Title

Integrative analysis of ultra-deep RNA-seq reveals alternative promoter usage as a mechanism of activating oncogenic programmes during prostate cancer progression

Journal

Nature Cell Biology, 26(7)

ISSN

1465-7392

Authors

Zhang, Meng
Sjöström, Martin
Cui, Xiekui
et al.

Publication Date

2024-07-01

DOI

10.1038/s41556-024-01438-3

Peer reviewed

Integrative analysis of ultra-deep RNA-seq reveals alternative promoter usage as a mechanism of activating oncogenic programs during prostate cancer progression

Meng Zhang[1,2], Martin Sjöström[1,2], Xiekui Cui[1,2,3], Adam Foye[1,2], Kyle Farh[4], Raunak Shrestha[1,2], Arian Lundberg[1,2], Ha X. Dang[5,6,7,8], Haolong Li[1,2], Phillip G. Febbo[4], Rahul Aggarwal[2,9], Joshi Alumkal[10], Eric J. Small[2,9], The SU2C/PCF West Coast Prostate Cancer Dream Team*, Christopher A. Maher[5,6,7,11,14], Felix Y. Feng[1,2,9,12,14], David A. Quigley[2,12,13,14]

[1] Department of Radiation Oncology, University of California at San Francisco, San Francisco, CA, USA
[2] Helen Diller Family Comprehensive Cancer Center, University of California at San Francisco, San Francisco, CA, USA
[3] Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA
[4] Illumina, Inc., San Diego, CA, USA
[5] McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO, USA
[6] Department of Internal Medicine, Washington University in St. Louis, St. Louis, MO, USA
[7] Alvin J. Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO, USA
[8] Current affiliation: Bristol Myers Squibb, San Diego, CA, USA.
[9] Division of Hematology and Oncology, Department of Medicine, University of California at San Francisco, San Francisco, CA, USA
[10] Division of Hematology and Oncology, University of Michigan Rogel Cancer Center, Ann Arbor, MI.
[11] Department of Biomedical Engineering, Washington University School of Medicine, St. Louis, MO, USA
[12] Department of Urology, University of California at San Francisco, San Francisco, CA, USA
[13] Department of Epidemiology & Biostatistics, University of California at San Francisco, San Francisco, CA, USA
[14] These authors jointly supervised the work: Christopher A. Maher, Felix Y. Feng, David A. Quigley
* A list of members and their affiliations appears at the end of the manuscript

Corresponding author:
David A. Quigley (david.quigley@ucsf.edu)

ABSTRACT

Transcription factor (TF) proteins regulate gene activity by binding to regulatory regions, most importantly at gene promoters. Many genes have alternative promoters (APs) bound by distinct TFs. The role of differential TF activity at alternative promoters during tumor development is poorly understood. Here we show, using deep RNA sequencing in 274 biopsies of benign prostate tissue, localized prostate tumors, and metastatic castration-resistant prostate cancer (mCRPC), that AP usage increases as tumors progress and APs are responsible for a disproportionate amount of tumor transcriptional activity. Expression of the androgen receptor *(AR)*, the key driver of prostate tumor activity, is correlated with elevated AP usage. We identified *AR*, *FOXA1* and *MYC* as potential drivers of AP activation. DNA methylation is a likely mechanism for AP activation during tumor progression and lineage plasticity. Our data suggest that prostate tumors activate alternative promoters to magnify the transcriptional impact of tumor drivers including *AR* and *MYC*.

Abstract word count: 151

INTRODUCTION

Over half of human genes are under the control of multiple promoters[1]. Each alternative promoter can initiate transcription of one or more distinct isoforms of the same gene with partially overlapping transcribed sequences. Alternative promoters (APs) are a tissue-specific mechanism for transcriptome regulation during development[2]. A recent pan-cancer study identified numerous cancer-associated alternative promoters, implicating them in tumorigenesis[3]. Alternative promoters harbor distinct transcription factor (TF) binding sites, contributing to complex gene regulation[2,4]. However, the impact of alternative promoters during tumor progression remains poorly understood, as does the impact of genomic and epigenomic changes on promoter usage. Increased methylation at canonical promoters generally represses transcription[5]. We have shown that during prostate tumorigenesis, methylation changes preferentially impact prostate tumor-specific genes[6]. Systematic analysis of methylation changes, TF binding, and promoter activity during tumor progression could elucidate how epigenetic factors cooperate to influence alternative promoter usage.

Prostate tumors are mainly driven by the androgen receptor *(AR)*, a hormone-responsive transcription factor, making androgen signaling the primary target for prostate cancer (PCa) therapy. Progression from localized hormone-sensitive PCa to lethal metastatic castration-resistant prostate cancer (mCRPC) is accompanied by genomic and epigenomic alterations affecting *AR* and other TFs[6-10]. A recent discovery linked a PCa risk SNP-mediated promoter switch in lncRNA *PCAT19* to PCa initiation and progression[11], highlighting the impact of alternative promoter usage. Here, we systematically investigated alternative promoter usage in benign prostate tissue, localized PCa, and mCRPC. Combining ultra-deep whole transcriptome sequencing with whole genome methylation sequencing, we define how transcriptional programs are activated during PCa progression and reveal a link between alternative promoter usage, TF binding, and DNA methylation.

