

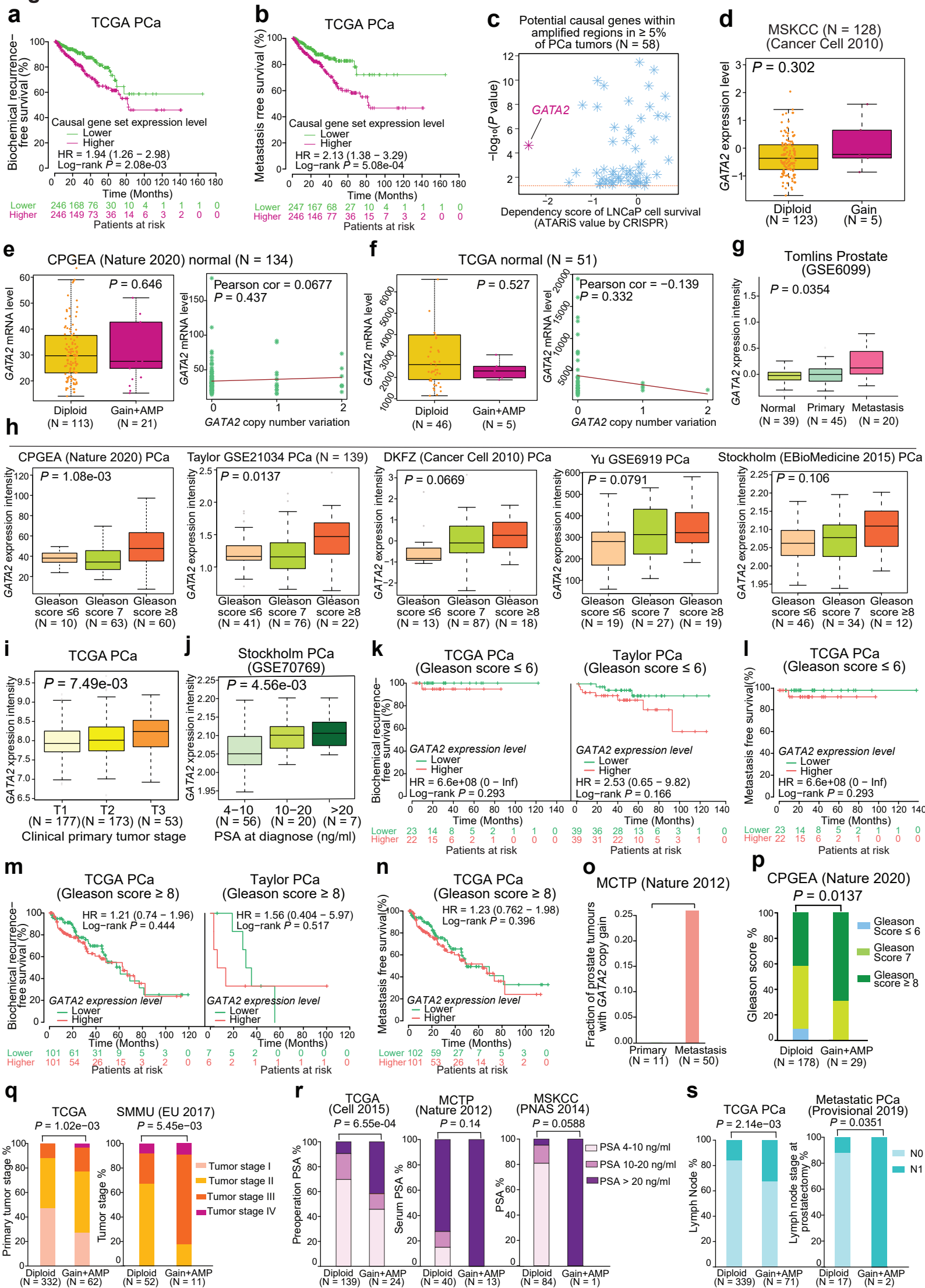
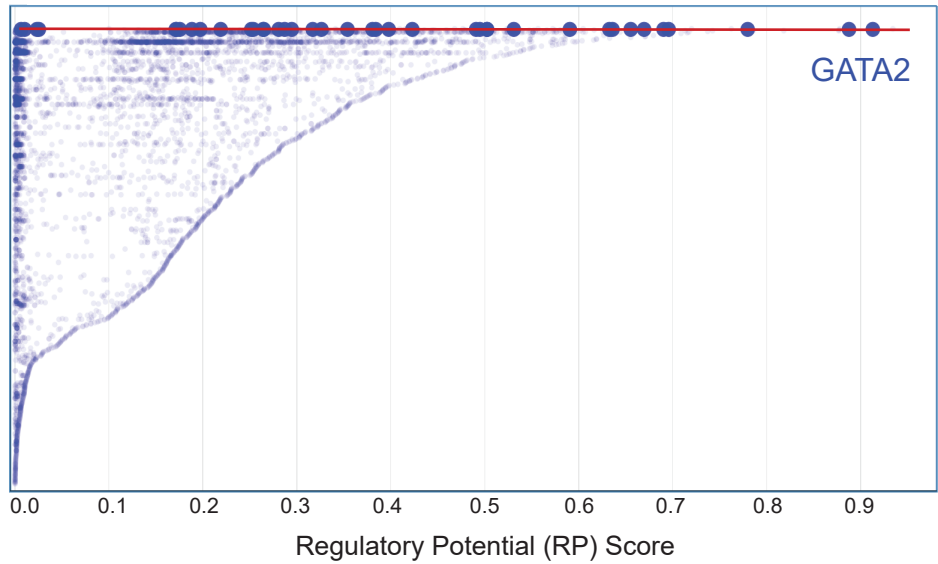
Fig. S1

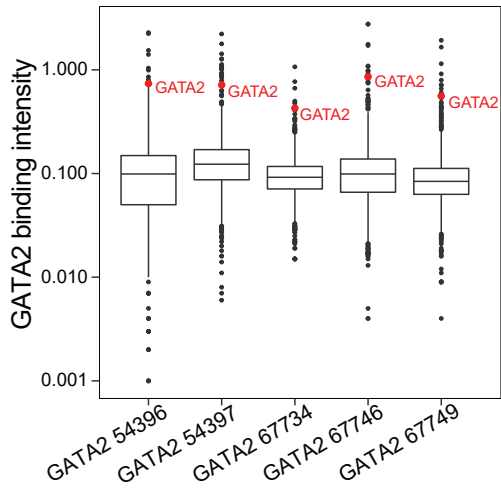
Fig. S1 GATA2 amplification and upregulation are correlated with tumor progression and poor prognosis in PCa patients. **a-b** PCa patients showing higher expression levels of the amplified causal gene set were associated with shorter biochemical recurrence-free (**a**) and metastasis-free survival (**b**). **c** Volcano representation of causal genes with the genomic amplification rates over 5%. *P* values in y-axis were calculated with Mann–Whitney U test for examining the expression association with copy number gain levels. Lowest *P* values from the CPGEA or EU PCa cohort were represented. **d** GATA2 expression was higher in PCa patient group with GATA2 copy number gain. *P* value was assessed by the Mann–Whitney U test. **e-f** GATA2 mRNA level in normal prostate tissues with GATA2 wild type (diploid) or with matched tumors harboring copy number gain and gene amplification (left panel). Pearson correlation between GATA2 mRNA expression and copy number gain/amplification (right panel). Note that GATA2 copy number gain/amplification in the matched PCa tumors showed no apparent correlations with its upregulation in normal prostate samples. Gain: with presence of one copy; AMP, amplification: with the presence of two copies. *P* values determined by the Mann-Whitney U test or Pearson correlation. **g** Box plot showing GATA2 upregulation in human primary and metastasis PCa compared to normal prostate tissues. *P* values determined by the Kruskal-Wallis H test. **h** GATA2 was upregulated in PCa tumors with higher Gleason score. *P* values calculated by the Kruskal-Wallis H test. **i-j** Elevated GATA2 expression level correlated with higher tumor stage (**i**), and elevated PSA level (**j**). *P* values determined by the Kruskal-Wallis H test. **k-n** GATA2 expression levels cannot stratify PCa patients with lower (Gleason score ≤ 6 , **k-l**) or higher risks (Gleason score ≥ 8 , **m-n**) in multiple independent PCa cohorts. *P* values assessed by the log-rank test. **o** Fraction of tumors harboring GATA2 copy number gain/AMP was markedly higher in metastatic tumors compared to primary PCa. *P* values were examined by the Fisher's exact test. **p** Proportion of PCa patients with advanced Gleason scores was higher in group with GATA2 copy number gain/AMP in comparison with wild type (diploid). *P* values examined by the Fisher's exact test. **q-s** PCa patients with GATA2 copy gain/AMP were markedly associated with advanced tumor stage (**q**), elevated PSA levels (**r**) and higher stage of lymph node metastasis (**s**). *P* values examined by Fisher's exact test.

