

Supplementary Information

Extensive germline-somatic interplay contributes to prostate cancer progression through HNF1B co-option of TMPRSS2-ERG

Nikolaos Giannareas^{1#}, Qin Zhang^{1#}, Xiayun Yang^{1#}, Rong Na^{2#}, Yijun Tian³, Yuehong Yang¹, Xiaohao Ruan⁴, Da Huang⁴, Xiaoqun Yang⁵, Chaofu Wang⁵, Peng Zhang⁶, Aki Manninen¹, Liang Wang³, Gong-Hong Wei^{1,6*}

¹ Disease Networks Research Unit, Faculty of Biochemistry and Molecular Medicine & Biocenter Oulu, University of Oulu, Oulu, Finland.

² Division of Urology, Department of Surgery, Li Ka Shing Faculty of Medicine, the University of Hong Kong, Hong Kong.

³ Department of Tumour Biology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA.

⁴ Department of Urology, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China.

⁵ Department of Pathology, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China.

⁶ Fudan University Shanghai Cancer Center & MOE Key Laboratory of Metabolism and Molecular Medicine and Department of Biochemistry and Molecular Biology of School of Basic Medical Sciences, Shanghai Medical College of Fudan University, Shanghai, China.

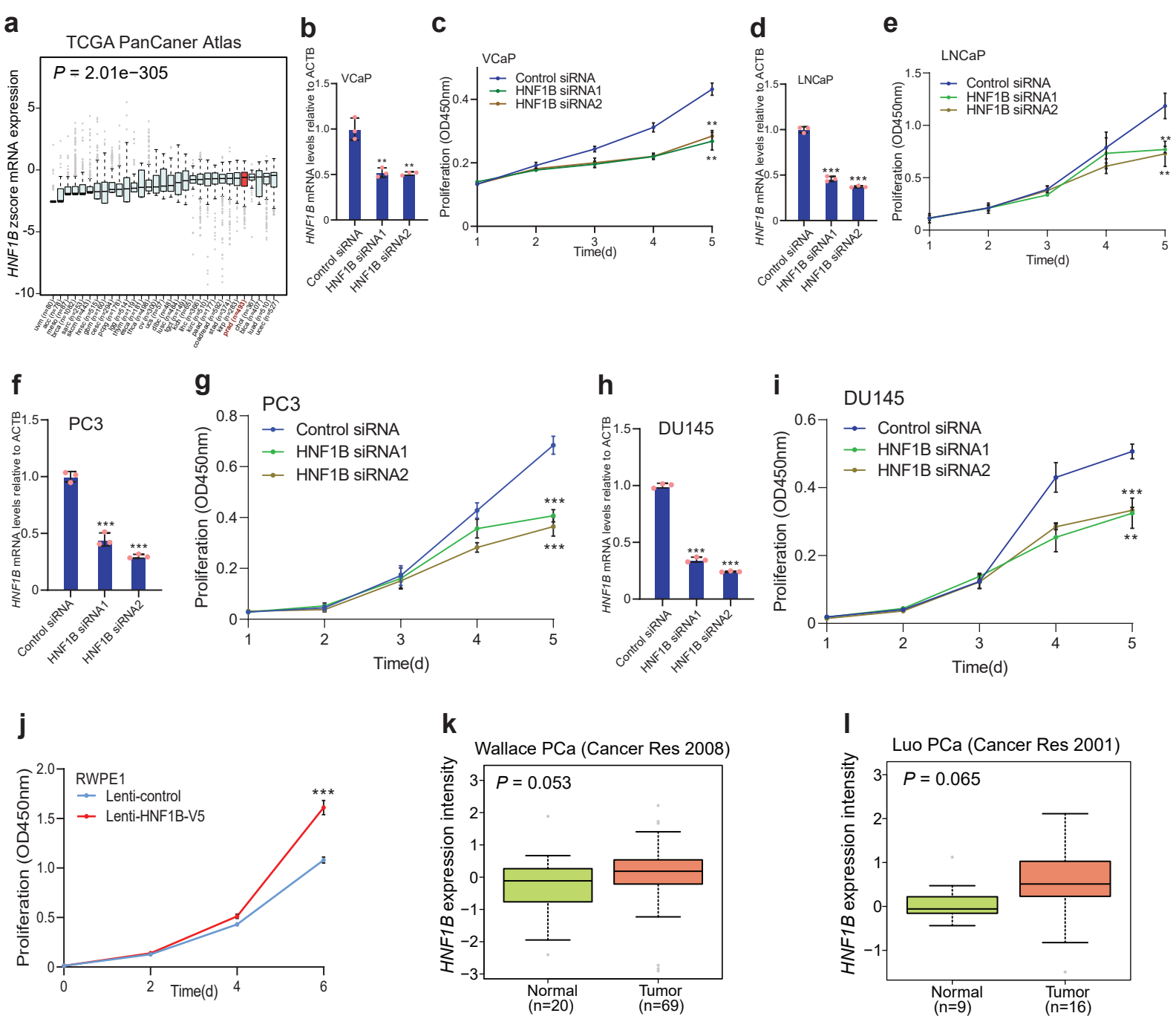
These authors contributed equally.

* Correspondence: gonghong_wei@fudan.edu.cn

This supplementary information file includes the following items:

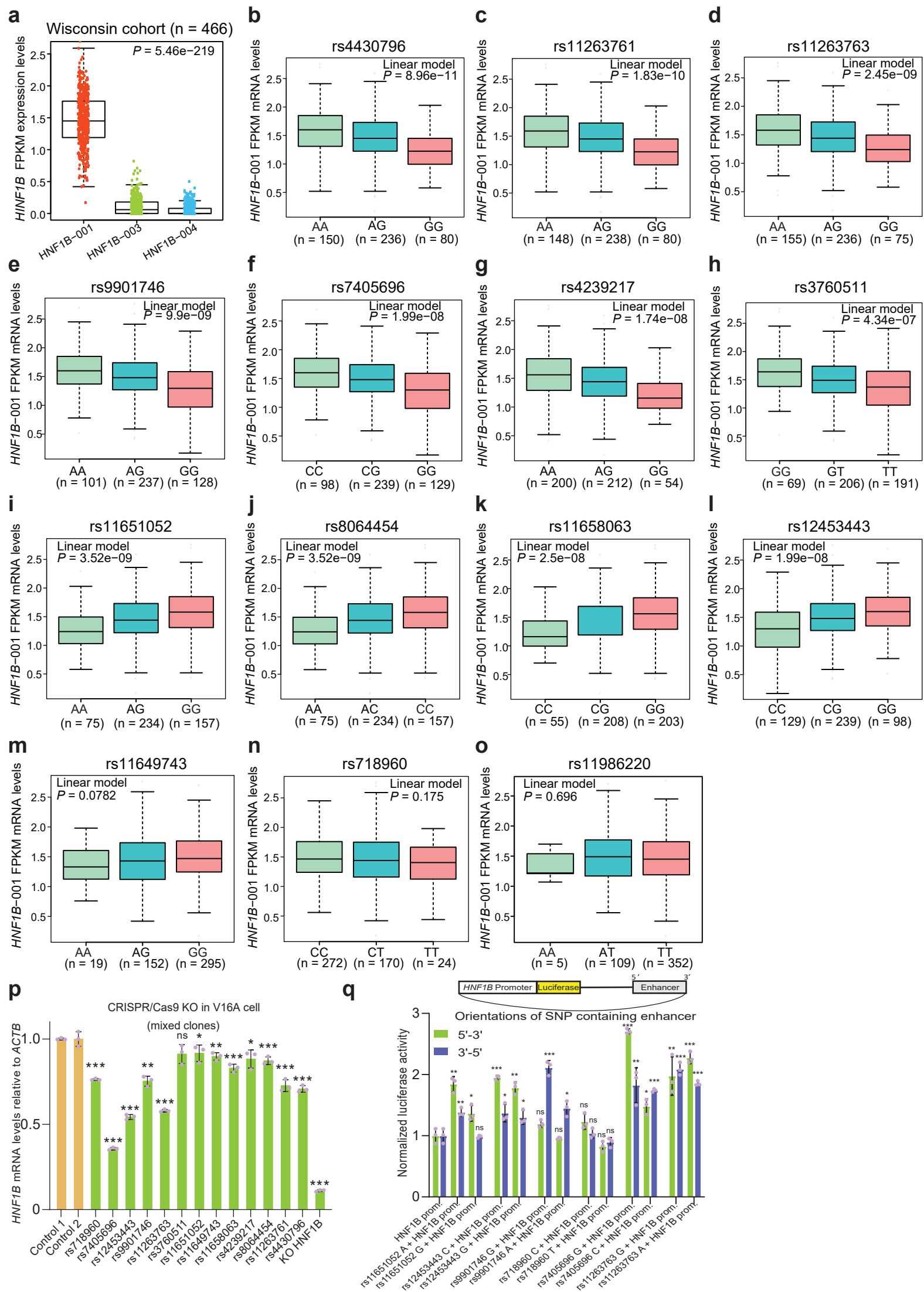
Supplementary Figures 1-10

Supplementary Tables 1-15



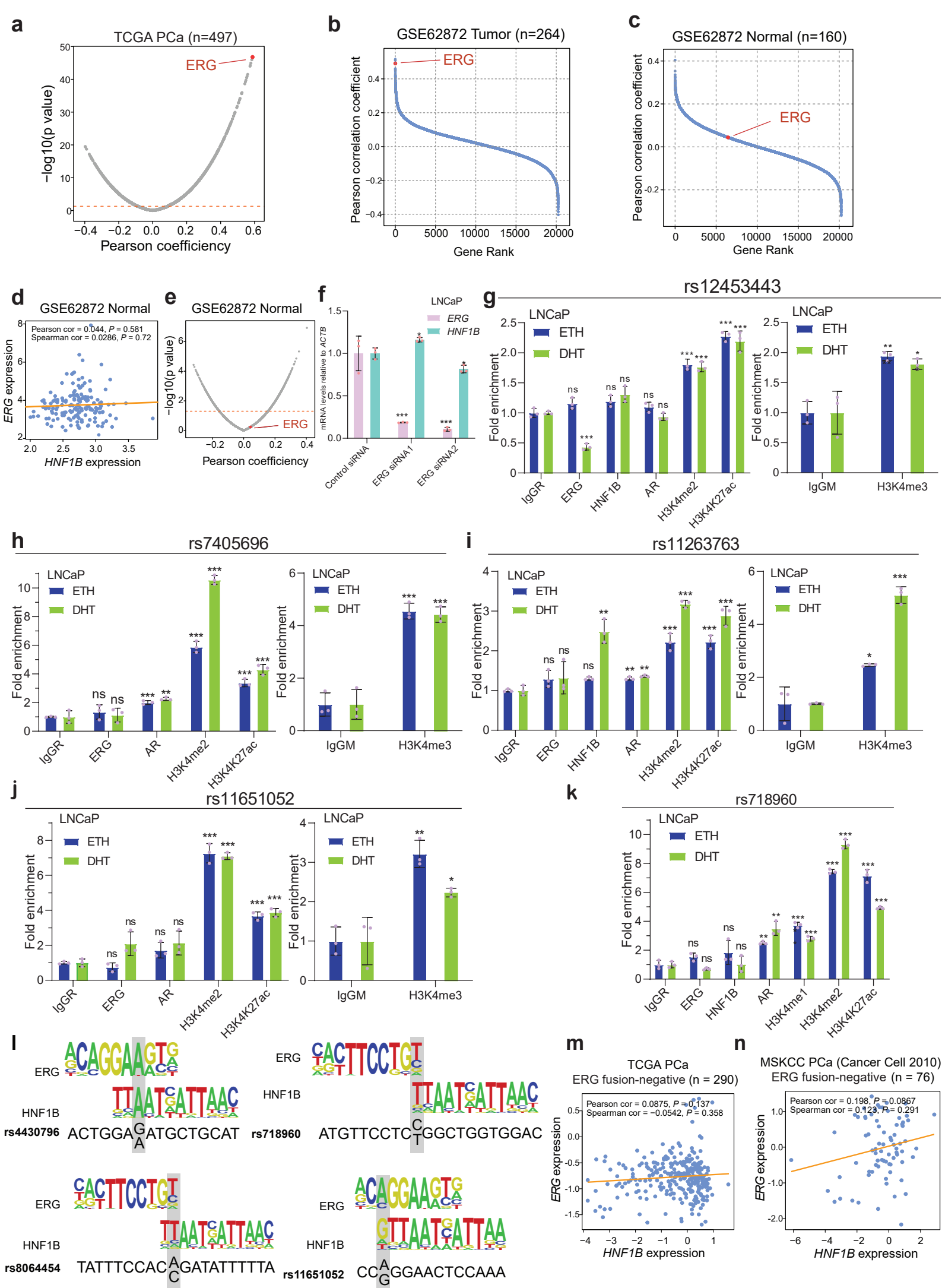
Supplementary Fig. 1: Association of HNF1B with PCa.

a, Expression profiles of HNF1B across 31 cancer types using The Cancer Genome Atlas (TCGA) RNA-seq data ($n=10,967$). P value was evaluated by Kruskal-Wallis test. **b-i**, HNF1B knockdown in VCaP (**b**), LNCaP (**d**), PC3 (**f**) and DU145 (**h**) diminish cell proliferation of VCaP (**c**), LNCaP (**e**), PC3 (**g**) and DU145 (**i**) measured by XTT colorimetric assays. (**b-i**) $n=3$ samples; (**b**) P values based on the order of appearance: 0,003, 0,002; (**c**) P values HNF1B siRNA1: 1,2E-03, HNF1B siRNA2: 1,4E-03; (**d**) P values based on the order of appearance: 2,5E-05, 5,5E-06; (**e**) P values HNF1B siRNA1: 4,6E-03, HNF1B siRNA2: 9,9E-03; (**f**) P values based on the order of appearance: 2E-04, 2E-05; (**g**) P values HNF1B siRNA1: 3,6E-04, HNF1B siRNA2: 4,2E-04; (**h**) P values based on the order of appearance: 2,7E-06, 3,8E-07; (**i**) P values HNF1B siRNA1: 3,1E-03, HNF1B siRNA2: 2,2E-04. **j**, Ectopic expression of *HNF1B* promotes RWPE1 cell proliferation measured by XTT colorimetric assays. $n=2$ samples; P values lenti-HNF1B-V5: 1,04E-05. **k-l**, *HNF1B* expression level is elevated in PCa tumours compared to normal prostate tissues in the Wallace ($n=89$) and Luo ($n=25$) cohorts. P values were evaluated by the two-sided Mann-Whitney U test. In **a** and **k-l**, the interquartile range (IQR) is depicted by the box with the median represented by the center line. Whiskers maximally extend to $1.5 \times$ IQR (with outliers shown). In **b-j**, $n=3$ technical replicates, error bars, mean $\pm$ SD, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, P values were assessed using two-tailed Student's t tests. Source data are provided in Source Data file.