RESULTS

AP use increases as localized PCa progresses to mCRPC

To characterize how alternative promoter usage changes during prostate tumor progression, we compiled a consensus set of promoters from all annotated transcripts and used deep RNA sequencing to assess their activity across different PCa stages. We enumerated 99,589 candidate promoters for 37,885 genes (**Supplementary Table 1, Methods**). Our transcriptomic dataset comprised 274 patient biopsies from benign tissue, localized PCa, and mCRPC (**Supplementary Table 2, Extended Data Figure 1A**). We sequenced 104 mCRPC biopsies to a median of 453M [million] (± 130M) reads/sample (WCDT cohort, with previously characterized DNA and bisulfite sequencing[6,8,12,13]), and obtained RNA-seq data from 8 benign tissue biopsies (median of 89M ± 9M reads/sample, PAIR cohort)[14], and 162 localized PCa biopsies (median of 190M ± 68M reads/sample, PAIR and CPC-GENE cohorts)[15]. **Extended Data Figure**

We then quantified transcriptional activity at each promoter in each sample using a modified version of the proActiv tool[3]. We enhanced proActiv to measure internal promoters (promoters that overlap with internal exons, **Methods**) using split read subtractions, DNA methylation data, and publicly available H3K4me3 ChIP-seq data[16] (**Methods**, **Extended Data Figure 1B and S1C**). This produced 72,399 promoters (9,015 high confidence internal promoters, 63,384 non-internal promoters, **Extended Data Figure 1D**). Ultra-deep RNA sequencing data improved promoter activity detection by an average of 26.31% compared to typical sequencing depths (42,766±1,547 promoters at 500M reads vs 33,883±744 promoters at subsets of 31.25M reads). Saturation analysis with down-sampled reads indicated that our sequencing depth captured most active promoters (**Extended Data Figure 1E**).

We next analyzed promoter activity changes during prostate tumor progression. Active promoters per gene were similar across disease stages, with median values of 2.18, 2.16, and

2.20 in benign, localized, and mCRPC respectively (**Extended Data Figure 2A**). However, compared to benign prostate tissue, genes upregulated in localized PCa or mCRPC and known oncogenes were more likely to switch from a single active promoter in benign tissue to multiple active promoters in tumors (**Figure 1A**). Conversely, genes downregulated in localized PCa or mCRPC tended to switch from multiple promoters in benign to a single promoter in localized or mCRPC (**Extended Data Figure 2B**).

A gene's most active promoter was not necessarily its canonical promoter, defined as the region around the transcription start site (TSS) of the canonical transcript listed in the Ensembl database[17]. To identify APs, we compared promoter activity across disease stages, relying solely on differential absolute and relative promoter activities measured by RNA-seq (**Methods**). We identified 463, 3,237, and 2,326 APs comparing localized PCa vs. benign, mCRPC vs. benign, and mCRPC vs localized PCa, respectively (**Figure 1B, Supplementary Tables 3-5**). Principal component analysis of promoter activity showed separate clustering of benign and localized PCa samples within the PAIR cohort by disease stage, while same-stage samples from the PAIR and CPC-GENE cohorts overlapped (**Figure 1C**). The largest source of variation separated mCRPC samples from other disease stages, consistent with the larger number of differential promoters in mCRPC compared to other stages (**Figure 1B**). We did not detect 5' to 3' bias of RNA-seq coverage (**Extended Data Figure 2C**). In summary, progression was not associated with total APs counts, but promoter activity was linked to gene expression changes during progression.

AP use is linked to upregulation of PCa-relevant genes

We next focused on APs upregulated in tumors. Most upregulated APs were non-canonical promoters: 78.9% in localized (146 of 185) and 66.7% in mCRPC (827 of 1,239). We tested whether AP upregulation could be explained by uniform increase in total expression from all promoters. The sum of each gene's promoter activity was positively correlated with that gene's expression level in localized PCa and mCRPC (**Figure 1D, Extended Data Figure 3A**). However, upregulated APs were more strongly correlated with gene expression, whereas downregulated APs exhibited a weaker correlation (**Figure 1D, Extended Data Figure 3A**). We hypothesized that upregulated APs contribute a disproportionately high percentage of the total transcriptional increase for genes upregulated during progression. This was the case in both localized PCa and mCRPC (**Figure 1E**), with a stronger effect in mCRPC (median contribution = 98.8% vs 75.5%, Student's t-test p value = 1.04×10^{-6}) (**Figure 1E**). These observations support a model where APs are a primary source of increased tumor transcriptional activity among upregulated genes, particularly in advanced tumors.

We next examined whether the most active promoters switched to upregulated APs in tumors. We labeled the promoter with the highest activity for each gene as the major promoter and the other active promoters for that gene as minor promoters. A minority of upregulated APs in both localized (17.8%, 33 out of 185) and mCRPC (35.8%, 444 out of 1,239) switched from being minor or inactive in benign to major in tumors. Notably, the switching frequency was twice as high in mCRPC, consistent with our observation of increased AP usage in advanced PCa compared to localized disease. For instance, the upregulated non-canonical AP of the *RALBP1* gene replaced the canonical promoter to become the dominant promoter mCRPC (**Figure 1F and 1G**). RALBP1 mediates endocytosis, binding to RalA to orchestrate human tumorigenesis[18].

We then investigated whether genes with upregulated APs were associated with biological functions that drive PCa progression. Genes with APs upregulated in mCRPC compared to benign prostate tissue were significantly enriched for roles in mitosis, cell cycle regulation, E2F and MYC targets, the P53 pathway and androgen signaling, as well as developmental pathways related to other cancers like pancreatic and hepatocellular carcinoma

(**Extended Data Figure 3B**). The MYC target, P53 pathway, androgen signaling, and pancreatic/hepatocellular cancer pathways were unique to the upregulated AP analysis and not enriched in differentially expressed genes (**Extended Data Figure 3B, Extended Data Figure 3C**), indicating that AP usage analysis may provide additional biological insights beyond standard differential expression. Our findings reveal a substantial increase in the number of APs exhibiting differential activity as localized PCa progresses to mCRPC. These observations support the hypothesis that AP activation constitutes an important mechanism driving the upregulation of genes relevant to disease progression.