Fig. S2

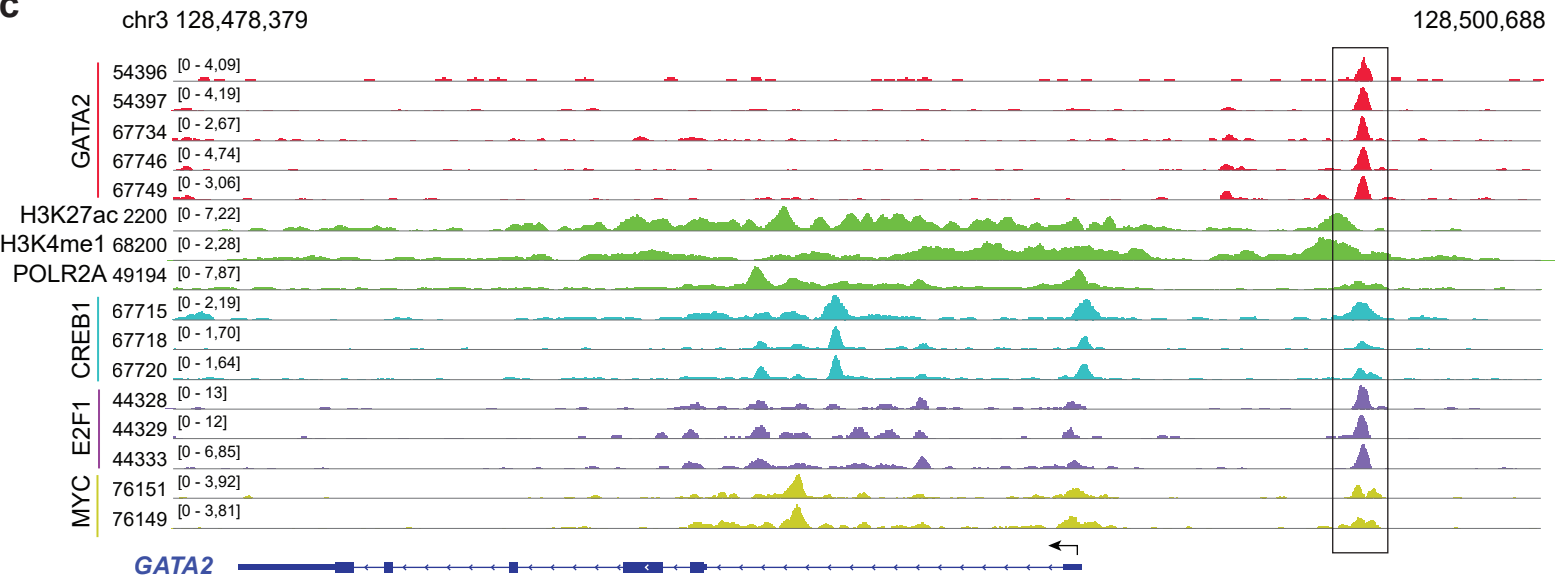
a



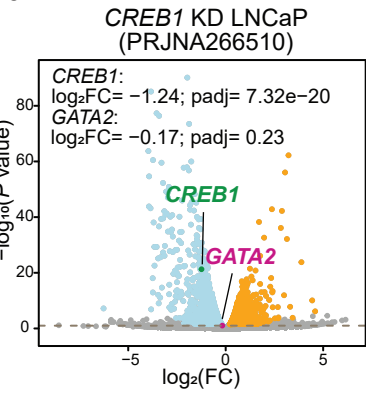
b



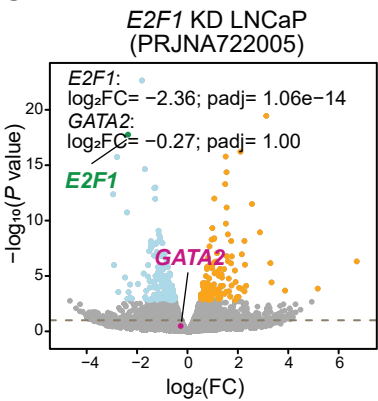
c



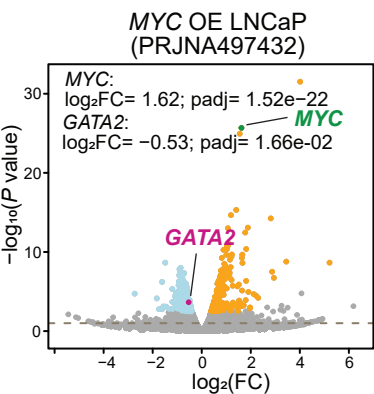
d



e



f



g

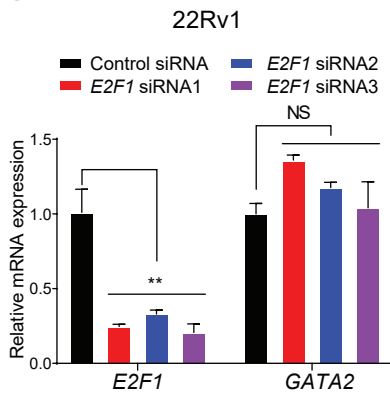


Fig. S2 GATA2 shows a most autoregulatory potential over its expression. **a** Regulatory potential of transcription factors regulating GATA2 expression. Big blue dots represented GATA2 transcription factor. X-axis represented the RP score while Y-axis indicated different transcription factors. Regulatory potential (RP) represented the estimate of how possible a transcription factor can regulate a gene. **b** Box plots depicting the distribution of GATA2 signal over active enhancers defined by the overlap of H3K27ac, H3K4me1, and POLR2A in LNCaP cells. The active enhancer in the GATA2 locus was top ranked among GATA2 targets. Red dots represented GATA2 that ranked number 10, 20, 6, 12 and 11 in five illustrated GATA2 ChIP-seq profiles in LNCaP cells, respectively. **c** ChIP-seq tracks demonstrating the enrichment of active enhancer marks (H3K27ac and H3K4me1), RNA polymerase II (POLR2A), and transcription factors, including GATA2, CREB1, E2F1, and MYC at the 3q21.3 *GATA2* locus in the LNCaP cells. The CistromeDB ID of each ChIP-seq experiment were mentioned beside. The enhancer region was highlighted by the orange square. **d-f** RNA-seq analysis indicated the impact of knockdown of CREB1 (**d**) or E2F1 (**e**) or overexpression of MYC (**f**) to GATA2 expression in LNCaP cells. Differential gene expression was assessed by the DESeq2. **g** Downregulation of E2F1 showed little influence on GATA2 mRNA expression.

