicates: R1, R2) and analysed as in **b**.

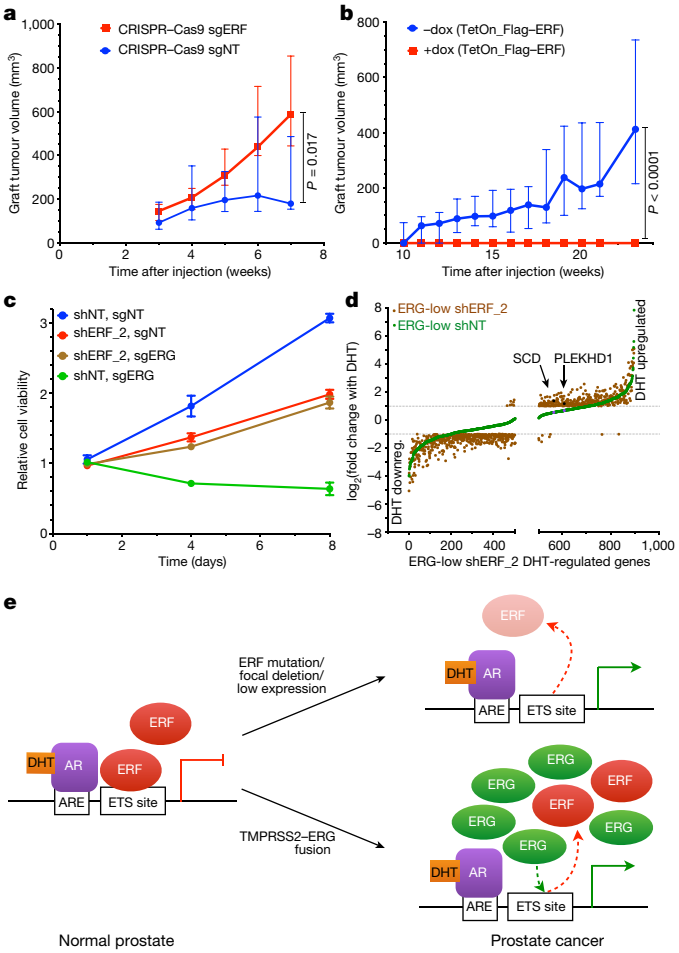


Figure 4 | TMPRSS2–ERG activity is mediated, in part, by inactivation of ERF function. a, Tumour volumes of grafts derived from $Pten^{-/-}$;$R26^{ERG/ERG}$ organoids infected with CRISPR–Cas9 targeting ERF (sgERF). Data are median ± interquartile range; $n = 10$ tumours per condition, P value via Mann–Whitney exact two-tailed test. **b**, Similar to **a** but infected instead with doxycycline-inducible Flag–ERF. **c**, Cell viability assay in pooled VCaP cells infected first with shRNA targeting ERF (shERF_2) followed by sgERF. Data are mean ± s.e.m.; $n = 3$ biological replicates. **d**, Androgen-regulated genes (as in Extended Data Fig. 4b) in ERG-low VCaPs and shERF_2 analysed by magnitude of change; $n = 3$ biological replicates. DHT downreg., DHT downregulated. **e**, ERF inactivation model for genes with androgen receptor and ETS binding sites.

that possess the *TMPRSS2–ERG* fusion (Fig. 2d), suggesting that ERF may have an androgen receptor repressive function in this subtype as well. To investigate this possibility, the expression of *ERG* or *ERF* was separately inhibited via shRNA (Extended Data Fig. 6a, b) in the ERG-fusion-positive VCaP cell line and the androgen transcriptome analysed as before. Consistent with earlier work[15], inhibition of *ERG* expression ('ERG-low') resulted in a contracted androgen transcriptome compared with the wild-type state ('ERG-high'). Conversely, *ERF* inhibition increased the change in expression of androgen receptor target genes and doubled the size of the androgen transcriptome (Fig. 3a and Extended Data Fig. 7). *ERF* and *ERG* knockdown had no effect on each other's expression, on androgen receptor protein levels, or on androgen receptor subcellular localization (Extended Data Fig. 6c, d).

Given the similarity of their ETS domains[12] (Extended Data Fig. 1b), we postulated that the opposing effects of ERF and ERG on androgen signalling could be explained by competition for androgen-receptor-associated ETS binding sites, which we investigated by ERF chromatin immunoprecipitation followed by sequencing (ChIP–seq) in the ERG-high and ERG-low states (Fig. 3b and Extended Data Fig. 8). *De novo* motif analysis identified the canonical ETS motif as the primary ERF binding site, with 2,793 binding sites in the ERG-high condition ((1) and (2) in Fig. 3b). Remarkably, an additional 26,714 ERF binding sites were observed in the ERG-low state, 76% of which were bound by ERG before ERG knockdown ((1) and (3) in Fig. 3b, and Extended Data Fig. 8d). Furthermore, the ChIP–seq signal intensity of the smaller number of ERF peaks observed in the ERG-high state was increased in the ERG-low state in almost all cases (Extended Data Fig. 8c, e). Finally, ERF peaks in the ERG-low state largely overlap with androgen receptor binding peaks ((3) in Fig. 3b), as previously reported for ERG[15], and are exemplified by *PLEKHD1* and *SCD*[20] ((4) in Fig. 3b, and Extended Data Fig. 8f). In the ERG-high state, both genes have prominent androgen-receptor- and ERG-associated binding but only limited ERF binding. However, in the ERG-low state, ERF binding is substantially increased in parallel with a decrease in their androgen-induced expression (Fig. 3a).

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Given the role that ERF also plays in benign prostate cells, we performed ChIP–seq to confirm that ERF binds ETS sites in the normal prostate organoids (Fig. 3c). Moreover, transient overexpression of ERG led to a significant decrease of ERF chromatin occupancy (Fig. 3c and Extended Data Fig. 9), consistent with the competition for binding also seen in VCaP cancer cells. We explored the overlap between androgen receptor and ERF sites in normal prostate organoids by androgen receptor ChIP–seq and found 28% of the ERF sites overlap with androgen receptor binding sites ((3) in Fig. 3c). This was lower than observed in the ERG-positive VCaP tumour cells, consistent with differences in androgen receptor chromatin occupancy between normal prostate and cancer[21].

We next asked whether *ERF* could modulate oncogenicity mediated by *ERG* expression. First, we used CRISPR–Cas9 (clustered regularly interspaced short palindromic repeats and CRISPR-associated protein 9) to partially delete *ERF* (sgErf) in pooled mouse $Pten^{-/-}$;$R26^{ERG/ERG}$ organoids (Extended Data Fig. 10a), which require ERG to form tumours[15]. The sgErf organoids formed tumours more rapidly than those infected with a non-targeting CRISPR (Fig. 4a). Likewise, induction of human ERF expression in these cells blocked androgen-dependent gene expression (Extended Data Fig. 10b, c) and prevented tumour formation (Fig. 4b). To address the role of ERF in transformation of human prostate cancer cells, we predicted that the anti-proliferative effect of ERG knockdown in ERG-positive cells[22] should be rescued by concurrent knockdown of ERF. Consistent with previous findings[22], partial *ERG* deletion via CRISPR–Cas9 led to a complete halt of proliferation (Fig. 4c and Extended Data Fig. 10d). Strikingly, stable shRNA knockdown of ERF before partial *ERG* deletion rescued cells from ERG-dependent proliferation and survival. Moreover, ERF knockdown restored the constricted androgen transcriptome conferred by *ERG* loss, including restoration of *SCD* and *PLEKHD1* mRNA upregulation (Fig. 4d and Extended Data Fig. 10e).

We propose that loss of ERF activity, either by rare genomic loss-of-function mutations or more commonly by competition with the *TMPRSS2–ERG* oncogenic gene product, leads to activation of the androgen receptor pathway and prostate cancer (Fig. 4e). The consequences of ERF loss and ERG gain have many common functional consequences, including regulation of luminal morphology, expanded activity of androgen receptor, and an ability to form tumours in a $Pten^{-/-}$ background. Such similarities are consistent with their mutual exclusivity and the ability of ERF loss to rescue ERG-positive prostate cancer cells from ERG dependency. On the other hand, the fact that *ERG* translocations are more common than *ERF* mutant tumours (46% versus 4%, respectively, in the TCGA-333 primary prostate cancer cohort) raises the possibility that ERG has additional gain-of-function activities that favour oncogenic transformation, such as an intrinsic androgen receptor reprogramming activity[15,21]. A related but unresolved question is whether the oncogenic phenotype of ERF loss is simply due to loss of repression, or to gain-of-function conferred by competition with an unknown endogenous 'positive' ETS factor. Understanding ETS competition has implications also for ETS-driven leukaemias and sarcomas, and perhaps more broadly for other oncogenic transcription factors for which corresponding loss-of-function mutations in cognate repressors might be considered.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 July 2016; accepted 4 May 2017.
Published online 14 June 2017.

1. Tomlins, S. A. *et al.* Recurrent fusion of *TMPRSS2* and ETS transcription factor genes in prostate cancer. *Science* **310,** 644–648 (2005).
2. Carver, B. S. *et al.* Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate. *Nat. Genet.* **41,** 619–624 (2009).
3. Robinson, D. *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* **161,** 1215–1228 (2015).
4. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* **163,** 1011–1025 (2015).
5. Gao, D. *et al.* Organoid cultures derived from patients with advanced prostate cancer. *Cell* **159,** 176–187 (2014).
6. Kumar, A. *et al.* Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. *Nat. Med.* **22,** 369–378 (2016).
7. Beltran, H. *et al.* Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat. Med.* **22,** 298–305 (2016).
8. International Cancer Genome Consortium. Prostate cancer - adenocarcinoma. *ICGC Cancer Genome Projects* https://icgc.org/icgc/cgp/70/508/71331 (2016).
9. Sgouras, D. N. *et al.* ERF: an ETS domain protein with strong transcriptional repressor activity, can suppress ets-associated tumorigenesis and is regulated by phosphorylation during cell cycle and mitogenic stimulation. *EMBO J.* **14,** 4781–4793 (1995).
10. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).
11. Twigg, S. R. F. *et al.* Reduced dosage of ERF causes complex craniosynostosis in humans and mice and links ERK1/2 signaling to regulation of osteogenesis. *Nat. Genet.* **45,** 308–313 (2013).
12. Hollenhorst, P. C., McIntosh, L. P. & Graves, B. J. Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu. Rev. Biochem.* **80,** 437–471 (2011).
13. Regan, M. C. *et al.* Structural and dynamic studies of the transcription factor ERG reveal DNA binding is allosterically autoinhibited. *Proc. Natl Acad. Sci. USA* **110,** 13374–13379 (2013).
14. Pires, D. E. V., Ascher, D. B. & Blundell, T. L. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res.* **42,** W314–W319 (2014).
15. Chen, Y. *et al.* ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss. *Nat. Med.* **19,** 1023–1029 (2013).
16. Karthaus, W. R. *et al.* Identification of multipotent luminal progenitor cells in human prostate organoid cultures. *Cell* **159,** 163–175 (2014).
17. Klezovitch, O. *et al.* A causal role for ERG in neoplastic transformation of prostate epithelium. *Proc. Natl Acad. Sci. USA* **105,** 2105–2110 (2008).
18. Smith, B. A. *et al.* A basal stem cell signature identifies aggressive prostate cancer phenotypes. *Proc. Natl Acad. Sci. USA* **112,** E6544–E6552 (2015).
19. Hieronymus, H. *et al.* Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* **10,** 321–330 (2006).
20. Nelson, P. S. *et al.* The program of androgen-responsive genes in neoplastic prostate epithelium. *Proc. Natl Acad. Sci. USA* **99,** 11890–11895 (2002).
21. Pomerantz, M. M. *et al.* The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat. Genet.* **47,** 1346–1351 (2015).
22. Wang, J. *et al.* Pleiotropic biological activities of alternatively spliced TMPRSS2/ERG fusion gene transcripts. *Cancer Res.* **68,** 8516–8524 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Heguy, H. Hieronymous, J. Li, Y. Liang, E. Peguero, M. Pirun, N. Socci, P. Watson, A. Viale, Y. Zhang, Memorial Sloan Kettering Cancer Center core facilities, and the members of the Sawyers laboratory for comments. R.B. was supported by an American Society of Clinical Oncology (ASCO) Young Investigator Award, a Department of Defense Physician Training Award, and a Prostate Cancer Foundation Young Investigator Award. W.A. was supported by an ASCO Young Investigator Award and Prostate Cancer Foundation Young Investigator Award. M.G.D. was supported by a Howard Hughes Medical Institute (HHMI) Summer Medical Fellowship. N.Sch. is supported by the Prostate Cancer Foundation. C.L.S. is an investigator of the HHMI and this project was supported by National Institutes of Health grants CA155169, CA19387, CA092629, and CA008748.

Author Contributions R.B. and C.L.S. conceived and oversaw the project, performed data interpretation, and co-wrote the manuscript. R.B., E.V.W., W.A., Z.Z., and P.S.S. performed immunoblots, R.B. and P.S.S. performed RNA analysis, W.R.K., P.J.I. and R.B. performed immunohistochemistry, J.W. performed the mouse grafting, R.B. and A.P. prepared experiments for ChIP–seq, and R.B. performed *in vitro* cell growth assays. W.R.K. made three-dimensional organoid cultures, E.V.W. made bacterially expressed proteins, and R.B., W.R.K., W.A., E.V.W., P.J.I., and M.G.D. cloned plasmid reagents. E.V.W. performed ETS stability analysis, J.A., N.Sch., and R.B. performed analysis of human prostate cancer cohorts, N.Sh. performed gene set enrichment analysis (GSEA), and P.W., Y.Z., and D.Z. performed ChIP–seq analysis. All individual authors made intellectual contributions and reviewed the manuscript. The International SU2C/PCF Cancer Dream provided four unpublished ERF point mutations. The International SU2C-PCF Dream Team is led by A.M.C. and C.L.S.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to C.L.S. (sawyersc@mskcc.org).

Reviewer Information *Nature* thanks J. Carroll, T. Sato, L. Trotman and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

The International SU2C/PCF Prostate Cancer Dream Team
Dan Robinson[9,10], Eliezer M. Van Allen[11,12], Yi-Mi Wu[9,10], Nikolaus Schultz[13,14], Robert J. Lonigro[9], Juan-Miguel Mosquera[15,16,17,18], Bruce Montgomery[19,20,21], Mary-Ellen Taplin[11], Colin C. Pritchard[22], Gerhardt Attard[23,24], Himisha Beltran[16,17,18,25], Wassim Abida[26], Robert K. Bradley[19], Jake Vinson[27], Xuhong Cao[9,28], Pankaj Vats[9], Lakshmi P. Kunju[9,10,29], Maha Hussain[29,30,31], Scott A. Tomlins[9,10,29,31], Kathleen A. Cooney[29,30,31], David C. Smith[29,30,31], Christine Brennan[9], Javed Siddiqui[9], Rohit Mehra[9,10], Yu Chen[9,25,26], Dana E. Rathkopf[25,26], Michael J. Morris[25,26], Stephen B. Solomon[32], Jeremy C. Durack[32], Victor E. Reuter[33], Anuradha Gopalan[33], Jianjiong Gao[14], Massimo Loda[11,12,34,35], Rosina T. Lis[11,34], Michaela Bowden[11,34,35], Stephen P. Balk[36], Glenn Gaviola[37], Carrie Sougnez[12], Manaswi Gupta[12], Evan Y. Yu[20,21], Elahe A. Mostaghel[19,20,21], Heather H. Cheng[19,20,21], Hyojeong Mulcahy[38], Lawrence D. True[39], Stephen R. Plymate[20,21], Heidi Dvinge[19], Roberta Ferraldeschi[23,24], Penny Flohr[23,24], Susana Miranda[23,24], Zafeiris Zafeiriou[23,24], Nina Tunariu[23,24], Joaquin Mateo[23,24], Raquel Perez-Lopez[23,24], Francesca Demichelis[16,40], Brian D. Robinson[15,16,17,18], Marc Schiffman[15,18,41], David M. Nanus[16,17,18,25], Scott T. Tagawa[16,17,18,25], Alexandros Sigaras[16,42,43], Kenneth W. Eng[16,42,43], Olivier Elemento[42], Andrea Sboner[15,16,18,42], Elisabeth I. Heath[44,45], Howard I. Scher[25,26], Kenneth J. Pienta[46], Philip Kantoff[3], Johann S. de Bono[23,24], Mark A. Rubin[15,16,17,18], Peter S. Nelson[18,20,21,47,48], Levi A. Garraway[11,12], Charles L. Sawyers[49] & Arul M. Chinnaiyan[9,10,28,29,31]

[9]Michigan Center for Translational Pathology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. [10]Department of Pathology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. [11]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. [12]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts 02142, USA. [13]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [14]Marie-Josée and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [15]Department of Pathology and Laboratory Medicine, Weill Medical College of Cornell University, New York, New York 10021, USA. [16]Institute for Precision Medicine, Weill Medical College of Cornell University, New York, New York 10021, USA. [17]New York Presbyterian Hospital, New York, New York 10021, USA. [18]Meyer Cancer, Weill Medical College of Cornell University, New York, New York 10021, USA. [19]Computational Biology Program, Public Health Sciences Division and Basic Science Division, Fred Hutchinson Cancer Center, University of Washington, Seattle, Washington 98109, USA. [20]Department of Medicine, University of Washington, Seattle, Washington 98109, USA. [21]Veterans Affairs Puget Sound Health Care System, University of Washington, Seattle, Washington 98109, USA. [22]Department of Laboratory Medicine, University of Washington, Seattle, Washington 98195, USA. [23]The Royal Marsden, Downs Road, Sutton SM2 5PT, UK. [24]The Institute of Cancer Research, Downs Road, Sutton SM2 5PT, UK. [25]Department of Medicine, Weill Medical College of Cornell University, New York, New York 10021, USA. [26]Genitourinary Oncology Service, Department of Medicine, Sidney Kimmel Center for Prostate and Urologic Cancers, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [27]Prostate Cancer Clinical Trials Consortium, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [28]Howard Hughes Medical Institute, University of Michigan, Ann Arbor, Michigan 48109, USA. [29]Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. [30]Department of Internal Medicine, Division of Hematology Oncology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. [31]Department of Urology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. [32]Interventional Radiology, Department of Radiology Service, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [33]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [34]Center for Molecular Oncologic Pathology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. [35]Department of Pathology, Brigham & Women's Hospital, Boston, Massachusetts 02115, USA. [36]Division of Hematology-Oncology, Department of Medicine, Beth Israel Deaconess Cancer Center, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA. [37]Department of Musculoskeletal Radiology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. [38]Department of Radiology, University of Washington, Seattle, Washington 98109, USA. [39]Department of Pathology, University of Washington Medical Center, Seattle, Washington 98109, USA. [40]Laboratory of Computational Oncology, CIBIO, Centre for Integrative Biology, University of Trento, 38123 Mattarello TN, Italy. [41]Division of Interventional Radiology, Department of Radiology, New York-Presbyterian Hospital/Weill Cornell Medical Center, New York, New York 10021, USA. [42]Institute for Computational Biomedicine, Department of Physiology and Biophysics, Weill Medical College of Cornell University, New York, New York 10021, USA. [43]Department of Physiology & Biophysics, Weill Medical College of Cornell University, New York, New York 10021, USA. [44]Department of Oncology, Wayne State University School of Medicine, Detroit, Michigan 48201, USA. [45]Molecular Therapeutics Program, Barbara Ann Karmanos Cancer Institute, Detroit, Michigan 48201, USA. [46]The James Buchanan Brady Urological Institute and Department of Urology, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA. [47]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [48]Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [49]Howard Hughes Medical Institute, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Cell lines. Mouse prostate organoids were isolated and grown as described previously[16] from the prostates of normal (*Pten*$^{+/+}$), *Pten*$^{-/-}$, and *Pten*$^{-/-}$; *R26*$^{ERG/ERG}$ mice[15]. Isolation and growth of MSK-PCa3 cells were described previously[5]. LNCaP cells were obtained from American Type Culture Collection (ATCC) and maintained as previously described[23]. CWR22Pc were a gift from M. Nevalainen and maintained as described previously[23]. For CWR22Pc cells specifically, charcoal-stripped media was used for transcriptome analysis. VCaP cells were obtained from ATCC (CRL-2876) and maintained as previously described[23]. Cells were confirmed to be free of mycoplasma using a Lonza detection kit (LT07-318). Cell line authentication was confirmed by SNP fingerprinting and hallmark gene-fusions/mutations were identified by deep sequencing.

Prostate cancer tumour profiling. Profiles of the various cohorts[3–7] outlined in Extended Data Fig. 1a can be explored in the cBioPortal for Cancer Genomics (http://www.cbioportal.org). The TCGA data can also be accessed through the Broad Institute FireBrowse portal (http://firebrowse.org/?cohort=PRAD), and the ICGC–CRUK Prostate Adenocarcinoma data can be accessed through the ICGC Portal (https://icgc.org/icgc/cgp/70/508/71331). In Fig. 1a and Extended Data Fig. 2b, we report four previously unpublished mutations in ERF from a larger Stand Up to Cancer[3] (SU2C) cohort that will be described in the future.

Organoid histology and xenografts. Organoids were infected with the indicated lentivirus, selected with puromycin and further selected with FACS. For histology and immunohistochemistry, organoids were processed as described previously[5,16]. Immunohistochemistry was performed using a Ventana BenchMark ULTRA. The anti-CK8/18 antibody was purchased from Abcam (ab53280). *In vivo* xenograft experiments were performed as described previously[23], using 7-week-old male C.B17 SCID mice (Taconic): one million cells were injected into the flank for a total of ten tumours per shRNA and cell type. Once tumours were palpable, tumour volume was measured weekly using a Peira TM900 system (Peira, Belgium). The maximal tumour volume permitted by our Institutional Animal Care and Use Committee under protocol 06-07-012 was 2 cm^3, beyond which mice were euthanized. All animal experiments were performed in compliance with the guidelines of the Research Animal Resource Center of Memorial Sloan Kettering Cancer Center.

Inhibition of gene expression, and overexpression. ERF shRNA knockdown experiments were performed by infection and puromycin selection of cells with lentivirus containing the miR-E based *SGEP* or *SGEN* vectors[24] gifted by J. Zuber containing the following guide sequences: shErf_m: TTGAACTTGTAGGTGAACCGTT, shERF_2: TTGTTTTGAATACATTCTCCAG, shERF_1: TTGAAATTGAACT TGTAGGTGA, shNT was previously described as Ren.713 targeting *Renilla* luciferase[24]. ERG shRNA knockdown experiments were performed by infection and neomycin selection of cells with lentivirus containing the *Tet-pLKO*-neo vector gifted by D. Wiederschain (Addgene plasmid 21916) using the previously published ERG shRNA and non-targeting sequences[25]. For Tet-responsive reporters, either dimethylsulfoxide (DMSO) vehicle or doxycycline (Sigma) was added at 100 ng ml^{-1}. CRISPR–Cas9 experiments were performed by infection and puromycin selection of cells with lentivirus containing the *lentiCRISPRv2* vector gifted by F. Zhang (Addgene plasmid 52961) containing the following guide sequences chosen via the http://www.genome-engineering.org website: sgERG (GATAACTCTGCGCTCGTTCG), sgERFm (CCTGCCAAGCGATGACGCCC), and previously described sgNT[26]. Overexpression of cDNAs was achieved by the constitutive or Tet-inducible *pLV*-based lentiviral expression system.

Transcription analysis. RNA was extracted from cell lines using an RNeasy Kit (Qiagen). For quantitative PCR with reverse transcription (RT–qPCR) and RNA-seq experiments, cells were plated in triplicate per condition/infected construct at the beginning of the assay (duplicate for mouse organoid RNA-seq), and thereafter replicates were processed independently. Error bars for bar graphs indicate s.e.m. for the biological replicates. DHT (Sigma) (or DMSO vehicle) was added at 1 nM, and treatments were performed for 16 h. For RT–qPCR, cDNA was generated with a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Data were quantified relative to β-actin or GAPDH expression, and relative expression was plotted. Primers for human and mouse ERF were purchased from Qiagen. Other qPCR primers were as follows: androgen receptor (F: CCATCTTGTCGTCAATGTTATGAAGC, R: AGCTTCTGGGTTGTCTCCTCAGTGG), ERG (F: CAAAACTCTCCA CGGTTAATGC, R: ACCGGTCCAGGCTGATCT).

For RNA-seq, library preparation and sequencing were performed by the Memorial Sloan Kettering Cancer Center Integrated Genomics Operation Core using Illumina HiSeq with 50 base pair (bp) paired-end reads, with approximately 30 million reads generated for each sample. The output data (FASTQ files) were mapped to the target genome (UCSC HG19 or UCSC MM10) using the rnaStar aligner, which mapped reads genomically and resolved reads across splice junctions. The 2-pass mapping method was used in which the reads were mapped twice. The first mapping pass used a list of known annotated junctions from Ensembl. Novel junctions found in the first pass were then added to the known junctions and a second mapping pass was done (on the second pass the RemoveNoncanonical flag was used). After mapping, we post-processed the output SAM files using the PICARD tools to add read groups, with AddOrReplaceReadGroups, which sorted the file and converted it to the compressed BAM format. We then computed the expression count matrix from the mapped reads using HTSeq (http://www-huber.embl.de/users/anders/HTSeq). The raw count matrix generated by HTSeq was then processed using the R/Bioconductor package DESeq (http://www-huber.embl.de/users/anders/DESeq/), which was used both to normalize the full dataset and to analyse differential expression between sample groups. Androgen-regulated genes were defined as a twofold difference, false discovery rate (FDR) < 0.05 with 1 nM DHT treatment for 16 h DHT. For GSEA, statistical analysis was performed with publicly available software from the Broad Institute (http://www.broadinstitute.org/gsea/index.jsp).