AP use is linked to FOXA1 binding and androgen activity

Having shown that upregulated APs contribute disproportionately to gene expression increases during prostate tumor progression, and that activated promoters preferentially involve genes linked to tumor progression, we asked which transcription factors might drive the activation of these promoters. PCa depends on androgen signaling[19], and we hypothesized that elevated *AR* activity during PCa progression might correlate with increased AP usage. Indeed, we found a significant positive correlation between *AR* gene expression and the number of upregulated APs in both localized PCa and mCRPC tumors (**Figure 2A**, **FigureS4A**).

We then hypothesized that upregulated APs would be more likely to harbor binding sites for AR and FOXA1, a key AR pioneer factor[20]. To test this, we analyzed published AR and FOXA1 ChIP-seq data from benign prostate tissue, localized prostate tumors, and mCRPC patient-derived xenograft (PDX) models[16]. As expected, the canonical promoters of Hallmark AR targets were enriched for AR binding in localized PCa and mCRPC (**Figure 2B**). Consistent with our hypothesis, upregulated APs in localized PCa and mCRPC were enriched for tumor-specific AR and FOXA1 binding (**Figure 2B, 2C**). Additionally, canonical promoters of Hallmark AR target genes and upregulated APs in mCRPC both showed enrichment for FOXA1 binding (**Figure 2C**). Although the percentage of AP sites with AR binding was lower than that at canonical AR target promoters, FOXA1 binding enrichment was comparable at these sites, suggesting that FOXA1 pioneers the activation of upregulated APs in tumors. A significantly lower proportion of AR-FOXA1 co-binding was present in FOXA1-bound upregulated APs in mCRPC compared to localized PCa, implying that other AR co-factors may be more pertinent than FOXA1 in this state (**Extended Data Figure 4B**).

We hypothesized that reducing *FOXA1* expression in a PCa cell line would diminish FOXA1 binding at AP sites. We assessed this using previously published experiments conducted in LNCaP cells, a PCa model[20]. FOXA1 binding measured by ChIP-seq in LNCaP cells was enriched at upregulated APs containing FOXA1 ChIP-seq peaks in localized PCa and mCRPC (**Figure 2D**). Knockdown of *FOXA1* mRNA in LNCaP cells caused a decrease in FOXA1 binding peaks at all sites, including upregulated APs (**Figure 2D**). Furthermore, most upregulated APs in localized PCa (71.4%) or in mCRPC (64.2%) bound by FOXA1 showed decreased expression in LNCaP cells following *FOXA1* knockdown (**Figure 2E**), suggesting close association between FOXA1 binding and AP regulation.

In summary, we observed a significant association between *AR* expression levels and alternative promoter activation during prostate cancer development and a direct impact of the binding sites of AR and pioneer factor FOXA1 on these upregulated alternative promoters.

MYC is a potential driver of AP activation in mCRPC

To systematically map associations between driver genes altered during PCa progression and altered AP activity, we adopted an unbiased approach to identify TFs that bind upregulated APs. Using UniBind[21], we identified TFs with experimentally defined ChIP-seq peaks enriched at upregulated APs. In localized PCa upregulated APs, the TFs most enriched for binding were

AR, along with the AR co-factors FOXA1 and GATA2 (**Extended Data Figure 5A**). In mCRPC upregulated APs, while AR and FOXA1 binding were also enriched, the most significantly enriched TFs were MYC, E2F1 and HIF1A (**Figure 3A, Extended Data Figure 5B**). This finding aligned with known drivers of prostate tumor progression to mCRPC, including *MYC* and *HIF1A* overexpression, as well as loss of the E2F repressor *RB1*[7,8,22-24]. MYC induces transcriptional amplification and is oncogenic in various tumors[25]. Genes with upregulated mCRPC APs overlapping with MYC ChIP-seq peaks[26] were more likely to be MYC and E2F targets, and they were enriched for roles in the cell cycle and P53 signaling (**Extended Data Figure 5C**). This supports a model where MYC promotes proliferation during tumor progression in part by activating APs.

MYC gene amplification occurs in 30% of mCRPC tumors[8] and is correlated with increased *MYC* expression. We hypothesized that AP sites in samples with higher *MYC* expression would be more likely to be bound by proteins relevant to MYC signaling. We divided mCRPC samples into tertiles by *MYC* expression and performed differential promoter analysis in samples with high versus low *MYC* expression (**Supplementary Table 6**). Indeed, top TFs enriched for binding at upregulated APs in *MYC* high samples included MYC along with its interactors MAX, MXI1 and MYCN (**Figure 3B**).

We next investigated potential mechanisms behind MYC activation of APs in mCRPC. The histone methyltransferase enhancer of zeste homolog 2 (EZH2) can directly bind MYC, collaborating to enhance gene expression[27]. We hypothesized that EZH2 drives PCa in part by facilitating MYC binding at AP sites in mCRPC. We observed stronger enrichment of MYC binding at upregulated mCRPC APs that overlap EZH2 ChIP-seq peaks in LNCaP cells (**Figure 3C**). Since EZH2 is not a TF, it isn't present in the UniBind TF ChIP-seq database. We analyzed previously published EZH2 ChIP-seq data in LNCaP cells[28] and found EZH2 binding was enriched at upregulated mCRPC APs, particularly in those overlapping with MYC ChIP-seq peaks in LNCaP cells (**Figure 3D**). Furthermore, the co-occurrence frequency of MYC and EZH2 ChIP-seq peaks was higher in upregulated APs and canonical promoters of upregulated genes in mCRPC than background or all canonical promoters (**Figure 3E**). These findings support the model that MYC and EZH2 cooperatively activate both canonical and alternative promoters.