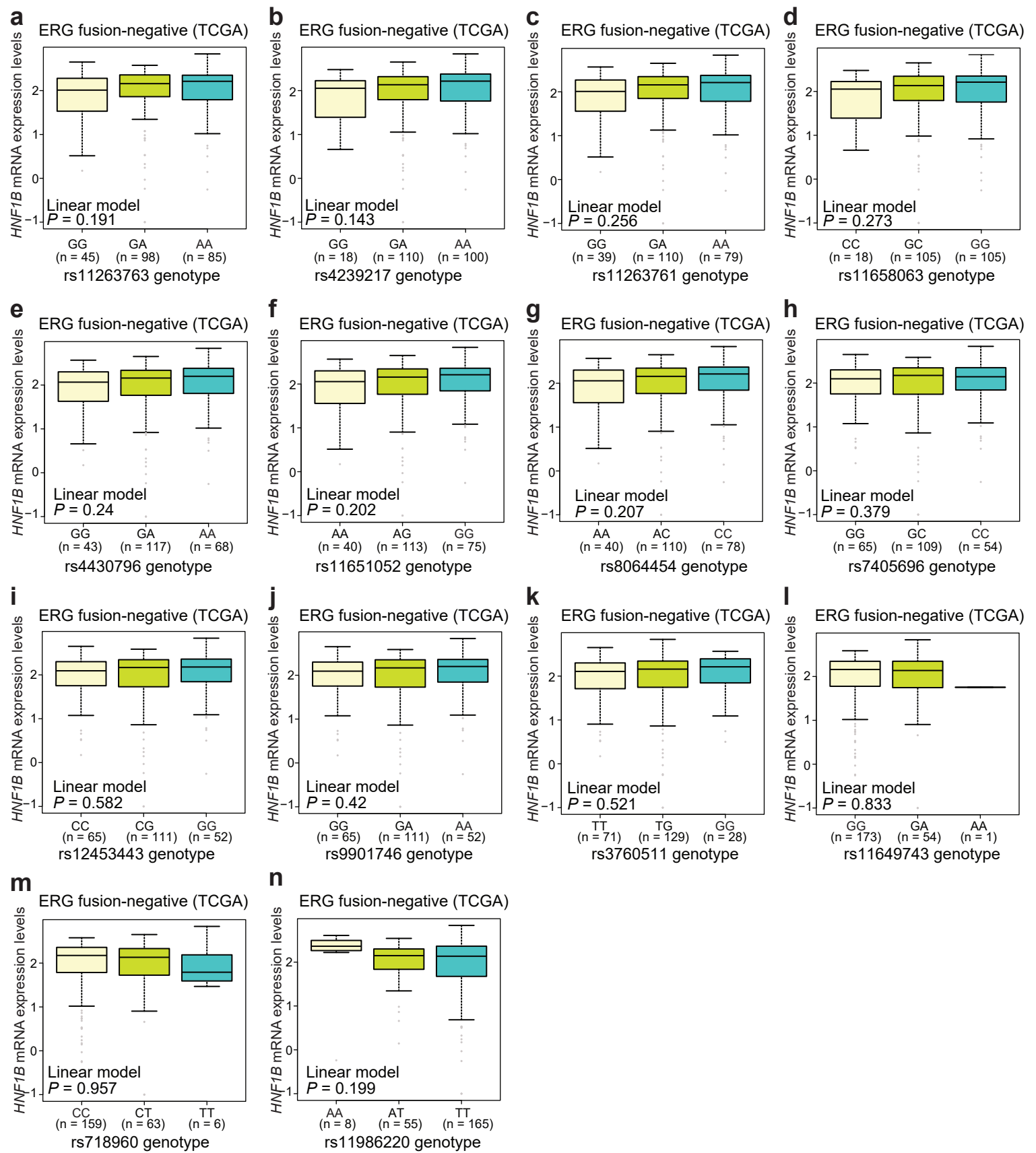
Supplementary Fig. 2: An eQTL analysis of the 17q12 locus variants in association with HNF1B expression.

a, HNF1B isoform 1 is dominantly expressed in normal prostate specimens in the Wiscousin cohort ($n=466$). P value was evaluated by Kruskal-Wallis Test. **b-o**, 11 of the 13 highly associated SNPs were found to have significant eQTL associations with HNF1B isoform 1 in the Wiscousin cohort ($n=466$). P values were assessed by the linear regression and adjusted by the false discovery rate. **p**, RT-qPCR analysis to determine the *HNF1B* mRNA expression levels in V16A cells with partial knockout for each individual SNP region or dampened *HNF1B* via CRISPR/Cas9 genome editing technology. $n=16$ samples; P values based on the order of appearance: $7E-07$, $3E-08$, $9E-07$, $1E-04$, $2E-07$, $0,053$, $0,041$, $2E-03$, $2E-04$, $0,019$, $7E-04$, $2E-04$, $2E-05$, $4E-09$. **q**, Enhancer reporter assays on six SNP regions located in *HNF1B* locus with different alleles and in both orientations (5'-3', 3'-5') cloned along with *HNF1B* promoter in LNCaP. Prom: Promoter. $n=25$ samples; P values based on the order of appearance: $1E-03$, $8,3E-03$, $0,022$, $0,86$, $3E-04$, $0,034$, $1,4E-03$, $0,048$, $0,097$, $4,8E-04$, $0,65$, $0,016$, $0,105$, $0,71$, $0,15$, $0,32$, $3,9E-05$, $3,1E-03$, $0,012$, $9,4E-04$, $8E-03$, $4,2E-04$, $2E-04$, $5,2E-04$. In **a-o**, the interquartile range (IQR) is depicted by the box with the median represented by the center line. Whiskers maximally extend to $1.5 \times$ IQR (with outliers shown). In **p**, **q**, $n=3$ technical replicates, error bars, mean $\pm$ SD, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, ns: non-significant, P values were assessed using two-tailed Student's t tests. Source data are provided in Source Data file.



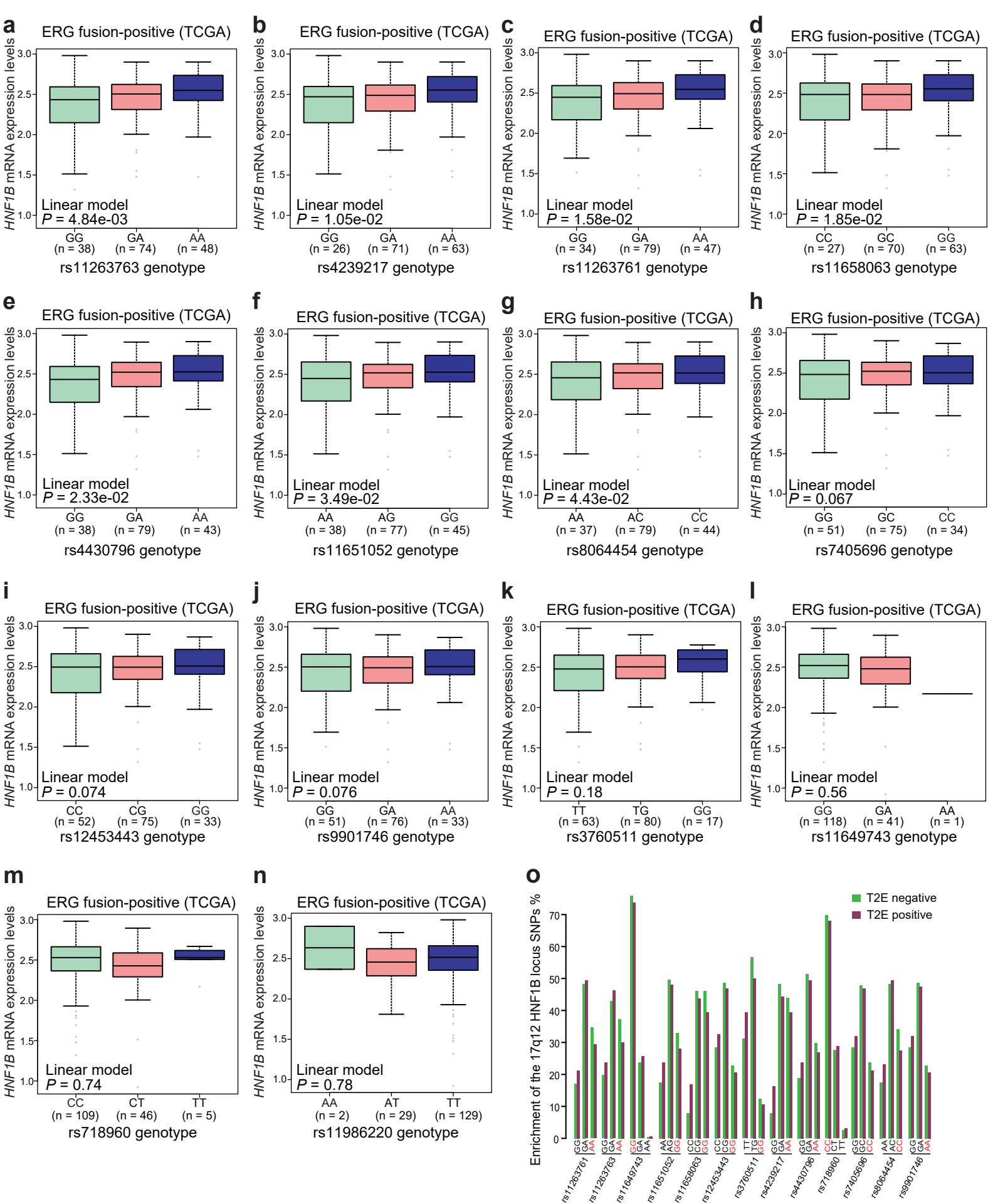
Supplementary Fig. 3: ERG is the top-ranking gene showing the highest coexpression correlation with HNF1B in PCa tumours.

a, Pearson correlation displaying *ERG* as the most co-expressed gene with *HNF1B* ($n=497$). **b**, *ERG* is the most co-expressed gene with *HNF1B* in an additional PCa tumour cohort GSE62872 ($n=264$). **c-e**, *ERG* shows no significant co-expression with *HNF1B* in normal prostate glands ($n=160$). In **a**, **e**, P values were assessed by the two-sided Pearson's product-moment correlation test. **f**, Decreased mRNA levels of *ERG* have no influence in *HNF1B* mRNA levels in *ERG* non fusion cell line, LNCaP. $n=3$ samples; P values based on the order of appearance: $2,3E-03$, $0,016$, $1,7E-03$, $0,016$. **g-k**, ChIP-qPCR determination of chromatin enrichment for *ERG*, *HNF1B*, AR, H3K4me1, H3K4me2, H3K27ac, and H3K4me3 at rs12453443 (**g**), rs7405696 (**h**), rs11263763 (**i**), rs11651052 (**j**) and rs718960 (**k**) enhancers of the 17q12/*HNF1B* regions in LNCaP cells treated with 100nM DHT and without (ETH-treated). **l**, SNPs rs4430796, rs718960, rs8064454 and rs11651052 reside within *HNF1B* and *ERG* DNA-binding motifs. (**g**, **i**) $n=16$ samples, (**h**, **j**, **k**) $n=14$ samples; (**g**) P values based on the order of appearance: $0,084$, $7,9E-05$, $0,05$, $0,24$, $0,16$, $0,45$, $3E-04$, $1E-04$, $3,8E-05$, $2,7E-04$, $1,2E-03$, $0,02$; (**h**) P values based on the order of appearance: $0,309$, $0,748$, $2E-04$, $8,6E-03$, $3E-05$, $7E-06$, $1E-04$, $5E-04$, $3E-04$, $7E-04$; (**i**) P values based on the order of appearance: $0,084$, $0,26$, $0,46$, $1,3E-03$, $7E-03$, $8,6E-03$, $6,7E-04$, $1,7E-05$, $2,4E-04$, $2,6E-04$, $0,017$, $2,2E-05$; (**j**) P values based on the order of appearance: $0,162$, $0,055$, $0,052$, $0,053$, $4,1E-05$, $3,5E-06$, $3,9E-05$, $1E-04$, $1,5E-03$, $0,025$; (**k**) P values based on the order of appearance: $0,078$, $0,083$, $0,17$, $0,92$, $1,1E-03$, $1,1E-03$, $1,7E-04$, $3,1E-04$, $4,8E-06$, $3,1E-06$, $3,5E-05$, $7,8E-06$. **m-n**, Expression levels of *ERG* are not significantly correlated with *HNF1B* in *TMPRSS2-ERG* fusion-negative group in the TCGA ($n=290$) or MSKCC cohort ($n=76$). In **d** and **m-n**, P values were assessed by the two-sided Pearson's product-moment correlation and Spearman's rank correlation rho tests. In **f-k**, $n=3$ technical replicates, error bars, mean $\pm$ SD, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, ns: non-significant, P values were evaluated using two-tailed Student's t -tests. Source data are provided in Source Data file.