ChIP. Chromatin processing, as well as anti-androgen receptor, anti-ERG, and immunoglobulin-G ChIP were described previously[15]. ERF ChIP was performed with anti-ERF antibody (Pierce PA5-30237). For ChIP–qPCR, the ETS2 promoter primers (forward: TTACTTCCTCCAGAGACTGACGA; reverse: CGCCGGCCAGAGACGAT) were used. The PSA upstream sequence lacking the ERF motif or androgen receptor binding motif was described previously[27]. For ChIP–seq, library preparation and RNA-seq were performed by the Memorial Sloan Kettering Cancer Center Integrated Genomics Operation Core using Illumina HiSeq with 50-bp paired-end reads. The reads were aligned to the human genome (hg19, build 37) or the mouse genome (mm10, build 38) using the program BWA (default parameters) within the PEMapper. The software MACS2 (ref. 28) (-q 0.1) was used for peak identification with data from ChIP input DNAs as controls. Peaks of sizes >100 bp and with at least one base pair covered by >15 reads were selected as the final peaks. Peaks from different conditions were merged to obtain non-overlapping genomic regions, which were then used to determine conditional specific binding. Overlapped peaks were defined as those sharing at least one base pair. Two replicates were performed per condition, and peaks called in both replicates were used as the final peaks for each condition. To generate heat maps depicting ERF ChIP–seq read density in ± 2 kilobase regions of the ERF peak summits, the same number of ChIP–seq reads from different conditions were loaded into the software seqMINER[29], and the resulting read density matrices were sorted by the read densities in the ERG-low condition in VCaP cells or wild-type condition for the mouse prostate organoids, before colouring. The criteria for assigning peaks to genes have been described previously[30,31]. The MEME-ChIP software[32] was applied to 300-bp sequences around the peak summits for motif discovery.

Immunoprecipitation and western blot analysis. Protein was extracted from cell lines using M-PER Reagent (Thermo Scientific) and quantified by BCA Protein Assay (ThermoFisher Scientific). Nuclear/cytoplasmic fractionation was achieved with an NE-PER kit (ThermoFisher Scientific). ERF immunoprecipitation was performed with anti-ERF antibody (Pierce PA5-30237). Western blots were imaged using the fluorometric-based LiCor system, using primary antibodies against ERF (Santa Cruz sc-15435 or Abcam ab61108) and PTEN (Cell Signaling 9188). Western blotting for androgen receptor, ERG, GAPDH, and actin was described previously[15].

***In vitro* growth assay.** Cells were plated in triplicate and assayed at the time points indicated using CellTiter-Glo (Promega). Viability was plotted normalized to day 1. Error bars indicate s.e.m.

Statistics. Mutual exclusivity in the cBio Oncoprints was calculated using a two-tailed Fisher's exact test. For RT–qPCR and RNA-seq experiments, cells were plated in triplicate per condition/infected construct at the beginning of the assay (duplicate for mouse organoid RNA-seq), and thereafter replicates were processed independently. RT–qPCR bar graphs are plotted as mean $\pm$ s.e.m. For RNA-seq, differentially expressed genes were defined as a twofold difference, FDR < 0.05 of DESeq-normalized expression. For GSEA, statistical analysis was performed with publicly available software from the Broad Institute (http://www.broadinstitute.org/gsea/index.jsp). For the correlation of the androgen transcriptional signatures with *ERF* mRNA in human tumours, statistical significance of correlations was calculated by Spearman's test. Mouse graft experiments consisted of ten tumours per condition, and a Mann–Whitney test was used in Fig. 4 to analyse differences in tumour volume. The sample size estimate was based on our experience with previous experiments[5,15,23]. No formal randomization process was used to assign mice to a given organoid injection, and experimenters were not blinded.

Data availability. All data are available from the authors upon reasonable request. The described RNA-seq and ChIP–seq data have been deposited in the Gene Expression Omnibus under accession number GSE83653.

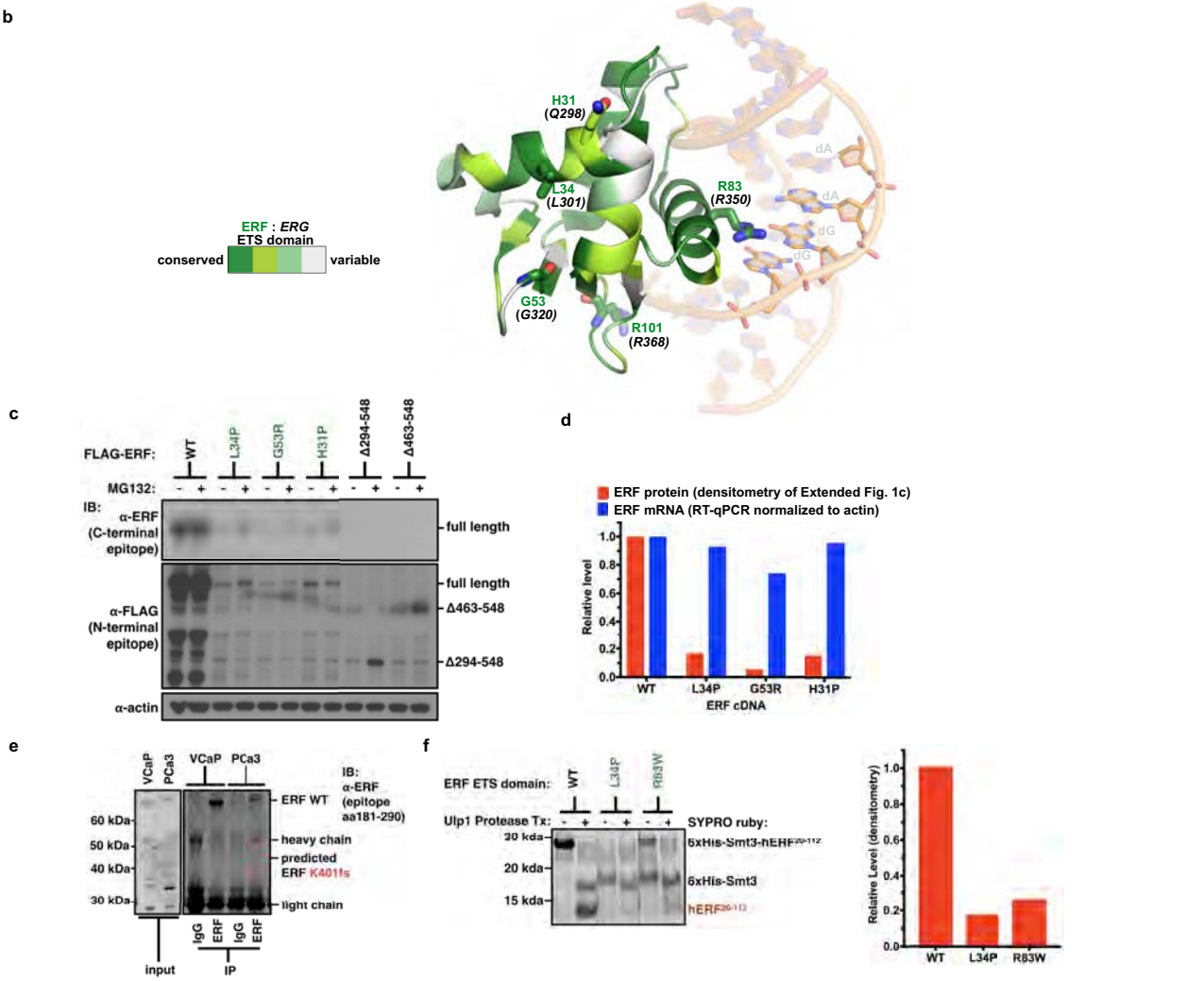
© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

23. Balbas, M. D. *et al.* Overcoming mutation-based resistance to antiandrogens with rational drug design. *eLife* **2,** e00499 (2013).
24. Fellmann, C. *et al.* An optimized microRNA backbone for effective single-copy RNAi. *Cell Reports* **5,** 1704–1713 (2013).
25. Mounir, Z. *et al.* TMPRSS2:ERG blocks neuroendocrine and luminal cell differentiation to maintain prostate cancer proliferation. *Oncogene* **34,** 3815–3825 (2015).
26. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR–Cas9 system. *Science* **343,** 80–84 (2014).
27. Wang, Q., Carroll, J. S. & Brown, M. Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Mol. Cell* **19,** 631–642 (2005).
28. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP–seq enrichment using MACS. *Nat. Protoc.* **7,** 1728–1740 (2012).
29. Ye, T. *et al.* seqMINER: an integrated ChIP–seq data interpretation platform. *Nucleic Acids Res.* **39,** e35 (2011).
30. Arora, V. K. *et al.* Glucocorticoid receptor confers resistance to antiandrogens by bypassing androgen receptor blockade. *Cell* **155,** 1309–1322 (2013).
31. Rockowitz, S. & Zheng, D. Significant expansion of the REST/NRSF cistrome in human versus mouse embryonic stem cells: potential implications for neural development. *Nucleic Acids Res.* **43,** 5730–5743 (2015).
32. Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27,** 1696–1697 (2011).

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

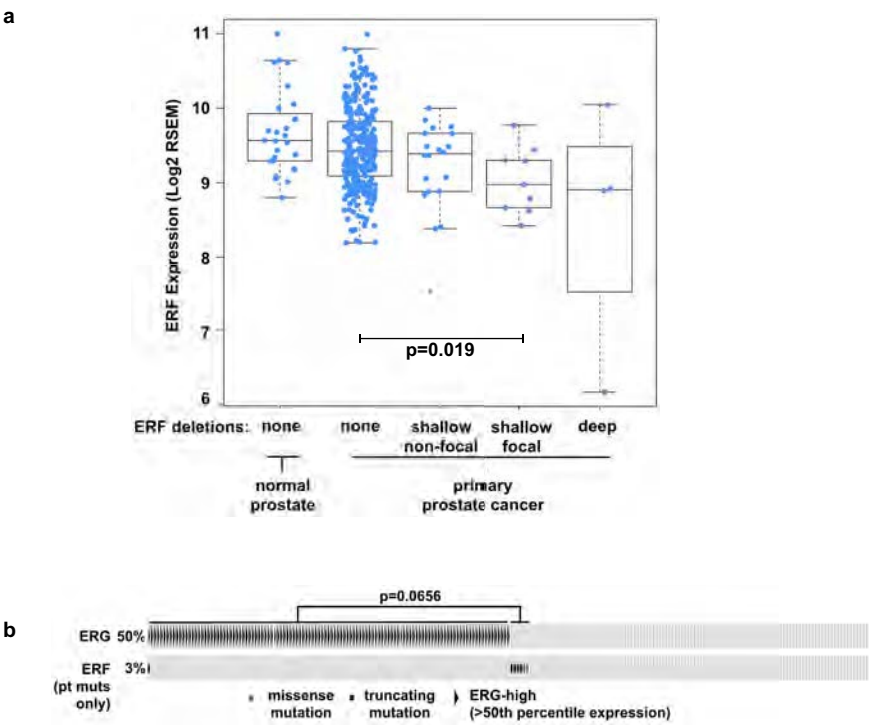
a

Cohorts	Source of specimen	Clinical State	Cohort size	# of patients with ERF point mutations	ERF Point Mutation Rate	Reference	Notes
TCGA-333	Primary site (prostate)	Localized, treatment naive	333	1	0.3%	Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. Cell 163, 1011–1025 (2015).	
TCGA-498	Primary site	Localized, treatment naive	498	4	0.8%	Broad Institute FireBrowse portal: (http://firebrowse.org/?cohort=PRAD)	Includes TCGA-333
CRUK	Primary site	Mixed	108	1	0.9%	ICGC Portal: https://icgc.org/icgc/cgp/70/508/71331	
SU2C-150	Metastatic site	Castration-resistant, metastatic prostate cancer	150	4	2.7%	Robinson, D. et al. Integrative clinical genomics of advanced prostate cancer. Cell 161, 1215– 1228 (2015).	
SU2C-294	Metastatic site	Castration-resistant, metastatic prostate cancer	294	8	2.7%	Pending publication from International SU2C/PCF Consortium.	Includes SU2C-150
FHCRC	Primary or metastatic site	Castration-resistant, metastatic prostate cancer	54	2	3.7%	Kumar, A. et al. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. Nature Medicine 1–13 (2016).	
Trento/Cornell/ Broad 2016	Metastatic site	Castration-resistant, metastatic prostate cancer	77	1	1.3%	Beltran, H. et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. Nature Medicine 22, 298–305 (2016).	
MSK Organoids	Local recurrence or metastatic site or circulating tumor cell	Castration-resistant, metastatic prostate cancer	7	1	14.3%	Gao, D. et al. Organoid Cultures Derived from Patients with Advanced Prostate Cancer. Cell (2014).	



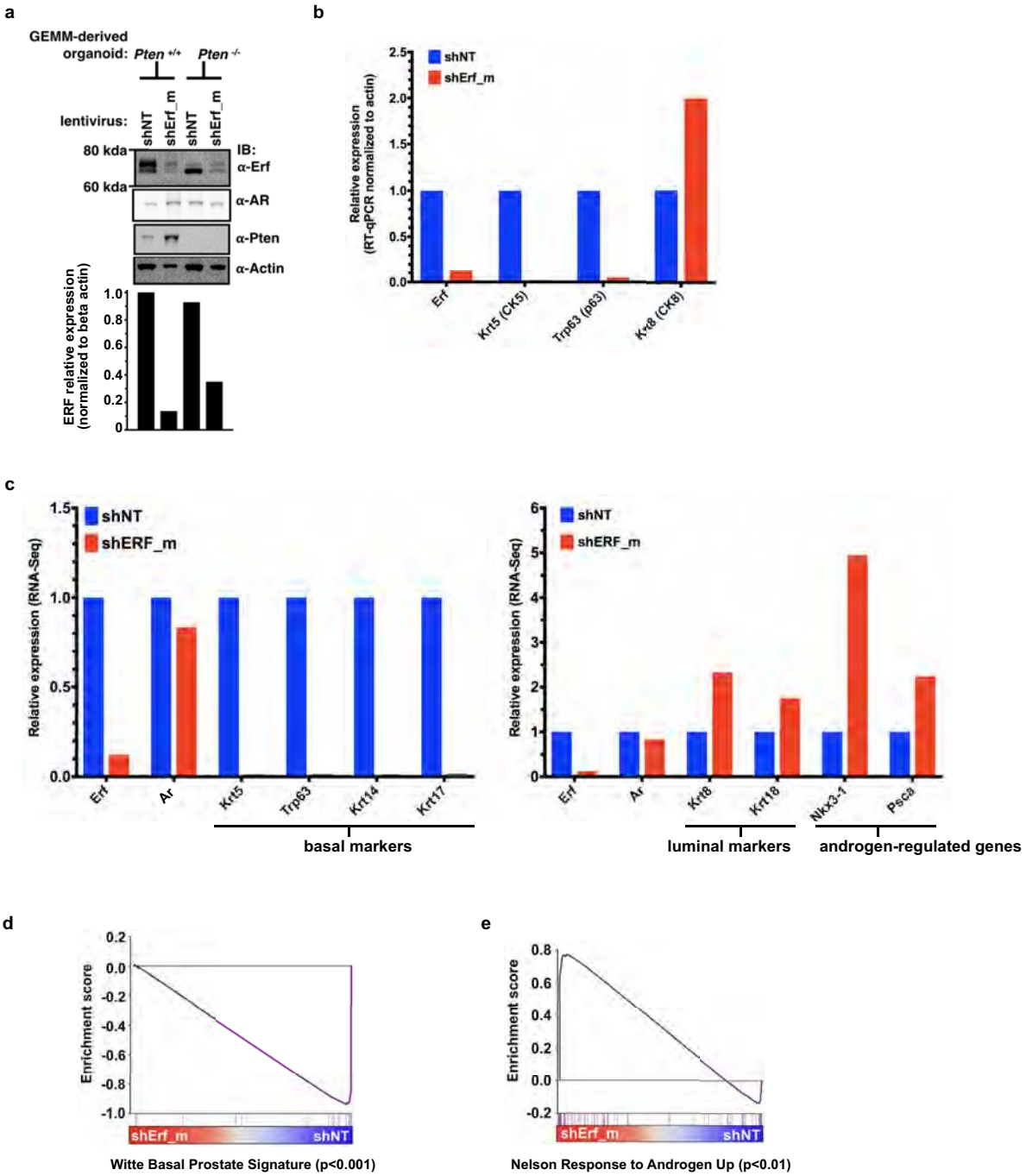
Extended Data Figure 1 | Recurrent ERF loss-of-function mutations are found in prostate cancer and are destabilizing. Related to Fig. 1. **a**, Description of the prostate cancer patient cohorts[3–8]. **b**, ERG ETS domain (Protein Data Bank accession number 4IRI) illustrating sequence conservation with ERF. ERF missense mutations indicated by sticks. **c**, Expression of ERF mutant proteins in LNCaP cells. For gel source data, see Supplementary Fig. 1. **d**, ERF mutant protein densitometry from immunoblot in Extended Data Fig. 1c compared with mRNA RT–qPCR. **e**, ERF immunoprecipitation in VCaP and MSK-PCa3 cells. **f**, Left, bacterially expressed ETS domain ($hERF^{20–112}$) with ERF tumour mutations before (−) or after (+) Ulp1 protease cleavage. Right, densitometry of *in vitro* ETS domains.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



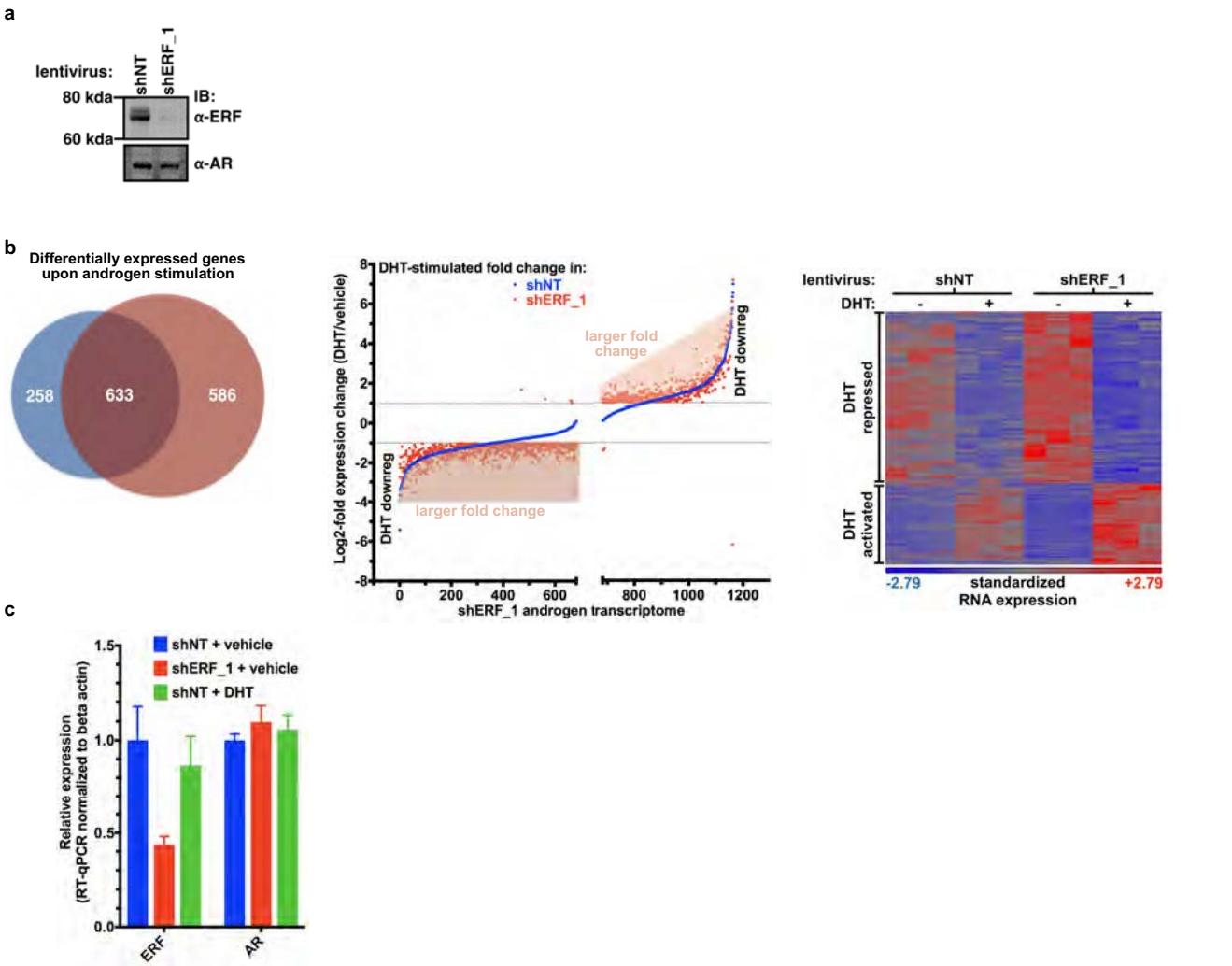
Extended Data Figure 2 | Recurrent ERF loss-of-function mutations and focal deletions are found in prostate cancer and are mostly exclusive to tumours without TMPRSS2–ERG. Related to Fig. 1. **a**, ERF expression in TCGA-333 cohort[4] ($n = 333$ patients) segregated by copy number loss. Data are Tukey box-and-whisker plots; *P* value calculated by two-tailed *t*-test of $\log_2$(RSEM values). **b**, cBio Oncoprint of the SU2C-294 ($n = 294$) metastatic prostate cancer cohort[3]. Unlike Fig. 1c, focal ERF deletions could not be ascertained because of differences between TCGA and SU2C copy number data[3,4]. *P* value calculated by Fisher's exact two-tailed test.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



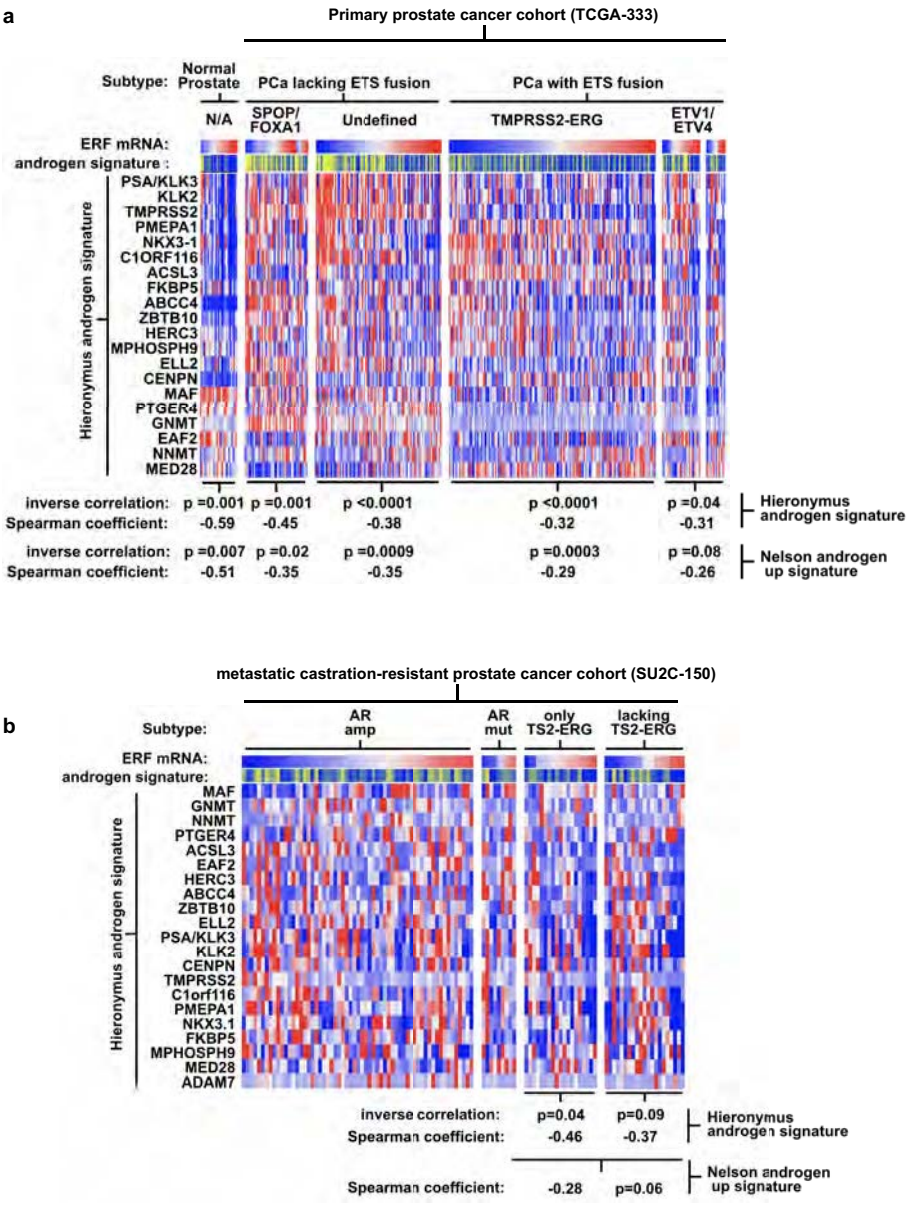
Extended Data Figure 3 | ERF is a negative regulator of androgen signalling. Related to Fig. 2. **a**, Mouse prostate organoids infected with non-targeting shRNA (shNT) or targeting ERF (shErf_m), grown in three-dimensional culture. For RT–qPCR, $n = 2$ biological replicates. For gel source data, see Supplementary Fig. 1. **b**, RT–qPCR analysis of the $Pten^{+/+}$ organoids; $n = 2$ biological replicates. **c**, RNA-seq analysis of organoids derived from $Pten^{+/+}$ and $Pten^{-/-}$ mouse prostates infected with non-targeting shNT or shErf_m; $n = 2$ biological replicates. **d**, $Pten^{+/+}$ organoid RNA-seq ($n = 2$ biological replicates) interrogated by GSEA for expression signature of Witte basal prostate cancer[18]. **e**, Same data interrogated by GSEA for Nelson androgen up expression signature[20].