Genes bound by EZH2 but not by MYC were enriched for epithelial-mesenchymal transition (EMT) and developmental pathways (**Extended Data Figure 6A**), consistent with EZH2's reported role in promoting EMT[29]. Conversely, genes co-bound by EZH2 and MYC were enriched for mTOR signaling, cell proliferation, MYC targets and androgen signaling (**Extended Data Figure 6C**), resembling those bound by MYC alone (**Extended Data Figure 6B**). This suggests that when EZH2 co-binds with MYC, it contributes to a pro-proliferation AR-responsive program. In one example, MYC and EZH2 co-bound at an AP of *BMI1* (**Figure 3F**). BMI1, a core component of the PRC1 complex, is overexpressed in various tumors, including PCa[30], with reported roles in cancer cell proliferation, invasion, metastasis, and patient survival[31]. Notably, MYC and EZH2 specifically co-bound at the *BMI1* alternative promoter associated with a protein coding isoform overexpressed in mCRPC vs. benign (log2FC = 2.2, **Figure 3G**). These findings suggest that the co-binding of EZH2 and MYC at an AP of *BMI1* plays a role in BMI1 protein upregulation in mCRPC. In summary, these observations demonstrate that elevated MYC activity in advanced PCa contributes to AP activity of genes with tumor-promoting functions.

AP usage reflects lineage plasticity in response to therapy

Prostate tumors that progress on therapy targeting androgen signaling can develop expression phenotypes resembling the neuroendocrine lineage[9,12,13]. These tumors lack androgen signaling and have poor prognosis[32]. We hypothesized that alternative promoter usage contributes to

lineage plasticity. Among 104 mCRPC tumors, three previously identified by our group as neuroendocrine (treatment-emergent Small Cell Neuroendocrine Carcinoma, t-SCNC)[8] were examined. We identified 523 APs differentially active in t-SCNC vs. adenocarcinoma samples (**Supplementary Table 7**). In UniBind ChIP-seq binding enrichment analysis, TFs enriched for binding at upregulated t-SCNC APs included HAND2 (critical for neurogenesis[33,34]), ASCL1 (a driver of small-cell neuroendocrine tumors[10]), and TFAP2A (associated with neural crest development[35]) (**Figure 4A**). Conversely, androgen-associated TFs AR, FOXA1, GRHL2, and the glucocorticoid receptor (NR3C1) were enriched at downregulated APs (**Extended Data Figure 7A**). The top enriched TF *HAND2* was overexpressed in t-SCNC tumors (**Figure 4B**). Upregulated APs overlapping with HAND2 ChIP-seq peaks were enriched for genes related to neuronal function and cellular development (**Figure 4C**). These findings suggest that lineage switching and the development of t-SCNC are associated with AP activity influencing genes in the neuronal lineage.

Previous studies by our group[36] and others[37] have demonstrated mCRPC tumors can exhibit a gastrointestinal (GI) expression phenotype. Using a previously published GI signature gene set[37], we identified 25 WCDT tumors with a GI signature (**Extended Data Figure 7B**). Differential promoter analysis comparing samples with the highest quartile GI signature to other samples revealed 127 upregulated and 86 downregulated APs (**Supplementary Table 8**). UniBind ChIP-seq enrichment analysis of the upregulated APs in GI-high samples showed significant enrichment for binding of HNF family members HNF1A, HNF1B, HNF4G and HNF4A, key TFs in liver development that are upregulated in GI tissues[38].(**Figure 4D**). The proto-oncogene *SRC* was an intriguing example of a specific active promoter in GI-high samples. A non-canonical promoter (P1) upstream of the canonical promoter (P2) overlapped HNF1A ChIP-seq peaks and has been experimentally shown to be regulated by HNF1A in the HepG2 liver tumor cell line[39]. We observed HNF1A-bound promoter P1 was upregulated in mCRPC samples with high GI scores, correlating with *SRC* gene upregulation (**Figure 4E-G**).

The observation that *SRC* transcription is linked to lineage-specific TFs that bind APs suggests AP upregulation can be linked to therapy-induced lineage plasticity. Consistent with this model, examination of the upregulated promoters in these lineages identified TFs linked to lineage differentiation.

AP activation is associated with DNA hypomethylation

DNA methylation affects whether genomic regions are accessible to TFs, and we hypothesized that it impacts AP activity. We assessed differential methylation at AP sites between t-SCNC and adenocarcinoma tumors and observed a strong negative correlation between AP activity and methylation levels at APs (Spearman's Rho = -0.7, p value = 2.1e-25, **Extended Data Figure 8A**)**.** Intriguingly, the negative correlation between gene-level expression and methylation was notably stronger at differentially active APs (Spearman's Rho = -0.31, p value= 4.4e-05, **Figure 5A**) compared to canonical promoters within the same genes (Spearman's Rho = -0.19, p value = 0.059 **Figure 5B**). This finding supports a model where methylation directly influences state-specific APs. For example, *CBX5*, encoding the heterochromatin protein HP1α, is reported to be upregulated and to facilitate the development of neuroendocrine prostate tumors[40]. The mechanism underlying *CBX5* upregulation remains unclear. The 5'-most *CBX5* promoter (P1) was methylated at similar levels in adenocarcinoma and t-SCNC. However, intragenic *CBX5* promoter P2 was hypomethylated in t-SCNC tumors with increased P2 use and elevated overall *CBX5* expression (**Figure 5C-E**). Two adenocarcinoma tumors also had hypomethylated P2 associated with higher P2 activity (**Figure 5F**). TFs predicted to bind at upregulated APs overlapping with differentially hypomethylated regions in t-SCNC were enriched for the neuroendocrine-associated TFs such as ASCL1, HAND2, NEUROD1, and SOX2 (**Extended Data Figure 8B**).