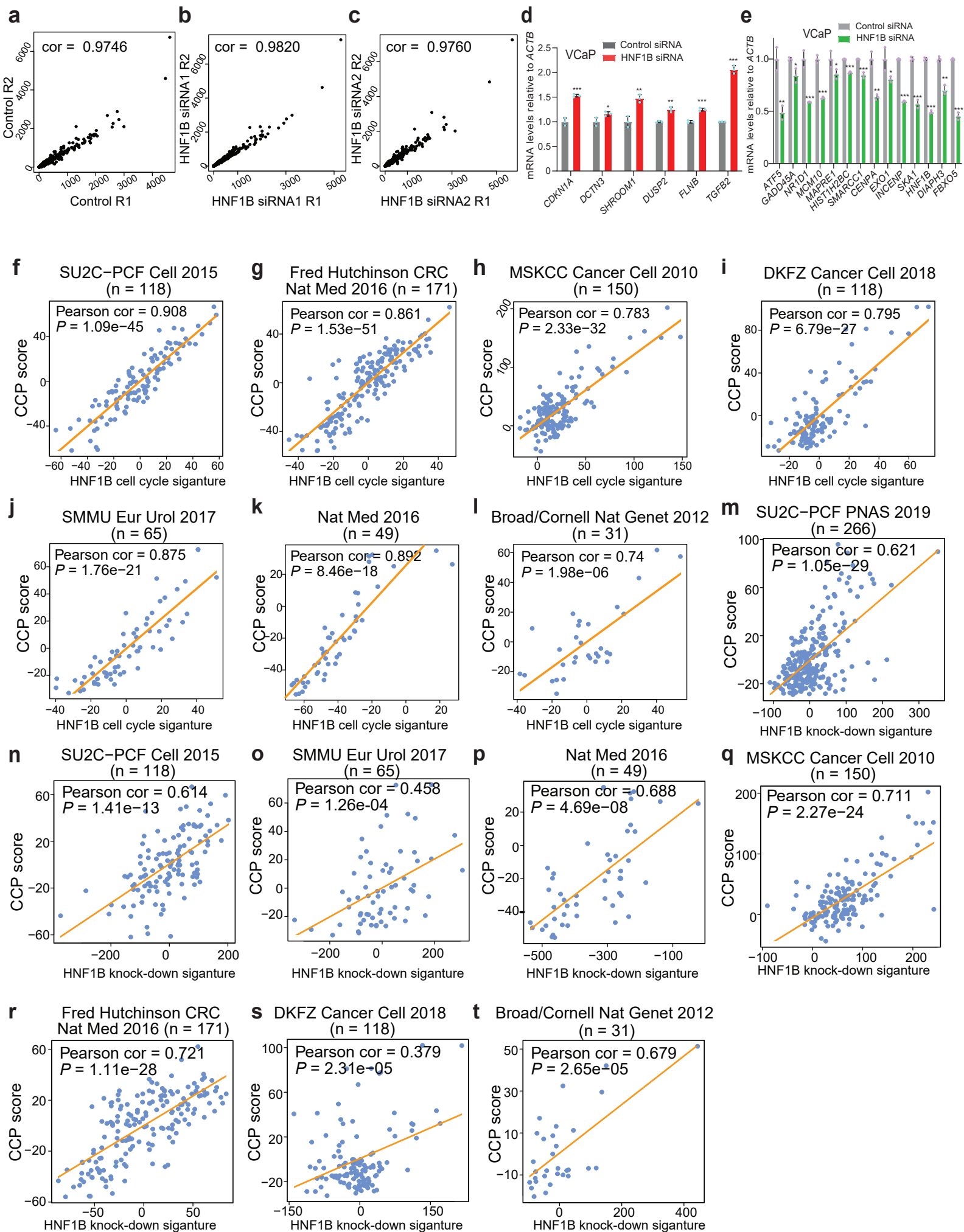
Supplementary Fig. 4: eQTL association with HNF1B in TMPRSS2-ERG fusion-negative group of PCa patients.

a-n, None of the 13 highly associated SNPs in the 17q12/*HNF1B* locus was found to have significant eQTL association with HNF1B in *TMPRSS2-ERG* fusion-negative PCa patients of TCGA cohort ($n=228$). Genotype to phenotype correlations were evaluated with linear regression. P values are adjusted by false discovery rate. The interquartile range (IQR) is depicted by the box with the median represented by the center line. Whiskers maximally extend to $1.5 \times$ IQR (with outliers shown). Source data are provided in Source Data file.



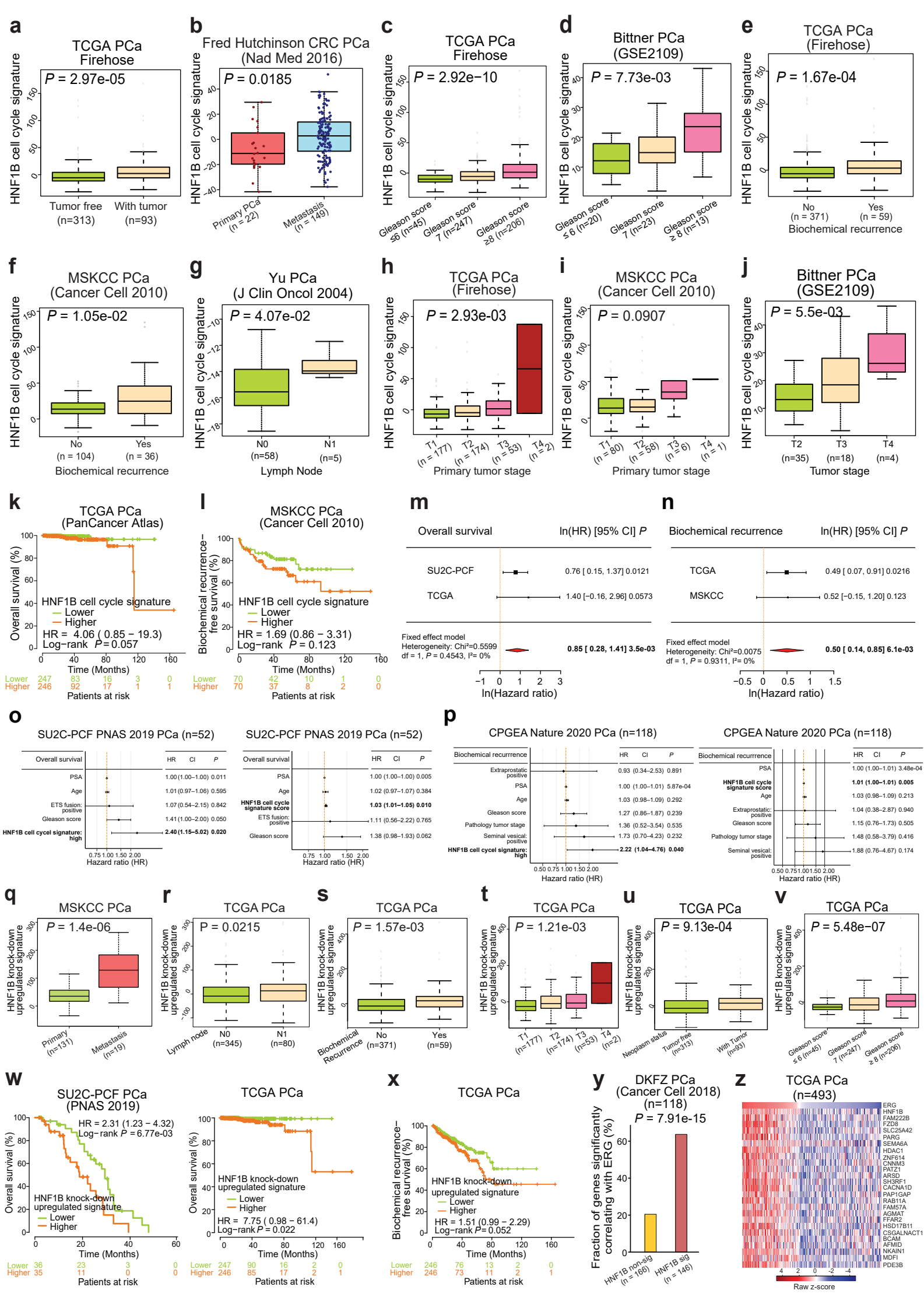
Supplementary Fig. 5: eQTL association with HNF1B in TMPRSS2-ERG fusion-positive group of PCa patients.

a-n, Four of the 13 highly associated SNPs at the 17q12/*HNF1B* locus were found to have significant eQTL association with *HNF1B* expression levels in PCa patient group with *TMPRSS2-ERG* fusion ($n=160$). Genotype to phenotype correlations were evaluated with linear regression. P values are adjusted by false discovery rate. The interquartile range (IQR) is depicted by the box with the median represented by the center line. Whiskers maximally extend to $1.5 \times$ IQR (with outliers shown). **o**, Enrichment of the 17q12 *HNF1B* locus SNPs in *TMPRSS2-ERG* fusion-positive and -negative tumours in the TCGA cohort ($n=388$). Source data are provided in Source Data file.



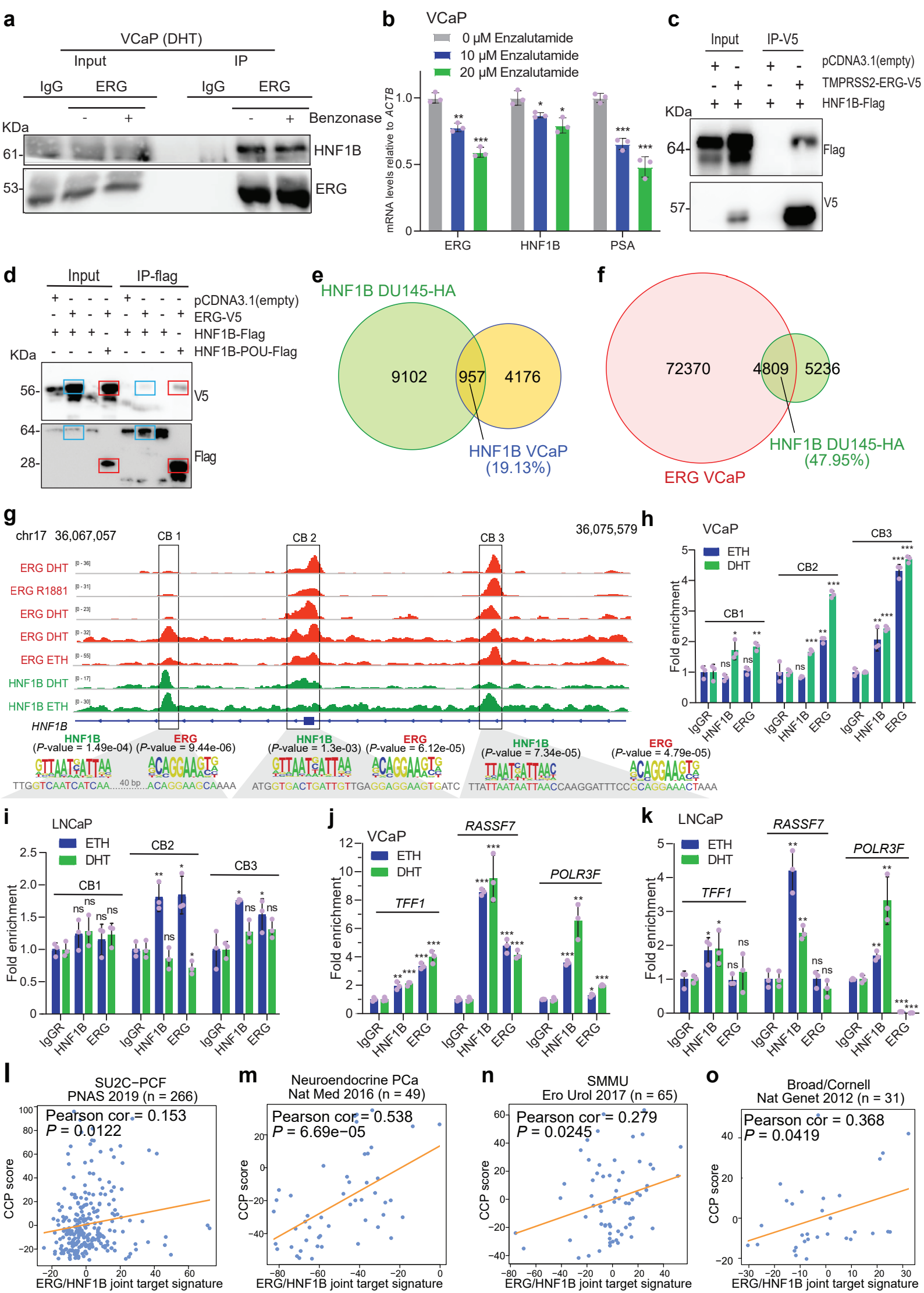
Supplementary Fig. 6: Experimental validation of HNF1B target genes and exploration of their clinical relevance.

a-c, RPKM expression correlation between two biological replicates of control siRNA, HNF1B siRNA1, or HNF1B siRNA2, respectively. **d-e**, RT-qPCR validation of HNF1B positively or negatively regulated target genes revealed by RNA-seq profiling. (**d**, **e**) $n=2$ samples; (**d**) P values based on the order of appearance: 4E-04, 0,04, 4E-04, 1E-04, 5E-04, 2E-05; (**e**) P values based on the order of appearance: 0,002, 0,016, 9E-04, 7E-05, 0,049, 2E-05, 6E-04, 1E-03, 0,042, 2E-07, 7E-05, 2E-06, 6E-04, 1E-05. **f-l**, HNF1B cell-cycle gene signature score indicates significant positive Pearson correlation with the Cell Cycle Progression (CCP) scores across several independent cohorts of PCa patients (**f**, $n=118$), (**g**, $n=171$), (**h**, $n=150$), (**i**, $n=118$), (**j**, $n=65$), (**k**, $n=49$), and (**l**, $n=31$). **m-t**, Scatter plots displaying significant positive linear correlation between HNF1B knock-down signature and CCP scores in eight independent cohorts of PCa patients (**m**, $n=266$), (**n**, $n=118$), (**o**, $n=65$), (**p**, $n=49$), (**q**, $n=150$), (**r**, $n=171$), (**s**, $n=118$), and (**t**, $n=31$). HNF1B knock-down signature score was calculated by the z-score sum of the 339 differentially expressed genes upon HNF1B siRNA knockdown followed by RNA-seq profiling. In **d**, **e**, $n=3$ technical replicates, error bars, mean $\pm$ SD, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, P values were assessed using two-tailed Student's t tests. In **f-t**, P values were assessed by the two-sided Pearson's product-moment correlation test. Source data are provided in Source Data file.