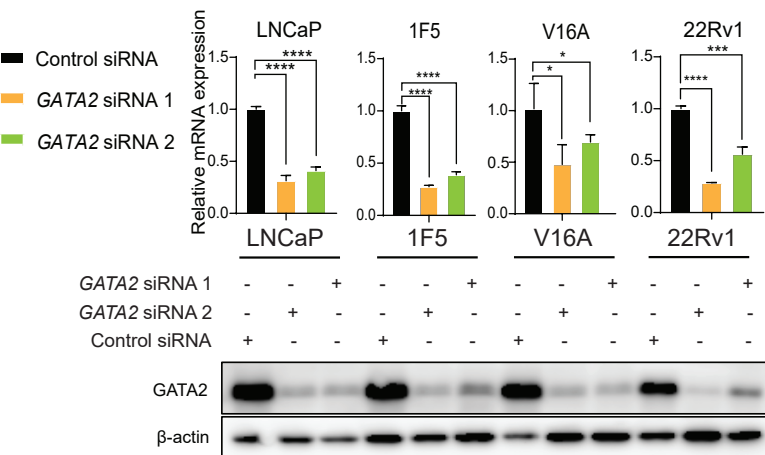
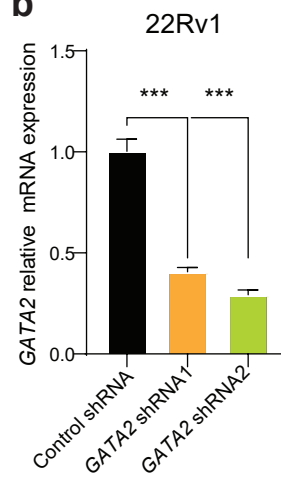
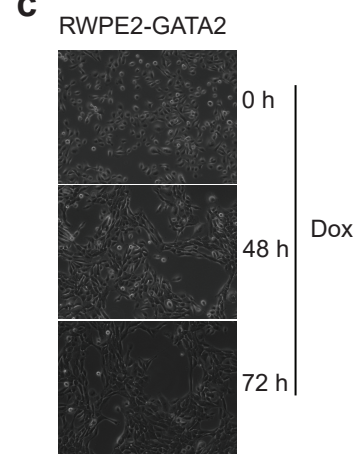
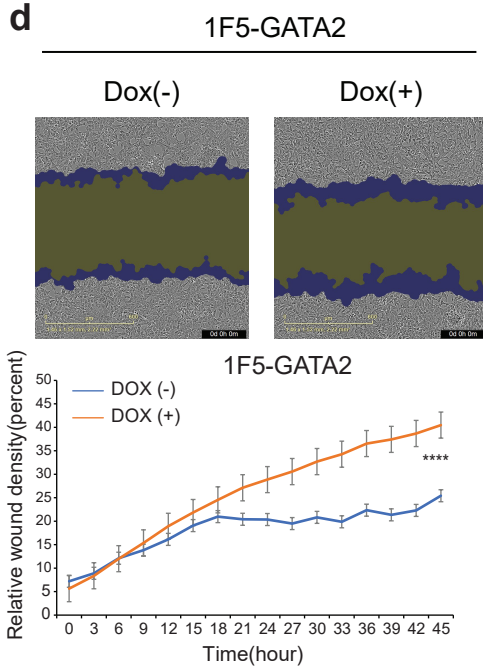
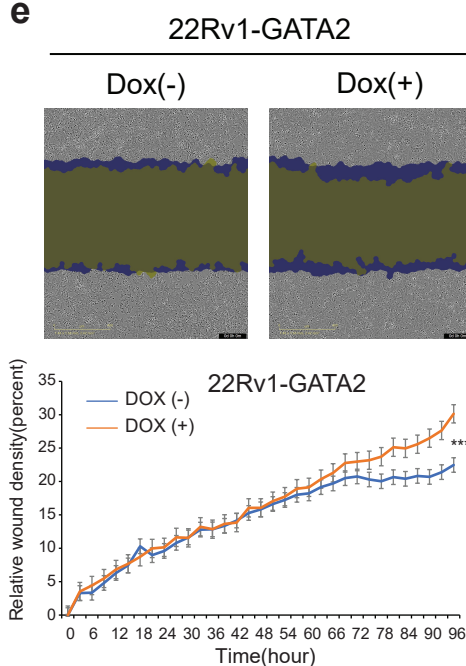
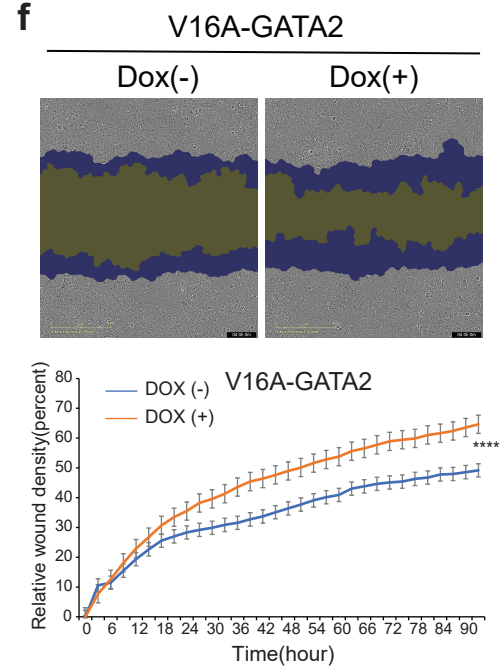
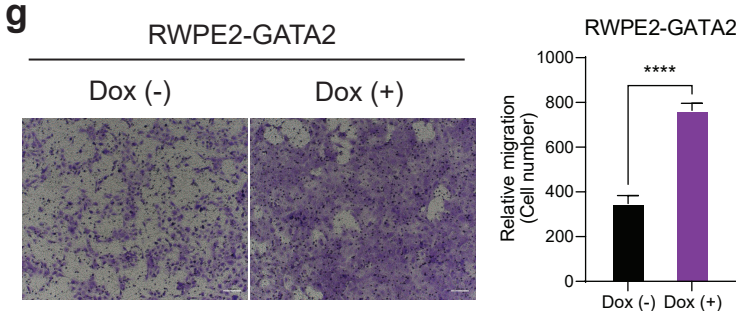
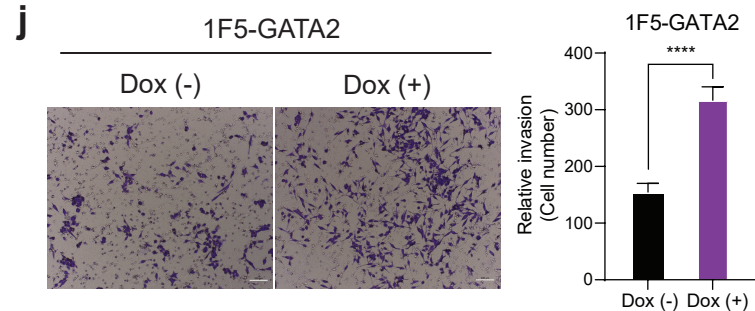
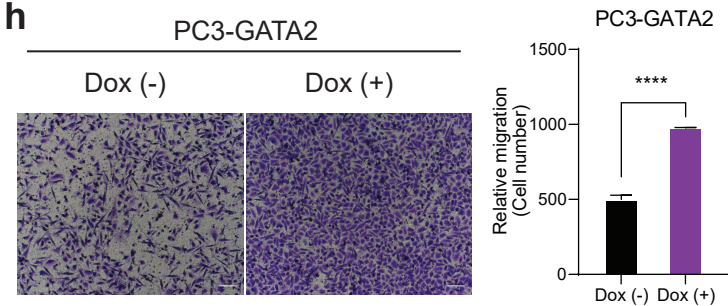
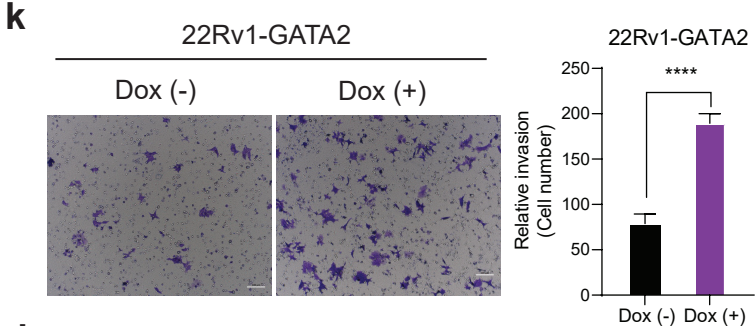
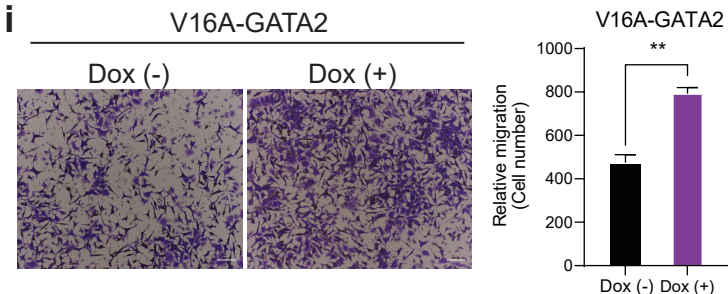
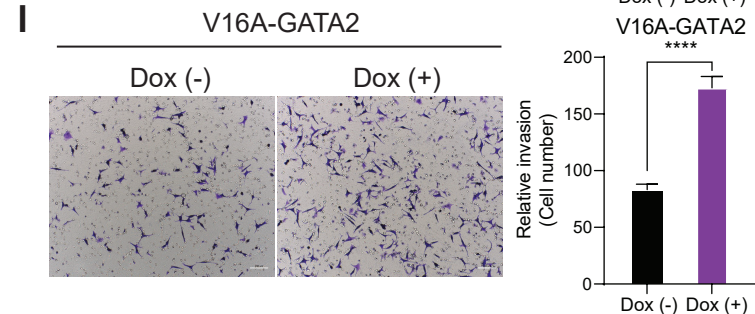
Fig. S3**a****b****c****d****e****f****g****j****h****k****i****l**