We then tested the hypothesis that DNA hypomethylation is a general mechanism for AP activation. Canonical promoter methylation has modest inverse correlation with gene expression[6]. We hypothesized that methylation levels at APs would exhibit higher variance and stronger association with gene expression. Indeed, APs displaying differential activity in mCRPC had higher methylation variance compared to either canonical promoters (defined in GENCODE) or major promoters (defined as the most active promoter for each gene in our data) (**Figure 5G**). Differential APs exhibited a significantly stronger negative correlation between promoter methylation and gene expression compared to those in the other groups (**Figure 5H**). This suggests promoter methylation at APs plays an important role in driving gene expression.

Collectively, these observations suggest methylation is linked to AP activity and may provide a mechanism to reinforce transcriptional phenotypic changes during disease progression.

DISCUSSION

Prostate cancer is primarily driven by aberrant TF activity, most importantly the androgen receptor. Comprehensive analysis of TF activity in mCRPC is challenging due to limited tissue availability for protein assays. Although TFs can bind to DNA at regulatory sites distant from a gene being targeted, they are most influential at gene promoters. Functional assays such as CAGE have revealed that many genes harbor multiple promoters[1]. Alternative promoter usage has been reported in various cancer types and is suggested to play a role in immune editing in GI tumors[41]. However, the extent to which TFs and epigenetic modifications exploit APs to modulate overall gene expression in cancer is unclear. We conducted ultra-deep RNA sequencing of mCRPC tumors to explore how disease progression and somatic changes in TF genes influence promoter selection and elucidate how alternative promoter usage contributes to gene regulation during progression from benign prostate tissue to advanced, treatment-resistant mCRPC.

We found that promoter activity patterns evolve throughout progression to mCRPC and are linked to epigenomic alterations that arise on therapy. The total number of genes employing multiple promoters didn't significantly vary across different tumor progression states, and the proportion of 481 oncogenes with multiple active promoters was not significantly different across progression stages. However, a notable trend emerged: oncogenes more frequently switched from using single promoters to multiple promoters during progression. We then sought to identify promoters with differential usage by comparing their absolute and relative activities across progression states. Canonical promoters were not always the most active promoters in benign prostate tissue. For genes where the non-canonical promoter predominated in benign tissue, and where the canonical promoter exhibited a significant increase in both relative and absolute activity, that canonical promoter was categorized an upregulated alternative promoter. We identified both canonical and non-canonical promoters as APs in localized PCa and mCRPC.

A minority of APs switched from inactive or minor status in benign tissue to major promoters in tumors. Nevertheless, APs significantly contributed to increases in mRNA abundance. In mCRPC, upregulated APs accounted for nearly all of the total increased transcriptional activity in genes featuring an AP. This suggested that the increase in promoter activity disproportionately favored alternative promoters during progression, highlighting alternative promoter usage as a vital mechanism for gene upregulation in this context. While alternative promoter usage was not the dominant mechanism for elevating gene expression, it contributed to transcriptional activity associated with disease progression.

Our investigation of AP regulation mechanisms first focused on *AR* and *FOXA1* because of their central roles in prostate tumor biology. Increased *AR* expression in both localized and metastatic PCa correlated with increased numbers of upregulated APs in individual samples. APs upregulated during tumor progression were enriched for binding sites of both AR and

FOXA1. Only four genes in the Hallmark androgen response pathway harbored upregulated APs in localized PCa, with 13 in mCRPC. Despite increased AR signaling during PCa progression, upregulated APs were not restricted to canonical AR target genes, and AR may bind to and activate promoters of both canonical and non-canonical AR targets. Our results were consistent with the model that AR and FOXA1 binding is enriched at upregulated APs. APs with experimentally observed FOXA1 binding had a lower proportion of AR-FOXA1 co-binding in mCRPC compared to localized PCa. Since FOXA1 is a pioneer factor which opens chromatin to facilitate binding of other TFs, this may indicate that FOXA1 serves as a pioneer factor for non-AR TFs in mCRPC tumors. This aligns with observations that enzalutamide-induced reprogramming of FOXA1 favors active *cis*-regulatory elements that drive pro-survival signals via non-AR TFs such as ARNTL[42,43].

We also observed less enrichment for AR binding in upregulated mCRPC APs than localized APs, implicating the involvement of additional TFs driving AP activation in mCRPC. Indeed, APs containing MYC and E2F binding were the most enriched in advanced disease. We observed enrichment of EZH2 in upregulated APs in mCRPC. EZH2 is the primary enzymatic catalytic subunit of polycomb repressive complex 2 (PRC2), a histone methyltransferase complex frequently upregulated in aggressive advanced PCa[44,45]. EZH2 is overexpressed in PCa, and it directly interacts with MYC[27]. While EZH2's best-studied function involves gene repression, it has a non-canonical activation role in cancer[28,45,46]. Specifically, EZH2 has been identified as a transcription co-activator for AR[45] and has been shown to mediate gene activation directly by binding to MYC, thereby promoting oncogenesis[27].