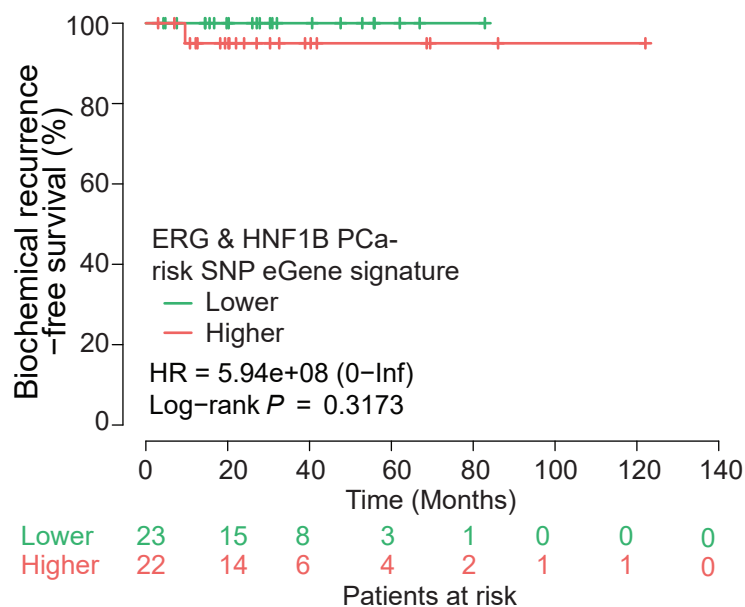
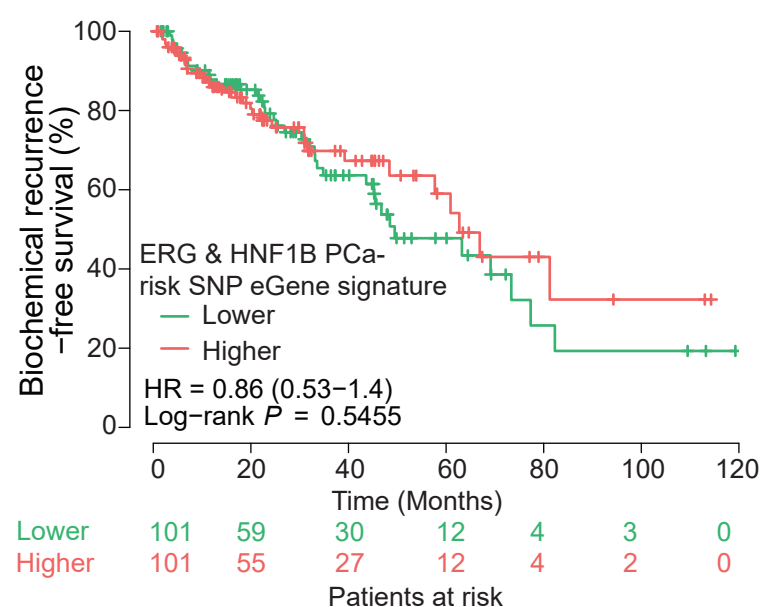
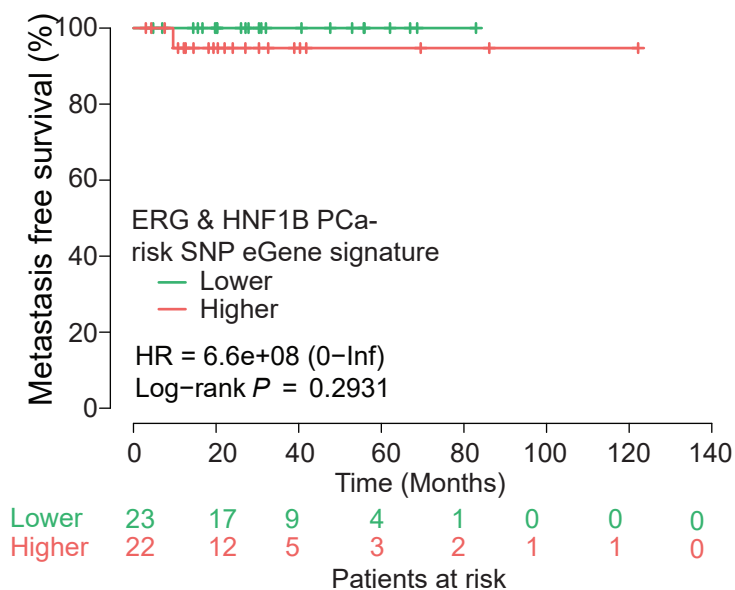
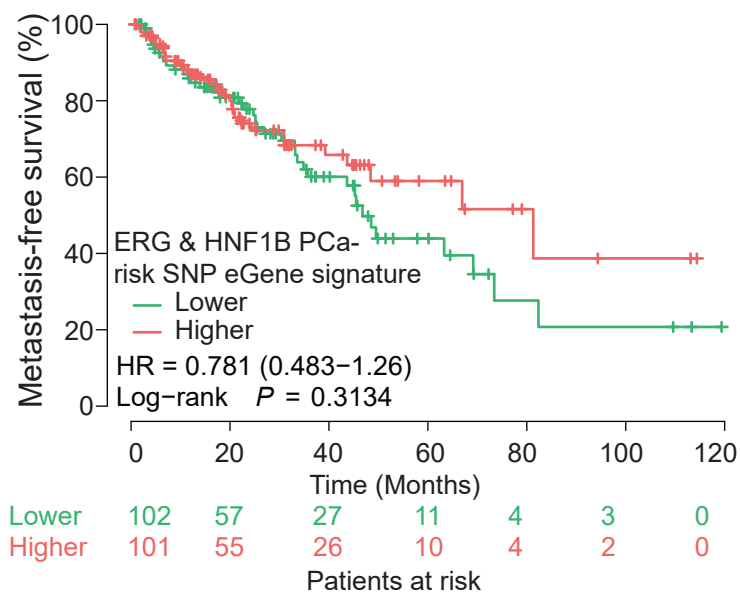
Supplementary Fig. 7: Characterization of clinical relevance for HNF1B target gene signatures.

a-j, The HNF1B cell cycle signature was upregulated upon PCa tumour development and progression (**a**, $n=406$), (**b**, $n=171$), (**c**, $n=498$), (**d**, $n=56$), (**e**, $n=430$), (**f**, $n=140$), (**g**, $n=63$), (**h**, $n=406$), (**i**, $n=145$), and (**j**, $n=57$). P values were calculated using the two-sided Mann-Whitney U test or Kruskal-Wallis test. **k, l**, Kaplan-Meier plots showing the overall survival (**k**, $n=493$) and biochemical relapse (**l**, $n=140$) of PCa patients. P values were calculated using log-rank test. **m, n**, Forest plots displaying the meta-analysis of the hazard ratio estimates for overall survival in SU2C-PCF ($n=71$) and TCGA ($n=493$) cohorts (**m**), and biochemical recurrence in TCGA ($n=492$) and MSKCC ($n=140$) cohorts (**n**). The horizontal error bars represent the 95% CI with the measure of centre as HR. The HR and 95% CI were presented in the form of natural logarithm ($\ln$). P values were calculated by the two-way Fixed-Effects Model. **o, p**, Multivariate analysis of the risk for patient overall survival (**o**, $n=52$) and biochemical recurrence (**p**, $n=118$) in PCa tumours with several clinical variables and the categorical or continuous value of the HNF1B cell cycle signature. P values were evaluated by the Cox's proportional hazards regression. Forest plot displays the measurement of HR with horizontal error bars representing the 95% CI. **q-v**, Correlations between HNF1B knock-down upregulated signature and PCa severity (**q**, $n=150$), (**r**, $n=425$), (**s**, $n=430$), (**t**, $n=406$), (**u**, $n=406$), and (**v**, $n=498$). P values were calculated using the two-sided Mann-Whitney U test or Kruskal-Wallis test. **w, x**, Kaplan-Meier plots showing rates of overall survival (**w**, $n=71, 493$) and biochemical relapse (**x**, $n=492$) of PCa patients. P values were calculated using log-rank test. **y**, Differentially expressed genes that are significantly co-expressed with *HNF1B* indicate increased likelihood to be greatly co-expressed with *TMPRSS2-ERG* ($n=118$). P value was evaluated by the two-sided Fisher's exact test. **z**, A 25-gene co-expression signature with *HNF1B* displayed a similar expression pattern with *ERG* ($n=493$). In **a-j** and **q-v**, the interquartile range (IQR) is depicted by the box with the median represented by the center line. Whiskers maximally extend to $1.5 \times$ IQR (with outliers shown). HR, hazard ratio; CI, confidence interval. Source data are provided in Source Data file.



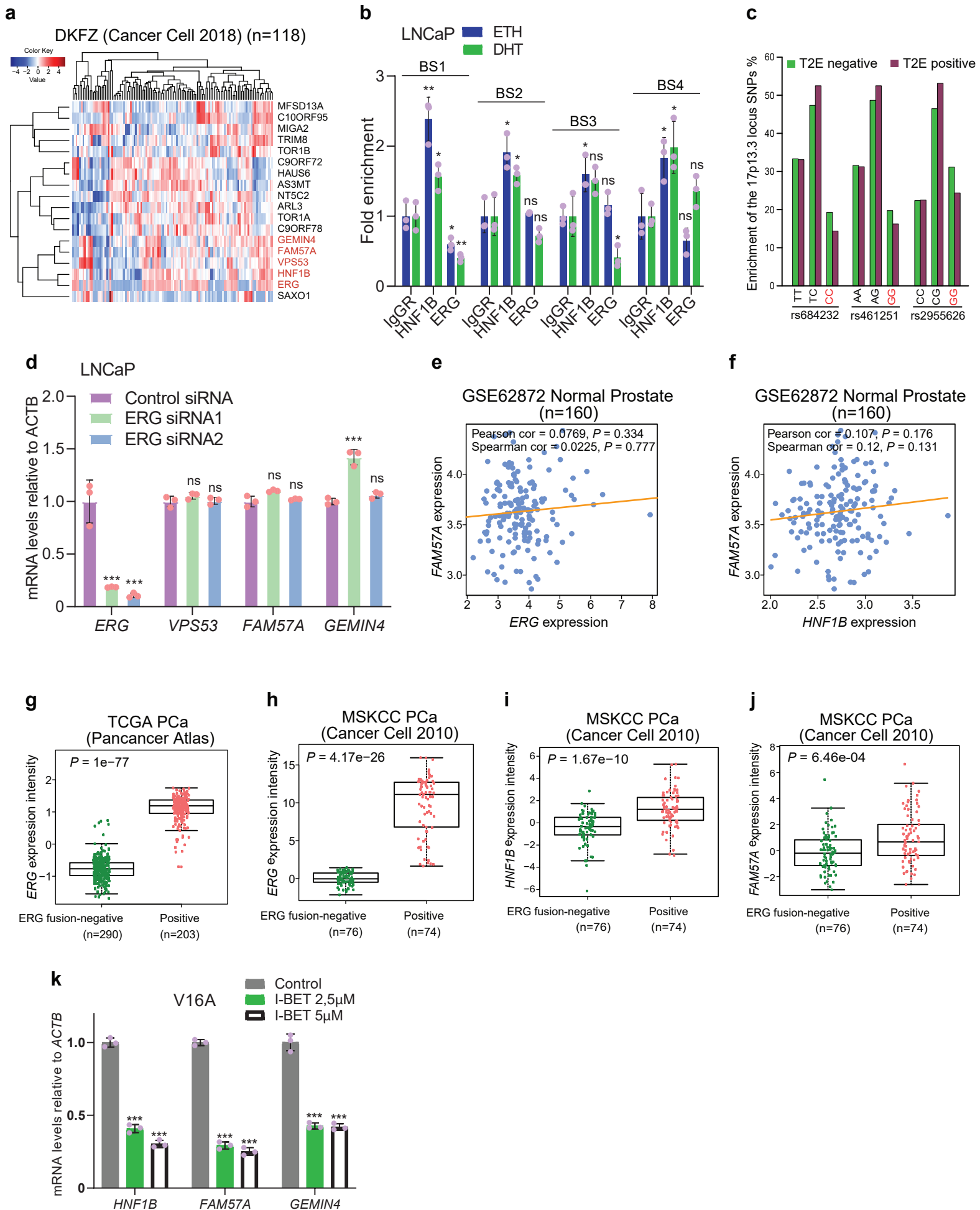
Supplementary Fig. 8: Physical interaction of HNF1B and ERG, and their direct target gene signature in the clinical setting.

a, Physical interaction of endogenous expression of HNF1B and ERG protein with benzonase treatment in VCaP treated with 100nM. **b**, RT-qPCR analysis of the mRNA levels of *ERG*, *HNF1B* and *PSA* with enzalutamide treatment for 72hrs. $n=3$ samples; P values based on the order of appearance: 2E-03, 2E-04, 0,018, 0,011, 4E-04, 4E-04. **c**, Exogenous interaction of recombinant HNF1B and TMPRSS2-ERG proteins in 293T cells. **d**, Interaction of recombinant HNF1B protein with exogenous ERG in vitro via the POU domain defined by co-IP with anti-flag antibody in 293T cells. **e**, HNF1B in VCaP cells shares 19.13% of common chromatin binding sites with that of HNF1B in DU145 cells. **f**, 47.95% of HNF1B binding sites in DU145 cells are co-occupied by that of ERG in VCaP cells. **g**, HNF1B and ERG chromatin binding sites of at the *HNF1B* locus. Upper: Common binding regions of HNF1B and ERG. Genomic coordinates in hg19 assembly. Lower: FIMO sequence scan for ERG and HNF1B motifs at common binding regions. **h**, **i**, ChIP-qPCR validation for chromatin enrichment of ERG and HNF1B at common binding (CB) sites at *HNF1B* locus in VCaP and LNCaP cells treated with 100nM DHT and without (ETH-treated). (**h,i**) $n=6$ samples; (**h**) P values based on the order of appearance: 0,23, 0,036, 0,71, 7E-03, 0,45, 6E-04, 6E-03, 7E-06, 9E-03, 7E-06, 2E-05, 5E-07; (**i**) P values based on the order of appearance: 0,18, 0,14, 0,37, 0,13, 5E-03, 0,31, 0,01, 0,042, 0,011, 0,089, 0,075, 0,047. **j**, **k**, ChIP-qPCR validation of **Fig. 5g** for chromatin enrichment of ERG and HNF1B at *TFF1*, *RASSF7* and *POLR3F* loci in VCaP and LNCaP cells treated with 100nM DHT and without (ETH-treated). (**j,k**) $n=6$ samples; (**j**) P values based on the order of appearance: 5E-03, 6E-04, 8E-05, 4E-04, 8E-07, 9E-04, 2E-04, 8E-05, 1E-05, 2E-03, 0,01, 2E-04; (**k**) P values based on the order of appearance: 0,03, 0,039, 0,87, 0,57, 1E-03, 1E-03, 0,96, 0,22, 1E-03, 5E-03, 1E-07, 1E-04. **l-o** Pearson correlation tests demonstrate significant positive linear correlation between HNF1B&ERG direct target gene signature scores and Cell Cycle Progression signature scores in four independent PCa cohorts (**l**, $n=266$), (**m**, $n=49$), (**n**, $n=65$), and (**o**, $n=31$). P values were assessed by the two-sided Pearson's product-moment correlation test. In **a**, **c** and **d**, representative experimental results of three independent co-immunoprecipitation assessments are shown. In **b** and **h-k**, $n=3$ technical replicates, error bars, mean $\pm$ SD, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, ns: non-significant, P values were assessed using two-tailed Student's t tests. Source data are provided in Source Data file.