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



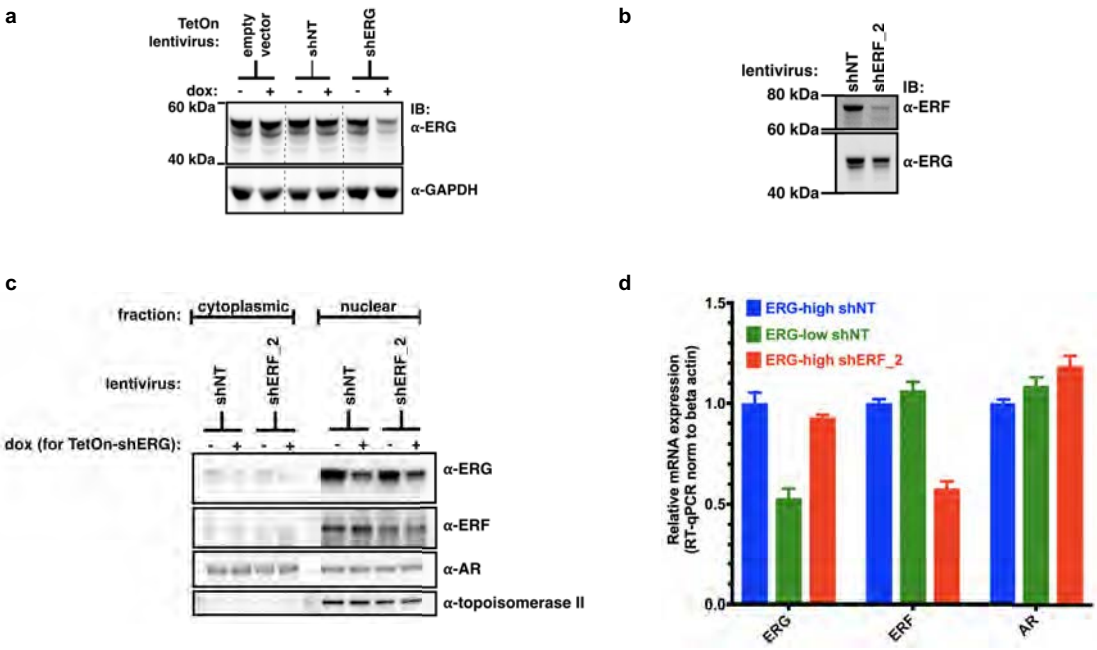
Extended Data Figure 4 | ERF is a negative regulator of androgen signalling. Related to Fig. 2. **a**, CWR22Pc cells infected with shRNA targeting human ERF (shERF_1) or a non-targeting shRNA (shNT). For gel source data, see Supplementary Fig. 1. **b**, Androgen-regulated genes (at least a twofold change, FDR < 0.05 by RNA-seq with 1 nM DHT for 16 h) in CWR22PC cells infected with either shERF_1 or shNT, analysed by the number (left, Venn diagram), the magnitude of expression change (centre, graph), and heat map (right); $n = 3$ biological replicates. **c**, RT–qPCR from CWR22Pc cells infected with shERF_1 or shNT. Data are mean ± s.e.m.; $n = 3$ biological replicates.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



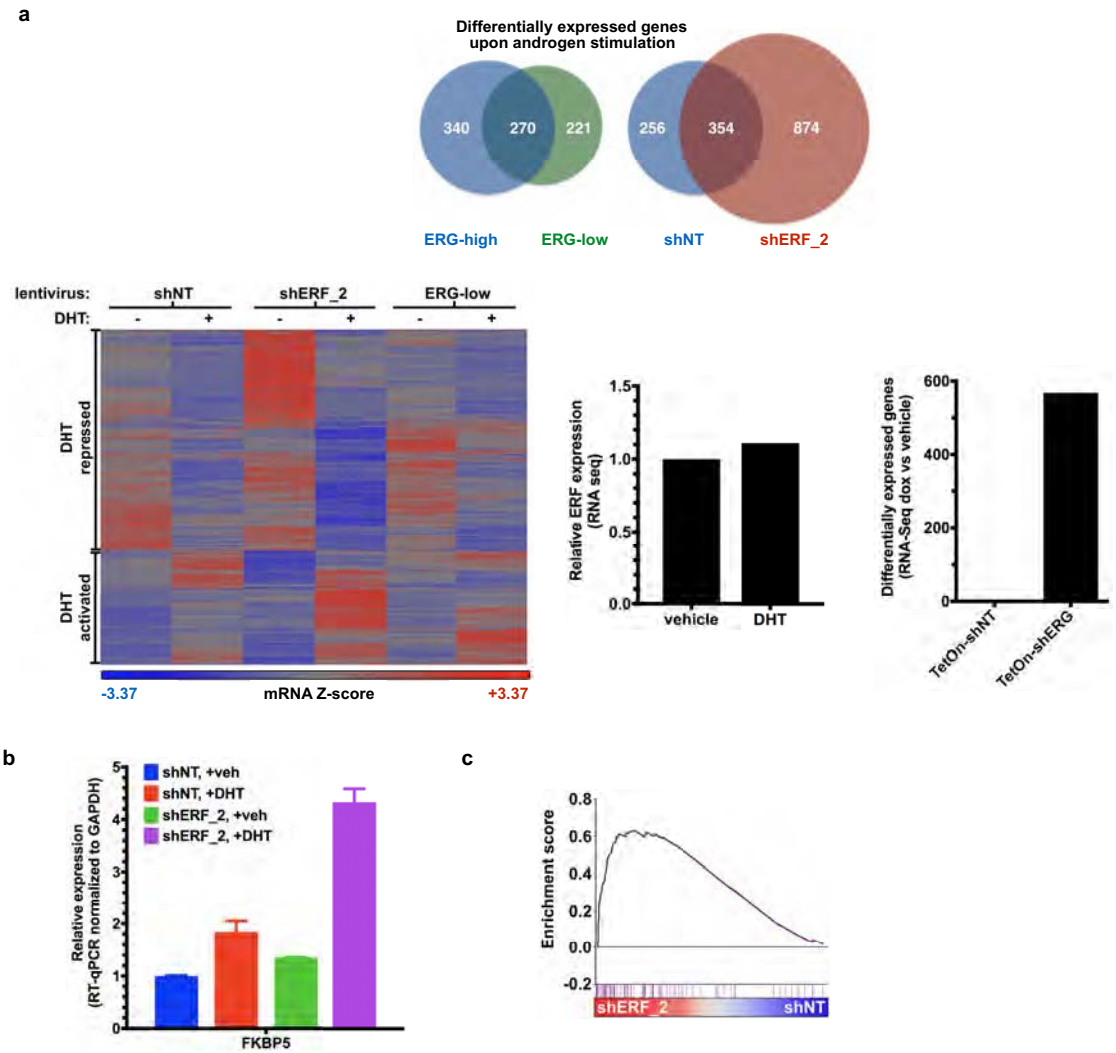
Extended Data Figure 5 | ERF is a negative regulator of androgen signalling. Related to Fig. 2. **a**, Full version of Fig. 2d. Expression profiles of the TCGA-333 primary prostate cancer cohort[4] ($n = 333$ patients) were interrogated for correlation between the ERF mRNA level and two androgen transcriptional activity signatures[4,19]. *P* values were calculated by the Spearman correlation test. **b**, The same analysis as **a** was applied to the SU2C-150 ($n = 150$ patients) metastatic castration-resistant prostate cancer cohort[3] (mCRPC).

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



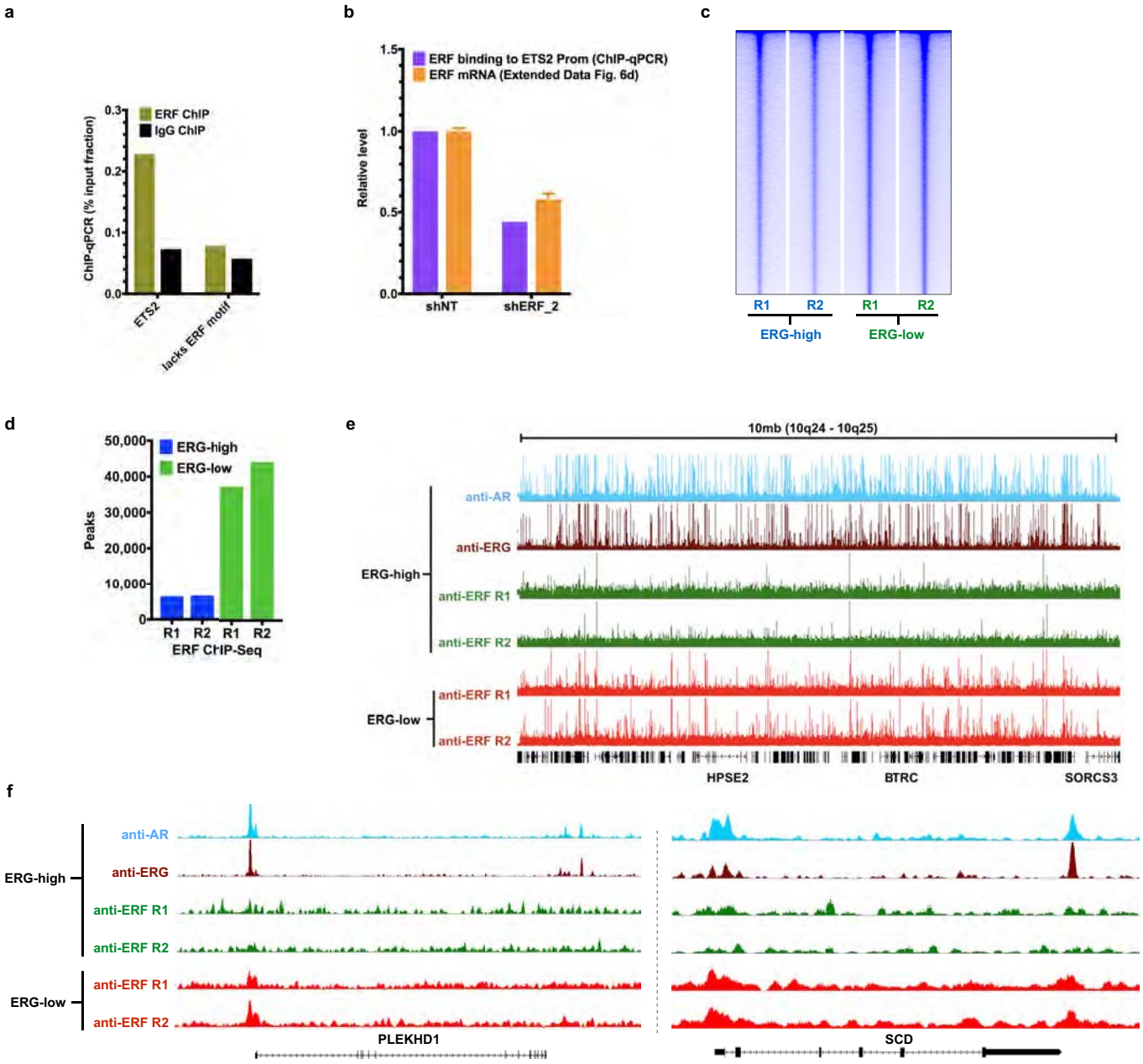
Extended Data Figure 6 | ERF and ERG knockdown do not affect androgen receptor levels or its subcellular localization. Related to Fig. 3. **a**, VCaP cells infected with doxycycline (dox)-inducible shRNA targeting ERG (+dox, ERG-low; −dox, ERG-high). For gel source data, see Supplementary Fig. 1. **b**, VCaP cells infected with shRNA targeting ERF (shERF_2) or a non-targeting shRNA (shNT). **c**, Nuclear/cytoplasmic fractionation of VCaP cells infected with shNT or shERF_2, and the doxycycline-inducible shRNA targeting ERG (shERG). **d**, RT–qPCR with ERG-low VCaP cells compared with those infected with shERF_2. Data are mean ± s.e.m.; $n = 3$ biological replicates.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



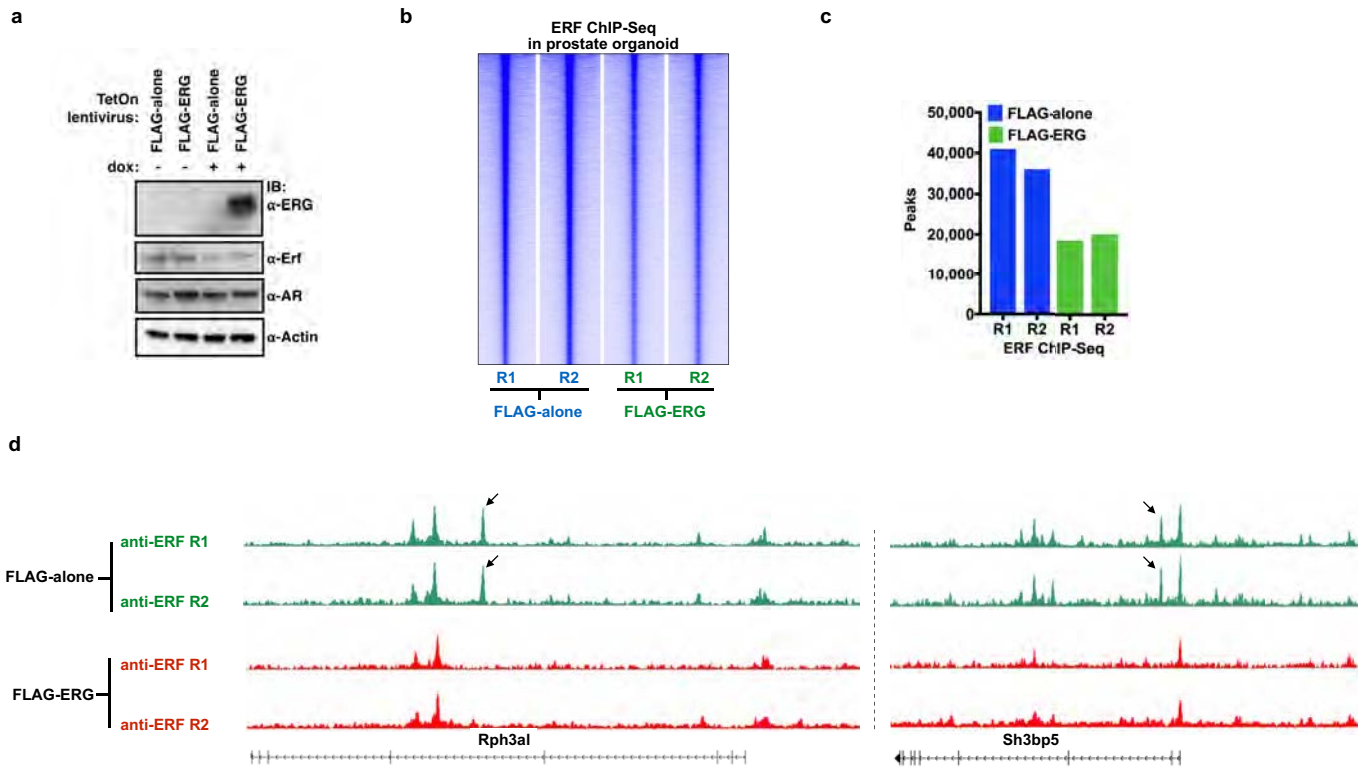
Extended Data Figure 7 | ERF and TMPRSS2–ERG have opposing effects on the androgen transcriptome. Related to Fig. 3. **a**, Androgen-regulated genes (at least a twofold change, FDR < 0.05 by RNA-seq with 1 nM DHT for 16 h) in VCaP cells infected with a doxycycline-inducible shRNA targeting ERG (with doxycycline, ERG-low; without doxycycline, ERG-high) or a constitutive shRNA targeting ERF (shERF_2) and analysed by number (top) and heat map (bottom left); $n = 3$ biological replicates. Bottom centre, RNA-seq analysis evaluating the effect of dihydrotestosterone on ERF expression. Bottom right, the effect of doxycycline alone on RNA-seq differential expression analysis. **b**, RT–qPCR of shERF_2-infected VCaP cells treated ± DHT. Data are mean ± s.e.m.; $n = 3$ biological replicates. **c**, Interrogation of RNA-seq in shERF_2 VCaP cells ($n = 3$ biological replicates) by GSEA for Nelson androgen up expression signature[20].

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



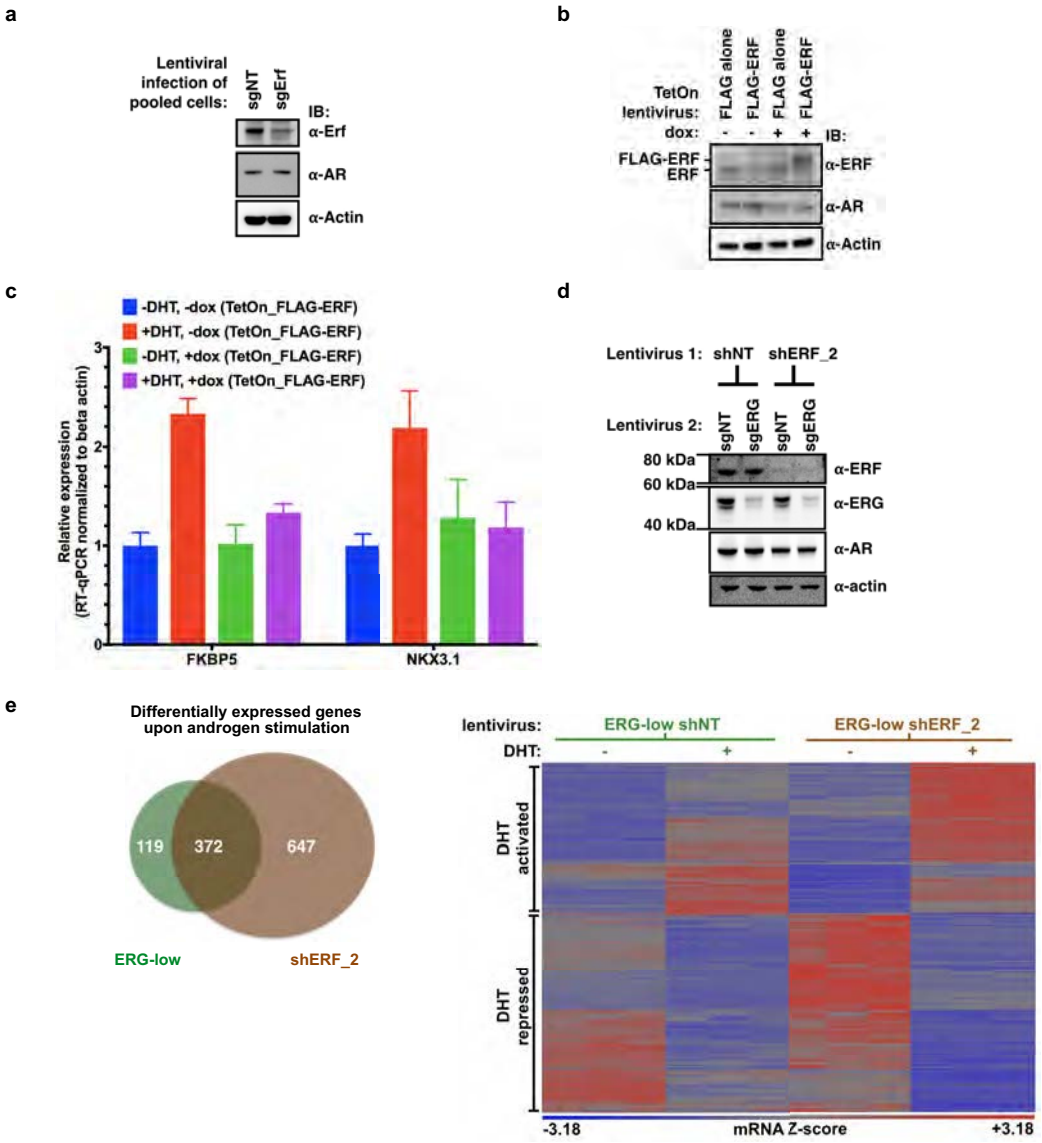
Extended Data Figure 8 | ERF and TMPRSS2–ERG have opposing effects on the androgen receptor cistrome. Related to Fig. 3. **a**, ChIP–qPCR ($n = 2$ biological replicates) in VCaP cells, amplifying either the ETS2 promoter region that contains a known ERF binding site[9], or an upstream element of PSA[27] lacking the ERF binding motif noted in (2) in Fig. 3b. **b**, The effect of ERF shRNA knockdown on its binding to the ETS2 promoter as assessed by ChIP–qPCR ($n = 2$ biological replicates), compared with its effect on *ERF* mRNA by RT–qPCR. Data are mean ± s.e.m.; $n = 3$ biological replicates. **c**, ERF ChIP–seq in ERG-high or ERG-low ($n = 2$ biological replicates: R1, R2) analysed by heat maps. **d**, Comparison of ERF ChIP–seq peak numbers, $n = 2$ biological replicates: R1, R2. **e**, A 10-Mb region illustrating ChIP–seq of ERF binding in ERG-high condition compared with ERG-low. **f**, ChIP–seq signals for the SCD and PLEKHD1 loci. In both **e** and **f**, ChIP–seq signals at the *y* axis were normalized by read depths.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 9 | ERG expression decreases the ERF cistrome in normal prostate organoids. Related to Fig. 3. **a**, Mouse normal prostate organoids were infected with a TetOn doxycycline-inducible Flag–ERG or empty vector (Flag-alone) and treated with or without doxycycline. For gel source data, see Supplementary Fig. 1. **b**, ERF ChIP–seq in normal prostate organoids infected with either Flag-alone or Flag–ERG lentivirus, both treated with doxycycline ($n = 2$ biological replicates: R1, R2), and analysed with heat maps. **c**, Comparison of ERF ChIP–seq peak numbers, $n = 2$ biological replicates (R1, R2). **d**, ERF ChIP–seq signals for the Rph3a1 and Sh3bp5 loci, normalized by read depths.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 10 | TMPRSS2–ERG activity is mediated, in part, by inactivation of ERF function. Related to Fig. 4. **a**, Pooled mouse *Pten*$^{-/-}$; *R26*$^{ERG/ERG}$ organoids infected with CRISPR–Cas9 targeting AAVS1 (sgNT) or ERF (sgErf). For gel source data, see Supplementary Fig. 1. **b**, *Pten*$^{-/-}$; *R26*$^{ERG/ERG}$ organoids infected instead with a doxycycline-inducible Flag–ERF or empty vector (Flag alone). **c**, RT–qPCR of mRNA isolated from the Flag–ERF-infected organoids and treated with or without doxycycline, with or without dihydrotestosterone 1 nM for 16 h. Data are mean ± s.e.m.; $n = 3$ biological replicates. **d**, Pooled VCaP cells were first infected with non-targeting shRNA (shNT) or ERF (shERF_2), followed by sgNT or sgERF CRISPR–Cas9. **e**, Androgen-regulated genes (at least a twofold change, FDR < 0.05 by RNA-seq with 1 nM DHT for 16 h) in VCaP cells infected with a doxycycline-inducible shRNA targeting ERG (with doxycycline, ERG-low) and either a constitutive shRNA targeting ERF (shERF_2) or a non-targeting shRNA (shNT). Data analysed by number (left, Venn diagram) and heat map (right); $n = 3$ biological replicates.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

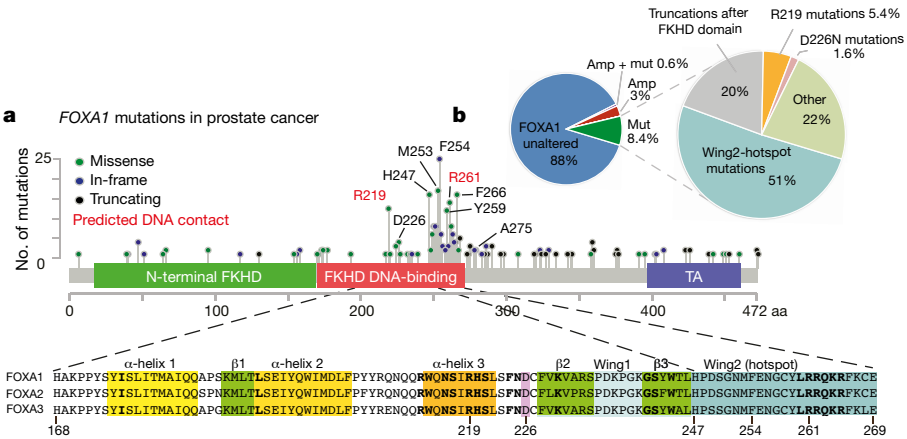
https://doi.org/10.1038/s41586-019-1318-9

FOXA1 mutations alter pioneering activity, differentiation and prostate cancer phenotypes

Elizabeth J. Adams[1], Wouter R. Karthaus[1], Elizabeth Hoover[1], Deli Liu[2,3,4], Antoine Gruet[5], Zeda Zhang[1,6], Hyunwoo Cho[7,8], Rose DiLoreto[8,9], Sagar Chhangawala[7,8], Yang Liu[10], Philip A. Watson[1], Elai Davicioni[10], Andrea Sboner[2,4,11], Christopher E. Barbieri[2,3,11], Rohit Bose[12], Christina S. Leslie[8] & Charles L. Sawyers[1,13]*

Mutations in the transcription factor FOXA1 define a unique subset of prostate cancers but the functional consequences of these mutations and whether they confer gain or loss of function is unknown[1–9]. Here, by annotating the landscape of *FOXA1* mutations from 3,086 human prostate cancers, we define two hotspots in the forkhead domain: Wing2 (around 50% of all mutations) and the highly conserved DNA-contact residue R219 (around 5% of all mutations). Wing2 mutations are detected in adenocarcinomas at all stages, whereas R219 mutations are enriched in metastatic tumours with neuroendocrine histology. Interrogation of the biological properties of wild-type FOXA1 and fourteen *FOXA1* mutants reveals gain of function in mouse prostate organoid proliferation assays. Twelve of these mutants, as well as wild-type FOXA1, promoted an exaggerated pro-luminal differentiation program, whereas two different R219 mutants blocked luminal differentiation and activated a mesenchymal and neuroendocrine transcriptional program. Assay for transposase-accessible chromatin using sequencing (ATAC-seq) of wild-type FOXA1 and representative Wing2 and R219 mutants revealed marked, mutant-specific changes in open chromatin at thousands of genomic loci and exposed sites of FOXA1 binding and associated increases in gene expression. Of note, ATAC-seq peaks in cells expressing R219 mutants lacked the canonical core FOXA1-binding motifs (GTAAAC/T) but were enriched for a related, non-canonical motif (GTAAAG/A), which was preferentially activated by R219-mutant FOXA1 in reporter assays. Thus, *FOXA1* mutations alter its pioneering function and perturb normal luminal epithelial differentiation programs, providing further support for the role of lineage plasticity in cancer progression.