Tumor lineage switching, a phenomenon observed in therapy-resistance mCRPC[47] that also arises in other tissue sites[47,48], is closely tied to lineage-specific TFs. We observed that APs in mCRPC tumors with neuroendocrine or gastrointestinal lineage phenotypes were preferably bound by lineage-specific TFs. The inverse correlation between gene expression and DNA methylation at canonical promoters is widely recognized[5]. Canonical promoters of genes expressed in prostate cancer are generally hypomethylated within prostate tumors. DNA methylation at intragenic CpGs is much more variable[49]. In PCa, tumor-specific DNA methylation has been linked to alternative promoters in certain genes like *RASSF1*, *NDRG2*, and *APC*[50]. By studying sample-matched DNA methylation and transcriptomes we observed that DNA methylation levels at alternative promoters were strongly associated with alternative promoter activity and overall gene expression. This supported a model of methylation as a mechanism of AP regulation during lineage switching.

Although we did not identify APs in known prostate cancer driver oncogenes such as *AR* and *MYC*, our results support alternative promoter usage as a mechanism by which these drivers alter the tumor transcriptome. Our alternative promoter analysis produced an unbiased view of transcriptional dysregulation and rewiring of TF binding profiles throughout PCa progression.

Extended Data Figure Extended Data Figure Extended Data Figure Extended Data Figure Extended Data Figure Extended Data Figure Extended Data Figure Extended Data Figure Extended Data Figure Extended Data Figure Extended Data Figure Extended Data Figure

Data availability

RNA sequencing and whole genome bisulfite sequencing data that support the findings of this study have been deposited in the European Genome-Phenome Archive (EGA) under accession code EGAS00001006275, and the SRA repository with Bioproject number PRJNA479544. Previously published RNA sequencing that were re-analyzed here are available under accession code GSE115414, EGAD00001004424, and

GSE119757. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Code and source data availability

The R scripts and source data used to reproduce all figures and tables in this manuscript are available on Github: https://github.com/DavidQuigley/WCDT_alternative_promoter. The source data to reproduce all the figures and tables are available on Zenodo: https://zenodo.org/records/10966958.

ACKNOWLEDGEMENTS

We thank the patients who selflessly contributed samples to this study and without whom this research would not have been possible. This research was supported by a Stand Up To Cancer-Prostate Cancer Foundation Prostate Cancer Dream Team Award (SU2C-AACR-DT0812 to E.J.S.) and by the Movember Foundation. Stand Up To Cancer is a division of the Entertainment Industry Foundation. This research grant was administered by the American Association for Cancer Research, the scientific partner of SU2C. R.A. and MS were funded by a Prostate Cancer Foundation Young Investigator Award. D.A.Q. was funded by a Young Investigator and Challenge awards from the PCF and by the UCSF Benioff Initiative for Prostate Cancer Research. F.Y.F. was funded by Prostate Cancer Foundation Challenge Awards. Additional funding was provided by a UCSF Benioff Initiative for Prostate Cancer Research award. F.Y.F. was supported by National Institutes of Health (NIH)/National Cancer Institute (NCI) 1R01CA230516-01. F.Y.F. was supported by NIH/NCI 1R01CA227025 and Prostate Cancer Foundation (PCF) 17CHAL06. F.Y.F. was supported by NIH P50CA186786.

AUTHOR CONTRIBUTIONS

The studies were conceptualized and designed by M.Z, F.Y.F, and D.A.Q. Data analysis was carried out by M.Z, M.S, R.S, A.L, and H.X.D. Validation experiments were performed by X.C and H.L. Biopsy samples were processed by A.F and K.F. Resources were contributed by P.G.F, R.A, J.A, E.J.S, and the SU2C/PCF West Coast Prostate Cancer Dream Team. Supervision was provided by C.A.M, F.Y.F, and D.A.Q. The first draft of the manuscript was written by M.Z, M.S, and D.A.Q. All authors revised and approved the manuscript.

COMPETING INTERESTS

J.J.A. has consulted for or held advisory roles at Astellas Pharma, Bayer and Janssen Biotech Inc. He has received research funding from Aragon Pharmaceuticals Inc., Astellas Pharma, Novartis, Zenith Epigenetics Ltd. and Gilead Sciences Inc. F.Y.F. has consulted for Astellas, Bayer, Blue Earth Diagnostics, BMS, EMD Serono, Exact Sciences, Foundation Medicine, Janssen Oncology, Myovant, Roivant, and Varian, and serves on the Scientific Advisory Board for BlueStar Genomics and SerImmune. F.Y.F. has patent applications with Decipher Biosciences on molecular signatures in prostate cancer unrelated to this work. F.Y.F. has a patent application licensed to PFS Genomics/Exact Sciences. F.Y.F. has patent applications with Celgene unrelated to this work. The remaining authors declare no competing interests.

FIGURE LEGENDS

Figure 1. Activation and upregulation of alternative promoters are associated with increased expression of disease related genes during prostate cancer progression.