aTCGA PCa
(Gleason score ≤ 6 , n=45)**b**TCGA PCa
(Gleason score ≥ 8 , n=202)**c**TCGA PCa
(Gleason score ≤ 6 , n=45)**d**TCGA PCa
(Gleason score ≥ 8 , n=203)

Supplementary Fig. 9: No prognostic values were found for the ERG&HNF1B binding sites enriched PCa risk eQTL gene signature in patient groups with Gleason score 6 or 8.

a-d, Kaplan-Meier curves depicting biochemical relapse and metastatic rates of PCa patients in TCGA cohort with Gleason score ≤ 6 ($n=45$) or ≥ 8 ($n=202$). Patients were stratified by median z-score sum of HNF1B&ERG eQTL gene (eGene) signature score. *P* values were examined by a log-rank test. Source data are provided in Source Data file.



Supplementary Fig. 10: Associations of the 17P13.3 locus risk genes with HNF1B and TMPRSS2-ERG fusion status in PCa clinical settings.

a, Heatmap displaying the expression levels of the HNF1B and ERG binding sites enriched PCa risk loci of the eQTL genes. Note that *FAM57A*, *GEMIN4* and *VPS53* show a similar expression pattern to HNF1B and ERG in the DKFZ cohort of PCa specimens ($n=118$). **b**, ChIP-qPCR validation of HNF1B and ERG binding at the regions within 17p13.3 in LNCaP cells treated with 100nM DHT and without (ETH-treated). $n=6$ samples; P values based on the order of appearance: $2.9E-03$, 0.024 , 0.034 , $7.9E-03$, 0.01 , 0.027 , 0.74 , 0.18 , 0.021 , 0.064 , 0.25 , 0.036 , 0.031 , 0.013 , 0.18 , 0.082 . **c**, Enrichment of the 17p13.3 locus SNPs in *TMPRSS2-ERG* fusion-positive and -negative tumours in the TCGA cohort. **d**, siRNA-ERG knockdown assays in LNCaP cells. $n=3$ samples; P values based on the order of appearance: $2.3E-03$, $1.7E-03$, 0.19 , 0.84 , 0.29 , 0.52 , $8.3E-04$, 0.069 . **e, f**, The expression levels of *FAM57A* and *ERG* or *HNF1B* are not significantly correlated in normal prostate glands ($n=160$). P values were assessed by the two-sided Pearson's product-moment correlation and Spearman's rank correlation rho tests. **g-j**, Expression levels of *ERG*, *HNF1B* and *FAM57A* are significantly elevated in the *TMPRSS2-ERG* fusion-positive PCa specimens compared to the *TMPRSS2-ERG* fusion-negative groups in the TCGA ($n=493$) and MSKCC ($n=150$) cohorts. P values were examined by the two-sided Mann–Whitney U test. The interquartile range (IQR) is depicted by the box with the median represented by the center line. Whiskers maximally extend to $1.5 \times$ IQR (with outliers shown). **k**, The mRNA levels of *HNF1B*, *FAM57A* and *GEMIN4* were downregulated in V16A cells treated with BRD4 antagonist I-BET for 24hrs. $n=3$ samples; P values based on the order of appearance: $1.6E-05$, $7.4E-06$, $2.7E-06$, $2.5E-06$, $8.5E-05$, $8.4E-05$. In **b**, **d** and **k**, $n=3$ technical replicates, error bars, mean $\pm$ SD, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, ns: non-significant; two-tailed Student's t tests. Source data are provided in Source Data file.

Supplementary Table 1.
Descriptive characteristics of Ruijin Biopsy cohort.

Characteristics	PCa Biopsy Group	Prostatectomy Group
Total number of patients	907	162
Age, yrs		
Median (IQR)	68 (63-75)	68.5 (64-73)
tPSA, ng/mL		
Median (IQR)	12.36 (7.58-25.05)	12.83 (7.48-27.92)
Prostate volume, mL		
Median (IQR)	81.55 (58.50-111.54)	63.95 (48.30-95.47)
ERG		
Negative	826 (91.1%)	123 (75.9%)
Positive	81 (8.9%)	39 (24.1%)
Family history of PCa		
Negative	884 (97.5%)	159 (98.1%)
Positive	23 (2.5%)	3 (1.9%)
Missing	0	0
Biopsy result		
Negative	489 (53.9%)	0
Positive	418 (46.1%)	162 (100%)
Grade Group		
1 (GS≤3+3)	95 (10.5%)	15 (9.3%)
2 (GS=3+4)	73 (8.0%)	55 (34.0%)
3 (GS=4+3)	80 (8.8%)	45 (27.8%)
4 (GS=8)	96 (10.6%)	10 (6.2%)
5 (GS=9 or 10)	60 (6.6%)	19 (11.7%)
Missing	8	18

Abbreviations: IQR, Interquartile range; tPSA, total prostate-specific antigen; PCa, prostate cancer; GS, Gleason Score.

Supplementary Table 2.**Association between PCa status and SNPs (791 PCa vs 752 non-PCa).**

SNP	BP	Alt allele	Ref allele	Freq (alt)	Freq in gnomAD [#]	OR (95%CI)	P
rs11649743	36074979	A	G	0.32	0.34	0.78 (0.66-0.91)	1.89E-03
rs718960*	36077279	T	C	0.29	0.34	N/A	N/A
rs11263761	36097775	G	A	0.30	0.32	0.91 (0.77-1.07)	0.25
rs4430796	36098040	G	A	0.25	0.28	0.87 (0.74-1.03)	0.11
rs4239217	36098987	G	A	0.25	0.29	0.85 (0.72-1.01)	0.06
rs8064454	36101586	A	C	0.24	0.28	0.80 (0.68-0.95)	0.01
rs7405696	36102035	G	C	0.32	0.36	0.80 (0.68-0.95)	8.89E-03
rs11651052	36102381	A	G	0.24	0.28	0.82 (0.69-0.97)	0.02
rs9901746	36103149	G	A	0.32	0.38	0.81 (0.69-0.96)	0.01
rs11263763	36103565	G	A	0.24	0.26	0.80 (0.67-0.95)	0.01
rs11658063	36103872	C	G	0.23	0.26	0.81 (0.68-0.97)	0.02
rs12453443	36104121	C	G	0.32	0.38	0.82 (0.70-0.97)	0.02
rs3760511	36106313	T	G	0.31	0.36	0.86 (0.73-1.01)	0.06
Additional							
rs79882976	36083710	C	A	0.05	0.06	0.93 (0.65-1.33)	0.69
rs11651496	36084244	C	T	0.05	0.06	0.93 (0.65-1.33)	0.69
rs3744764	36090396	C	T	0.07	0.06	1.12 (0.84-1.50)	0.43

*This SNP failed to pass quality control (80%<genotype call <90%) in the new dataset.

[#]Frequency of alt allele in East Asian population by gnomAD dataset (<https://gnomad.broadinstitute.org/>).

Association results were performed by logistic regression model (additive effect) adjusted for patient's age at diagnosis. SNPs in HNF1B region could be grouped into 2-3 clusters based on the linkage disequilibrium. A two-tailed $p < 0.05/2$ or $0.05/3$ was considered statistically significant for multiplicity.

Abbreviation: N/A, not applicable; Freq, frequency; PCa, prostate cancer.

Supplementary Table 3.**Association between ERG expression and SNPs in PCa samples (136 ERG+ vs 655 ERG-).**

SNP	BP	Alt allele	Ref allele	Freq (alt)	Freq in gnomAD [#]	OR (95%CI)	P
rs11649743	36074979	A	G	0.30	0.34	0.84 (0.63-1.12)	0.23
rs718960*	36077279	T	C	0.29	0.34	N/A	N/A
rs11263761	36097775	G	A	0.29	0.32	1.07 (0.80-1.43)	0.64
rs4430796	36098040	G	A	0.24	0.28	0.97 (0.72-1.31)	0.84
rs4239217	36098987	G	A	0.24	0.29	0.95 (0.71-1.29)	0.76
rs8064454	36101586	A	C	0.22	0.28	0.96 (0.70-1.30)	0.77
rs7405696	36102035	G	C	0.30	0.36	1.10 (0.82-1.47)	0.52
rs11651052	36102381	A	G	0.23	0.28	0.94 (0.69-1.28)	0.68
rs9901746	36103149	G	A	0.30	0.38	1.06 (0.78-1.42)	0.72
rs11263763	36103565	G	A	0.22	0.26	0.89 (0.64-1.23)	0.47
rs11658063	36103872	C	G	0.22	0.26	0.88 (0.64-1.22)	0.45
rs12453443	36104121	C	G	0.30	0.38	1.06 (0.78-1.42)	0.72
rs3760511	36106313	T	G	0.30	0.36	0.86 (0.73-1.01)	0.78
Additional							
rs79882976	36083710	C	A	0.05	0.06	2.00 (1.15-3.48)	0.01
rs11651496	36084244	C	T	0.05	0.06	2.00 (1.15-3.48)	0.01
rs3744764	36090396	C	T	0.07	0.06	1.67 (1.08-2.60)	0.02

*This SNP failed to pass quality control (80%<genotype call <90%) in the new dataset.

[#]Frequency of alt allele in East Asian population by gnomAD dataset (<https://gnomad.broadinstitute.org/>).

Association results were performed by logistic regression model (additive effect) adjusted for patient's age at diagnosis. SNPs in HNF1B region could be grouped into 2-3 clusters based on the linkage disequilibrium. A two-tailed $p < 0.05/2$ or $0.05/3$ was considered statistically significant for multiplicity.

Abbreviation: N/A, not applicable; Freq, frequency; PCa, prostate cancer.

Supplementary Table 4.
Experimental Models: Cell Lines.

Name	Source	Identifier
22Rv1	ATCC	Cat#CRL-2505
V16A	Hua et al. Cell 2018	
LNCaP	ATCC	Cat#CRL-1740
RWPE1	ATCC	Cat#CRL-11609
VCaP	ATCC	Cat#CRL-2876
293T	ATCC	Cat#CRL-11268
A549	ATCC	Cat#CCL-185
DU145	ATCC	Cat#HTB-81
PC3	ATCC	Cat#CRL-1435

Supplementary Table 5.
Gene cloning and related primers.

primer	Sequence 5' to 3'	target	restriction site	vector	experiment
rs718960.312BamHI-F	CGCGGATCCTCCTGGGCAATGTGGGGCTC	rs718960 region	BamHI/BamHI	pGL4.10[luc2]	Luciferase reporter assay
rs718960.312BamHI-R	CGCGGATCCGACCCAACAGGGCTGGGCTG				
rs7405696.403BamHI-F	CGCGGATCCAGCCCAACAATTGGGGAGGTTT	rs7405696 region	BamHI/BamHI	pGL4.10[luc2]	Luciferase reporter assay
rs7405696.403BamHI-R	CGCGGATCCTTAACGAACGCGCTGCGGAC				
rs11651052.359BamHI-F	CGCGGATCCGTCCGCAGCGCGTTCGTAAAG	rs11651052 region	BamHI/BamHI	pGL4.10[luc2]	Luciferase reporter assay
rs11651052.359BamHI-R	CGCGGATCCCGGGAGCGCCGGACTTGATT				
rs9901746.383BamHI-F	CGCGGATCCGCCGCGTTTACAACCTCACCA	rs9901746 region	BamHI/BamHI	pGL4.10[luc2]	Luciferase reporter assay
rs9901746.383BamHI-R	CGCGGATCCAGGCGGAAGGTGTGAGTCGC				
rs11263763.383.BamHI-F	CGCGGATCCGGGAGAAATCACCTCCGGGCG	rs11263763 region	BamHI/BamHI	pGL4.10[luc2]	Luciferase reporter assay
rs11263763.383.BamHI-R	CGCGGATCCCCAGGGCCGGGATCGTTTGT				
rs12453443.291BamHI-F	CGCGGATCCATAATTCCGGGCCTTGCATG	rs12453443 region	BamHI/BamHI	pGL4.10[luc2]	Luciferase reporter assay
rs12453443.291BamHI-R	CGCGGATCCAGCCCTTTGCCGGGAGGTTT				
HNF1Bpromoter(1)915.EcoRV-F	CGCTTGATATCGGAGCTTCACTTTGGCTACTCC	HNF1B promoter	EcoRV/HindIII	pGL4.10[luc2]	Luciferase reporter assay
HNF1Bpromoter(1)915.HindIII-R	CCCAAGCTTTTTCCAAGGACGGAAGAAAGAAG				
rs12453443G-F	CAGACCCGAGGCAGgGCGCAGGAAAAATG	rs12453443 region	NA	pGL4.10[luc2]	site-directed mutagenesis C→G
rs12453443G-R	CATTTTTCCTGCGCcCTGCCTCGGGTCTG				
rs9901746A-F	CGCCTGCGGCCGaGTGGGAGGGCCT	rs9901746 region	NA	pGL4.10[luc2]	site-directed mutagenesis G→A
rs9901746A-R	AGGCCCTCCCACtCGGCCGCAGGCG				
rs11263763A-F	CAGCCCCGAAATCGCaGCGCTTCAGTGTCGCG	rs11263763 region	NA	pGL4.10[luc2]	site-directed mutagenesis G→A
rs11263763A-R	CGGACACTGAAGCGCtGCGATTTTCGGGGCTG				
rs718960T-F	CAGGAATGTTCTCtGGCTGGTGGACCTC	rs718960 region	NA	pGL4.10[luc2]	site-directed mutagenesis C→T
rs718960T-R	GAGGTCCACCAGCCaGAGGAACATTCTG				
rs7405696C-F	AGGTTCCCTAACACcGAGAAAGGCACCGG	rs7405696 region	NA	pGL4.10[luc2]	site-directed mutagenesis G→C
rs7405696C-R	CCGGTGCCTTTCTCgGTGTTAGGGAACCT				
rs11651052G-F	GGTTTACTAGAACTCcGGAACCTCAAAAACGCG	rs11651052 region	NA	pGL4.10[luc2]	site-directed mutagenesis A→G
rs11651052G-R	CGCGTTTTGGAGTTCCcGGAGTTCTAGTAACACC				
GEMIN4pr(1)526EcoRV-F	CGCTTGATATCGGAACGCTGCTGTCTGGAGG	GEMIN4 promoter	EcoRV/HindIII	pGL4.10[luc2]	Luciferase reporter assay
GEMIN4pr(1)526HindIII-R	CCCAAGCTTGGCGGCGACGCCGGC				