To investigate the role of mutant and wild-type *FOXA1* in prostate cancer, we examined the landscape of *FOXA1* mutations across a cohort of 3,086 patients with primary or metastatic disease. The overall frequency of *FOXA1* mutation in these patients is around 11% (Fig. 1a, b), 3% of which are genomic amplifications and 8.4% are somatic point mutations, with less than 1% having both types of mutations (Fig. 1b). More than 50% of *FOXA1* mutations map to a specific hotspot in the Wing2 region of the forkhead (FKHD) DNA-binding domain, often as missense mutations or indels in Wing2 (mainly between H247 and



c

MSK-IMPACT 504 mutant frequencies by stage	R219	Wing2-hotspot alterations	Others	Total
Locoregional (*n* = 30)	6.7% (2)	66.7% (20)	26.7% (8)	100%
Metastatic non-castration-resistant (*n* = 13)	0.0% (0)	38.5% (5)	61.5% (8)	100%
Castration-resistant (*n* = 18)	11.1% (2)	44.4% (8)	44.4% (8)	100%

d

Point mutation frequencies in adenocarcinoma vs NEPC	*FOXA1* mutant cases			Non-*FOXA1* mutant cases
	R219 point mutants	Other point mutants	Total	
Adenocarcinoma (*n* = 1,742)	9.5% (8)	90.5% (76)	100% (84)	1,658
NEPC (*n* = 80)	75% (3)	25% (1)	100% (4)	76

***

Fig. 1 | Recurrent *FOXA1* mutations in prostate cancer cluster in the FKHD DNA-binding domain. a, Top, Distribution of *FOXA1* mutations from a pan-prostate cancer analysis of 3,086 patients along the protein sequence, depicting the various alterations seen in patients. Bottom, The amino acid sequence of the conserved FKHD DNA-binding domain, with secondary structural elements indicated. Residues in red (top) or bold (bottom) are predicted to make contacts with DNA[10]. **b**, Classification of observed *FOXA1* alterations. Mutations can be subdivided into several classes on the basis of their location in the FOXA1 protein. Amp, amplification; mut, somatic point mutation. **c**, Frequency of the various classes of *FOXA1* alterations in the three clinical stages reported in MSK-IMPACT 504. Data are expressed as percentage of the total number of samples with *FOXA1* mutations at a given clinical stage. **d**, Prevalence of R219 mutations compared to all other point mutations found in *FOXA1* in adenocarcinoma versus NEPC. Cases pooled from the Trento–Cornell–Broad[12] dataset and MSK-IMPACT 1708. ***$P = 0.0059$, Fisher's exact test, two-sided.

[1]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [2]Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA. [3]Department of Urology, Weill Cornell Medicine, New York, NY, USA. [4]HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY, USA. [5]Center for Epigenetics Research, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [6]Louis V. Gerstner Jr Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [7]Computational and Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [8]Physiology, Biophysics, and Systems Biology Program, Weill Cornell Graduate School, New York, NY, USA. [9]Tri-Institutional Training Program in Computational Biology & Medicine, Weill Cornell Medicine, New York, NY, USA. [10]GenomeDx Bioscience, Vancouver, British Columbia, Canada. [11]Englander Institute for Precision Medicine of Weill Cornell Medicine and NewYork-Presbyterian Hospital, New York, NY, USA. [12]Departments of Anatomy, Medicine and Urology, University of California, San Francisco, San Francisco, CA, USA. [13]Howard Hughes Medical Institute, Chevy Chase, MD, USA. *e-mail: sawyersc@mskcc.org

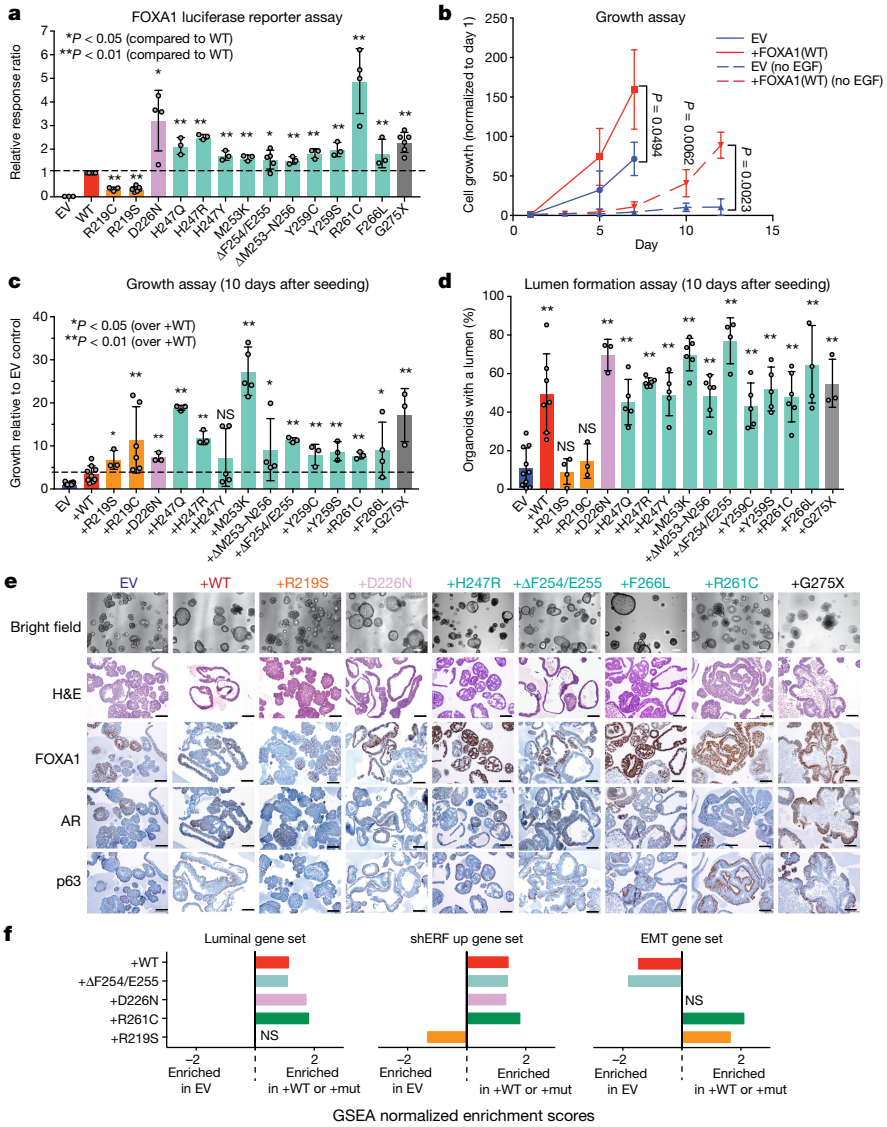


Fig. 2 | Expression of *FOXA1* mutants promotes growth and reveals distinct morphologies for different classes of alterations. a, FOXA1 luciferase reporter assay with results normalized to level of FOXA1(WT) activity. Colours indicate the position of the altered amino acid within the FKHD DNA-binding domain depicted in Fig. 1a. Grey indicates truncation. **b**, Overexpression of FOXA1 promotes growth in prostate organoids in standard medium conditions (solid lines, $n = 3$ biological replicates) and in restrictive medium conditions (dashed lines, no EGF, $n = 8$ biological replicates). EV, pCW empty vector control. **c**, Overexpression of wild-type (+WT) or various *FOXA1* mutants promotes growth ten days after seeding in medium lacking EGF. **d**, Quantification of lumen-containing organoids for each line in the *Foxa1* allelic series. All values of *P* are relative to empty vector control, calculated using unpaired, two-tailed Student's *t*-test. **e**, Histology and immunohistochemistry of organoid lines overexpressing variants of FOXA1 via the doxycycline-inducible pCW vector ten days after seeding. Images from a single biological experiment. H&E, haematoxylin and eosin staining. Scale bars: top, 200 μm; other rows, 100 μm. **f**, Summary of GSEA comparing *Foxa1* wild-type or mutant organoid lines to empty vector control for a basal low (luminal) gene set, the hallmark EMT gene set and a gene set of the top 100 genes induced following ERF knockdown (using short hairpin RNA directed against ERF, shERF) in organoids. Data are from RNA-sequencing (RNA-seq) analysis of three biological replicates for each organoid line. Only comparisons with an FDR of <0.25 are shown with the corresponding normalized enrichment score. Gene sets with a positive normalized enrichment score are enriched in organoids that have either *Foxa1* wild-type or mutant alleles. Data in **a**–**d** are mean ± s.d. Numbers of biological replicates (indicated as dots) and specific *P* values are presented in the source data. $*P < 0.05$, $**P < 0.01$, NS, not significant. All *P* values are relative to overexpression of wild-type FOXA1 unless otherwise noted; unpaired, two-tailed Student's *t*-test.

F266), some of which are predicted to be sites of direct DNA contact[10] (Fig. 1a, Extended Data Fig. 1). Another mutational hotspot is at R219, a DNA-contact site in α-helix 3, which is a highly conserved fold of the FKHD domain that sits in the major groove of target DNA (Extended Data Fig. 1). Finally, 20% of *FOXA1* mutations encode truncations just downstream of the FKHD DNA-binding domain, resulting in the loss of the C-terminal transactivating domain. Annotation of all *FOXA1* mutations in the MSK-IMPACT 504 cohort[11] revealed that Wing2-hotspot mutations—the most common subclass—are found across all disease stages, but are more prevalent in primary locoregional cases (Fig. 1c). There are only four cases of R219 mutation in *FOXA1* in this cohort but, notably, two of these had castration-resistant disease. We therefore expanded the analysis to 1,822 patients by including a larger cohort from MSK-IMPACT and a published cohort from Weill Cornell[12], which is enriched for neuroendocrine prostate cancer (NEPC), and observed significant enrichment ($P < 0.006$) of R219 mutations versus other *FOXA1* mutations in NEPC (3 out of 4) versus adenocarcinoma (8 out of 84) (Fig. 1d).

We next investigated whether *FOXA1* mutation in patients is associated with clinical outcome. In the absence of appropriate longitudinal

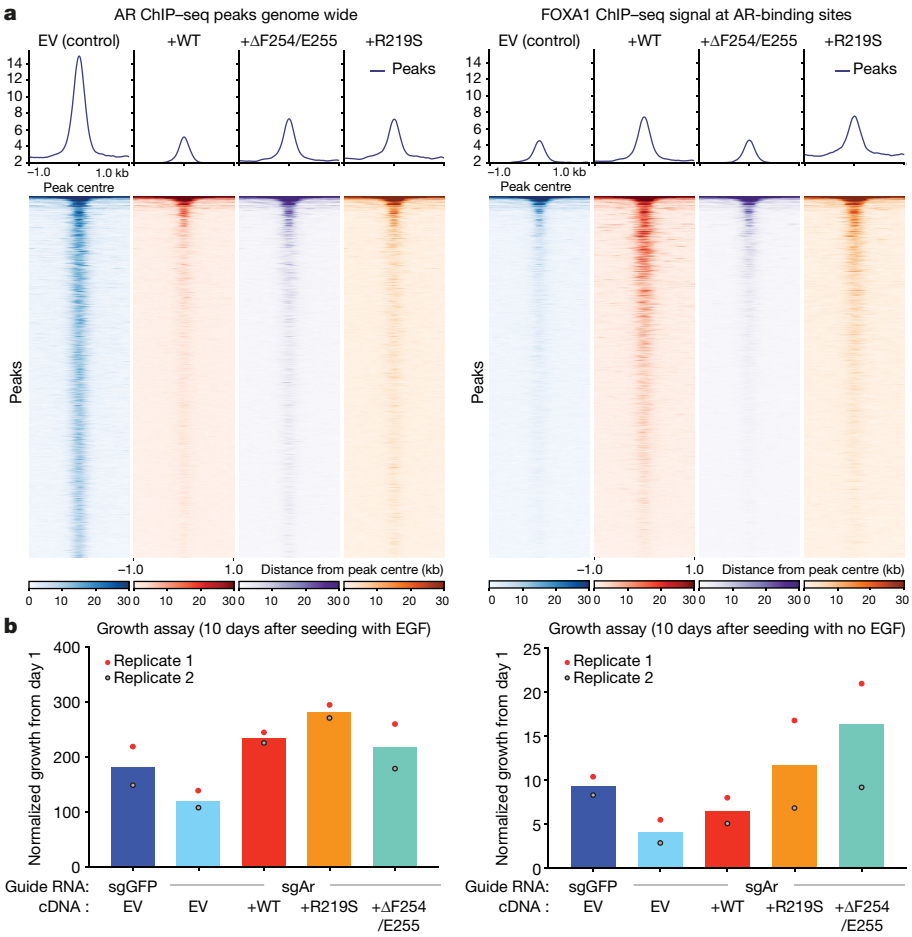


Fig. 3 | FOXA1 expression constricts the AR cistrome and promotes AR-independent growth programs. a, Left, AR ChIP–seq in organoids overexpressing wild-type or mutant FOXA1 compared to control shows significant changes in the AR cistrome in response to FOXA1 expression. Right, FOXA1 ChIP–seq showing FOXA1 binding at same loci. ChIP–seq data are from two biological replicates. Statistical analysis of peaks is shown in Extended Data Fig. 6. **b**, Overexpression of FOXA1 promotes growth in prostate organoids in the setting of significantly reduced AR (CRISPR-mediated depletion in a bulk population), in both standard medium conditions (left) and in the absence of EGF (right). Two independent experiments result in the same growth trends for biological replicates 1 and 2.

data, we generated an RNA signature using mutant *FOXA1* status of The Cancer Genome Atlas (TCGA) samples to interrogate the Decipher GRID cohort of 1,626 primary prostate cancer patients[13] and found that tumours predicted to be *FOXA1* mutant were significantly associated with higher Gleason Scores, shorter time to biochemical recurrence and more rapid progression to metastatic disease than unaltered cases (Extended Data Fig. 1b, c). Together with recent evidence[14], these data suggest that patients with *FOXA1* mutations have a less favourable prognosis.

To characterize a large panel of the most recurrent alterations seen in prostate cancer, including truncating mutations, we generated a FOXA1 reporter construct (Extended Data Fig. 2), and found that all Wing2 mutations, D226N (a mutation in spatial proximity to Wing2 in the protein[10]) and the truncation mutant G275X result in increased transcriptional activity (around twofold) compared to the wild type, whereas mutations at R219 (R219S and R219C) cause impaired activity (around 50% of wild-type activity) (Fig. 2a). To investigate the consequences of *FOXA1* mutations on the growth of prostate cells, we used primary mouse prostate organoid culture (previously used to model tumour initiation)[15] by introducing a series of wild-type or mutant mouse *Foxa1* alleles using doxycycline-inducible lentiviral constructs (Extended Data Fig. 3a–c). Increased expression of wild-type (WT) FOXA1 resulted in a 2–3-fold increase in growth compared to vector control. This relative difference was substantially greater (about 50-fold) after removal of epidermal growth factor (EGF), a critical growth factor for normal organoid proliferation (Fig. 2b). In this setting, nearly all mutants tested showed an increase in growth relative to overexpression of FOXA1(WT), including the two α-helix 3 mutants (R219S and R219C) that had reduced reporter activity, as well as the truncation mutant G275X (Fig. 2c). All 14 mutants promoted growth relative to the empty vector control.

We next examined the histological features of the resulting organoids. We observed that increased expression of FOXA1(WT), FOXA1(D226N) or the Wing2-hotspot mutations all promote exaggerated lumen formation and size (Fig. 2d, e, Extended Data Fig. 3d). By contrast, organoids expressing FOXA1(R219S), and to a lesser extent those expressing FOXA1(R219C), were unable to form measurable lumens, and the bilayer orientation of basal ($p63^+$) and luminal (androgen receptor-positive (AR^+)) cell layers appeared disrupted (Fig. 2e, Extended Data Fig. 3e). This phenotype resembles that of FOXA1-deficient organoids generated using CRISPR–Cas9 (Extended Data Fig. 4a–c), consistent with mouse models[16]. We also repeated the overexpression studies in endogenous-*Foxa1*-deleted organoids using CRISPR-resistant cDNAs encoding two pro-luminal FOXA1 mutants (ΔF254/E255 and D226N) and found that the pro-luminal phenotype was unchanged (Extended Data Fig. 4d–g). Findings from RNA sequencing were consistent with these histologies. Mutants conferring a pro-luminal phenotype showed similarity to ETS-mutant luminal organoids[17] by gene-set enrichment analysis (GSEA), with the notable exception of FOXA1(R219S), which instead showed enrichment of an epithelial–mesenchymal-transition (EMT) program and a repression of the ETS-mutant gene set (Fig. 2f), consistent with

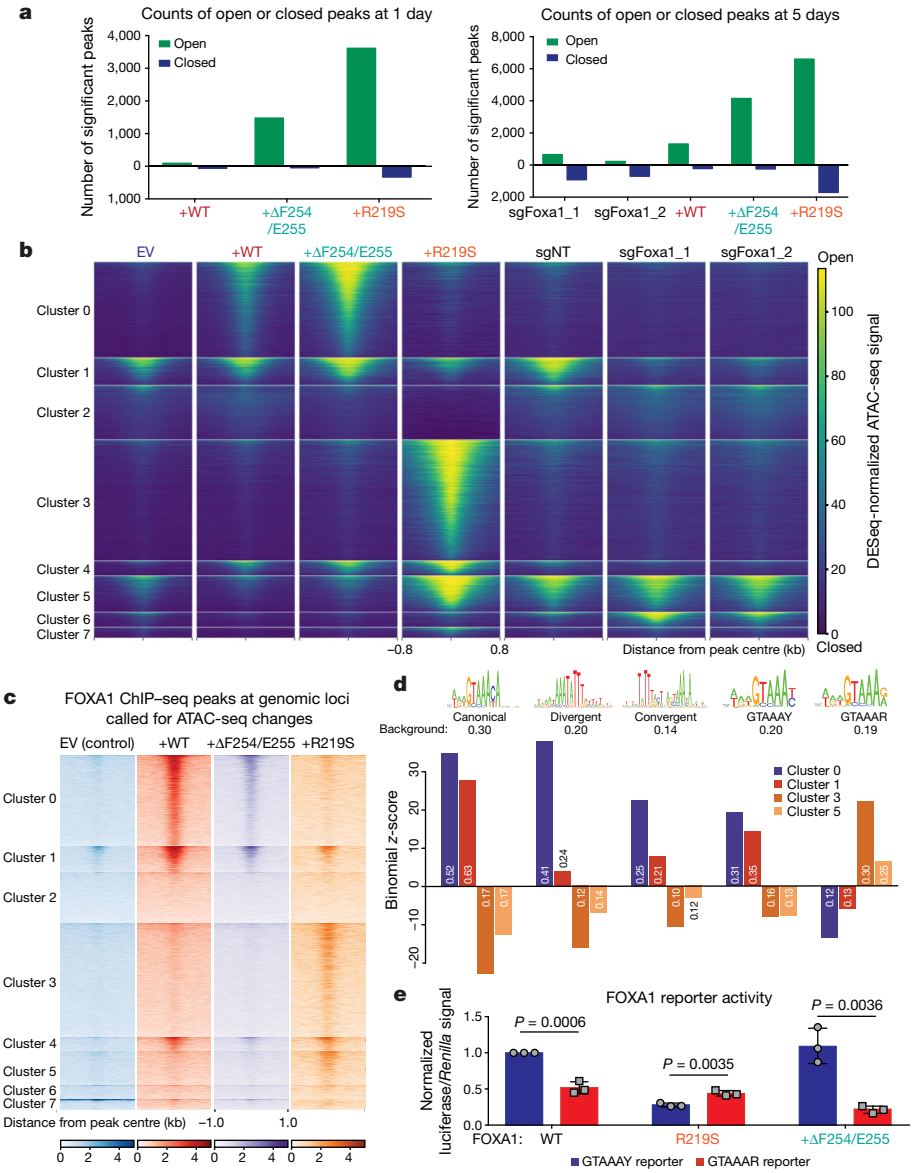


Fig. 4 | *FOXA1* mutations cause marked shifts in the chromatin landscape. a, Number of significant peaks open or closed ($\log_2$(fold change) > 2 for open peaks, $\log_2$(fold change) < -2 for closed peaks) after doxycycline treatment of organoids transfected with pCW-FOXA1 for expression of wild-type or mutant FOXA1 relative to empty vector. Right, includes counts for organoids following CRISPR-mediated *Foxa1* deletion five days after trypsinization, relative to the control sgRNA. Data are from three biological replicates, with FDR <0.05 using two-sided Wald test, with Benjamini–Hochberg FDR correction for multiple observations. **b**, ATAC-seq peak heat maps comparing organoids with (sgFoxa1_1, sgFoxa1_2) or without (sgNT) CRISPR-mediated deletion of *Foxa1* or expression of wild-type or mutant FOXA1 after five days of doxycycline treatment, with eight clusters defined by hierarchical clustering. **c**, FOXA1 ChIP–seq signal at genomic loci matching ATAC-seq clusters defined in **b** shows a similar pattern of peaks correlating FOXA1 binding with open chromatin. Data are from two biological replicates. **d**, Enrichment or depletion of FOXA1 motif variants in clusters that gain accessibility in organoids overexpressing FOXA1(ΔF254/E255) or FOXA1(R219S), including the canonical motif, divergent and convergent dimer motifs, and altered versions of the FOXA1 motif (GTAAAY, similar to canonical, and GTAAAR, non-canonical), expressed as a binomial *Z*-score computed from the number of cluster peaks with >1 motif occurrence relative to background occurrence in all heat map peaks. Occurrence within a given cluster is reported in the bar graph. Positive scores indicate enrichment; negative scores indicate depletion. **e**, Luciferase reporter assay showing activity of FOXA1 variants on GTAAAY (blue) or GTAAAR (red) DNA templates. Luciferase (ratio over *Renilla* luciferase signal, see Methods) signal normalized to signal from FOXA1(WT) on GTAAAY reporter. Data are from three biological replicates, mean ± s.d. Unpaired, two-tailed Student's *t*-test. No significant difference between activity of wild type and R219S on the GTAAAR reporter ($P = 0.2314$). FOXA1(ΔF254/E255) has significantly less activity on GTAAAR than either FOXA1(WT) ($P = 0.0059$) or FOXA1(R219S) ($P = 0.0033$).

its distinct morphology. We also examined the activity of FOXA1 in an in vivo setting[18,19] and observed increased proliferation across all lines, an increase in subcutaneous tumour size in two of four lines (FOXA1(WT) and FOXA1(G275X)), and an increased prevalence of invasive, intraductal basal disease (defined by the loss of AR expression) in tumours derived from FOXA1(R219S) organoids transduced with single-guide RNA (sgRNA) directed against PTEN—consistent with FOXA1(R219S) histology in vitro (Extended Data Fig. 4h–j).

As FOXA1 is a cofactor for AR and *FOXA1* mutant cases in the TCGA cohort have higher AR scores than either normal samples or other subtypes[6], we examined the AR cistrome. The number of AR binding peaks (defined by AR chromatin immunoprecipitation with sequencing (ChIP–seq)) is markedly reduced in organoids overexpressing wild-type or mutant FOXA1 (Fig. 3a, left, Extended Data Fig. 6a). However, FOXA1 binding is enhanced at the sites where AR binding is lost (Fig. 3a, right, $P < 1 \times 10^{-300}$, Extended Data Fig. 5a).