A. Upregulated and downregulated genes were identified by differential gene expression analysis. Oncogenes and upregulated genes are enriched for switching from having a single promoter active in benign prostate to multiple promoters active in localized PCa (left) or mCRPC (right). The total number of genes in each category (T) and the number of genes that switched from SP to MP (N) are labeled next to the bars. SP: single-promoter active, MP: multiple-promoter active. (Fisher's exact tests, two-sided).
B. Differentially used alternative promoters were identified based on statistically significant differences in both absolute and relative activities by running the DEXseq differential exon usage analysis using promoter counts, and proActiv in corresponding comparisons (see Methods for details). AP: alternative promoter.
C. Principal component analysis of all samples of different disease stages from three cohorts. PAIR: from the Henri Mondor institution, CPCG: Canadian Prostate Cancer Genome Network, WCDT: West Coast Dream Team, t-SCNC: treatment-emergent small cell neuroendocrine carcinoma.
D. Density plot of the correlation between absolute promoter activity and corresponding gene expression levels for upregulated APs (red), downregulated APs (blue) and non-differential promoters (gray) in genes with differential APs in mCRPC vs benign.
E. Density plot of the percentage of increased activity of upregulated APs over total increased activity of all promoters of the AP-containing genes in localized PCa and mCRPC. The other promoters from the AP-containing genes were plotted as controls. (Student's t-tests, two-sided).).
F. Tracks plot showing the mean normalized RNA-seq coverage of benign and mCRPC samples over the *RALBP1* gene on chromosome 18. Two annotated promoters (P1 and P2) are highlighted by shadows. CPM: counts per million reads.
G. Box plot showing the relative activity of *RALBP1* P1 and P2 in individual samples grouped by benign and mCRPC (Student's t-test) (n = 8 for benign, n = 101 for mCRPC adeno). Box plots show data from the 25th to the 75th percentile, with the median as a line inside the box. Whiskers extend to 1.5 times the interquartile range (IQR) from the lower and upper quartiles.

Figure 2. FOXA1 binding and androgen signaling are associated with alternative promoter usage in PCa.

A. Correlation between the number of upregulated APs in individual localized PCa samples and AR expression levels. 95% confidence interval for the predictions from a linear model is displayed. (Spearman's correlation test, two-sided)
B. Left: The percentage of upregulated APs in localized PCa and canonical promoters of Hallmark AR targets that overlap with localized PCa_specific AR ChIP-seq peaks; Right: The percentage of upregulated APs in mCRPC and canonical promoters of Hallmark AR targets that overlap with mCRPC PDX-specific AR ChIP-seq peaks. (Fisher's exact test, two-sided).
C. Left: The percentage of upregulated APs in localized PCa and canonical promoters of Hallmark AR targets that overlap with localized PCa_specific FOXA1 ChIP-seq peaks; Right: The percentage of upregulated APs in mCRPC and canonical promoters of Hallmark AR targets that overlap with mCRPC PDX-specific FOXA1 ChIP-seq peaks. (Fisher's exact test, two-sided).
D. The percentage of overlapping with FOXA1 ChIP-seq peaks in control and *FOXA1* knockdown (shFOXA1) LNCaP cells in upregulated APs in localized PCa (middle) and mCRPC (right) with evidence of FOXA1 binding by FOXA1 ChIP-seq used in Figure 2C. (Fisher's exact test one-sided).

E. The percentage of FOXA1-bound upregulated APs in localized and mCRPC showing downregulated activity upon *FOXA1* knockdown (shFOXA1) in LNCaP cells. (Fisher's exact test, two-sided).

Figure 3. MYC is a potential driver of alternative promoter activation in mCRPC.

A. UniBind results for top three TFs in localized PCa and mCRPC. Each dot represents one ChIP-seq dataset (n = 194, 70, 4, 4, 11, and 5 for AR, FOXA1, GATA2, MYC, E2F1, and HIF1A). TFs were ranked by the ChIP-seq dataset with the most significant overlap with upregulated APs. P values were calculated using Fisher's exact test. Y axis shows the p values without multi-test adjustments, but the horizontal dashed line shows the corresponding Benjamini Hochberg (BH)-adjusted p value 0.05.
B. UniBind results showing significance of overlap between TF ChIP-seq peaks and upregulated APs in *MYC* expression high vs. low mCRPC samples. Each dot represents one ChIP-seq dataset. TFs were ranked by the most significant ChIP-seq dataset. P values were calculated using Fisher's exact tests. Y axis shows the p values without multi-test adjustments, but the horizontal dashed line shows the corresponding BH-adjusted p value 0.05.
C. The percentage of upregulated APs and EZH2 bound upregulated APs in mCRPC that overlapped with MYC ChIP-seq peaks in LNCaP cells. (Fisher's exact test, two-sided).
D. The percentage of upregulated APs and MYC bound upregulated APs in mCRPC that overlapped with EZH2 ChIP-seq peaks in LNCaP cells. (Fisher's exact test, two-sided).
E. The percentage of upregulated APs in mCRPC and canonical promoters of upregulated genes in mCRPC that overlapped with both MYC and EZH2 ChIP-seq peaks in LNCaP cells. (Fisher's exact test, two-sided).
F. Tracks plot showing the mean normalized RNA-seq coverage of benign and mCRPC samples over the *BMI1* gene on chromosome 10. Two annotated promoters (P1 and P2) are highlighted by shadows. EZH2 and MYC ChIP-seq peaks in LNCaP cells are displayed. CPM: counts per million reads.
G. Box plot showing the absolute activity of *BMI1* P1 and P2 in individual samples grouped by benign and mCRPC (Student's t-test) (n = 8 for benign, n = 101 for mCRPC). Box plots show data from the 25^{th} to the 75^{th} percentile, with the median as a line inside the box. Whiskers extend to 1.5 times the interquartile range (IQR) from the lower and upper quartiles.

Figure 4. Alternative promoter usage reflects lineage plasticity in response to therapy.