VPS53pr(1)603EcoRV-F	CGCTTGATATCCTTCTGCGCAGAGCCCGG	VPS53 promoter	EcoRV/HindII	pGL4.10[luc2]	Luciferase reporter assay
VPS53pr(1)603HindIII-R	CCCAAGCTTTCCGCCACCCGGCCC				
FAM57Apr(1)604EcoRV-F	CGCTTGATATCGGAGAGGCCACATCCTGAGG	FAM57A promoter	EcoRV/HindII	pGL4.10[luc2]	Luciferase reporter assay
FAM57Apr(1)604HindIII-R	CCCAAGCTTCGGGGCTGCGGGTCCG				
rs684232 and rs461251.338BamHI-F	CGCGGATCCAAAGCCAGATGACGCAGC	rs684232 and rs461251 region	BamHI/BamHI	pGL4.10[luc2]	Luciferase reporter assay
rs684232 and rs461251.338BamHI-R	CGCGGATCCGGAGTTTCCCGGTTACCT				
rs2955626.221BamHI-F	CGCGGATCCCAGCACAGCAACTCCCTC	rs2955626	BamHI/BamHI	pGL4.10[luc2]	Luciferase reporter assay
rs2955626.221BamHI-R	CGCGGATCCCCATCCTAAGGACGCGCTGA				
HNF1B-D-BamHI-F	CGGGATCCatggtgtccaagctcacgtcg	HNF1B sequence of D domain	BamHI/BstZ17I	pcDNA3.1/flag	Co-immunoprecipitation assay
HNF1B-D-BstZ17I-R	GCGTATACtgtgtcatagtcgtcgccgtcc				
HNF1B-POU-BamHI-F	CGGGATCCatgcctccatcctcaaggagct	HNF1B sequence of POU domain	BamHI/BstZ17I	pcDNA3.1/flag	Co-immunoprecipitation assay
HNF1B-POU-BstZ17I-R	GCGTATACcagcttttgcggaatgcctc				
HNF1B-T-BamHI-F	CGGGATCCatggccatggacgcctatagctc	HNF1B sequence of T domain	BamHI/BstZ17I	pcDNA3.1/flag	Co-immunoprecipitation assay
HNF1B-T-BstZ17I-R	GCGTATACccaggctttagaggacactgt				
HNF1B-BamHI-F	CGGGATCCatggtgtccaagctcacgtcg	HNF1B	BamHI/BstZ17I	pcDNA3.1/flag	Co-immunoprecipitation assay
HNF1B-BstZ17I-R	GCGTATACccaggctttagaggacactgt				
ERG-KpnI-F	GGGGTACCatgattcagactgtcccgagc	ERG	KpnI/XbaI	pcDNA3.1/V5	Co-immunoprecipitation assay
ERG-XbaI-R	GCTCTAGAgtagtaagtcccagatgagaagg				
TMP-ERGKpnI-F	GGGGTACCtaggcgcgagctaagcaggag	TMPRSS2-ERG	KpnI/XbaI	pcDNA3.1/V5	Co-immunoprecipitation assay
TMP-ERG-XbaI-R	GCTCTAGAgtagtaagtcccagatgagaagg				
HNF1BV5-F	AGCTTTGTTTAAACatggtgtccaagctcacgtc	HNF1B-V5	PmeI/BamHI	pLVET-IRES-GFP	Cell proliferation
HNF1BV5-R	CGGGATCCTCACGTAGAATCGAGACCG				

Supplementary Table 6.
Antibodies.

Name	Catalog #	Manufacturer	Application
Rabbit polyclonal anti- HNF1B	Cat#sc-22840X	Santa Cruz Biotechnology	ChIP/WB
Rabbit polyclonal IgG	Cat#sc-2027X	Santa Cruz Biotechnology	ChIP
Mouse monoclonal anti-V5	Cat#R960-25	Invitrogen (ThermoFisher)	co-IP/WB
Mouse monoclonal anti-V5-HRP	Cat#R961-25	Invitrogen (ThermoFisher)	WB
Mouse monoclonal anti-ERG	Cat#sc-376293	Santa Cruz Biotechnology	WB
Mouse monoclonal anti- HNF-1B	Cat#sc-130407	Santa Cruz Biotechnology	WB
Rabbit monoclonal IgG	Cat#ab172730	Abcam	ChIP/co-IP
Rabbit monoclonal anti-ERG	Cat#ab92513	Abcam	ChIP/co-IP/WB
Mouse monoclonal anti-FLAG	Cat#F1804	Sigma-aldrich	co-IP/WB
Goat anti-mouse IgG (H+L) secondary antibody, HRP	Cat#32430	Thermo Fisher	WB
Goat anti-rabbit IgG (H+L) secondary antibody, HRP	Cat#32460	Thermo Fisher	WB
Anti-rabbit Androgen Receptor	Cat#ab108341	Abcam	ChIP
H3K4me1	Cat#ab8895	Abcam	ChIP
H3K4me2	Cat#39141	ACTIVE MOTIF	ChIP
H3K4me3	Cat#ab12209	Abcam	ChIP
H3K27ac	Cat#ab4729	Abcam	ChIP
Mouse polyclonal IgG	Cat#sc-2025	Santa Cruz Biotechnology	ChIP

Supplementary Table 7.
Target sequences of siRNA and shRNA.

Name	Catalog #	Manufacturer	Sequence
Control-shRNA1	SHC002	Merck	CAACAAGATGAAGAGCACCAA
HNF1B-shRNA1	TRCN0000017508	Merck	CCGTACTGTCTATGTTGTGAT
HNF1B-shRNA2	TRCN0000017511	Merck	CCGACAATTCAACCAGACAGT
VPS53-shRNA1	TRCN0000180068	Merck	CCAGAAGTACCTCCGAGAATA
VPS53-shRNA2	TRCN0000148267	Merck	GATCAACCAAAGAAGCCTAAA
FAM57A-shRNA1	TRCN0000134609	Merck	GTGGTTTCTGATTCCATACAT
FAM57A-shRNA2	TRCN0000136978	Merck	CCTCAGACTTTGGGTATTGAT
GEMIN4-shRNA1	TRCN0000007892	Merck	GCAACCGTGTATCTGGACAAA
GEMIN4-shRNA2	TRCN0000007895	Merck	GTTTGTTTACACCCAGGTGTT
AllStars Hs Cell Death Control siRNA	SI04381048	Qiagen	
HNF1B-siRNA1	SI03040765	Qiagen	
HNF1B-siRNA2	SI03056956	Qiagen	
control siRNA	1027280	Qiagen	
ERG-siRNA1	SI03064726	Qiagen	
ERG-siRNA2	SI03089443	Qiagen	

Supplementary Table 8.
Oligonucleotides used in this study.

Experiment	primer description	Sequence 5' to 3'
RT-qPCR	hactin188RT-f	AGAAAATCTGGCACCACACC
RT-qPCR	hactin188RT-r	AGAGGCGTACAGGGATAGCA
RT-qPCR	GEMIN4-RT156f	ATGGACCTAGGACCCTTGAACAT
RT-qPCR	GEMIN4-RT156r	CTCCCTTAAGGCCTCCACGA
RT-qPCR	FAM57A-RT70f	AGCCTGCTCCAAGTACCCTT
RT-qPCR	FAM57A-RT70r	GAGGAGCTACGAGGAAGGCA
RT-qPCR	VPS53-RT69f	CCCTTGCTGCTGCTACAGTC
RT-qPCR	VPS53-RT69r	GAGCAAGAGTCGTCACGCAT
RT-qPCR	ERG exon56-F	CGCAGAGTTATCGTGCCAGCAGAT
RT-qPCR	ERG exon56-R	CCATATTCTTTCACCGCCCACTCC
RT-qPCR	HNF1BRT145-F	ACCAACAAGAAGATGCGCCG
RT-qPCR	HNF1BRT145-R	GCAAACATTCTGCCCTGTTGC
RT-qPCR	HNF1BRT92-F	AGGGCACCCCTATGAAGACC
RT-qPCR	HNF1BRT92-R	TGTCTGGTTGAATTGTCGGAGG
RT-qPCR	ATF5RT153-F	CCTGGCAGGTGATGGCTTCT
RT-qPCR	ATF5RT153-R	TCCTTCTTGAGGAGGGAGGC
RT-qPCR	GADD45ART92-F	CTCCTGCTCTTGAGACCGA
RT-qPCR	GADD45ART92-R	TGTGGATTTCGTCACCAGCAC
RT-qPCR	CENPART90-F	TCCTTAGGCGCTTCCTCCC
RT-qPCR	CENPART90-R	CAAGAGGTGTGTGCTCTTCTGA
RT-qPCR	EXO1RT244-F	GCTCGGCTAGGAATGTGCAG
RT-qPCR	EXO1RT244-R	TGGCCCGAATAAACCCTTG
RT-qPCR	NR1D1RT224-F	CTGCCCAGCGTCATAACGAG
RT-qPCR	NR1D1RT224-R	G TTCATAGGACATGCCAGCAGAA
RT-qPCR	DIAPH3RT190-F	GCTTTTAAGTCTCAGTTTGGTGCC
RT-qPCR	DIAPH3RT190-R	GACCACTGAATGGCATCCGC
RT-qPCR	MCM10RT118-F	TGAAGAGTTGCAAGAGGAATTAAGG
RT-qPCR	MCM10RT118-R	AGGGGATTTTTGCAGACGGG
RT-qPCR	FBXO5RT242-F	AGATCGGGAGATGCTGAAGGAA
RT-qPCR	FBXO5RT242-R	TGTACAAC TGGAATGCCCCC
RT-qPCR	SKA1RT146-F	ACGAAGGATACCAAAGGTCGT
RT-qPCR	SKA1RT146-R	GTAAGTCCTCCCCCTCGGAC
RT-qPCR	MAPRE1RT139-F	GCTCCTTCCCTTGTTGCTCC
RT-qPCR	MAPRE1RT139-R	GGTTCTTTCGCACCACACCA
RT-qPCR	SMARCC1RT226-F	GAAAGAAGCTCTAGAACAACAGAGG
RT-qPCR	SMARCC1RT226-R	G TAGGGAGGGGGCTGTTGAT
RT-qPCR	INCENPRT106-F	GAGCTGATGCCCAAACACCT
RT-qPCR	INCENPRT106-R	TGCGGGATAACCTTCTCCTGAT
RT-qPCR	FLNBRT124-F	TGGTCCAGTCGGAGATTGGT
RT-qPCR	FLNBRT124-R	ATGCCACCATAACCTGCATCC
RT-qPCR	CDKN1ART171-F	AGGGGACAGCAGAGGAAGAC