This suggests that FOXA1 may replace AR function at these sites, supported by retention of the increased growth advantage conferred by FOXA1 despite CRISPR-mediated deletion of *Ar* (Fig. 3b, Extended Data Fig. 5b). To reconcile the high AR scores seen in TCGA with this AR-independent growth program, we examined expression levels of the mouse orthologues of the human AR gene signature[20] and found that the majority are induced by FOXA1 (Extended Data Fig. 5c). Thus, while the number of AR-binding sites is substantially reduced, a core set of AR-target genes are maintained in the setting of increased FOXA1 activity. We also investigated whether transcriptomic changes observed in the *Foxa1* mutant mouse organoids were similar to those observed in *FOXA1* mutant human tumours. Remarkably, the human orthologues of differentially expressed genes in FOXA1(ΔF254/E255) mouse organoids were sufficient to cluster *FOXA1* mutant tumours within the TCGA cohort ($P = 2.1 \times 10^{-8}$, Extended Data Fig. 5d).

Given the role of FOXA1 as a pioneer transcription factor, we conducted a genome-wide analysis of changes in open and closed chromatin using ATAC-seq. Expression of FOXA1(WT) led to an increase in open chromatin after five days (more than 1,000 open peaks with significant change in accessibility, false discovery rate (FDR) < 0.05, log fold change of 2 in peak read coverage compared to control) whereas deletion of *Foxa1* led to the opposite, with the closing of around 1,000 peaks. Organoids expressing FOXA1(ΔF254/E255) and FOXA1(R219S) also had increased peak numbers, but these changes occurred substantially faster (in one day) and involved many more peaks (Fig. 4a), consistent with altered pioneering activity.

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

Motif analysis revealed enrichment of FOXA1-binding motifs in clusters 0 and 1 (FOXA1(WT) and FOXA1(ΔF254/E255)) (Extended Data Fig. 7a) but not in clusters 3 and 5 (FOXA1(R219S)), despite evidence of FOXA1(R219S)–DNA binding and associated gene-expression changes. However, de novo motif analysis of cluster 3 peaks identified a motif with similarities to the core GTAAA(C/T) FOXA1 binding motif but with substitution of (G/A) for (C/T) at position 6 (Extended Data Fig. 7b). This impression was confirmed by selective enrichment of the (G/A) motif in clusters 3 and 5 versus the (C/T) motif in clusters 0 and 1 (Fig. 4d). To determine whether this motif is functional, we repeated the reporter assays described in Fig. 2a with FOXA1(R219S) and found that FOXA1(R219S) preferentially activates a DNA template modified to reflect the (G/A) bias at position 6, whereas FOXA1(WT) and FOXA1(ΔF254/E255) exhibit substantially higher activity on the canonical (C/T) sequence (Fig. 4e, Extended Data Fig. 7c–e), suggesting a mechanism by which FOXA1(R219S) selectively targets novel genomic loci. Finally, two motifs recently associated with FOXA1 dimers—termed convergent and divergent[21]—were relatively enriched in cluster 0 versus cluster 1, potentially explaining the novel pioneering activity of FOXA1(ΔF254/E255) (Fig. 4d).

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

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1318-9.

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

1. Pomerantz, M. M. et al. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat. Genet.* **47**, 1346–1351 (2015).
2. Grasso, C. S. et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239–243 (2012).
3. Gerhardt, J. et al. FOXA1 promotes tumor progression in prostate cancer and represents a novel hallmark of castration-resistant prostate cancer. *Am. J. Pathol.* **180**, 848–861 (2012).
4. 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).
5. Barbieri, C. E. et al. Exome sequencing identifies recurrent *SPOP*, *FOXA1* and *MED12* mutations in prostate cancer. *Nat. Genet.* **44**, 685–689 (2012).
6. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* **163**, 1011–1025 (2015).
7. Robinson, D. et al. Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015).
8. Annala, M. et al. Frequent mutation of the *FOXA1* untranslated region in prostate cancer. *Commun. Biology* **1**, 122 (2018).
9. Wedge, D. C. et al. Sequencing of prostate cancers identifies new cancer genes, routes of progression and drug targets. *Nat. Genet.* **50**, 682–692 (2018).
10. Ciriello, G. et al. Comprehensive molecular portraits of invasive lobular breast cancer. *Cell* **163**, 506–519 (2015).
11. Abida, W. et al. Prospective genomic profiling of prostate cancer across disease states reveals germline and somatic alterations that may affect clinical decision making. *JCO Precis. Oncol.* **2017**, https://doi.org/10.1200/PO.17.00029 (2017).
12. Beltran, H. et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat. Med.* **22**, 298–305 (2016).
13. Liu, D. et al. Impact of the SPOP mutant subtype on the interpretation of clinical parameters in prostate cancer. *JCO Precis. Oncol.* **2018**, 1–13 (2018).
14. Armenia, J. et al. The long tail of oncogenic drivers in prostate cancer. *Nat. Genet.* **50**, 645–651 (2018).
15. Karthaus, W. R. et al. Identification of multipotent luminal progenitor cells in human prostate organoid cultures. *Cell* **159**, 163–175 (2014).
16. Gao, N. et al. Forkhead box A1 regulates prostate ductal morphogenesis and promotes epithelial cell maturation. *Development* **132**, 3431–3443 (2005).
17. Bose, R. et al. ERF mutations reveal a balance of ETS factors controlling prostate oncogenesis. *Nature* **546**, 671–675 (2017).
18. King, J. C. et al. Cooperativity of TMPRSS2–ERG with PI3-kinase pathway activation in prostate oncogenesis. *Nat. Genet.* **41**, 524–526 (2009).
19. Blattner, M. et al. *SPOP* mutation drives prostate tumorigenesis in vivo through coordinate regulation of PI3K/mTOR and AR signaling. *Cancer Cell* **31**, 436–451 (2017).
20. Hieronymus, H. et al. Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* **10**, 321–330 (2006).
21. Wang, X. et al. DNA-mediated dimerization on a compact sequence signature controls enhancer engagement and regulation by FOXA1. *Nucleic Acids Res.* **46**, 5470–5486 (2018).

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

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

METHODS

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

Pan-prostate cancer mutation analysis. The 12 cohorts used for analysis (total of 3,086 samples) included published datasets as well as unpublished data from the MSK-IMPACT 1708 cohort (frozen 25 May 2018), across all stages of prostate cancer (see Supplementary Table 1). Samples were compiled and duplicate samples were pruned to generate a master list of 3,086 prostate cancer cases, which were then stratified on the basis of their *FOXA1* alteration status and the class of mutation in the samples. The Wing2 hotspot includes cases with mutations or indels between H247 and E269. Truncations after the FKHD domain were defined as any frameshift alteration distal to residue E269. Any mutation that did not specifically fall into one of the distinct classes was called 'other'. Sample analysis was performed in part using the CBioPortal for Cancer Genomics[22,23].

3D modelling. Three-dimensional representation of the FKHD domain of FOXA3 complexed with DNA was generated using PyMOL (PDB: 1VTN).

Constructs. To create pCW-Flag-2A-dsRED (pCW-EV), sequences for p2A and DsRED were cloned in the pCW-Cas9 plasmid (Addgene Plasmid #50661) using in-fusion cloning (Takara Bio). To generate pCW-Flag-Foxa1-2A-dsRED (pCW-Foxa1), mouse *Foxa1* cDNA was cloned into pCW-Flag-2A-dsRED using in-fusion cloning (Takara Bio). All primers and sequences are listed in Supplementary Table 2. To generate the sgRNA vector CRISPR-Zeo, GFP from pLKO5.sgRNA.EFS.GFP (a gift from B. Ebert, Addgene plasmid no. 57822) was excised with BamHI and MluI. The zeomycin-resistance gene was removed from lenti sgRNA(MS2)_zeo backbone (a gift from F. Zhang, Addgene plasmid no. 61427) using BsrGI and EcoRI. ZeoR was ligated into the pLKO5.sgRNA.EFS backbone in a four-way ligation using BamHI–BsrGI and EcoRI–MluI adaptors. To create the LVX-UbC-EGFP-Luc2_Hygro construct for visualization of injected cells by live imaging or immunohistochemistry, we first generated the plasmid LVX-UbC-EGFP-Luc2_Puro as follows: 0.72 kb EGFP cDNA from pQCXIP-EGFP[24] was cloned into the BamHI and NotI sites of pLVX-TRE3G-IRES (Clontech, 631362) via a EcoRI–NotI cloning adaptor to make pLVX-TRE3G-EGFP-IRES. The TRE3G promoter was then removed with an XhoI and BamHI digestion, and replaced with the 1.26-kb UbC promoter obtained from Duet011 (Addgene) with a PacI and BamHI digest and using a XhoI–PacI cloning adaptor to make pLVX-UbC-EGFP-IRES. pLVX-UbC-EGFP-Luc2 was then constructed by cloning the 1.7 kb-Luc2 cDNA derived from pGL4.10(luc2) (Promega) with a HindIII and XbaI digest into the MluI and EcoRI sites of pLVX-UbC-EGFP-IRES via MluI–HindIII and XbaI–EcoRI cloning adaptors. The puromycin cassette was replaced with the hygromycin cassette to generate LVX-UbC-EGFP-Luc2_Hygro.

Generation of *FOXA1* mutant cDNA. Site directed mutagenesis was carried out on pCW-Flag-Foxa1-2A-dsRED to induce patient mutations in the cDNA using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent), according to the manufacturer's protocol. Primers were designed using Agilent's QuikChange Primer Design tool (https://www.genomics.agilent.com/primerDesignProgram.jsp). To prevent CRISPR–Cas9 targeting by sgFOXA1_1 sgRNA mutagenesis was used to introduce three silent mutations in the sgRNA recognition sequence (see Extended Data Fig. 8a).

Guide RNA design. Guide RNAs targeting mouse *Foxa1*, *Ar* and *Pten* were generated using the CRISPR design tool (http://crispr.mit.edu). sgFoxa1_1 targets the cDNA near the 5′ end, whereas sg_Foxa14 and sgFoxa1_15 target the FKHD DNA-binding domain. Control guides sgNT (targeting safe harbour locus AAVS1[25]) and sgGFP were used. All guide RNAs were cloned into lentiCRISPRv2 (Addgene no. 52961), lentiGuide-Puro (Addgene no. 52963) or CRISPR-Zeo using BsmbI digest, according to the Zhang laboratory protocol. For cells carrying CRISPR-Zeo or lentiGuide-Puro, lentiCas9-Blast (a gift from F. Zhang, Addgene plasmid no. 52962) was used as the Cas9 source.

FOXA1 luciferase reporter pGL-6xFBS-Luc. Oligonucleotide fragments containing six tandem FKHD-consensus (canonical or non-canonical) motifs with 5-bp spacers (Supplementary Table 2) were cloned into pGL4.28 luc2CP/minP/hygro (Pomega) between HindIII and XhoI restriction sites. Oligonucleotide sequences were verified using Sanger sequencing. Canonical FOXA1-binding sites were based on the top binding motifs predicted on the basis of ChIP–seq results in HepG2 cells[26], whereas the non-canonical FOXA1-binding site was based on the top hit of de novo motif analysis of ATAC-seq cluster 3 using HOMER (Extended Data Fig. 10). The pGL-5xFBS-Luc was transiently transfected using Lipofectamine 2000 (Thermo Fisher) into lentiX293T cells (Clonetech) along with CMV-Renilla (pRL-CMV *Renilla*, Promega) as an internal control. Response ratios are expressed relative to signal obtained for the positive-control wells transfected with 170 ng of pCMV6-mFOXA1mycDDK (Origene MR225487), which was set to 1, and the negative-control well receiving 170 ng of 'stuffer' DNA (pCW-Flag-2A-dsRED (pCW-EV), no exogenous FOXA1), which was set to 0. To test the response of these reporters to varying levels of FOXA1 introduced into the system, ratios of pCMV6-mFOXA1mycDDK and pCW-EV constructs were altered, keeping the total amount of DNA transfected into each well constant. In evaluating the relative response ratios between FOXA1(WT) and various mutants, one concentration of cDNA (170 ng per well) was used and relative response ratios reflect activity of given variant on the reporter. Luminescence measurements were taken 24 h after transfection. All results are means and standard deviations from experiments performed in biological triplicates (as a minimum, *n* ranging from 3 to 7) (see source data for Fig. 2 for details), and Firefly luciferase activity of individual wells was normalized against *Renilla* luciferase activity.

Organoid lines. All parental organoid lines were established in our laboratory as previously described[15]. The blue red organoids (BRO) line was established from mice harbouring red fluorescent protein (RFP) driven by a composite human keratin 18 promoter and cerulean fluorescent protein (CFP) driven by a bovine keratin 5 promoter[27]. BROs were transduced with lentiCrisprv2 carrying either sgNT or sgFoxa1_1 and selected using puromycin. BRO lines were maintained in standard mouse organoid media conditions[15]. K14-1 organoids were derived from mice harbouring an actin–GFP fusion protein driven by a human keratin 14 promoter[28]. K14-1 organoids were transduced with the allelic series of pCW-Foxa1 wild-type or mutant constructs, as well as pCW-EV as a control. Bulk cells were selected using puromycin. K14-1 organoids were maintained in standard mouse organoid media conditions[15], with 2.5 ng/ml EGF supplementation. For rescue experiments of either *Foxa1* or *Ar* deletion, K14-1 organoids carrying pCW-Foxa1 constructs were subsequently transduced with lentiCas9-Blast, bulk selected with blasticidin, and then transduced with either CRISPR-Zeo-sgFoxa1_1 or sgGFP, or sgAr and bulk selected with zeocin. Rosa26-Cas9-sgPTEN-luc2-pCW-FOXA1 organoids were derived from a homozygous Rosa26 Lox-stop-Lox Cas9 mouse (C67BL/6J background, Jackson Laboratory 026175) and transduced with adenoCre-GFP in vitro for expression of Cas9. These cells were then transduced with lentiGuide-Puro-sgPten and bulk selected with puromycin, transduced with LVX-UbC-EGFP-Luc2_Hygro and bulk selected with hygromycin, transduced with the allelic series of pCW-Foxa1 wild-type or mutant constructs or pCW-ERG, as well as pCW-EV as a control, and sorted for dsRED expression to enrich for transduced cells.

Organoid culture. Mouse organoids were sorted, cultured in 3D and transduced with lentiviruses as described previously[15,29]. Organoids infected with pCW-EV, pCW-FOXA1, or LentiCrispV2 constructs were selected with 2 μg/ml puromycin for 5 days, 3–4 days after transduction, while those infected with CRISPR-Zeo were selected with 30 μg/ml for 7 days, 3–4 days after transduction. Transduction with Lenti-Cas9-Blast was followed by five days of selection in 10 μg/ml blasticidin. Preparation of 3D organoids for histology was carried out as previously described[15]. Haematoxylin and eosin staining and immunohistochemistry was carried out by the MSKCC Molecular Cytology Core. Cells were confirmed to be free of mycoplasma using the Lonza MycoAlert Mycoplasma Detection Kit (LT07-318).

Growth assays. Organoids were treated with doxycycline (500 ng/ml) to induce expression of the FOXA1-2A-DsRED fusion, then sorted two days later to enrich for $DsRED^+$ cells. Cells were seeded at a density of 100 cells per μl (2,000 cells per 20-μl dome, three domes per line per time point, each dome in a single well of a 48-well plate) and maintained on doxycycline for the duration of the assay, refreshing media every 2–3 days. Y-27632 was supplemented for the first feeding at 10 μM. To measure proliferation, matrigel domes were washed with PBS, and then resuspended in 100 μl of PBS, and CellTiter-Glo 2.0 Assay (Promega) was used, following the manufacturer's instructions. Triplicate values for each time point were averaged, and all values on subsequent days were normalized to the day 1 reading. Experiments were repeated at least three times independently and each line was normalized to the empty vector control readings for a given replicate.

Lumen formation assays. Organoids were treated with doxycycline (500 ng/ml) to induce expression of the FOXA1-2A-DsRED fusion. Doxycycline-treated cells were sorted two days later to enrich for $DsRED^+$ cells. Sorted cells were seeded in matrigel at a density of 3 cells per μl (eight 25-μl domes per organoid line) and maintained on doxycycline for the duration of the assay, with fresh medium every 2–3 days. Y-27632 was supplemented for the first feeding at 10 μM. After 10 days, organoids were scored for the presence or absence of a visible lumen by bright-field microscopy, and the percentage of the total number of organoids that possessed a lumen was determined from examining ~50–200 organoids in a typical experiment. In CRISPR organoid lines, sorting was not performed for the lumen formation assay. Instead cells were trypsinized to a single-cell suspension, counted using trypan blue exclusion, and then seeded as described above. Experiments were repeated three times independently.

Lumen area measurements. Organoids were treated with doxycycline (500 ng/ml) to induce expression of the FOXA1-2A-DsRED fusion. Doxycycline-treated cells were sorted two days later to enrich for $DsRED^+$ cells. Sorted cells were seeded in matrigel in a dilution series of densities ranging from 32 cells per μl down to 4 cells per μl (five domes per density per line) and maintained on doxycycline for the duration of the assay, with the medium refreshed every 2–3 days. Y-27632 was

supplemented for the first feeding at 10 μM. After 10 days, the area of each visible lumen was measured using light microscopy and Nikon NIS Elements software. In a typical experiment, ~30–50 organoids were measured by an observer blinded to organoid genotypes.

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

Immunohistochemistry. Organoids and tumours were processed and stained as described previously[15]. The following antibodies were used for staining on mouse organoids and organoid-derived xenografts: HNF-3-α/FOXA1 antibody (5 μg/ml, 3B3NB, Novus Biologicals), AR (1:1,000, N-20, Santa Cruz), p63 (1:800, 4A4, Ventana). Staining was visualized with BrightVision (Dako), Ki67 (Abcam ab15580 at 1 μg/ml).

In vivo experiments. In vivo xenograft experiments were performed by subcutaneous injection of 2×10^6 dissociated organoid cells (Rosa26-Cas9-sgPTEN-luc2-pCW-FOXA1 or ERG) resuspended in 100 μl of 50% matrigel (BD Biosciences) and 50% growth medium into the flanks of five 8-to-12-week-old male NOD. Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ mice (005557, The Jackson Laboratory) to yield ten tumours per group. As soon as palpable, tumour volume was measured weekly using the tumour-measuring system Peira TM900 (Peira). Tumours were then collected at given time points for histology using 4% paraformaldehyde. Histological assessment was carried out by an observer blinded to tumour genotypes. All animal experiments were performed in compliance with the guidelines of the Research Animal Resource Center of the Memorial Sloan Kettering Cancer Center. In accordance with our IACUC and our approved protocol, none of the mice exceeded the maximum tumour burden allowed (total for both sides) of 2,000 mm^3.

RNA isolation and sequencing. RNA was extracted from organoids using an RNeasy Kit (Qiagen). Freshly sorted dsRED$^+$ cells were seeded in triplicate per infected construct at the start of the assay, and moving forward, replicates were processed independently, collected at the appropriate time points. Library preparation and sequencing were performed by the New York Genome Center, where RNA-seq libraries were prepared using the TruSeq Stranded mRNA Library Preparation Kit (Illumina) in accordance with the manufacturer's instructions. In brief, 100 ng of total RNA was used for purification and fragmentation of mRNA. Purified mRNA underwent first and second strand cDNA synthesis. cDNA was then adenylated, ligated to Illumina sequencing adapters, and amplified by PCR (using 10 cycles). Final libraries were evaluated using fluorescent-based assays including PicoGreen (Life Technologies) or Qubit Fluorometer (Invitrogen) and Fragment Analyzer (Advanced Analytics) or BioAnalyzer (Agilent 2100), and were sequenced on an Illumina HiSeq2500 sequencer (v.4 chemistry, v.2 chemistry for Rapid Run) using 2×50-bp cycles. Reads were aligned to the mm10 mouse reference genome using STARaligner[30] (v.2.4.2a). Quantification of genes annotated in Gencode vM2 was performed using featureCounts (v.1.4.3) and transcripts were quantified using Kalisto[31]. Quality control statistics were collected with Picard (v.1.83) and RSeQC[32] (http://broadinstitute.github.io/picard/). Normalization of feature counts was done using the DESeq2 package[33].

Analysis of RNA-seq from mouse organoids and patient samples. The gene read count data of TCGA primary prostate cancer were downloaded using the GDC tool. The mouse and human homologous genes were downloaded from Mouse Genome Informatics from The Jackson Laboratory (http://www.informatics.jax.org/homology.shtml). Differential expression analyses were performed using DESeq2 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html) based on the gene read count data. Multiple-hypothesis testing was considered by using Benjamini–Hochberg correction. The statistical significance of the overlap between two groups of genes was tested using Fisher's exact test. GSEA was performed using the JAVA program (http://www.broadinstitute.org/gsea) and run in pre-ranked mode to identify enriched signatures. The GSEA plot, normalized enrichment score and FDR and q values were derived from GSEA output. The following gene sets were used: Hallmark gene sets, Neuroendocrine high[12], Basal low[34], and shERF up[17].

Prostate cancer tumour samples and microarray data. A total of 1,959 radical prostatectomy tumour expression profiles were used for training and testing. For training and testing, we used RNA-seq expression and DNA mutation data from TCGA prostate cancer project[6] ($n = 333$). For testing, the expression profiles of retrospective ($n = 1{,}626$) were derived from the Decipher GRID registry (NCT02609269). The retrospective GRID cohort was pooled from seven published microarray studies: Cleveland Clinic[35] (CCF), Erasmus MC[36], Johns Hopkins[37] (JHMI), Memorial Sloan Kettering[38] (MSKCC), Mayo Clinic[39,40] (Mayo I and Mayo II), and Thomas Jefferson University[41] (TJU). Associated accession numbers are: GSE79957, GSE72291, GSE62667, GSE62116, GSE46691, GSE41408, and GSE21032. DNA and RNA from the TCGA cohort were extracted from fresh frozen radical prostatectomy tumour tissue, as previously described[6]. RNA from the GRID cohorts was extracted from routine formalin-fixed, paraffin embedded (FFPE) radical prostatectomy tumour tissues, amplified and hybridized to Human Exon 1.0 ST microarrays (Thermo Fisher).

***FOXA1* mutant transcriptional signature.** By following the similar strategy as previously reported for *SPOP* mutants[13], we developed the *FOXA1* mutant transcriptional signature that includes 67 genes differentially expressed between *FOXA1* mutant and wild-type samples from TCGA prostate cancer RNA-seq data. The low-expressed genes (mean RNA-seq by expectation maximization (RSEM) <1) were filtered before the analysis. Specifically, we identified significantly differentially expressed genes by comparing *FOXA1* mutants within FKHD DNA-binding domain and wild-type cases as determined from DNA mutational analyses among TCGA samples lacking *ETS* family gene fusions (*ERG*, *ETV1*, *ETV4* and *FLI1*), using Wilcoxon rank-sum test and controlled for false discovery using Benjamini–Hochberg adjustment (FDR ≤ 0.05).

SCaPT development based on *FOXA1* mutant transcriptional signature and SVM model. To predict tumours in the *FOXA1* mutant subclass in the absence of DNA sequencing data (that is, microarray datasets), we developed the subclass predictor based on transcriptional data (SCaPT) model based on the support vector machine (SVM) model. Given a set of training data marked with two categories, SVM builds a model that assigns testing data into one category or the other, making it a non-probabilistic binary linear classifier. In our SCaPT model, the training data were defined as the transcriptional scores of *FOXA1* mutant signature from TCGA cohort. The testing data would be the transcriptional z scores from RNA-seq or microarray expression data of *FOXA1* mutant signature.

Prostate cancer molecular subclass prediction by decision tree. In each individual study of retrospective and prospective GRID cohorts, the *FOXA1* mutant subclass was first predicted using the SCaPT model. Next, using a decision tree and previously developed microarray-based classifiers for the ERG$^+$ and ETS$^+$ subtypes, we classified the remaining cases in each cohort. Some cases with both predicted *FOXA1* mutant and ERG$^+$ETS$^+$ status were classified as conflict subclass, and the rest without *FOXA1* mutant calling and outlier expression were considered as 'other' subclass.

Statistical analysis of human data. Statistical analyses were performed in R v.3.4.0 (R Foundation). All statistical tests were two-sided with a significance level of $P < 0.05$. Univariate logistic regression analyses were performed on the combined cohort to test the statistical association between *FOXA1* mutant status and clinical variables, including age, race, pre-operative prostate-specific antigen (PSA), Gleason score, lymph node invasion (LNI), surgical margin status (SMS), extracapsular extension (ECE) and seminal vesicle invasion (SVI). We evaluated the associations between *FOXA1* mutant status and patient outcomes including biochemical recurrence (BCR), metastasis (MET) and prostate cancer specific mortality (PCSM) on the basis of Kaplan–Meier analysis.

ATAC-seq. Freshly sorted cells carrying pCW constructs (dsRED$^+$) were seeded in triplicate per infected construct at the start of the assay, and moving forward, replicates were processed independently, collected at the appropriate time points. CRISPR cell lines carried LentiCRISPRv2 with either the control guide (sgNT), guide 14 for FOXA1 (sgFOXA1_1) or guide 15 for FOXA1 (sgFOXA1_2). At time of collection, cells were trypsinized, and 50,000 cells (counted by using trypan blue exclusion) were processed for ATAC-seq as follows. After a wash step in cold cell wash buffer (CWB; 10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$), outer membranes were disrupted in lysis buffer (CWB + 0.1% NP40) for 2 min on ice. The lysis reaction was stopped with the addition of 1 ml CWB. After a centrifugation step at 1,500*g* for 10 min, pelleted nuclei were kept for the next step. In a 50-μl final volume, tagmentation was performed for 30 min at 37 °C, using the Nextera DNA library prep kit (Illumina FC-121-1030). After addition of SDS to 0.2% final concentration, DNA was purified on AMPure XP beads (Beckman Coulter A63881) using a 2:1 ratio (v/v) of beads:tagmented DNA. Freshly eluted DNA was barcoded and amplified in 110-μl PCR volume (NEB Next Q5 Hot Start HiFi PCR, M0543L) to generate a library with the following PCR program: 65 °C, 5 min; 98 °C, 30 s; 11 cycles of (98 °C, 10 s ramping to 65 °C, 30 s), 4 °C hold. Quality control of the libraries was performed with a Bioanalyzer 2200 (Agilent Technologies, D1000 screentapes and reagents, 5067-5582) to assess size range of amplified DNA fragments and with Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher P11496) to quantify the DNA fragments generated. ATAC libraries were then pooled at equimolar concentration and were sequenced multiplexed on the Illumina HiSeq with 50-bp paired-ends.