Fig. S3 GATA2 potentiates PCa proliferation and metastasis. **a** Knockdown efficiency of GATA2 by siRNAs in different PCa cells was measured by qRT-PCR and western blot. **b** Knockdown efficiency of GATA2 by shRNAs in 22Rv1 cells was measured by qRT-PCR. **c** Cell morphology was observed under optical microscope (20x) in RWPE2-GATA2 cells with or without Dox (1 µg/mL) in indicated time. **d-f** Overexpression GATA2 facilitated PCa cell migration. 1F5-GATA2, 22Rv1-GATA2 and V16A-GATA2 cells were treated with 1 µg/mL Dox and then wound density was measured at indicated time points in 1F5 (**d**), 22Rv1 (**e**) and V16A cells (**f**). **g-i** Overexpression of GATA2 promoted cell Transwell migration in RWPE2-GATA2 (**g**), PC3-GATA2 (**h**) and V16A-GATA2 (**i**) cells. **j-l** Overexpression of GATA2 potentiated cell invasion in 1F5-GATA2 (**j**), 22Rv1-GATA2 (**k**) and V16A-GATA2 (**l**) cells. All error bars represent s.e.m, n=3 independent experiments. ** $P<0.05$, *** $P<0.001$, **** $P<0.0001$, determined by unpaired Student's t -test.