RT-qPCR	CDKN1ART171-R	AGAAGATCAGCCGGCGTTTG
RT-qPCR	DCTN3RT151-F	GCTCACGGAAGGTGGCTGA
RT-qPCR	DCTN3RT151-R	GGTATGGCAATGCGGTCGAT
RT-qPCR	SHROOM1RT221-F	GCCCTTGGAGTTCCAGCATC
RT-qPCR	SHROOM1RT221-R	GGGGAACAGCCTGGACAATG
RT-qPCR	DUSP2RT187-F	CAGTCACTCGTCAGACCTGC
RT-qPCR	DUSP2RT187-R	CTTCACCCAGTCAATGAAGCCT
RT-qPCR	HIST1H2BCRT129-F	GCTCCAAGAAGGCAGTGACC
RT-qPCR	HIST1H2BCRT129-R	AAGAGATGCCAGTGTGCGGA
RT-qPCR	TGFB2RT119-F	CAACCCAGCGCTACATCGAC
RT-qPCR	TGFB2RT119-R	TCCCAGGTTTCCTGTCTTTATGGT
ChIP-qPCR	TFF1RT85-F	TATGTAAAACAGTGGCTCCTGGC
ChIP-qPCR	TFF1RT85-R	CCAGACAGGTAAGGCGTG
ChIP-qPCR	RASSF7RT88-F	GAGCCAGGTGAGGCGAGTA
ChIP-qPCR	RASSF7RT88-R	CACCAGCTTCCTCTTGCCAC
ChIP-qPCR	POLR3FRT84-F	CCGAAAGGCAAGGAAGGACC
ChIP-qPCR	POLR3FRT84-R	CGGGGAACCAGTGGAGGATA
ChIP-qPCR	CB1RT105-F	TTGCTAACGTTGAGGTGTGCAG
ChIP-qPCR	CB1RT105-R	CTGTGGACACGCTTTTGTCTC
ChIP-qPCR	CB2RT88-F	GGCCCAGGTGTACTCACCATT
ChIP-qPCR	CB2RT88-R	CGGTTTTACAGCAAGTCTCCC
ChIP-qPCR	CB3RT89-F	TCACAAGGAGGCTTCTCCCC
ChIP-qPCR	CB3RT89-R	AAGCCTTTTGTTCCTGCGG
ChIP-qPCR	BS1-140-F	GCAGCGACCTGGTGAGC
ChIP-qPCR	BS1-140-R	CACCTGACGTGACAGGAACG
ChIP-qPCR	BS2-127-F	CGCGTCGATGTGGAGAGGAT
ChIP-qPCR	BS2-127-R	GGGTGATCGGTGATCTGGCT
ChIP-qPCR	BS3-83-F	GCCCAGGGCGGAAAAGG
ChIP-qPCR	BS3-83-R	GAGGGGCTTCGGAAACTCG
ChIP-qPCR	BS4-149-F	GCTTAGGCCTGCTCACAACC
ChIP-qPCR	BS4-149-R	TAGTGAGCGACGTCCGGGTA
ChIP-qPCR	N.C-159-F	CTGGGCATGATGAGTGAGCG
ChIP-qPCR	N.C-159-R	TCGGGCCCTGTGACTGTTTA
ChIP-qPCR	rs12453443-215F	GAACCTCGCCTCCCAGATCCC
ChIP-qPCR	rs12453443-215R	GACTGAAGGTGCGAGTGGGT
ChIP-qPCR	rs7405696-297F	AGCCGGGAGAGCTGGTTATT
ChIP-qPCR	rs7405696-297R	CGGGGGCGTCTAGAAAGTGG
ChIP-qPCR	rs718960-220F	CCACGATGTGGCTTTTGGCA
ChIP-qPCR	rs718960-220R	GAGCCCCAGCCAGTTGAGTA
ChIP-qPCR	rs11263763-272F	GGCTAGAGGGCAAACCGC
ChIP-qPCR	rs11263763-272R	CAGGGCCGGGATCGTTTGTG
ChIP-qPCR	rs11651052-276F	TCGAGGCCGCGAGAAAGG
ChIP-qPCR	rs11651052-276R	GGGAGCGCCGACTTGATTA
ChIP-qPCR	CHIPneg-94F	GGCAGCAGAACATCACTGGG

ChIP-qPCR	CHIPneg-94R	ATGTCCTCCCCACCAAGAG
3C-qPCR	ERCC3-Tagman probe	5'FAM-tagctctgatgtctgggagc -3'TAM
3C-qPCR	ERCC3-3C reverse primer	aaatgcctcagccagtgaag
3C-qPCR	ERCCC-3C constant primer	aggccttagttgcctcagtt
3C-qPCR	ERCC3-F	acgtttattcctggtaactaca
3C-qPCR	ERCC3-R	gagaattcagggatgcgcta
3C-qPCR	PROBE-R	5'FAM-GGCGCACGGGGAATTTCTAT-3'TAM
3C-qPCR	VPS,GEM,FAMCut site1-F	CCTAGCTGCTGGCTCTGTCT
3C-qPCR	VPS,GEM,FAMCut site1-R	CTCTGTGAGGTGGTTCCCTCT
3C-qPCR	VPS,GEM,FAMCut site2-F	AGAGGTAAGTCCCCGTCATA
3C-qPCR	VPS,GEM,FAMCut site2-R	AGCAGTTTTACAGAAGGCTCAG
3C-qPCR	VPS,GEM,FAMCut site3-F	CTGTGAGACCTCAGGCT
3C-qPCR	VPS,GEM,FAMCut site3-R	GTTAGCCAGGTGTAGTAGTGTG
3C-qPCR	VPS,GEM,FAMCut site4-F	GACACTCACGAATTGGCCCTG
3C-qPCR	VPS,GEM,FAMCut site4-R	AAGCATATGAATGGCCCGTGAA
3C-qPCR	VPS,GEM,FAMCut site5-F	ACGTCCACCACCAAGAGAGA
3C-qPCR	VPS,GEM,FAMCut site5-R	CAAGTGTGAAGTGCCAGCCA
3C-qPCR	VPS,GEM,FAMCut site6-F	TTCTGCATGGCAGATGGTGG
3C-qPCR	VPS,GEM,FAMCut site6-R	TGTTCTTCTCGGTGGGCAAC
3C-qPCR	VPS,GEM,FAMCut site7-F	TCGGGAATGGAAAGGGCTTCTT
3C-qPCR	VPS,GEM,FAMCut site7-R	CACCGAGTTTCTCCTTTCCAACA
3C-qPCR	VPS,GEM,FAM-n.c.RT100-F	GAGCTTACCTCGAGATGGTGC
3C-qPCR	VPS,GEM,FAM-n.c.RT100-R	GTGCAAACACCAGCTACACC
3C-qPCR	VPS,GEM,FAMconstant-F1	CCCTCGGTCAGTGGCAG
3C-qPCR	VPSGEMFAM-R	GGCTGCAGTCACCTCGAAAA
CRISPR-Cas9	rs718960-gd1.1top	caccgGGCTCAGCTTGCCACGATG
CRISPR-Cas9	rs718960-gd1.1bottom	aaacCATCGTGGGCAAGCTGAGCCc
CRISPR-Cas9	rs718960-gd1.2top	caccgAGTACTGGGTCCATCAAGCC
CRISPR-Cas9	rs718960-gd1.2bottom	aaacGGCTTGATGGACCCAGTACTc
CRISPR-Cas9	rs7405696-gd2.1top	caccgAACAAGACGCCTCGAACCTT
CRISPR-Cas9	rs7405696-gd2.1bottom	aaacAAGGTTTCGAGGCGTCTTGTTc
CRISPR-Cas9	rs7405696-gd2.2top	caccgGCTCCTTAACGAACGCGCTG
CRISPR-Cas9	rs7405696-gd2.2bottom	aaacCAGCGCGTTCGTTAAGGAGCc
CRISPR-Cas9	rs11263763-gd3.1top	caccgGGGAGCCTGGTTCATCCGCC
CRISPR-Cas9	rs11263763-gd3.1bottom	aaacGGCGGATGAACCAGGCTCCCCc
CRISPR-Cas9	rs11263763-gd3.2top	caccgAGGGTCCGGGTGGACACGCG
CRISPR-Cas9	rs11263763-gd3.2bottom	aaacCGCGTGTCCACCCGACCCCTc
CRISPR-Cas9	rs11651052-gd4.1top	caccgCAGGTGTCAAACGCCGCGCCG
CRISPR-Cas9	rs11651052-gd4.1bottom	aaacCGGCCGCGCTTTGACACCTGc
CRISPR-Cas9	rs11651052-gd4.2top	caccgGCCGGAGACCCTCGCCGAGT
CRISPR-Cas9	rs11651052-gd4.2bottom	aaacACTCGGCGAGGGTCTCCGGCc
CRISPR-Cas9	HNF1B-gd1.1top	caccgTGCGCCGAACCGGTTCAAA
CRISPR-Cas9	HNF1B-gd1.1bottom	aaacTTTGAACCGGTTGCGGCGCAc
CRISPR-Cas9	HNF1B-gd1.2top	caccgGTTGGAGCTATAGGCGTCCA

CRISPR-Cas9	HNF1B-gd1.2bottom	aaacTGGACGCCTATAGCTCCAACc
CRISPR-Cas9	rs4430796-gd1.1.1top	caccgAATTCTCATTGAATACAGAG
CRISPR-Cas9	rs4430796-gd1.1.1bottom	aaacCTCTGTATTCAATGAGAATTc
CRISPR-Cas9	rs4430796-gd1.1.2top	caccgATGCTGCATAAAGCTTAAAT
CRISPR-Cas9	rs4430796-gd1.1.2bottom	aaacATTTAAGCTTTATGCAGCATc
CRISPR-Cas9	rs4430796-gd1.2.1top	caccgGGAAACTTCTCAAAAGACAG
CRISPR-Cas9	rs4430796-gd1.2.1bottom	aaacCTGTCTTTTGAGAAGTTTCCc
CRISPR-Cas9	rs4430796-gd1.2.2top	caccgTCCAAAGACCCAACAACGCT
CRISPR-Cas9	rs4430796-gd1.2.2bottom	aaacAGCGTTGTTGGGTCTTTGGAc
CRISPR-Cas9	rs11263761-gd1.1.1top	caccgGCCTGTATAATAGTTACAGA
CRISPR-Cas9	rs11263761-gd1.1.1bottom	aaacTCTGTAACTATTATACAGGCc
CRISPR-Cas9	rs11263761-gd1.1.2top	caccgGAAAGTTTCTCAAGATGCTG
CRISPR-Cas9	rs11263761-gd1.1.2bottom	aaacCAGCATCTTGAGAACTTTCCc
CRISPR-Cas9	rs11263761-gd1.2.1top	caccgTCCTTCTGTAAGTATTATAC
CRISPR-Cas9	rs11263761-gd1.2.1bottom	aaacGTATAATAGTTACAGAAGGAc
CRISPR-Cas9	rs11263761-gd1.2.2top	caccgATAACAACAATCTCTTGAAG
CRISPR-Cas9	rs11263761-gd1.2.2bottom	aaacCTTCAAGAGATTGTTGTTATc
CRISPR-Cas9	rs8064454-gd1.1.1top	caccgACGAGTTACAGTTTCAAACG
CRISPR-Cas9	rs8064454-gd1.1.1bottom	aaacCGTTTGAAACTGTAAGTCGTc
CRISPR-Cas9	rs8064454-gd1.1.2top	caccgCAAGACGCCTCGAACCTTAG
CRISPR-Cas9	rs8064454-gd1.1.2bottom	aaacCTAAGGTTGAGGCGTCTTGc
CRISPR-Cas9	rs8064454-gd1.2.1top	caccgAATGATCTGTGATGTTTACA
CRISPR-Cas9	rs8064454-gd1.2.1bottom	aaacTGTAACATCACAGATCATTc
CRISPR-Cas9	rs8064454-gd1.2.2top	caccgTAGGCATACAAAATGAACAC
CRISPR-Cas9	rs8064454-gd1.2.2bottom	aaacGTGTTCATTTTGTATGCCTAc
CRISPR-Cas9	rs4239217-gd1.1.1top	caccgTCCCCTATCCTCCAAGAAC
CRISPR-Cas9	rs4239217-gd1.1.1bottom	aaacGTTCTTGAGGATAGTGGGAc
CRISPR-Cas9	rs4239217-gd1.1.2top	caccgTGAAGTATGGTACAGTATCA
CRISPR-Cas9	rs4239217-gd1.1.2bottom	aaacTGATACGTGACCATCAGTCAc
CRISPR-Cas9	rs4239217-gd1.2.1top	caccgCCCCTATCCTCCAAGAACA
CRISPR-Cas9	rs4239217-gd1.2.1bottom	aaacTGTTCTTGAGGATAGTGGGc
CRISPR-Cas9	rs4239217-gd1.2.2top	caccgCTTCTTTGGAATTGACTGA
CRISPR-Cas9	rs4239217-gd1.2.2bottom	aaacTCAGTCAATTTCCAAAGAAGc
CRISPR-Cas9	rs11658063-gd1.1.1top	caccgTCTCAGACCTCAGTTCGCAG
CRISPR-Cas9	rs11658063-gd1.1.1bottom	aaacCTGCGAACTGAGGTCTGAGAc
CRISPR-Cas9	rs11658063-gd1.1.2top	caccgTGTTCCATGCAGGGTCCGGG
CRISPR-Cas9	rs11658063-gd1.1.2bottom	aaacCCCGGACCCTGCATGGAACAc
CRISPR-Cas9	rs11658063-gd1.2.1top	caccgACCAGATCCTCTGCGAACTG
CRISPR-Cas9	rs11658063-gd1.2.1bottom	aaacCAGTTCGCAGAGGATCTGGTc
CRISPR-Cas9	rs11658063-gd1.2.2top	caccgAGGGTCCGGGTGGACACGCG
CRISPR-Cas9	rs11658063-gd1.2.2bottom	aaacCGCGTGTCCACCCGGACCCTc
CRISPR-Cas9	rs3760511-gd1.1.1top	caccgTCCCCTCCTGCCCGTCACCA
CRISPR-Cas9	rs3760511-gd1.1.1bottom	aaacTGGTGACGGGCAGGAGGGGAc
CRISPR-Cas9	rs3760511-gd1.1.2top	caccgGACCTGACCTGTAGAAGAAC