ATAC-seq data and preprocessing. ATAC-seq data preprocessing was performed as previously described. Raw ATAC-seq reads were trimmed and filtered for quality using Trim Galore! v.0.4.5 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) powered by CutAdapt v.1.16 (https://doi-org.proxy.library.cornell.edu/10.14806/ej.17.1.200) and FastQC v.0.11.7 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Paired end reads were aligned to the mm10 genome using Bowtie2 v.2.3.4.1 in very sensitive local mode (-q –local –very-sensitive-local –no-discordant –no-mixed –dovetail -I 10 X 20), and paired reads that mapped

to different chromosomes or that mapped too far away were discarded. Unpaired reads, discordant reads, reads with mapQ <20, or SAM flags 0x4 and 0x400, as well as reads marked as optical or PCR duplicates using picard MarkDuplicates v.2.18.3-SNAPSHOT and reads overlapping the ENCODE mm10 functional genomics regions blacklist (at mitra.stanford.edu/kundaje/akundaje/release/blacklists/mm10-mouse/mm10.blacklist.bed.gz) were removed to improve the quality of the retained fragments. To correct for the fact that the Tn5 transposase binds as a dimer and inserts two adapters in the Tn5 tagmentation step, all positive-strand reads were shifted 4 bp downstream and all negative-strand reads were shifted 5 bp upstream to centre the reads on the transposase-binding event.

Overall mapping statistics confirmed high-quality ATAC-seq data, with a high alignment rate (over 76.8% in all samples) and high coverage (over 30 million aligned read pairs per sample) across experiments (Supplementary Table 13). As an additional quality control metric, we confirmed that all ATAC-seq libraries displayed the expected insert size distribution computed from aligned read pairs, with nucleosome-free, mono-nucleosomal, and di-nucleosomal modes (see Extended Data Fig. 8a for representative plots).

ATAC peak calling, reproducibility analysis and atlas creation. We then pooled the shifted reads by sample and identified peaks using MACS2 with a threshold of FDR-corrected $P < 0.01$ using the Benjamini–Hochberg procedure for multiple-hypothesis correction. As called peaks may be caused by noise in the assay and not reflect true chromatin accessibility, we calculated an irreproducible discovery rate (IDR) for all pairs of replicates across a cell type. Given two ranked lists of events from replicate experiments, in this case peak calls ranked by *P* value, IDR estimates a threshold at which events are no longer reproducible. Using this measure, we excluded peaks that were not reproducible (IDR <0.005) in at least one pair of replicates for at least one cell type or time point.

Reproducible peaks from each cell type were combined to create a genome-wide atlas of accessible chromatin regions. Reproducible peaks from different samples were merged if they overlapped by more than 75%. This produced an atlas of ~182,800 reproducible peaks of median width 586 bp. The numbers of reproducible peaks per time point and organoid line are provided in Supplementary Table 14. Track diagrams at specific loci visually confirm that replicate ATAC-seq experiments show reproducible accessible sites (Extended Data Fig. 8b).

Assignment of ATAC-seq peaks to genes. The RefSeq transcript annotations of the mm10 mouse genome were used to define the genomic location of transcription units. For genes with multiple gene models, the longest transcription unit was used for the gene locus definition. ATAC-seq peaks located in the body of the transcription unit, together with the 2-kb regions upstream of the transcription start site and downstream of the 3′ end, were assigned to the gene. If a peak was found in the overlap of the transcription units of two genes, one of the genes was chosen arbitrarily. Intergenic peaks were assigned to the gene with a transcription start site or 3′ end that was closest to the peak. In this way, each peak was unambiguously assigned to one gene. Peaks were annotated as promoter peaks if they were within 2 kb of a transcription start site. Non-promoter peaks were annotated as intergenic, intronic or exonic according to the relevant RefSeq transcript annotation. The atlas-wide distribution of promoter, intergenic or exonic peak assignment was consistent with high-quality ATAC-seq datasets (Extended Data Fig. 9), with 31.6% of peaks at promoters and the rest nearly equally divided between intergenic and intronic regions, with a small fraction annotated as exonic.

Differential peak accessibility. Reads aligning to the atlas peak regions were counted using htseq-count (-r pos s no). Differential accessibility of the peaks was assessed by applying DESeq2 v.1.18.1 to this count table, considering all pairwise comparisons of cell types. Peaks were defined as differentially accessible if they satisfied an FDR-corrected $P < 0.05$ and if the magnitude of the DESeq-normalized counts changed by a stringent factor of 4 or more between at least one pairwise comparison of organoid line to control (the comparisons used were EV day 1 vs FE255 day 1, EV day 1 vs R219 day 1, EV day 1 vs WT day 1, EV day 5 vs FE255 day 5, EV day 5 vs R219 day 5, EV day 5 vs WT day 5, WT day 1 vs FE255 day 1, WT day 1 vs R219 day 1, EV day 5 vs FE255 day 5, WT day 5 vs R219 day 5, sgNT day 5 vs sgFOXA1-sg1 day 5, and sgNT day 5 vs sgFOXA1-sg2 day 5) (two-sided Wald test, with Benjamini–Hochberg correction for multiple observations). MA plots for pairwise differential accessibility analyses confirmed that normalization was appropriate and that differential peaks displayed robust changes (see Extended Data Fig. 10 for representative plots and Supplementary Table 16 for numbers of differentially accessible peaks). These analyses produced a set of ~20,500 differentially accessible peaks of median width 410 bp; as expected, differential peaks were enriched for intergenic and intronic annotations and depleted for promoter annotations (Extended Data Fig. 9).

ATAC-seq peak clustering. The ATAC-seq peak heat maps were created using the DESeq size-factor normalized read counts, applying the variance-stabilizing transformation to the full peak atlas, selecting the differentially accessible peaks, and then clustering using hierarchical clustering with the ward.D distance metric. Clusters were defined by cutting the hierarchical clustering at the first 8 bifurcations of the dendrogram by ward.D distance. The number of clusters was chosen to be eight based on observation of biologically interesting patterns of accessibility, and then peaks were sorted within each cluster by maximum signal.

Peak heat maps. Heat maps (tornado plots) of peaks were generated by combining signals across replicates and binning the region ±750 bp around the peak summit in 1-bp bins after adjusting the reads for Tn5-induced bias, resulting in one signal track for each cell type or time point. Heat maps were generated using deeptools v.3.0.2.

De novo transcription factor motif analysis. The Homer v.4.10 utility findMotifsGenome.pl was used to identify the top ten transcription factor motifs enriched in each of the clusters produced by deeptools from each time point relative to genomic background. The top motifs were reported and compared to the Homer database of known motifs and then manually curated to restrict to transcription factors that are expressed based on RNA-seq data and to group similar motifs from transcription factors belonging to the same family.

FIMO motif search. Motif enrichment was performed relative to the 8 clusters defined by hierarchical clustering of 20,523 differentially accessible peaks (described above). Each ATAC-seq peak in the atlas was scanned for 718 transcription factor motifs in the *Mus musculus* CIS-BP database[42] using FIMO[43] of MEME suite[44], using the default *P* value cut-off of 1×10^{-4}. The background sequence distribution for motif analysis was based on nucleotide frequencies in the full set of 20,523 differentially accessible peaks (A = T = 0.2711, C = G = 0.2289). Of the 718 motifs in the database, 713 had a match within at least one peak among the differentially accessible peaks.

FIMO motif analysis. We restricted the analysis to 298 transcription factors that had a median RNA-seq expression across biological replicates of above 5 reads per kilobase of transcript per million mapped reads in at least one organoid line or time point. In addition, CTCF and CTCFL, DNA-binding proteins associated with 3D chromatin structure, were excluded. To rank the level of enrichment of transcription factor motifs in each cluster relative to the background, the number of peaks containing each motif was calculated for each cluster and for the full set of differentially accessible peaks. Enrichment–depletion scores for each motif in a cluster were reported as binomial *Z*-scores relative to the background of motif occurrences in the set of differential ATAC-seq peaks. Namely, if p represents the probability that a peak in the background set contains an occurrence of the motif, then the binomial *Z*-score for a cluster of size N with C peaks containing the motif is $\frac{C - Np}{\sqrt{Np(1-p)}}$. While these *Z*-scores do not incorporate a correction for multiple hypotheses, in practice the top-ranked motifs have such strong enrichments that they would still be highly significant after correction.

Non-canonical FOXA1 motif analysis. To examine enrichment–depletion of non-canonical FOXA1 motifs, we considered four additional motifs. First, we examined previously reported convergent and divergent FOXA1 dimer motifs. Second, we altered the canonical FOXA1 motif by replacing position 6 of the core GTAAAC/T pattern with either and equal probability of C/T (similar to canonical) or an equal probability of A/G (non-canonical). We used FIMO to search for hits of these motifs across differential peaks and reported enrichment–depletion within clusters as binomial *Z*-scores as before.

ChIP–seq. Freshly sorted cells carrying pCW constructs (dsRED$^+$) were seeded in duplicate per infected construct at the start of the assay, and moving forward, replicates were processed independently, collected following five days of doxycycline treatment. At time of collection, cells were trypsinized and 70,000 cells (counted by using trypan blue exclusion) were processed for ChIP–seq as follows. Cells were fixed with formaldehyde (1%) and the reaction was quenched with glycine 1.25 M and Tris 1 M pH 8. Fixed cells were lysed with SDS lysis solution containing protease inhibitors. Re-suspended pellets were sonicated, precipitated with antibodies (HNF-3 alpha/FoxA1 antibody (3B3NB) (Novus Biologicals), AR (ER179(2), Abcam) and protein A/G bead complex. The chromatin and immune complex were sequentially washed with a low-salt solution, high-salt solution, LiCl solution and Tris-NaCl solution. Chromatin was eluted from the complex with a solution containing 1% of SDS and 0.1 M $NaHCO_3$. Cross-linking between DNA and protein was reversed by adding NaCl solution and incubating at 65 °C overnight. Libraries were made using NEBNext Ultra II DNA library prep kit for Illumina (NEB E7645L). Quality control was performed with Bioanalyzer 2200 (Agilent Technologies, D1000 screentapes and reagents, 5067-5582) to assess size range of amplified DNA fragments, and with Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher P11496) to quantify the DNA fragments generated. ChIP libraries were then pooled at equimolar concentration and were sequenced multiplexed on the Illumina HiSeq with 50-bp paired-end sequencing.

Bioinformatic analysis of ChIP–seq. Raw reads were first trimmed with Trimmomatic[45] (v.0.35, options: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) to remove adapters and low-quality sequences. They were then aligned with Bowtie2[46] (v.2.2.6, options:–local–mm–no-mixed–no-discordant) using mm10 genome. After alignment, PCR duplicates were removed with Picard

tools (http://broadinstitute.github.io/picard/) (MarkDuplicates v.2.9.0) and peaks were called individually for each replicate with MACS2[47] (v.2.1.0.20151222,–options: keep-dup 1 -g mm -p 0.05). These called peaks between replicates were then used with IDR[48] (v.2.0.2) framework to identify reproducible peaks. Deeptools (v.3.1.3) was used for visualization and HOMER (v.4.10.3) was used for discovering de novo motifs.

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

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

Data availability

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

22. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).
23. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
24. Watson, P. A. et al. Constitutively active androgen receptor splice variants expressed in castration-resistant prostate cancer require full-length androgen receptor. *Proc. Natl Acad. Sci. USA* **107**, 16759–16765 (2010).
25. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR–Cas9 system. *Science* **343**, 80–84 (2014).
26. Motallebipour, M. et al. Differential binding and co-binding pattern of FOXA1 and FOXA3 and their relation to H3K4me3 in HepG2 cells revealed by ChIP–seq. *Genome Biol.* **10**, R129 (2009).
27. Peng, W., Bao, Y. & Sawicki, J. A. Epithelial cell-targeted transgene expression enables isolation of cyan fluorescent protein (CFP)-expressing prostate stem/progenitor cells. *Transgenic Res.* **20**, 1073–1086 (2011).
28. Vaezi, A., Bauer, C., Vasioukhin, V. & Fuchs, E. Actin cable dynamics and Rho/Rock orchestrate a polarized cytoskeletal architecture in the early steps of assembling a stratified epithelium. *Dev. Cell* **3**, 367–381 (2002).
29. Koo, B.-K. et al. Controlled gene expression in primary *Lgr5* organoid cultures. *Nat. Methods* **9**, 81–83 (2011).
30. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
31. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
32. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
33. Love, M. I., Huber, W., & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
34. Smith, B. A. et al. A basal stem cell signature identifies aggressive prostate cancer phenotypes. *Proc. Natl Acad. Sci. USA* **112**, E6544–E6552 (2015).
35. Klein, E. A. et al. A genomic classifier improves prediction of metastatic disease within 5 years after surgery in node-negative high-risk prostate cancer patients managed by radical prostatectomy without adjuvant therapy. *Eur. Urol.* **67**, 778–786 (2015).
36. Boormans, J. L. et al. Identification of *TDRD1* as a direct target gene of *ERG* in primary prostate cancer. *Int. J. Cancer* **133**, 335–345 (2013).
37. Ross, A. E. et al. Tissue-based genomics augments post-prostatectomy risk stratification in a natural history cohort of intermediate- and high-risk men. *Eur. Urol.* **69**, 157–165 (2016).
38. Taylor, B. S. et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11–22 (2010).
39. Erho, N. et al. Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. *PLoS ONE* **8**, e66855 (2013).
40. Karnes, R. J. et al. Validation of a genomic classifier that predicts metastasis following radical prostatectomy in an at risk patient population. *J. Urol.* **190**, 2047–2053 (2013).
41. Den, R. B. et al. Genomic prostate cancer classifier predicts biochemical failure and metastases in patients after postoperative radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.* **89**, 1038–1046 (2014).
42. Weirauch, M. T. et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
43. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
44. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
45. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
46. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
47. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP–seq enrichment using MACS. *Nat. Protoc.* **7**, 1728–1740 (2012).
48. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).

Acknowledgements We thank P. Iaquinta, B. Carver, Z. Cao, I. Ostrovnaya, H. Hieronymous, W. Abida, E. Wasmuth, K. Lawrence, T. Nadkarni, S. P. Gao and all members of the Sawyers laboratory for comments; Memorial Sloan Kettering Cancer Center core facilities; N. Fan and D. Yarilin from the MSKCC Molecular Cytology Core Facility; the MSKCC Integrated Genomics Operation; the New York Genome Center for RNA sequencing; and R. J. Karnes (Department of Urology, Mayo Clinic), R. B. Den (Department of Radiation Oncology, Thomas Jefferson University), E. A. Klein (Glickman Urological and Kidney Institute, Cleveland Clinic) and Bruce Trock (Department of Urology, Johns Hopkins University) for providing access to patient outcome data. Some of the results shown here are in part based on data generated by the TCGA Research Network (https://www.cancer.gov/tcga). E.J.A. was supported by an American Association for Cancer Research Basic Cancer Research Fellowship, the MSKCC Translational Research Oncology Training Program and the MSKCC Functional Genomics Initiative. R.D. was supported by NIH training grant 1T32GM083937. Z.Z. is supported by the NCI Predoctoral to Postdoctoral Fellow Transition Award (F99/K00 award ID: F99CA223063). R.B. was supported by grants from the Department of Defense (W81XWH1510277), NCI (1K08CA226348-01) and the Prostate Cancer Foundation. D.L., A.S. and C.E.B. were supported by the NCI (K08CA187417-01, R01CA215040-01 and P50CA211024-01 to C.E.B.), a Urology Care Foundation Rising Star in Urology Research Award (C.E.B.), a Damon Runyon Cancer Research Foundation MetLife Foundation Family Clinical Investigator Award (C.E.B.) and the Prostate Cancer Foundation (C.E.B). C.L.S. is an investigator of the Howard Hughes Medical Institute and this project was supported by NIH grants CA155169, CA193837, CA224079, CA092629, CA160001 and CA008748, the Starr Cancer Consortium grant I10-0062 and the Functional Genomics Initiative at MSKCC.

Author contributions E.J.A. and C.L.S. conceived and oversaw the project, performed data interpretation, and co-wrote the manuscript. E.J.A. and E.H. performed immunoblots, in vitro cell growth assays, lumen formation assays, lumen area quantification, processed organoids for immunohistochemistry and prepared experiments for RNA-seq and ATAC-seq. E.J.A., E.H. and W.R.K. made 3D organoid lines. E.J.A., W.R.K., E.H. and P.A.W. cloned plasmid reagents. E.J.A., E.H., W.R.K. and Z.Z. carried out in vivo experiments. E.J.A., R.B. and D.L. performed RNA-seq analysis and GSEA. E.J.A., R.B., D.L., A.S., Y.L., E.D. and C.E.B. performed analysis of human prostate cancer cohorts. A.G. optimized and carried out ATAC and ChIP protocols. R.D., S.C., H.C. and C.S.L. carried out ATAC-seq and ChIP–seq data analysis. All authors made intellectual contributions and reviewed the manuscript.

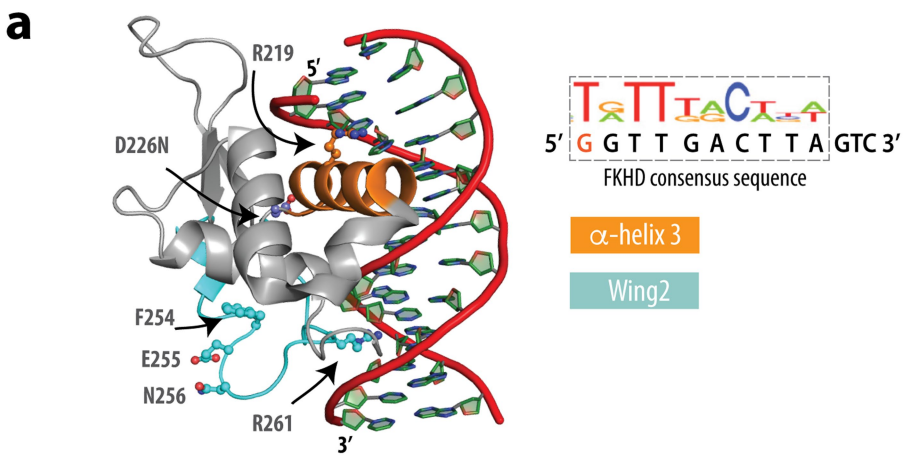
Competing interests C.L.S. serves on the board of directors of Novartis, is a co-founder of ORIC Pharm and co-inventor of enzalutamide and apalutamide. He is a science advisor to Agios, Beigene, Blueprint, Column Group, Foghorn, Housey Pharma, Nextech, KSQ, Petra and PMV. He was a co-founder of Seragon, purchased by Genentech/Roche in 2014. The other authors declare no competing interests.

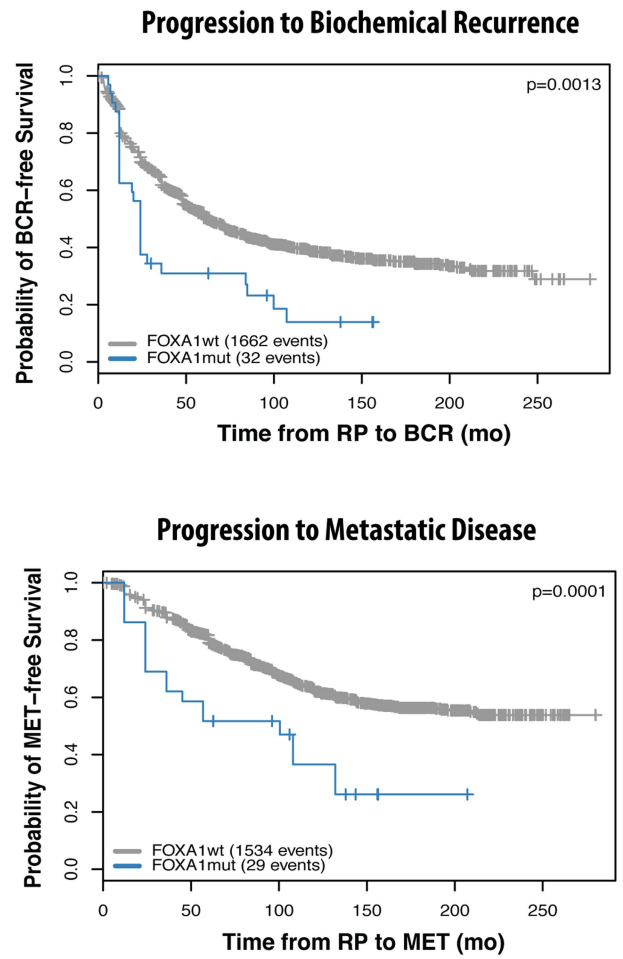
Additional information

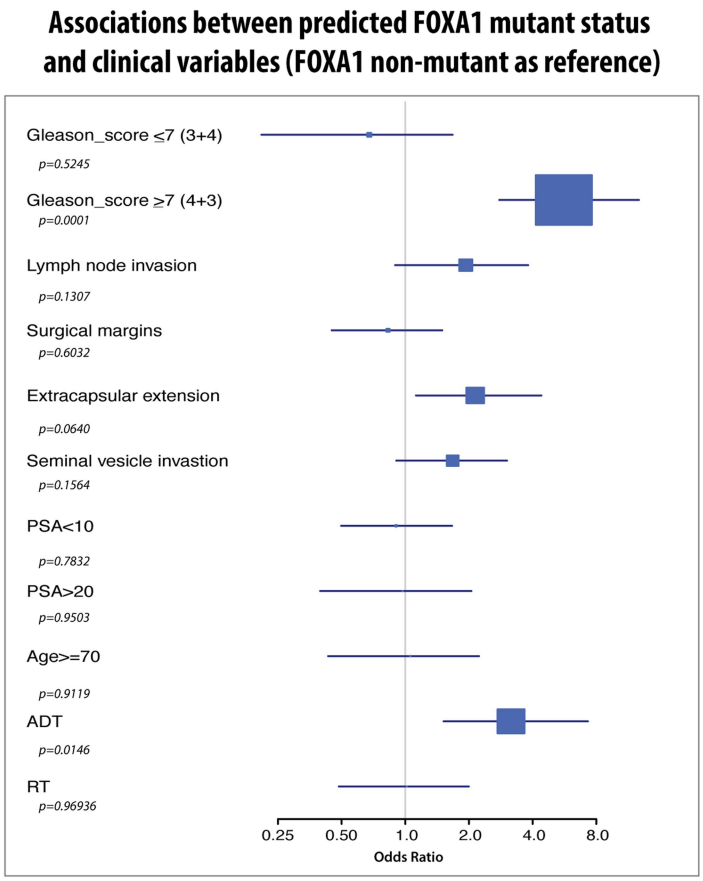
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1318-9.

Correspondence and requests for materials should be addressed to C.L.S.

Reprints and permissions information is available at http://www.nature.com/reprints.