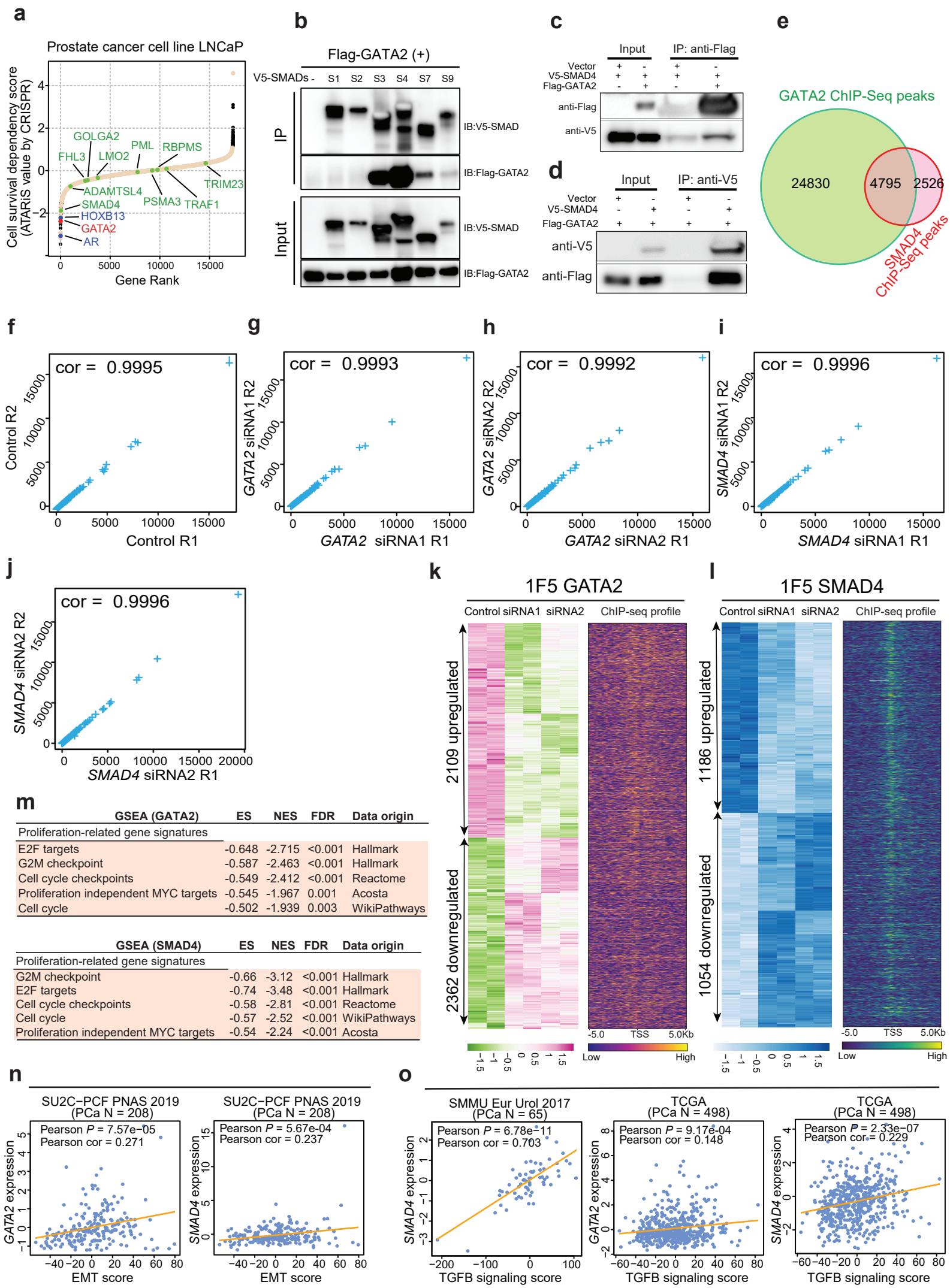
Fig. S4

Fig. S4 GATA2 physically interacts and is cooperative with SMAD4 for genome-wide chromatin co-occupancy and co-regulation of PCa genes and cancer metastasis pathways. **a** SMAD4 was identified as an essential gene for cell growth and survival in a genome-wide CRISPR loss-of-function screening in LNCaP cells. Green dots represented the 10 proteins interacting with GATA2 by a comprehensive PPI analysis. Note that only SMAD4 was comparable with GATA2 to be most essential for PCa cell survival. **b** 293T cells were transfected with or without GATA2 together with SMADs. Forty-eight hours later, cell lysates were collected and subjected to co-IP and western blotting analysis. **c-d** Reciprocal immunoprecipitation (IP) from 293T cells ectopically expressing Flag-GATA2 and V5-SMAD4. **e** Venn diagram illustrating the overlap of GATA2 and SMAD4 ChIP-seq binding peaks in 1F5 cells. **f-j** High linear correlation between each replicate from GATA2 or SMAD4 RNA-seq in 1F5 cells. **k-l** Heatmaps of genes differentially expressed following GATA2 (**k**) or SMAD4 (**l**) knockdown in 1F5 cells (FDR < 0.1). GATA2 (**k**) or SMAD4 (**l**) ChIP-seq signals in 1F5 cells were plotted for the promoters (TSS $\pm$ 5 kb) of the genes shown, where yellow indicated higher enrichment. **m** Additional enriched common proliferation-related gene signatures from GATA2 and SMAD4 target gene sets. **n-o** Additional data sets displaying that expression levels of GATA2 or SMAD4 were significantly correlated with the EMT score (**n**) or TGF β signaling score (**o**) in PCa tumors.

Fig. S5

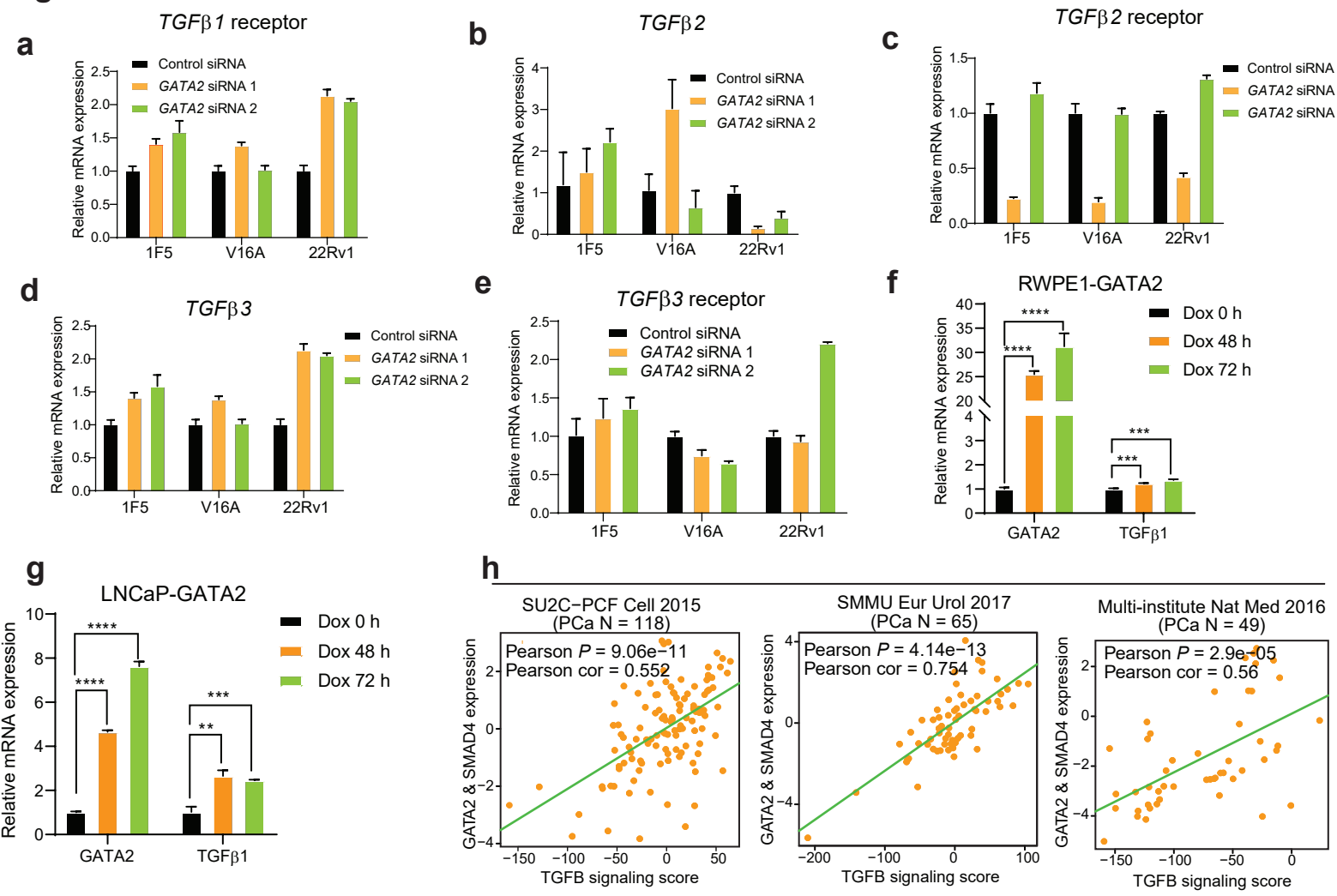
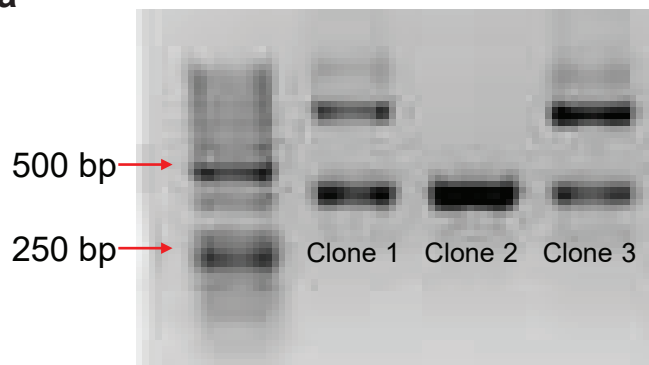


Fig. S5 GATA2 promotes the expression of *TGFβ1* but not other *TGFβ*s. **a-e** The effect of GATA2 knockdown on the mRNA expression levels of *TGFβ*s as shown in 1F5, V16A and 22Rv1 cells. **f-g** RWPE1-GATA2 and LNCaP-GATA2 cells, respectively, were treated with or without Dox (1μg/mL). Then mRNA expression of *TGFβ1* was determined by qRT-PCR. **h** GATA2 and SMAD4 expression levels positively correlated with *TGFβ* signaling scores in multiple independent PCa datasets. *P* values assessed by the Pearson's product-moment correlation test. All the error bars are mean s.e.m, n=3 independent experiments. All the error bars represent s.e.m, n=3 technical replicates. ***P*<0.001, ****P*<0.001, *****P*<0.0001, determined by unpaired Student's *t*-test.