CRISPR-Cas9	rs3760511-gd1.1.2bottom	aaacGTTCTTCTACAGGTCAGGTCc
CRISPR-Cas9	rs3760511-gd1.2.1top	caccgGAAACTCCAATAGTGCTGCA
CRISPR-Cas9	rs3760511-gd1.2.1bottom	aaacTGCAGCACTATTGGAGTTTCc
CRISPR-Cas9	rs3760511-gd1.2.2top	caccgTCCCTGTTCTTCTACAGGTC
CRISPR-Cas9	rs3760511-gd1.2.2bottom	aaacGACCTGTAGAAGAACAGGGAc
CRISPR-Cas9	rs11649743-gd1.1.1top	caccgGTTCTGCCAGTCTCAGGTGT
CRISPR-Cas9	rs11649743-gd1.1.1bottom	aaacACACCTGAGACTGGCAGAAAc
CRISPR-Cas9	rs11649743-gd1.1.2top	caccgTGAGAAGTGCATTAAGATTG
CRISPR-Cas9	rs11649743-gd1.1.2bottom	aaacCAATCTTAATGCACTTCTCAc
CRISPR-Cas9	rs11649743-gd1.2.1top	caccgAAAGCCCCAACACCTGAGAC
CRISPR-Cas9	rs11649743-gd1.2.1bottom	aaacGTCTCAGGTGTTGGGGCTTTC
CRISPR-Cas9	rs11649743-gd1.2.2top	caccgATCTTAATGCACTTCTCAAA
CRISPR-Cas9	rs11649743-gd1.2.2bottom	aaacTTTGAGAAGTGCATTAAGATc
CRISPR-Cas9	rs9901746-gd1.1top	caccgAACTGCAGCACGTGGAGTTG
CRISPR-Cas9	rs9901746-gd1.1bottom	aaacCAACTCCACGTGCTGCAGTTc
CRISPR-Cas9	rs9901746-gd1.2top	caccgTTTATGTTACGTCTCAGCTG
CRISPR-Cas9	rs9901746-gd1.2bottom	aaacCAGCTGAGACGTAACATAAAc
CRISPR-Cas9	rs9901746-gd2.1top	caccgTAGGAGCGGCTGGGGCGCCA
CRISPR-Cas9	rs9901746-gd2.1bottom	aaacTGGCGCCCCAGCCGCTCCTAc
CRISPR-Cas9	rs9901746-gd2.2top	caccgCAGGTCCTGAGCGACCTCCC
CRISPR-Cas9	rs9901746-gd2.2bottom	aaacGGGAGGTCGCTCAGGACCTGc
CRISPR-Cas9	rs12453443-gd1.1top	caccgGTTCTTTGCCGAAGGACAGG
CRISPR-Cas9	rs12453443-gd1.1bottom	aaacCCTGTCTTCGGCAAAGAACc
CRISPR-Cas9	rs12453443-gd1.2top	caccgGATCTGGGAGGCGAGTTCCC
CRISPR-Cas9	rs12453443-gd1.2bottom	aaacGGGAACTCGCCTCCCAGATCc
CRISPR-Cas9	rs12453443-gd2.1top	caccgGTGGGCAAACCTCCCGGCAA
CRISPR-Cas9	rs12453443-gd2.1bottom	aaacTTGCCGGGAGGTTTGCCACc
CRISPR-Cas9	rs12453443-gd2.2top	caccgCAGGAAAAATGCGCGGCTG
CRISPR-Cas9	rs12453443-gd2.2bottom	aaacCAGGCCGCGCATTTTTCTGc
CRISPR-Cas9	rs461251-gd1.1.1top	caccgTTATCAATGGTGGACTATTC
CRISPR-Cas9	rs461251-gd1.1.1bottom	aaacGAATAGTCCACCATTGATAAc
CRISPR-Cas9	rs461251-gd1.1.2top	caccgATCCTCTCCACATCGACGCG
CRISPR-Cas9	rs461251-gd1.1.2bottom	aaacCGCGTCGATGTGGAGAGGATc
CRISPR-Cas9	rs461251-gd1.2.1top	caccgCAGGTGAACCGGGAAACTCC
CRISPR-Cas9	rs461251-gd1.2.1bottom	aaacGGAGTTTCCCGGTTACCTGc
CRISPR-Cas9	rs461251-gd1.2.2top	caccgTGGTGGAGCGGGTCCCTCCTT
CRISPR-Cas9	rs461251-gd1.2.2bottom	aaacAAGGAGGACCCGCTCCACCAc
CRISPR-Cas9	rs2955626-gd2.1.1top	caccgACTCCCTCGCGGCAGCGACC
CRISPR-Cas9	rs2955626-gd2.1.1bottom	aaacGGTCGCTGCCGCGAGGGAGTc
CRISPR-Cas9	rs2955626-gd2.1.2top	caccgACAGAGCGGCTGACGGAGCC
CRISPR-Cas9	rs2955626-gd2.1.2bottom	aaacGGCTCCGTCAGCCGCTCTGTc
CRISPR-Cas9	rs2955626-gd2.2.1top	caccgGCGCGTCCTTAGGATGGGCG
CRISPR-Cas9	rs2955626-gd2.2.1bottom	aaacCGCCCATCCTAAGGACGCGCc
CRISPR-Cas9	rs2955626-gd2.2.2top	caccgCGCCGGGATACTGCGTTGCC

CRISPR-Cas9	rs2955626-gd2.2.2bottom	aaacGGCAACGCAGTATCCCGGCGc
CRISPR-Cas9	rs684232-gd3.1.1top	caccgAGCTGGGGAGACGTTTACTG
CRISPR-Cas9	rs684232-gd3.1.1bottom	aaacCAGTAAACGTCTCCCCAGCTc
CRISPR-Cas9	rs684232-gd3.1.2top	caccgCTCCAGGGCGCTGCGTCATC
CRISPR-Cas9	rs684232-gd3.1.2bottom	aaacGATGACGCAGCGCCCTGGAGc
CRISPR-Cas9	rs684232-gd1.1.1top	caccgTTATCAATGGTGGACTATTC
CRISPR-Cas9	rs684232-gd1.1.1bottom	aaacGAATAGTCCACCATTGATAAc
CRISPR-Cas9	rs684232-gd1.1.2top	caccgATCCTCTCCACATCGACGCG
CRISPR-Cas9	rs684232-gd1.1.2bottom	aaacCGCGTCGATGTGGAGAGGATc

Supplementary Table 9.
Recombinant DNA.

Name	Source	Identifier
pGEN-MCS-Renilla	Wei et al., 2010	N/A
pcDNA3.1/V5-His A	Invitrogen	Cat#V81020
pcDNA3.1/flag	designed for this study	N/A
pLVET-IRES-GFP	Zhang et al., 2017	N/A
pLKO.1 Puro	Stewart et al., 2003	Addgene #8453
pSpCas9 (BB)-2A-Puro (PX459)	Feng Zhang Lab at MIT	N/A
pGL4.10[luc2]	Promega	Cat#E665A
pVSVG-envelope	addgene	Cat#14888
pMDLg/pRRE-packaging	addgene	Cat#12251
pRSV-Rev-packaging	addgene	Cat#12253
pGL4.75[hRluc/CMV]	Promega	Cat#E693A

Supplementary Table 10.
Software and Algorithms.

Name	Source	Identifier
HaploReg v4.1	Ward and Kellis, 2012	http://archive.broadinstitute.org/mammals/haploreg/haploreg.php
FastQC	Babraham Bioinformatics Institute	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Matrix eQTL	R Bioconductor	https://bioconductor.org/
Trimmomatic	Bolger et al., 2014	http://www.usadellab.org/cms/?page=trimmomatic
Tophat2	Center for Computational Biology at Johns Hopkins University	https://ccb.jhu.edu/software/tophat/index.shtml
HTSeq	Bioconductor	https://htseq.readthedocs.io/en/release_0.9.1/
DESeq2	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
GSEA	Subramanian et al., 2005	http://software.broadinstitute.org/gsea/index.jsp
MACS2	Zhang et al., 2008	https://github.com/macs3-project/MACS
HOMER V4.10	Heinz et al., 2010	http://homer.ucsd.edu/homer/
IGV 2.4.10	Robinson et al., 2011	https://software.broadinstitute.org/software/igv/
ChIPseeker 1.18.0	Yu et al., 2015	https://www.bioconductor.org/packages/release/bioc/html/ChIPseeker.html
R Version 3.6.3	R	https://www.r-project.org/
CRISPR design tool	Feng Zhang Lab at MIT	http://crispr.mit.edu/

Supplementary Table 11.
Deposited Data.

Name	Source	Identifier
Raw RNA-seq data	This paper	
Oncomine database	Rhodes et al., 2004	https://www.oncomine.org/resource/login.html
cDNA microarray assay of gene expression profiling	Welsh et al., 2001	https://www.oncomine.org/resource/login.html
cDNA microarray assay of gene expression profiling	Vanaja et al., 2003	https://www.oncomine.org/resource/login.html
cDNA microarray assay of gene expression profiling	Glinsky et al., 2004	https://www.oncomine.org/resource/login.html
cDNA microarray assay of gene expression profiling	Luo et al., 2001	https://www.oncomine.org/resource/login.html
cDNA microarray assay of gene expression profiling	Yu et al., 2004	GEO: GSE6919
cDNA microarray assay of gene expression profiling	Taylor et al., 2010	GEO: GSE21032
cDNA microarray assay of gene expression profiling	Bittner et al., 2005	GEO: GSE2109
cDNA microarray assay of gene expression profiling	Wallace et al., 2008	GEO: GSE6956
cDNA microarray assay of gene expression profiling	Penney et al., 2014	GEO: GSE62872
cDNA microarray assay of gene expression profiling	Ross-Adams et al., 2015	GEO: GSE70769
cDNA microarray assay of gene expression profiling	Ross-Adams et al., 2015	GEO: GSE70768
cDNA microarray assay of gene expression profiling	Lapointe et al., 2004	GEO: GSE3933
Processed RNA-seq data	Cancer Genome Atlas Research Network, 2015	http://www.cbioportal.org/
Processed ChIP-seq data	Mei et al., 2017	http://cistrome.org/db/#/
GWAS catalog	Buniello et al., 2019	https://www.ebi.ac.uk/gwas/
Genome-wide CRISPR/Cas9 screening of cancer cell survival genes	DepMap, Broad, 2019	https://depmap.org/portal/achilles/

Supplementary Table 12.
Critical Commercial Assays.

Name	Source	Identifier
RNeasy Mini Kit	QIAGEN	Cat#74106
RNase-Free DNase	QIAGEN	Cat#79254
High-Capacity cDNA Reverse Transcription Kit	Applied Biosystems	Cat#4368814
GeneJET PCR Purification Kit	Thermo Scientific	Cat#K0701
GeneJET Gel Extraction Kit	Thermo Scientific	Cat#K0691
PureYield Plasmid Miniprep System	Promega	Cat#A1223
Dual-Glo Luciferase Assay System	Promega	Cat#E2940
MinElute PCR Purification Kit	QIAGEN	Cat#28006
iScript Reverse Transcription Supermix	Bio-rad	Cat#1708840

Supplementary Table 13.
SNP genomic locations in the 17q12 and 17p13.3 loci.

SNPs	Chr	Position	hg19 SNP-centered region	Gene
rs4430796	17	36098040	chr17:36097540-36098540	HNF1B
rs11263761	17	36097775	chr17:36097275-36098275	HNF1B
rs11263763	17	36103565	chr17:36103065-36104065	HNF1B
rs11651052	17	36102381	chr17:36101881-36102881	HNF1B
rs8064454	17	36101586	chr17:36101086-36102086	HNF1B
rs9901746	17	36103149	chr17:36102649-36103649	HNF1B
rs12453443	17	36104121	chr17:36103621-36104621	HNF1B
rs7405696	17	36102035	chr17:36101535-36102535	HNF1B
rs4239217	17	36098987	chr17:36098487-36099487	HNF1B
rs11658063	17	36103872	chr17:36103372-36104372	HNF1B
rs3760511	17	36106313	chr17:36105813-36106813	HNF1B
rs11649743	17	36074979	chr17:36074479-36075479	HNF1B
rs718960	17	36077279	chr17:36076779-36077779	HNF1B
rs2955626	17	618100	chr17:617600-618600	VPS53, FAM57A, GEMIN4
rs684232	17	618965	chr17:618465-619465	VPS53, FAM57A, GEMIN4
rs461251	17	619162	chr17:618162-619662	VPS53, FAM57A, GEMIN4

Supplementary Table 14.
Bacterial and Virus Strains.

Name	Source	Identifier
DH5 α Chemically Competent cells	This paper	N/A
Stbl3 Chemically Competent cells	This paper	N/A
Lentivirus	This paper	N/A

Supplementary Table 15.
Chemicals, Peptides, and Recombinant Proteins.

Name	Source	Identifier
5 α -Dihydrotestosterone (DHT) solution	From Olli A. Jänne lab (University of Helsinki)	N/A
BamHI-HF	New England Biolabs	Cat#R3136M
BstZ17I-HF	New England Biolabs	Cat#R3594S
KpnI-HF	New England Biolabs	Cat#R3142M
PmeI	New England Biolabs	Cat#R0560S
XbaI	New England Biolabs	Cat#R0145M
HindIII-HF	New England Biolabs	Cat#R3104M
EcoRV-HF	New England Biolabs	Cat#R3195M
BbsI	New England Biolabs	Cat#R0539S
CutSmart Buffer	New England Biolabs	Cat#B7204S
T4 DNA Ligase	New England Biolabs	Cat#M0202M
T4 DNA Ligase Reaction Buffer	New England Biolabs	Cat#B0202S
SYBR Master Mix	Applied Biosystems	Cat#4472920
Fetal Bovine Serum	Thermo Fisher	Cat#16000044
Penicillin-Streptomycin	Thermo Fisher	Cat#15140122
Keratinocyte-Serum Free Medium	Invitrogen	Cat#17005-042
Lipofectamine 3000	Thermo Fisher	Cat#L3000015
Lipofectamine 2000	Thermo Fisher	Cat#11668019
cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail	Roche	Cat#04693159001
Dynabead protein G	Invitrogen	Cat#10004D
X-treme GENE HP DNA Transfection Reagent	Roche	Cat#06366236001
HiPerFect Transfection Reagent	QIAGEN	Cat#301705
high glucose, DMEM	Thermo Fisher	Cat#11965092
Low glucose DMEM	Thermo Fisher	Cat#11885084
RPMI-1640 Medium	Sigma-Aldrich	Cat#R8758
F-12K	ATCC	Cat#30-2004
Trypsin-EDTA (0.25%), phenol red	Thermo Fisher	Cat#25200-072
Cell Proliferation Kit II	Roche	Cat#11465015001
Polybrene	Merck	Cat#H9268
Puromycin	Merck	Cat#P9620
2x Phusion Master Mix with HF Buffer	Thermo Fisher	Cat#F531
Quanti tech probe PCR mix	QIAGEN	Cat#204343
Pierce BCA Protein Assay Kit	Thermo Fisher	Cat#23225
Pierce lane marker reducing sample buffer	Thermo Scientific	Cat#39000
Opti-MEM I Reduced-Serum medium	Thermo Fisher	Cat#11058021
Lumi-Light Western Blotting Substrate	Merck	Cat#12015200001
SuperSignal West Femto Maximum Sensitivity Substrate	Thermo Fisher	Cat#34094
I-BET-CAS 1260907-17-2 -Calbiochem	Sigma-Aldrich	Cat#401010-5MG
EMEM	ATCC	Cat#30-2003
Unstained Protein Ladder (10 - 200 kDa)	abcam	Cat#ab234618
PageRuler™ Prestained Protein Ladder, 10 to 180 kDa	Thermo Scientific	Cat# 26616
Precision Plus Protein Dual Color Standards	BIO-RAD	Cat# 1610374