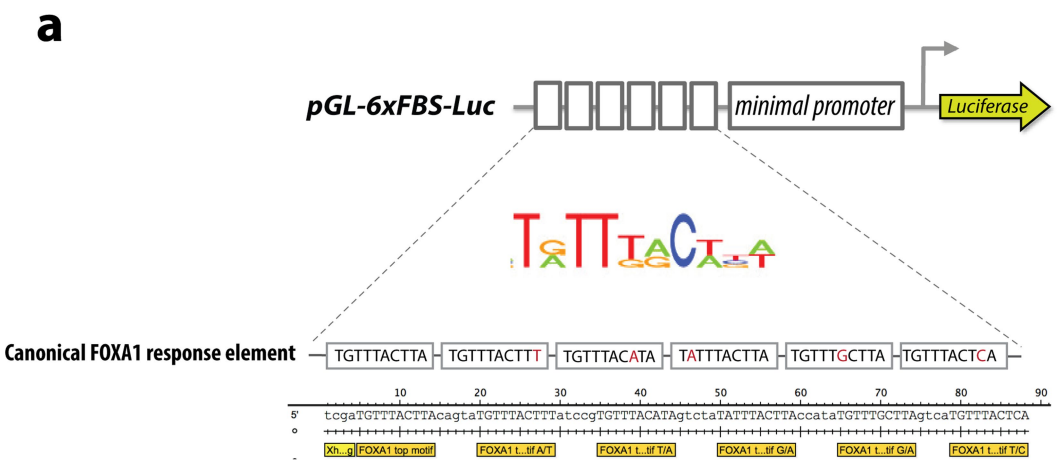


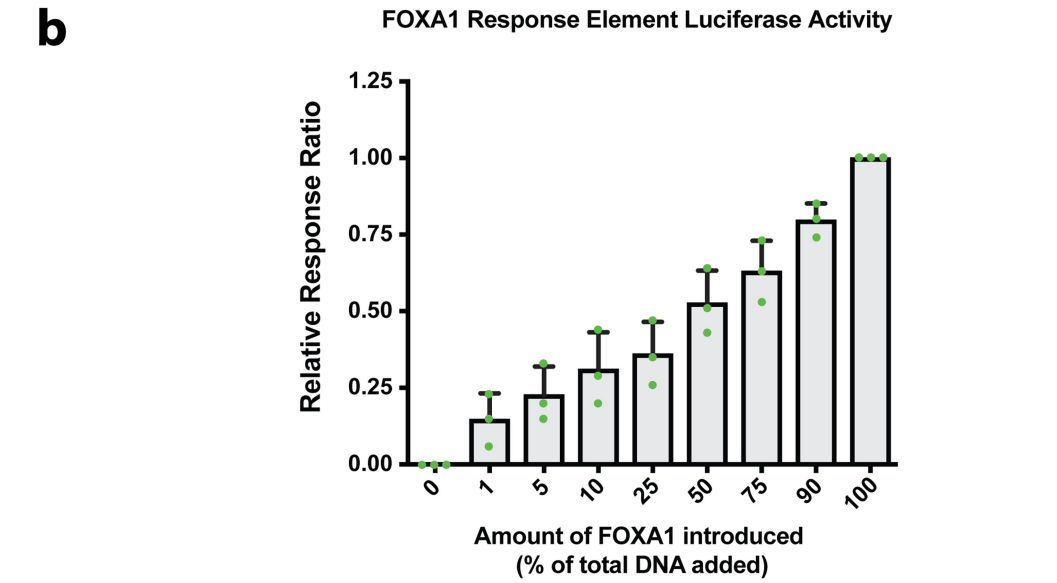


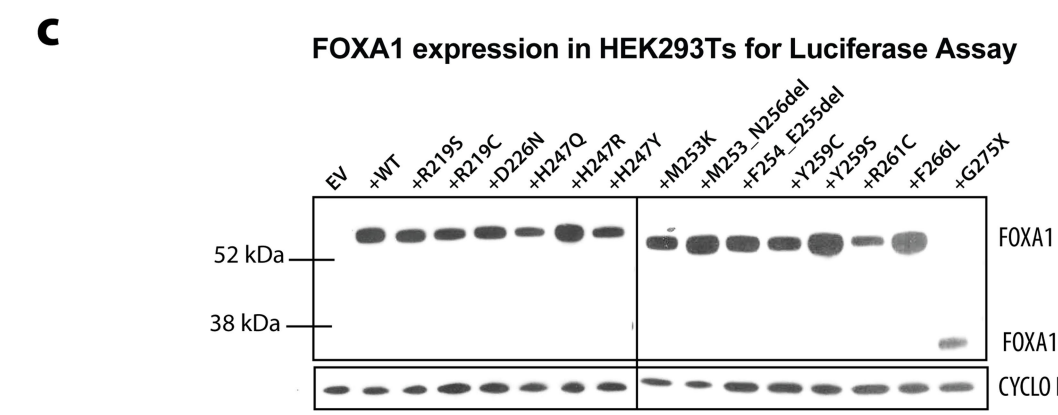


Extended Data Fig. 1 | See next page for caption.

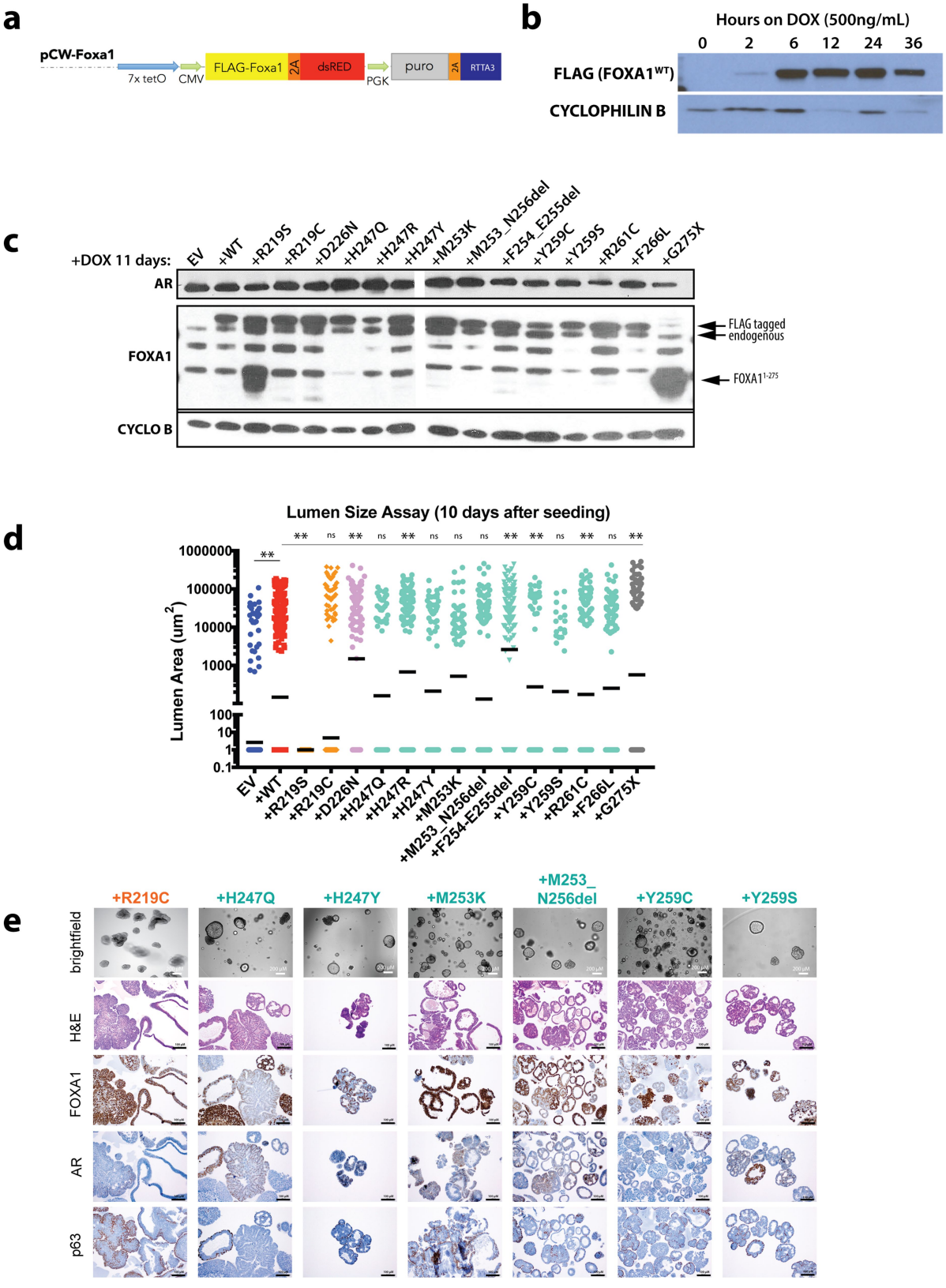
Extended Data Fig. 1 | Patients with predicted *FOXA1* mutant status have worse outcomes. a, Co-crystal structure of the FKHD domain of FOXA3 in complex with DNA resembling the FKHD consensus sequence (PDB 1VTN), with residues and folds of interest indicated, including α-helix 3 (orange), which sits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





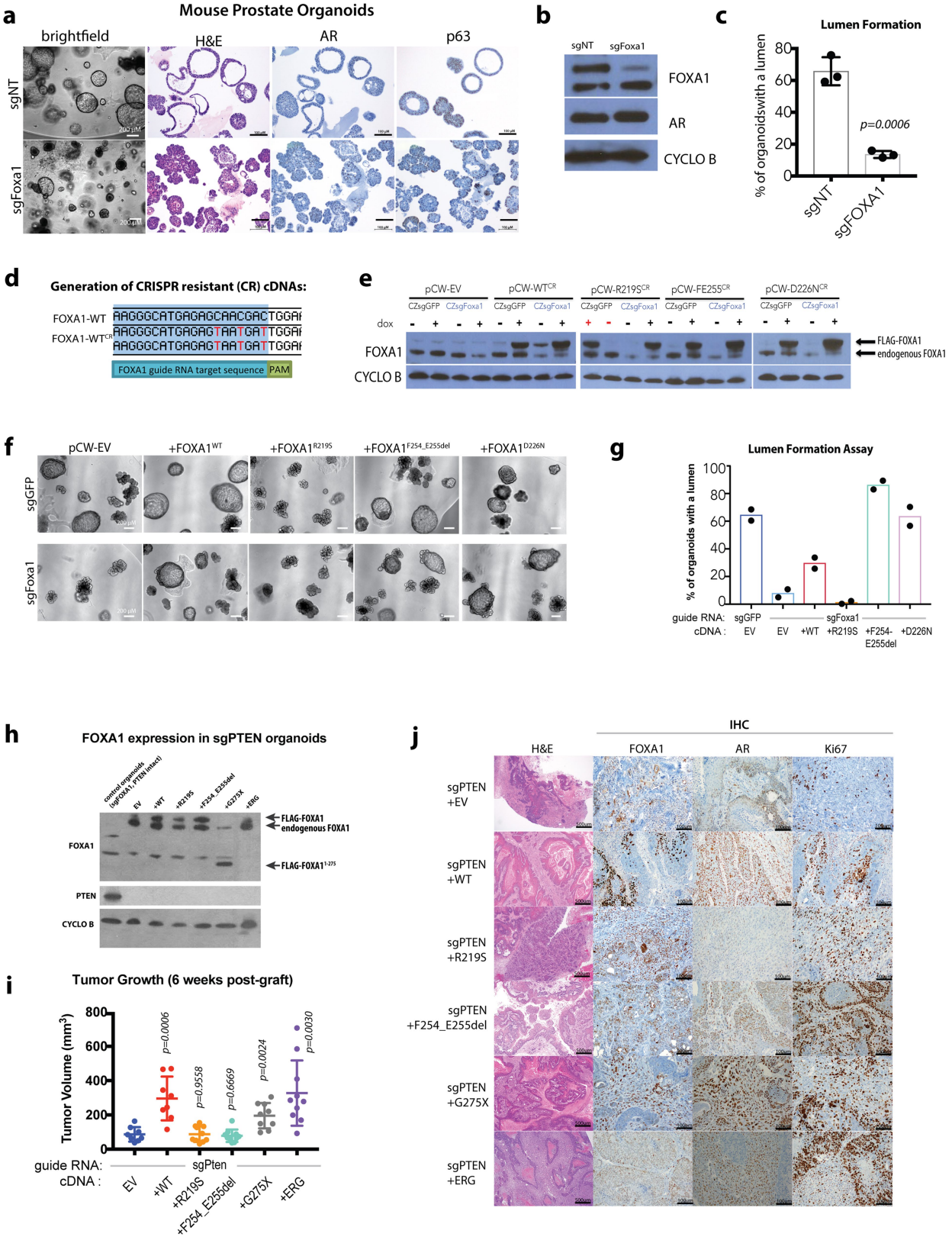


Extended Data Fig. 2 | Details of FOXA1 luciferase reporter assay. **a**, Schematic of FOXA1 luciferase reporter, depicting the modified response elements (at wobble positions within the canonical FOXA1 motif) cloned in tandem upstream of a minimal promoter driving luciferase expression. **b**, Dose–response curve of FOXA1 luciferase reporter activity in response to increased amounts of *Foxa1*^{WT} cDNA introduced into the system, expressed as a relative response ratio with 100% *Foxa1*^{WT} cDNA set to 1 and 0% *Foxa1*^{WT} cDNA (100% 'stuffer' DNA) set to 0. Data are from three biological replicates, central line and error bars represent mean ± standard deviation. **c**, Western blot of allelic series of *FOXA1* mutants in HEK293T cells 24 h after transfection with equal amounts of cDNA as used in FOXA1 luciferase reporter assay. The ΔF254/E255 and ΔM253–N256 mutants are shown as F254_E255del and M253_N256del, respectively, in the Extended Data. CYCLO B, loading control cyclophilin B. Representative blot, experiment repeated three independent times with similar results. For source gel data, see Supplementary Fig. 1.



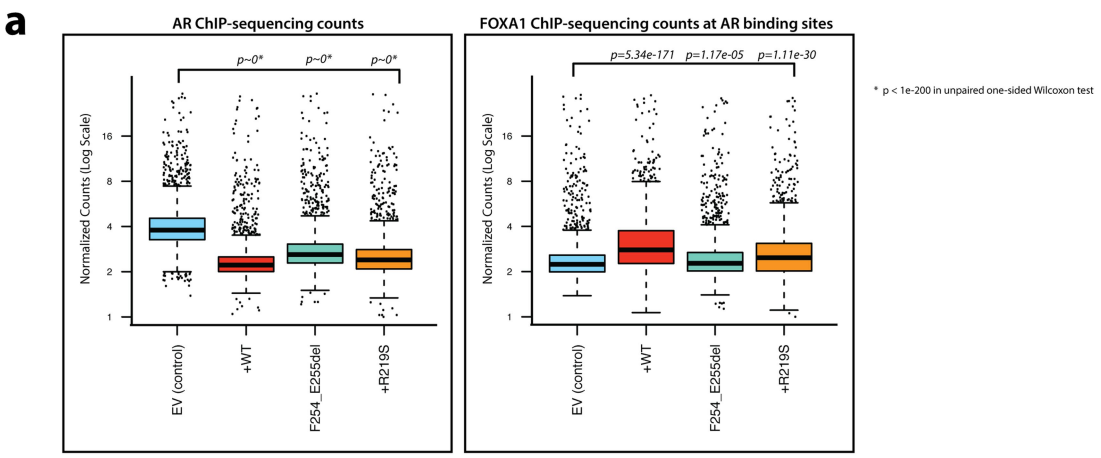
Extended Data Fig. 3 | See next page for caption.

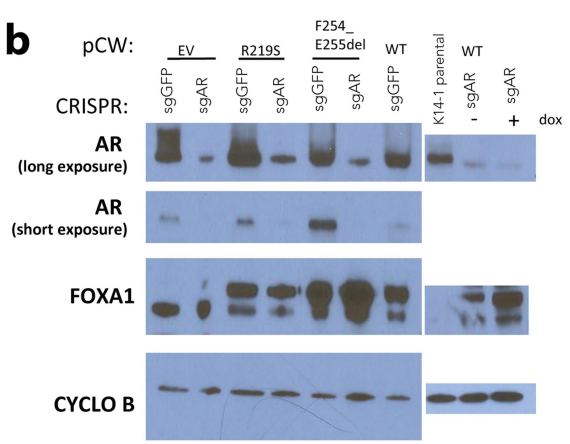
Extended Data Fig. 3 | Inducible overexpression of FOXA1 variants influences organoid lumen size and morphology. a, Schematic of doxycycline-inducible pCW-FOXA1 constructs used in the study. **b**, Western blot analysis of lysates from pCW-FOXA1(WT) organoids following acute doxycycline treatment. Representative blot, experiment repeated two independent times with similar results. For source gel data, see Supplementary Fig. 1. **c**, Western blot analysis of lysates from organoids following long-term doxycycline treatment. Sizes of endogenous and Flag-tagged FOXA1 are noted, as well as the smaller truncated form from G275X at the expected size ~38 kDa. Representative blot, experiment repeated three independent times with similar results. For source gel data, see Supplementary Fig. 1. **d**, Quantification of lumen areas measured at ten days after seeding. Solid black bar represents geometric mean. Values for sample size (indicated as dots) and *P* values are as follows: EV (292), +WT (284, $P < 0.0001$ over EV), +R219S (60, <0.0001), +F254_E255del (119, <0.0001), +D226N (120, <0.0001), +R261C (114, <0.0001), +R219C (333, 0.2915), +G275X (75, <0.0001), +M253_N256del (150, 0.2006), +M253K (63, 0.2343), +Y259S (32, 0.2045), +Y259C (45, 0.0082), +F266L (107, 0.1219), +H247Q (63, 0.8343), +H247R (180, <0.0001), +H247Y (71, 0.9104). All *P* values are relative to WT unless noted, calculated using unpaired, two-tailed Student's *t*-test. Colours represent location of mutation within FOXA1. **e**, Histology and immunohistochemistry of organoid lines overexpressing additional alleles of *FOXA1* via the doxycycline-inducible pCW vector 10 days after seeding. Images from a single biological experiment.

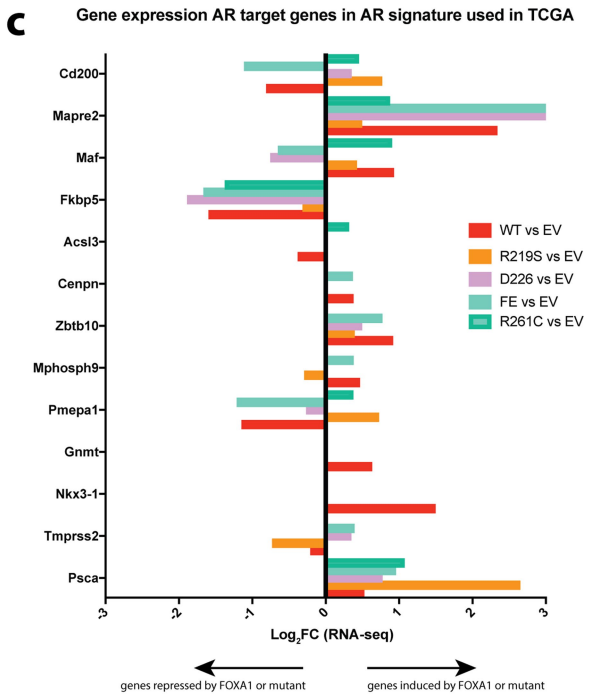


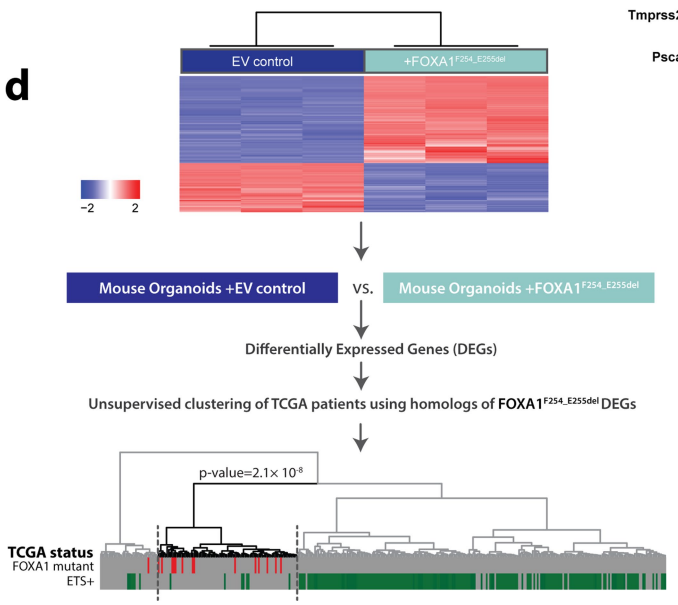
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Analysis of FOXA1 alterations in FOXA1-deleted or PTEN-deleted contexts. a, CRISPR–Cas9-mediated knockdown of *Foxa1* results in a markedly altered morphology. Organoids lacking *Foxa1* (sgFoxa1) have a reduced capacity to form lumens while maintaining expression of AR and the basal marker p63. sgNT (guide RNA targeting human gene AAVS1) serves as a negative control. **b**, Western blot analysis of lysates from organoids carrying control guide RNA (sgNT) or guide RNA targeting *Foxa1*. Representative blot, experiment repeated three times with similar results. For source gel data, see Supplementary Fig. 1. **c**, Quantification of organoids containing lumens, seven days after trypsinization in normal organoid media. Data are from three biological replicates, bars represent mean ± standard deviation, *P* value calculated using unpaired, two-tailed Student's *t*-test. **d**, Sequence indicating the location of three silent point mutations introduced upstream of the PAM sequence for *Foxa1*-targeting RNA sgFoxa1_1. **e**, Western blot analysis of lysates from organoids carrying either CRISPR-Zeo-sgGFP (CZsgGFP) or CRISPR-Zeo-sgFoxa1_1 (CZsgFoxa1) in addition to the pCW construct indicated, either EV or with a *Foxa1* allele present, plus or minus doxycycline treatment for ten days. Representative blot, experiment repeated twice with similar results. For source gel data, see Supplementary Fig. 1. **f**, Images of organoid lines carrying various combinations of guide RNA and cDNAs, ten days after doxycycline treatment. **g**, Quantification of lumen-containing organoids in lines with endogenous *Foxa1* deleted via CRISPR–Cas9 (sgFoxa1, sgGFP as control guide) and overexpression of CRIPSR-resistant $Foxa1^{WT}$ or mutant cDNA ten days after seeding. Data are from two biological replicates, bars represent mean. **h**, Western blot analysis of lysates from PTEN-deficient organoids grafted into mice, with doxycycline-induced overexpression of appropriate FOXA1 variants. Representative blot, experiment repeated twice with similar results. For source gel data, see Supplementary Fig. 1. **i**, Overexpression of FOXA1(WT) or FOXA1(G275X) in sgPTEN organoids promotes tumour growth in mice at six weeks after engraftment into the flank of NOD scid gamma mice. Data are from the following number of tumours: EV = 8, +WT = 8, +R219S = 10, +F254_E255del = 10, +G275X = 9, +ERG = 10. Data are mean ± s.d., *P* values calculated using unpaired, two-tailed Student's *t*-test vs EV. Colours represent location of mutation within FOXA1. **j**, Representative histology and immunohistochemistry (IHC) of a single tumour for given PTEN-deficient, FOXA1-expressing lines. Histology and immunohistochemistry were performed on 5–9 tumours per line, from a single in vivo experiment, with similar results.