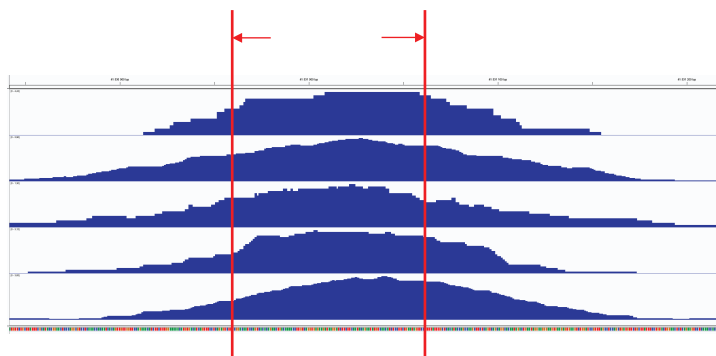
Fig. S6

a



b

Clone 1 and Clone 3
chr19: 41,530,965 - 41,531,062



Clone 2
chr19: 41,530,930 - 41,531,039

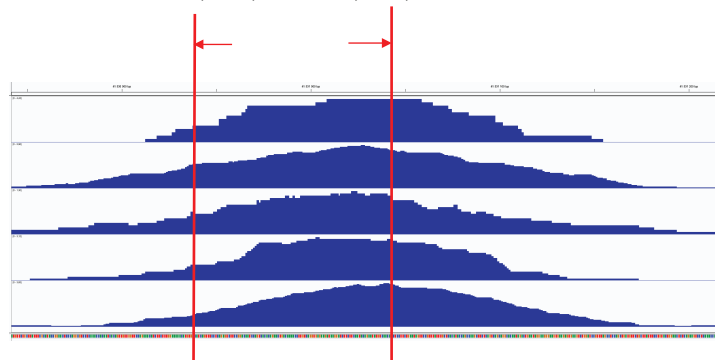


Fig. S6 Deletion of TGF β 1 enhancer by CRISPR/Cas9 genome editing system. **a** CRISPR/Cas9-mediated deletion of the GATA2-occupied TGF β 1 enhancer. Genomic DNA was subjected to PCR using specific primers and separated by agarose gel in three enhancer knockout clones. The top PCR band indicated the full-length WT PCR amplicons, whereas the lower band indicated the CRISPR-edited PCR amplicons. **b** Gel purification of the genomic DNA from lower bands in (**a**) followed by Sanger Sequencing confirmation.

Fig. S7

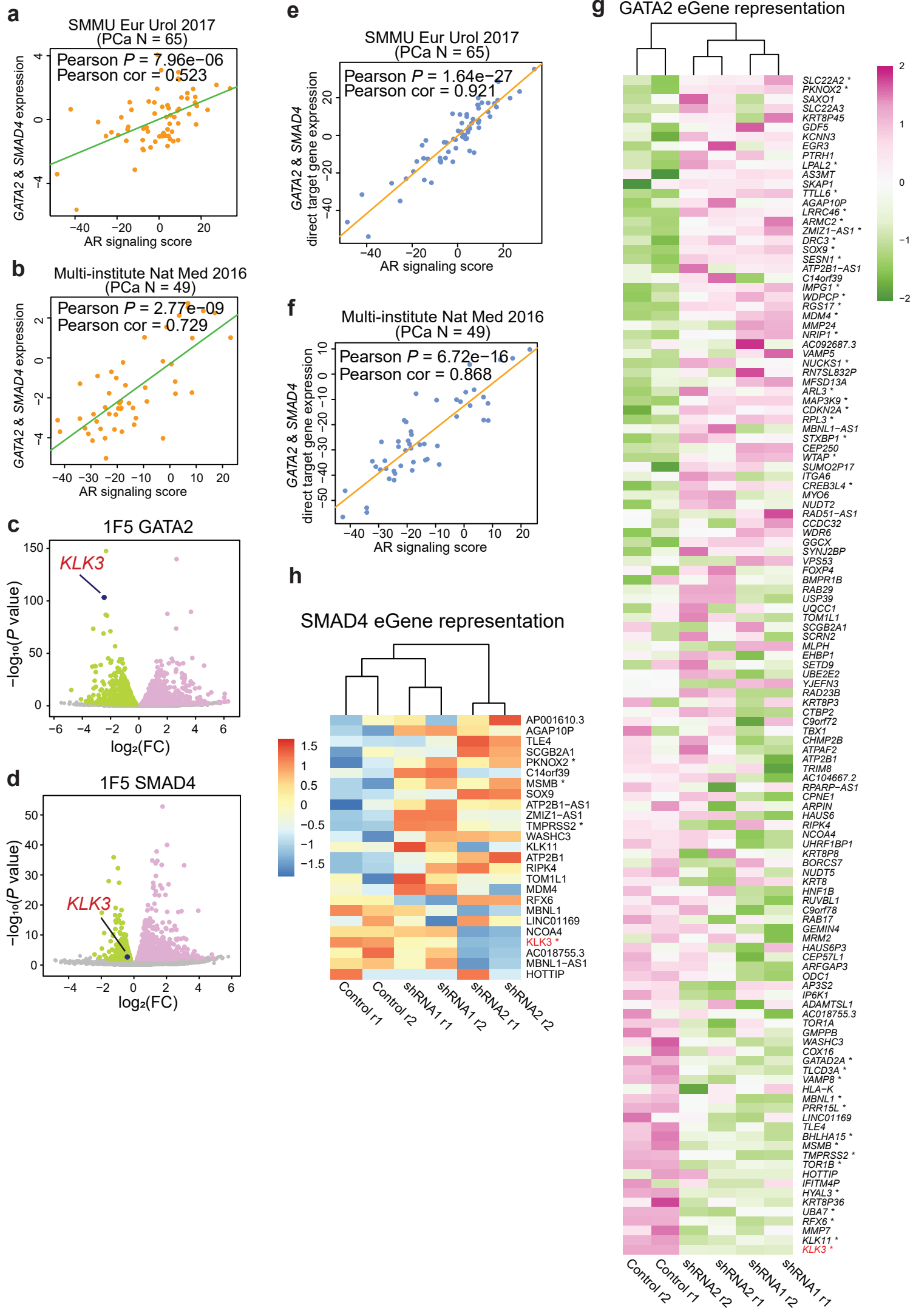


Fig. S7 GATA2 co-opts with SMAD4 to regulate AR signaling. **a-b** Scatter plots showed positive correlations between GATA2 & SMAD4 expression levels and AR signaling score in human PCa. **c-d** Volcano plots demonstrated that *KLK3* was significantly downregulated following knockdown of GATA2 (**c**) or SMAD4 (**d**) in 1F5 cells. Purple or green dots indicated upregulated or downregulated genes, respectively. **e-f** Expression values of GATA2 and SMAD4 joint direct targeting genes positively correlated with AR signaling score in PCa patients. *P* values assessed by the Pearson's product-moment correlation test. **g** Heatmap representation using RNA-seq data upon GATA2 knockdown in 1F5 cells for eQTL genes with the eQTL SNPs enriched in the GATA2 ChIP-seq peaks in multiple PCa cell lines. **h** Heatmap representation using RNA-seq data upon SMAD4 knockdown in 1F5 cells for eQTL genes with the eQTL SNPs enriched in the SMAD4 ChIP-seq peaks in 1F5 cells. Asterisks beside genes in the heatmaps denoted significant differentially expression genes. All the error bars represent s.e.m, n=3 technical replicates. *****P*<0.0001, determined by unpaired student's *t*-test.

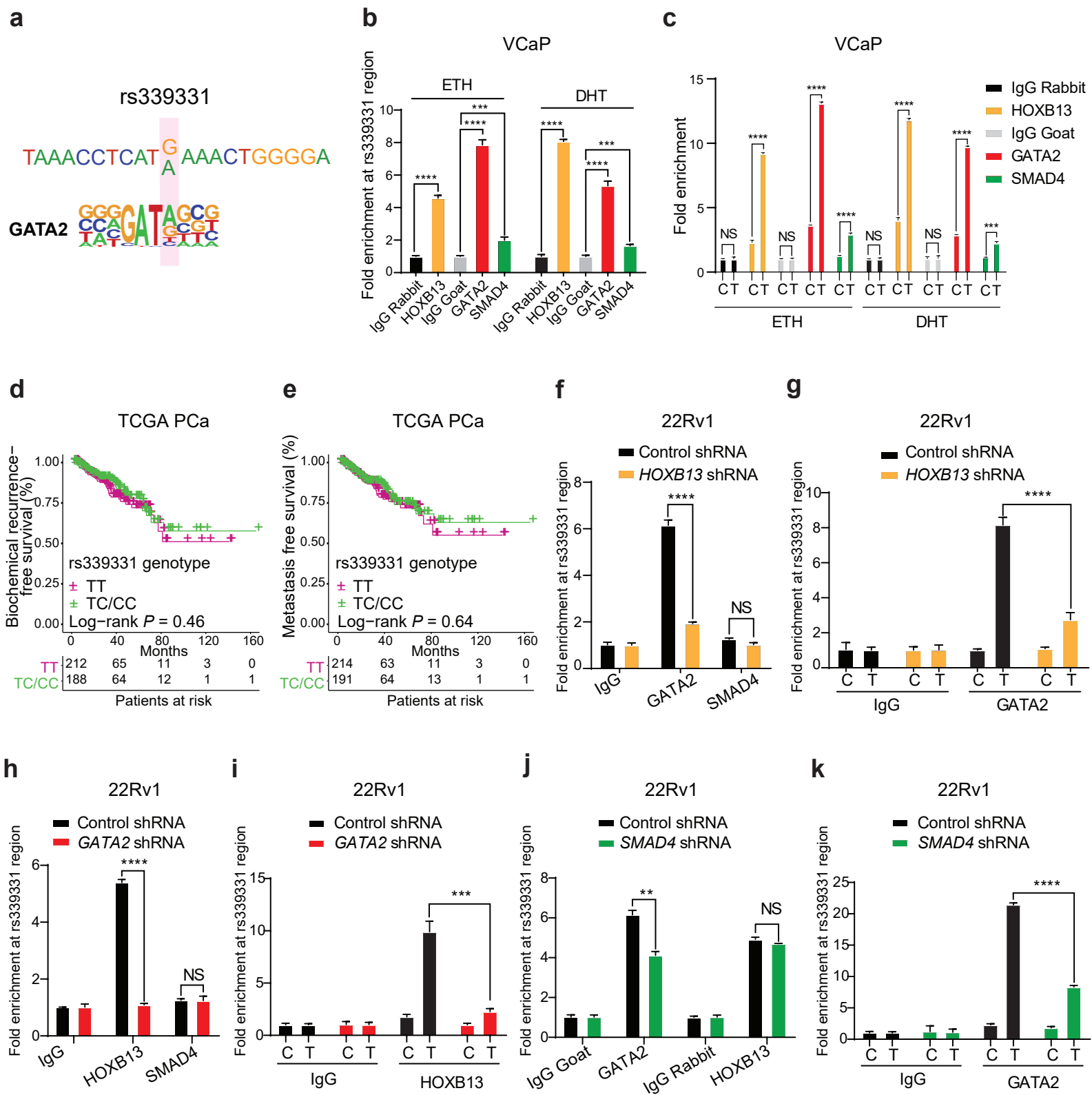
Fig. S8

Fig. S8 GATA2 and SMAD4 indicate a tendency of binding to the PCa-associated risk T allele at rs339331/6q22. **a** Representation of rs339331 located within the GATA2 DNA-binding motif in the reverse strand. **b** ChIP-qPCR for GATA2 and SMAD4 chromatin binding at the rs339331-containing region upon ETH or DHT treatment for 24h in VCaP cells. HOXB13 as a positive control. **c** GATA2 and SMAD4 favor binding to the T risk allele than C at rs339331 as determined by ChIP followed by AS-qPCR in ETH or DHT treated VCaP cells. **d-e** rs339331 could not stratify PCa patients with biochemical recurrence (**d**) or metastasis (**e**). *P* values assessed by the log-rank test. **f-g** Knockdown of HOXB13 decreased the capacity of GATA2, but not SMAD4, binding to rs339331 and its T allele. **h-i** Knockdown of GATA2 alleviated the capability of HOXB13, but not SMAD4, binding to rs339331 and its T allele. **j-k** Knockdown of SMAD4 reduced the potential of GATA2, but not HOXB13, binding to rs339331 and its T allele. ****P*<0.001, *****P*<0.0001, determined by unpaired student's *t*-test. N.S: Non-significant.