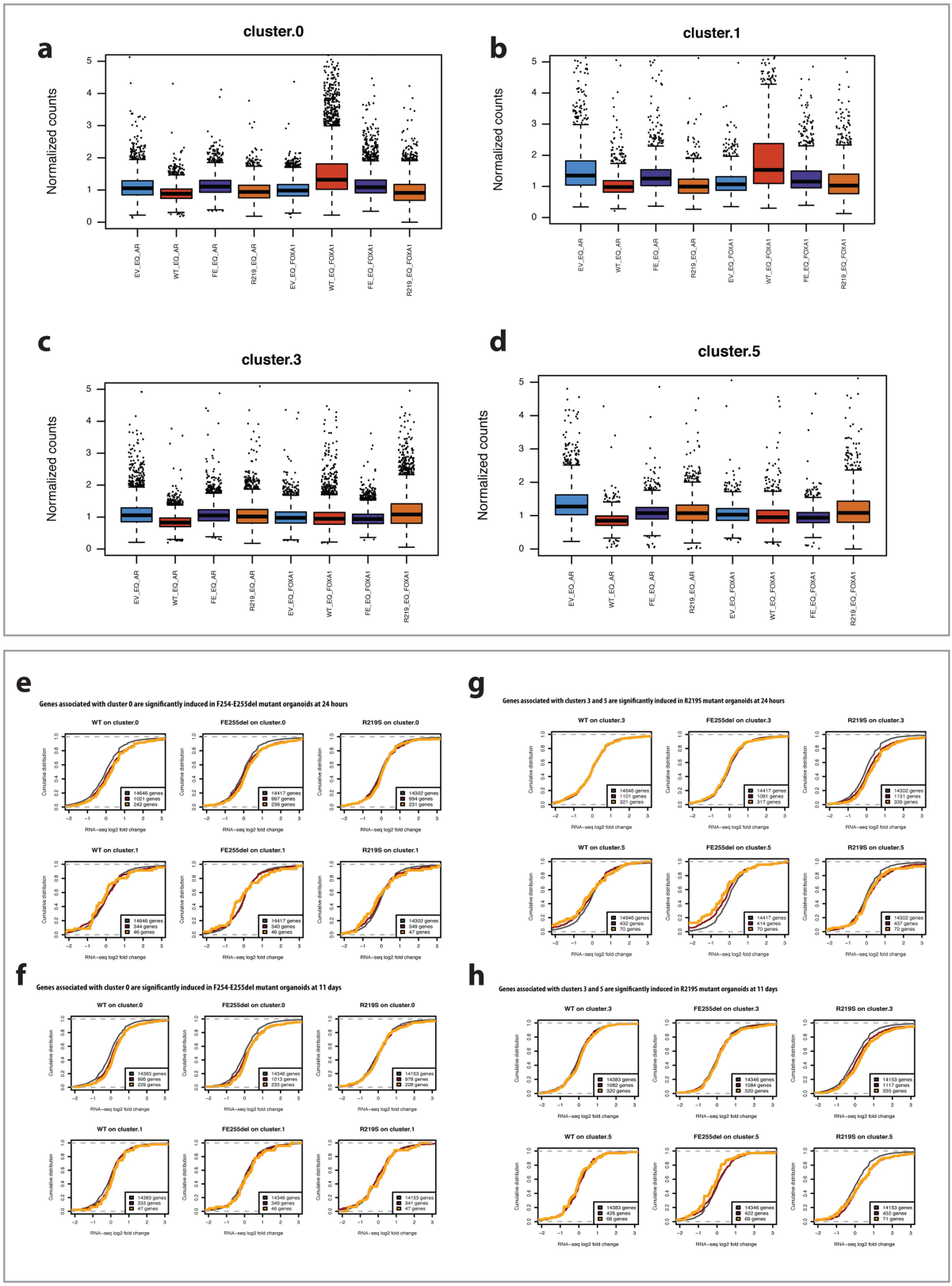






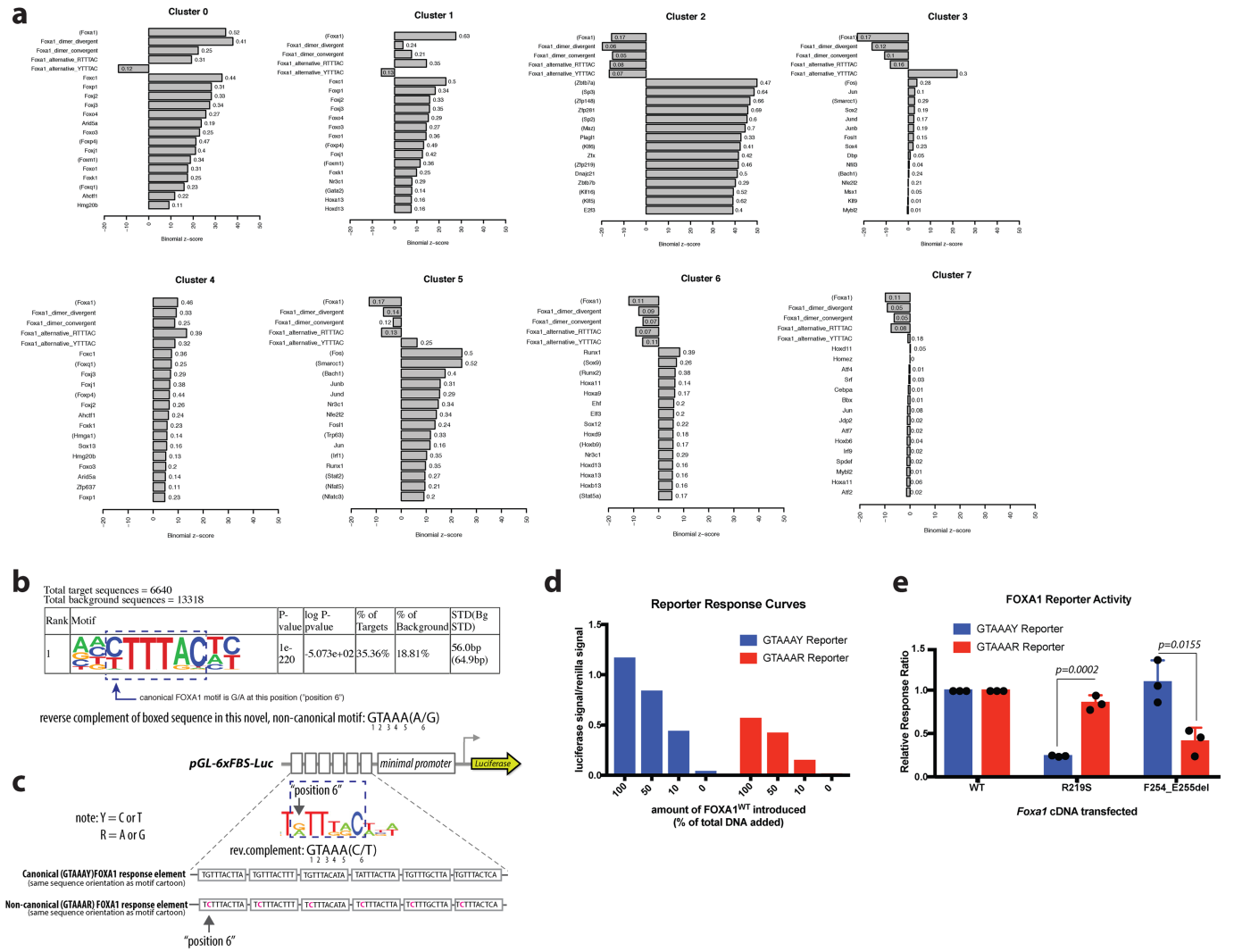
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Analysis of the interplay between AR and FOXA1 in mouse organoids expressing FOXA1 variants. a, Box plot representations of normalized counts from AR (left) and FOXA1 (right) ChIP–seq shown in Fig. 3a to quantify the reduction in AR binding following FOXA1 wild-type or mutant overexpression, and the increase in FOXA1 wild-type binding at those sites where AR is lost. Box: 25th to 75th percentile; band: median; top whisker: 75th percentile plus 1.5 times interquartile range; bottom whisker: 25th percentile minus 1.5 times interquartile range. Sample size = 2,914 peaks. *P* values calculated using an unpaired, one-sided Wilcoxon test. **b**, Western blot analysis of lysates from AR-deficient organoids generated using CRISPR–Cas9 carrying representative *Foxa1* alleles. Levels are significantly reduced but AR is not completely absent (as seen on the long exposure); this is a bulk population rather than single cell clones and thus a small number of cells escaped CRISPR–Cas9-mediated *Ar* deletion. Cells were treated with doxycycline for at least ten days. Representative blot, experiment repeated twice with similar results. For source gel data, see Supplementary Fig. 1. **c**, Expression of mouse orthologues of AR-target genes found in the AR signature used in TCGA cohort analysis based on mouse organoid RNA-seq. Genes depicted are those that have a mouse orthologue of the human gene found in the signature, and a significant expression change (DESeq2 adjusted $P < 0.05$) compared to EV control at 11 days of doxycycline treatment, as well as *Psca*, an AR-target gene expressed in mouse organoids. Data are from RNA-seq of three biological replicates. FE, F254_E255del. **d**, FOXA1(F254_E255del) signature can predict mutant tumours in TCGA. Hierarchical clustering and heat map of significantly differentially expressed genes between mouse FOXA1(F254_E255del) organoids and EV control (FDR $\leq 1 \times 10^{-10}$). Human homologues of differentially expressed genes (DEGs) from this analysis were used to cluster *FOXA1* mutant tumours ($n = 14$) and can detect nearly all *FOXA1* mutant human tumours ($P = 2.1 \times 10^{-8}$) out of the 333 TCGA samples, 199 of which are ETS^{+}. Two-sided Fisher's exact test was used to test the enrichment of *FOXA1* mutant samples within each sub-cluster, without adjustments for multiple comparisons.

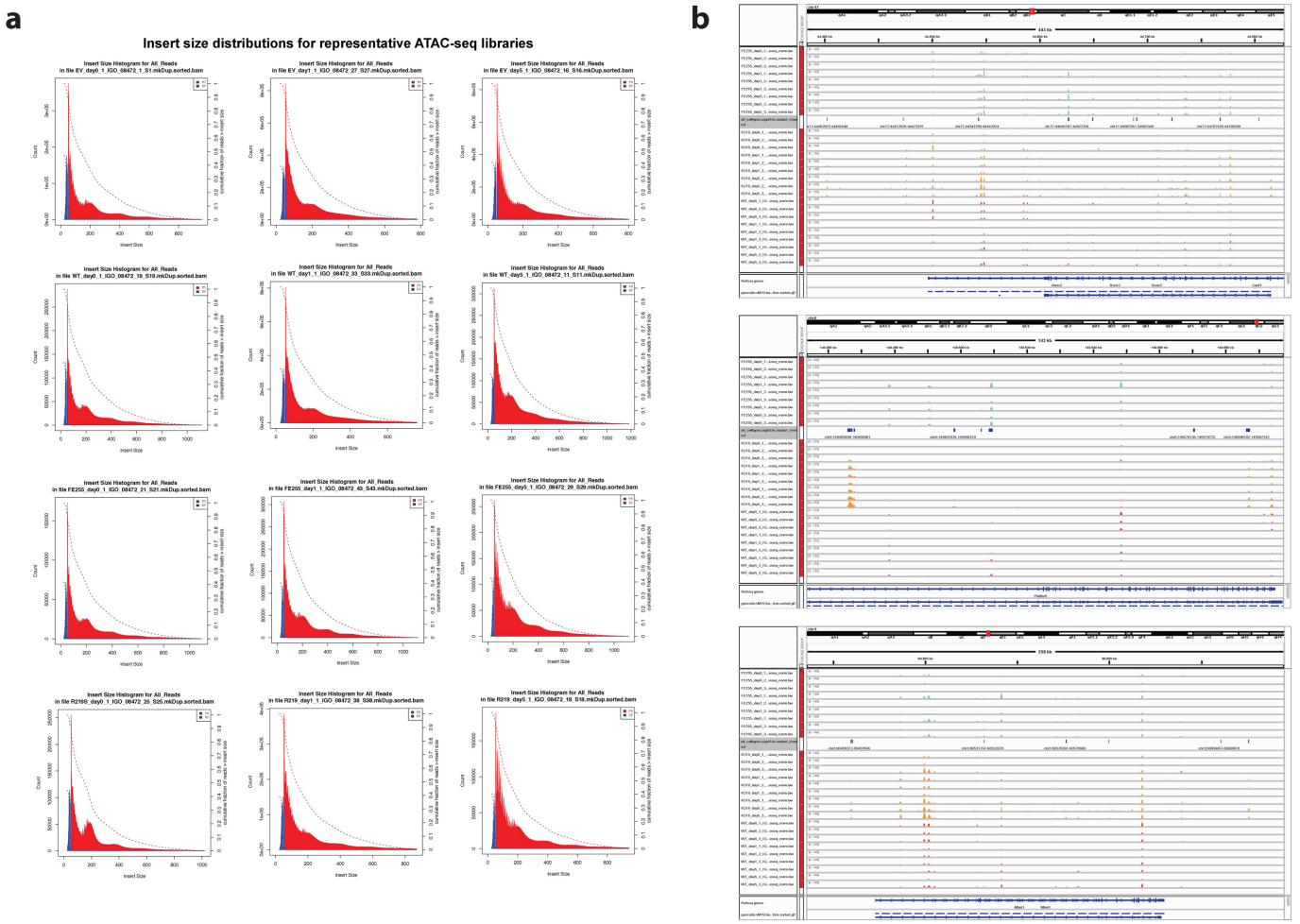


Extended Data Fig. 6 | See next page for caption.

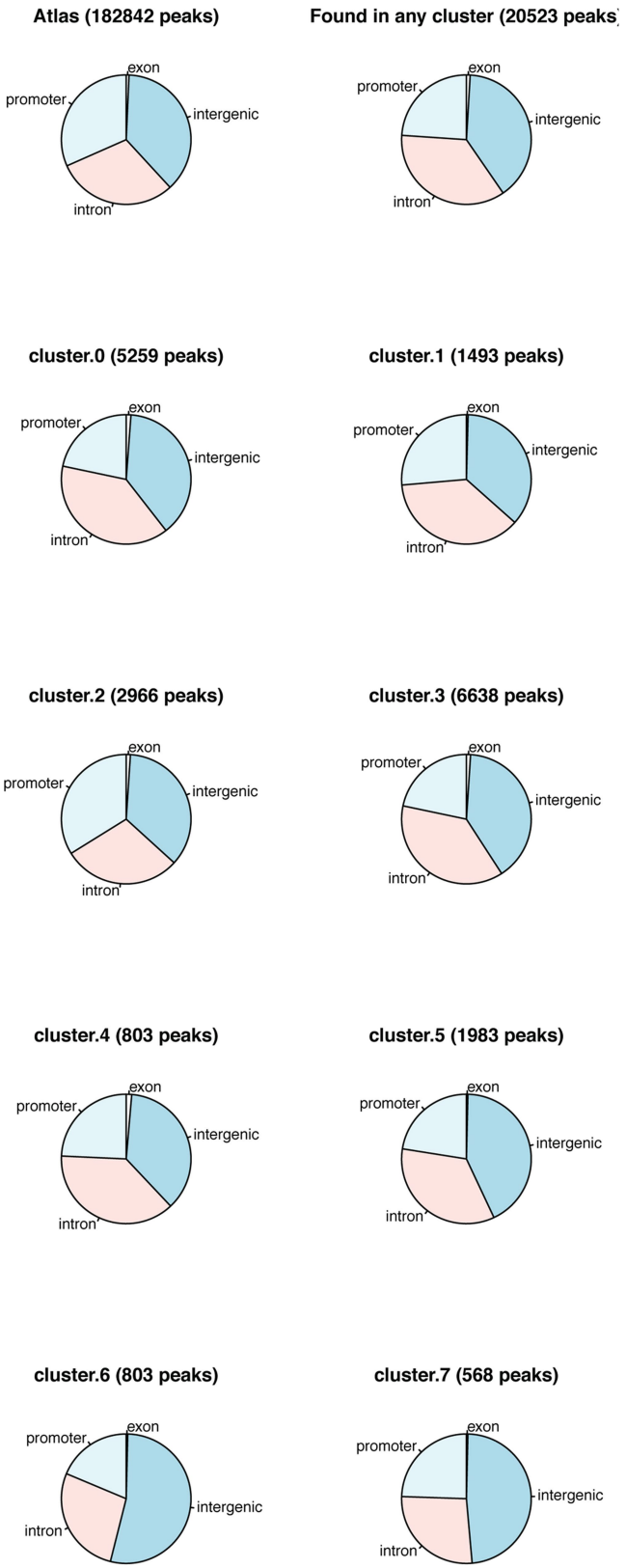
Extended Data Fig. 6 | Integrated analysis of ChIP–seq, ATAC-seq and RNA-seq data in *FOXA1* mutant organoid lines. a, Cluster 0 peaks have higher FOXA1 ChIP–seq signal in F254_E255del mutant organoid than empty vector control. Box plots show normalized day five AR ChIP–seq signal and FOXA1 ChIP–seq signal across different organoid lines at peaks from cluster 0, where normalization is based on background ChIP signal. FOXA1 ChIP signal is significantly higher in F254_E255del (FE) and in WT compared to EV control (*P* values are listed in Supplementary Table 11). Sample size = 5,260 peaks. **b**, Cluster 1 peaks have higher FOXA1 ChIP–seq signal and lower AR ChIP–seq signal in FOXA1(WT)-overexpressing organoids than in EV control. Box plots show normalized day five AR ChIP–seq signal and FOXA1 ChIP–seq signal across different organoid lines at peaks from cluster 1, where normalization is based on background ChIP signal. FOXA1 ChIP signal is significantly higher, and AR ChIP signal significantly lower, in WT compared to EV control. Sample size = 1,493 peaks. **c**, Cluster 3 peaks have higher FOXA1 ChIP–seq signal in R219S organoid than EV control. Box plots show normalized day five AR ChIP–seq signal and FOXA1 ChIP–seq signal across different organoid lines at peaks from cluster 3, where normalization is based on background ChIP signal. FOXA1 ChIP signal is significantly higher in R219S compared to EV control. Sample size = 6,641 peaks. **d**, Cluster 5 peaks have higher FOXA1 ChIP–seq signal and lower AR ChIP–seq signal in R219S organoid than EV control. Box plots show normalized day five AR ChIP–seq signal and FOXA1 ChIP–seq signal across different organoid lines at peaks from cluster 5, where normalization is based on background ChIP signal. FOXA1 ChIP signal is significantly higher, and AR ChIP signal significantly lower, in R219S compared to EV control. Sample size = 1,983 peaks. In **a**–**d**, box: 25th to 75th percentile; band: median; top whisker: 75th percentile plus 1.5 times interquartile range; bottom whisker: 25th percentile minus 1.5 times interquartile range. *P* values calculated using an unpaired, one-sided Wilcoxon test. **e**, Genes associated with cluster 0 are significantly induced in F254_E255del mutant organoids. Top, plots show empirical cumulative distribution of $\log_2$ expression changes at 24 h vs day 0 in WT (left), F254_E255del mutant (middle) and R219S mutant (right) organoids for all expressed genes (black), genes associated with at least one ATAC-seq peak in cluster 0 (cluster 0-associated genes, red), and the top quartile of these genes based on number of assigned cluster 0 peaks (strong cluster 0-associated genes, yellow). Cluster 0-associated genes show strong expression induction compared to all genes in F254_E255del as well as in WT (red vs black) but not in R219. Bottom, As a control, similar cumulative $\log_2$ expression changes for cluster 1-associated genes (red) or strong cluster 1-associated genes (yellow) do not show significant induction in F254_E255del. All *P* values are listed in Supplementary Table 12 and are one-sided Wilcoxon rank-sum tests. **f**, Genes associated with cluster 0 are significantly induced in F254–E255del mutant organoids. Top, plots show empirical cumulative distribution of $\log_2$ expression changes at 11 days vs day 0 in WT (left), F254_E255del mutant (middle) and R219S mutant (right) organoids for all expressed genes (black), genes associated with at least one ATAC-seq peak in cluster 0 (cluster 0-associated genes, red), and the top quartile of these genes based on number of assigned cluster 0 peaks (strong cluster 0-associated genes, yellow). Cluster 0-associated genes show strong expression induction compared to all genes in F254_E255del as well as in WT but not in R219. Bottom, As a control, similar cumulative $\log_2$ expression changes for cluster 1-associated genes (red) or strong cluster 1-associated genes (yellow) do not show significant induction in F254_E255del. All *P* values are listed in Supplementary Table 12 and are one-sided Wilcoxon rank-sum tests. **g**, Genes associated with clusters 3 and 5 are significantly induced in R219S mutant organoid. Top, plots show empirical cumulative distribution of $\log_2$ expression changes at 24 h vs day 0 in WT (left), F254_E255del mutant (middle) and R219S mutant (right) organoids for all expressed genes (black), genes associated with at least one ATAC-seq peak in cluster 3 (cluster 3-associated genes, red), and the top quartile of these genes based on number of assigned cluster 0 peaks (strong cluster 3-associated genes, yellow). Cluster 3-associated genes show strong expression induction compared to all genes in R219S but not in WT or F254_E255del. Bottom, similar analysis for cumulative $\log_2$ expression changes for cluster 5-associated genes (red) and strong cluster 5-associated genes (yellow). These genes are significantly induced in R219S and repressed in F254_E255del in WT for this time point. All *P* values are listed in Supplementary Table 12 and are one-sided Wilcoxon rank-sum tests. **h**, Genes associated with clusters 3 and 5 are significantly induced in R219S mutant organoid. Top, Plots show empirical cumulative distribution of $\log_2$ expression changes at day 11 vs day 0 in WT (left), F254_E255del mutant (middle) and R219S mutant (right) organoids for all expressed genes (black), genes associated with at least one ATAC-seq peak in cluster 3 (cluster 3-associated genes, red), and the top quartile of these genes based on number of assigned cluster 0 peaks (strong cluster 3-associated genes, yellow). Cluster 3-associated genes show strong expression induction compared to all genes in R219S but not in WT or F254_E255del. Bottom, similar analysis for cumulative $\log_2$ expression changes for cluster 5-associated genes (red) and strong cluster 5-associated genes (yellow). These genes are significantly induced in R219S and repressed in F254_E255del. All *P* values are listed in Supplementary Table 12 and are one-sided Wilcoxon rank-sum tests.



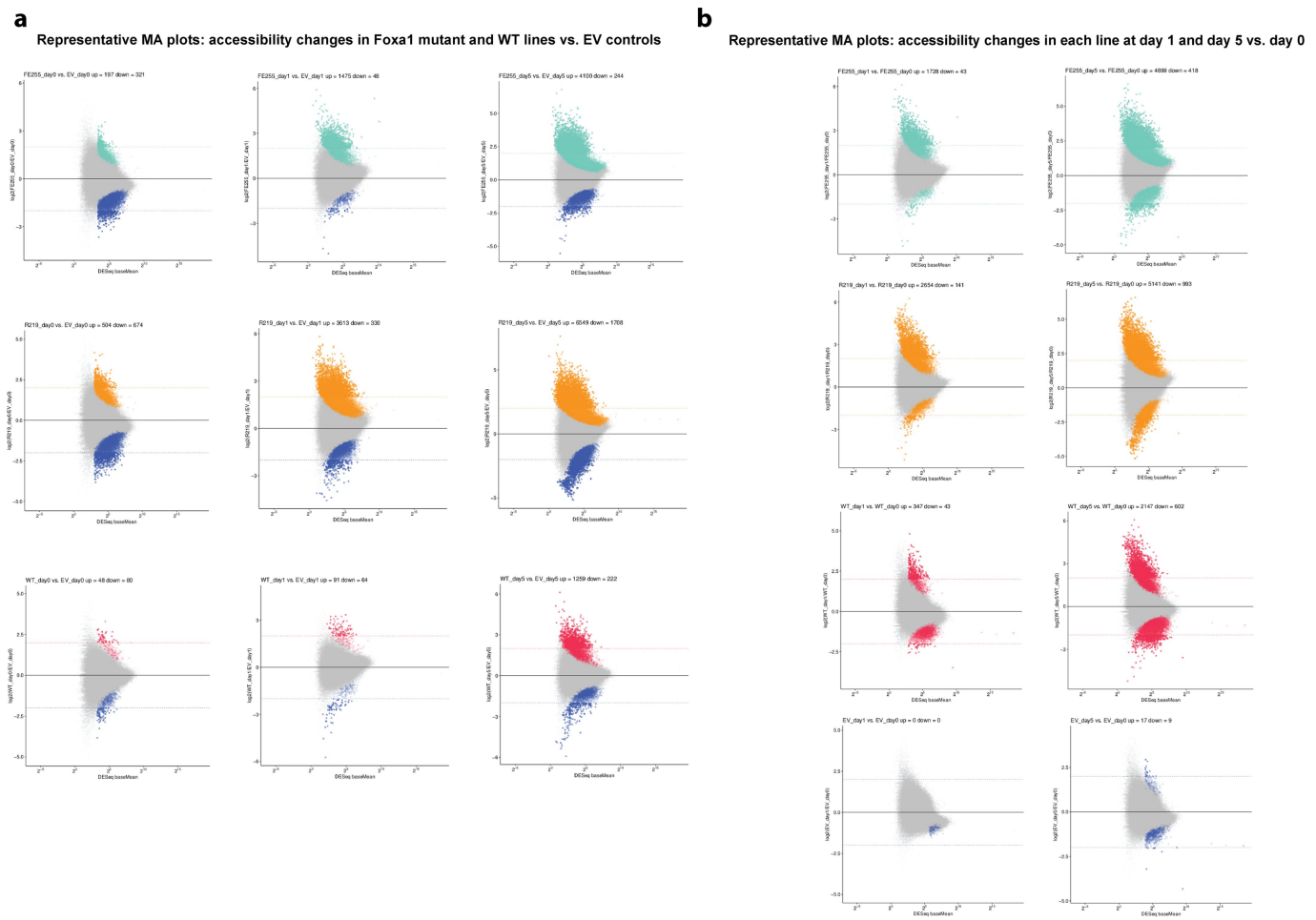
Extended Data Fig. 7 | Motif analysis of ATAC-sequencing and modification of FOXA1 reporter assay for evaluation of non-canonical FOXA1 motif. a, FIMO motif analysis of ATAC-seq clusters. Summary of motif enrichments and depletion results for each cluster relative to the background of all differentially accessible peaks, as reported by binomial *Z*-score. The top 15 enriched database motifs for expressed transcription factors are shown for each cluster. In addition, enrichment–depletion results for four additional FOXA1-related motifs are shown: convergent and divergent dimer motifs, and altered FOXA1 core binding motifs with either G/A or C/T at position 6. Transcription factors in parentheses represent motifs inferred from other species. Complete lists can be found in Supplementary Tables 3–10. **b**, Top motif identified de novo using HOMER on ATAC-seq cluster 3 (R219S-specific) with motif core indicated, and variation from canonical FOXA1 motif depicted. *P* values derived from one-sided binomial test. **c**, Schematic of reporter design. The canonical response element reporter is the same reporter used in Fig. 2, with various iterations of the canonical FOXA1 motif in tandem. The non-canonical motif has substitutions at position 6, indicated in pink, to reflect the newly identified motif enriched in cluster 3 of ATAC-seq. Note that the orientation of the upper motif cartoon and the sequence in the reporter schematic are the reverse complement of the motif identified by HOMER (GTAAAR). Modified base is noted in position 6. **d**, Dose–response curve for activity of both FOXA1 luciferase reporters in response to increased amounts of *Foxa1*^{WT} cDNA introduced into the system. Data shown are one representative biological replicate of three carried out, all showing same trends, but absolute luciferase/*Renilla* ratios vary from experiment to experiment. **e**, Results of reporter assays expressed as a relative response ratio, normalized to level of FOXA1(WT) activity for a given reporter. Data are from three biological replicates, mean ± s.d. Unpaired, two-tailed Student's *t*-test.



Extended Data Fig. 8 | Insert size distributions for ATAC-seq experiments, and track figures demonstrating peak reproducibility across ATAC-seq replicates. a, Representative insert size distributions computed from individual ATAC-seq experiments based on aligned read pairs, showing modes corresponding to nucleosome-free regions, mono-nucleosomal fragments, and di-nucleosomal fragments. **b**, Signal tracks for individual replicate ATAC-seq experiments at the *Runx2*, *Plekha5* and *Mbnl1* loci show reproducibility of accessibility events. DEseq scaling factors estimated from the atlas of IDR-reproducible peaks were used for library size normalization.



Extended Data Fig. 9 | ATAC-seq peak annotation distributions. Fraction of peaks annotated as promoter, intergenic, intronic and exonic for full atlas of reproducible peaks, differentially accessible peaks, and by ATAC-seq cluster. See Supplementary Table 15 for full annotation counts.



Extended Data Fig. 10 | MA plots for differential accessibility analysis. **a**, MA plots for differential accessibility analysis relative to EV controls. Representative MA plots (log (fold change) vs mean read counts) for differential peak accessibility analysis of FOXA1 mutant- and WT-expressing organoid lines vs empty vector controls at day 0, day 1, and day 5. Peaks that are significantly differential at FDR-corrected $P < 0.05$ are shown in colour. Dotted lines at log (fold change) = 2 and log (fold change) = −2 show cut-offs used for requiring robust accessibility changes in pairwise comparisons. **b**, MA plots for differential accessibility analysis at different time points relative to day 0. Representative MA plots (log (fold change) vs mean read counts) for differential peak accessibility analysis in each organoid line at day 1 vs day 0 and day 5 vs day 0. In **a**, **b**, all sample sizes are $n = 183{,}093$ (number of peaks in the atlas). Peaks that are significantly differential at FDR-corrected $P < 0.05$ are shown in colour, using two-sided Wald test with Benjamini–Hochberg correction.