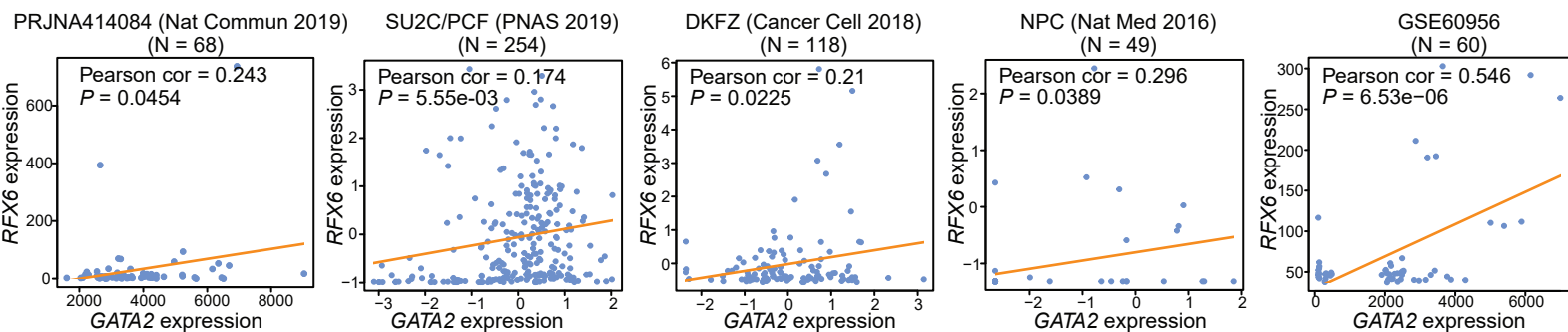
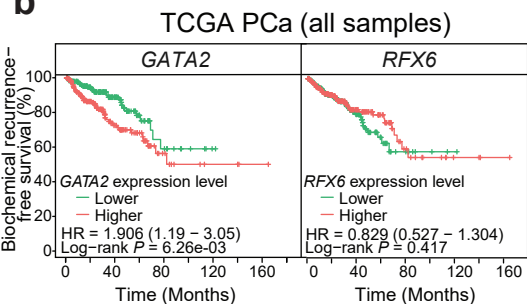
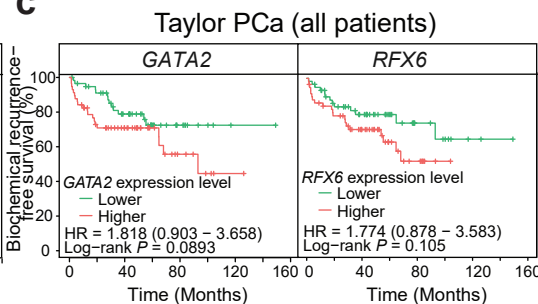
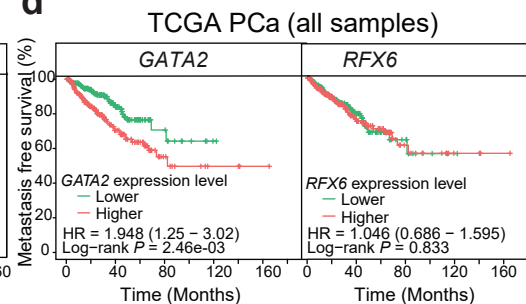
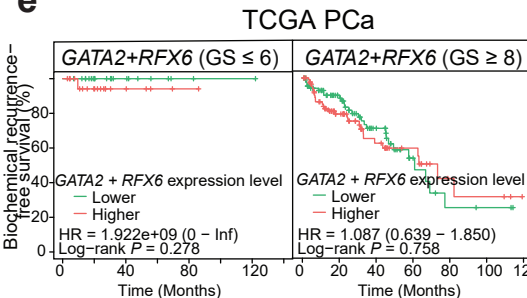
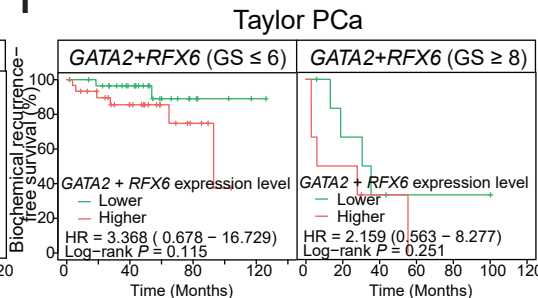
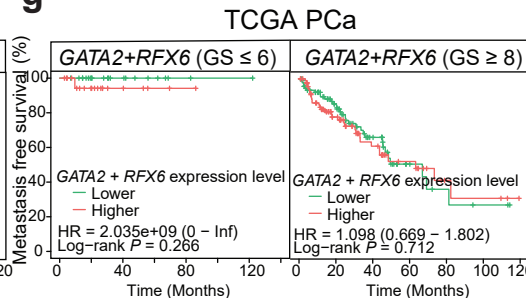
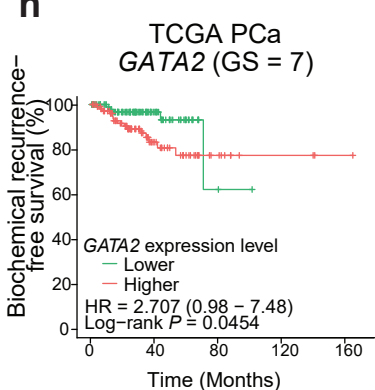
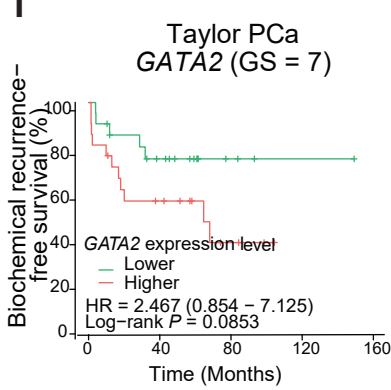
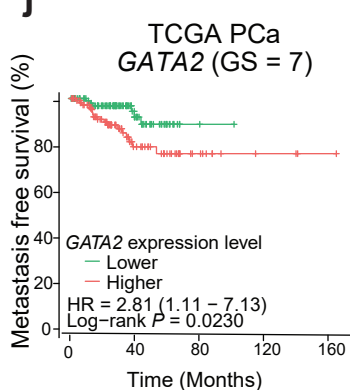
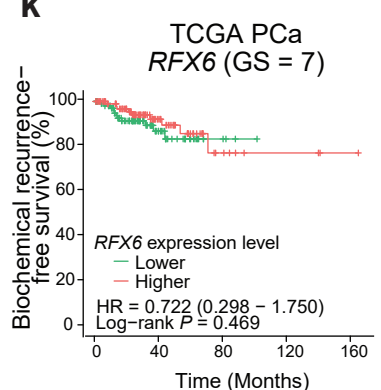
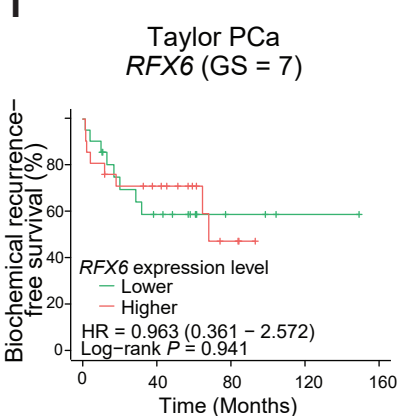
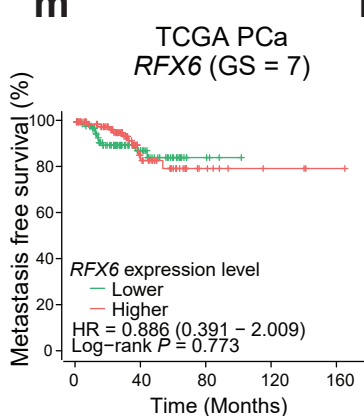
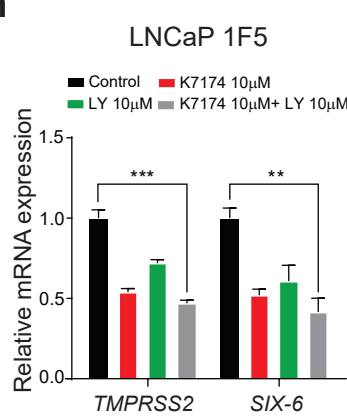
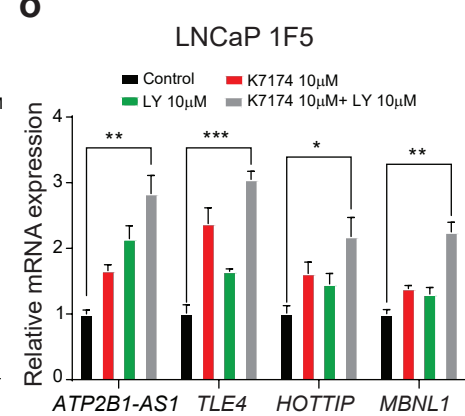
Fig. S9**a****b****c****d****e****f****g****h****i****j****k****l****m****n****o**

Fig. S9 The synergistic effects of GATA2 and RFX6 expression levels in PCa progression and prognosis. **a** Expression levels of RFX6 positively correlated with GATA2 in multiple independent PCa cohorts. The gene expression level in the PRJNA414084 dataset was measured by the RNA-seq and normalized by the median of ratios method from the DESeq2. **b-d** Kaplan Meier plots showed the values of biochemical recurrence and metastasis of PCa patients with high GATA2 or RFX6 expression level in two independent cohorts. **e-g** Joint expression levels of GATA2 and SMAD4 showed failure in stratifying PCa patients with lower (Gleason score 6) or higher risks (Gleason score ≥ 8). *P* values assessed by log-rank test. **h-m** GATA2 or RFX6 mRNA levels alone indicated less prognostic value in PCa patient group with an intermediate risk of Gleason Score 7 when compared to the synergistic expression levels of GATA2 and RFX6 (see **Fig. 8q-r**). PCa samples with RFX6 deep loss were excluded in the analysis. Patients were stratified by the median value of RFX6 and GATA2 expression levels. *P* values assessed by log-rank test. **n-o** The combination effect of TGF β signaling inhibitor and AR antagonist on the expression of the GATA2 and SMAD4 common targeting eQTL genes. **P*<0.05, ***P*<0.001, ****P*<0.001, determined by unpaired student's *t*-test.