A. UniBind results showing significance of overlap between TF ChIP-seq peaks and upregulated APs in treatment emergent small cell neuroendocrine carcinoma (t-SCNC) vs adenocarcinoma mCRPC samples. Each dot represents one ChIP-seq dataset. TFs were ranked by the most significant ChIP-seq dataset. P values were calculated using Fisher's exact tests. Y axis shows the p values without multi-test adjustments, but the horizontal dashed line shows the corresponding BH-adjusted p value 0.05.
B. Box plot showing *HAND2* expression in mCRPC adenocarcinoma (adeno) and t-SCNC tumors (Student's t-test) (n = 101 for adeno, n = 3 for tSCNC). Box plots show data from the 25^{th} to the 75^{th} percentile, with the median as a line inside the box. Whiskers extend to 1.5 times the interquartile range (IQR) from the lower and upper quartiles.
C. Pathway enrichment analysis of genes with upregulated APs in t-SCNC vs adenocarcinoma that overlapped with HAND2 ChIP-seq peaks. X axis shows the p values without multi-test adjustments, but the coloring was based on BH-adjusted p values. Dashed line shows unadjusted p value 0.05.
D. UniBind results showing significance of overlap between TF ChIP-seq peaks and upregulated APs in tumors with high gastrointestinal (GI) scores. Each dot represents one

ChIP-seq dataset. TFs were ranked by the most significant ChIP-seq dataset. P values were calculated using Fisher's exact tests. Y axis shows the p values without multi-test adjustments, but the horizontal dashed line shows the corresponding BH-adjusted p value 0.05.

E. Tracks plot showing the mean normalized RNA-seq coverage of mCRPC samples with high and low GI score over the 5' part of the *SRC* gene. Two annotated promoters (P1 and P2) are highlighted by shadows. CPM: counts per million reads.
F. Box plot showing the relative promoter activity of *SRC* P1 and P2 in individual samples grouped by GI score high and low (Student's t-test, two-sided) (n = 79 for GI low, n = 25 for GI high). Box plots show data from the 25th to the 75th percentile, with the median as a line inside the box. Whiskers extend to 1.5 times the interquartile range (IQR) from the lower and upper quartiles.
G. Box plot showing the gene expression of *SRC* in individual samples grouped by GI score levels (Student's t-test, two-sided) (n = 79 for GI low, n = 25 for GI high). Box plots show data from the 25th to the 75th percentile, with the median as a line inside the box. Whiskers extend to 1.5 times the interquartile range (IQR) from the lower and upper quartiles.

Figure 5. Activation of alternative promoters is associated with DNA hypomethylation.

A. Correlation between the gene expression fold change and methylation differences at alternative promoters differentially active between mCRPC t-SCNC and adenocarcinoma. 95% confidence interval for the predictions from a linear model is displayed. (Spearman's correlation test, two-sided)
B. Correlation between the gene expression fold change and methylation differences at canonical promoters of the genes harboring differential APs between mCRPC t-SCNC and adenocarcinoma. 95% confidence interval for the predictions from a linear model is displayed. (Spearman's correlation test, two-sided)
C. Tracks plot showing the mean normalized RNA-seq coverage of mCRPC t-SCNC and adenocarcinoma samples over the 5' region of the *CBX5* gene. Two annotated promoters (P1 and P2) are highlighted by shadows. CPM: counts per million reads. DMR: differentially methylated region. HMR: hypomethylated region. PC_TX: protein-coding transcript; NC_TX: non-coding transcript.
D. Box plot showing the activity of *CBX5* P1 and P2 in individual samples grouped by t-SCNC and adenocarcinoma (Student's t-test, two-sided) (n = 101 for adeno, n = 3 for tSCNC). Box plots show data from the 25th to the 75th percentile, with the median as a line inside the box. Whiskers extend to 1.5 times the interquartile range (IQR) from the lower and upper quartiles.
E. Box plot showing the gene expression of *CBX5* in individual samples grouped by t-SCNC and adenocarcinoma phenotype (Student's t-test, two-sided) (n = 101 for adeno, n = 3 for tSCNC). Box plots show data from the 25th to the 75th percentile, with the median as a line inside the box. Whiskers extend to 1.5 times the interquartile range (IQR) from the lower and upper quartiles.
F. Box plot showing the activity of *CBX5* P2 in individual samples grouped by harboring an hypomethylated region (HMR) at P2 or not (Student's t-test, two-sided) (n = 99 for No HMR, n = 5 for HMR). Box plots show data from the 25th to the 75th percentile, with the median as a line inside the box. Whiskers extend to 1.5 times the interquartile range (IQR) from the lower and upper quartiles.
G. Standard deviation of methylation levels at recurrent hypomethylated regions (rHMR) overlapping with canonical promoters (as defined in the GENCODE gene model), major promoters in mCRPC (the most active promoters of each gene), and alternative promoters with differential activity within the mCRPC cohort. (Student's t-test, two-sided) (n = 16,058 for Canonical promoters, n = 12,468 for Major promoters in mCRPC, n = 490 for Promoters

with alternative usage within mCRPC). Box plots show data from the 25th to the 75th percentile, with the median as a line inside the box. Whiskers extend to 1.5 times the interquartile range (IQR) from the lower and upper quartiles.

H. Correlation between promoter methylation and gene expression at canonical promoters, major promoters, and alternative promoters with differential activity within the mCRPC cohort. (Student's t-test, two-sided).

REFERENCES

1 Carninci, P. *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38**, 626-635 (2006). https://doi.org/10.1038/ng1789

2 Landry, J. R., Mager, D. L. & Wilhelm, B. T. Complex controls: the role of alternative promoters in mammalian genomes. *Trends Genet* **19**, 640-648 (2003). https://doi.org/10.1016/j.tig.2003.09.014

3 Demircioğlu, D. *et al.* A Pan-cancer Transcriptome Analysis Reveals Pervasive Regulation through Alternative Promoters. *Cell* **178**, 1465-1477.e1417 (2019). https://doi.org/10.1016/j.cell.2019.08.018

4 Davuluri, R. V., Suzuki, Y., Sugano, S., Plass, C. & Huang, T. H. The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet* **24**, 167-177 (2008). https://doi.org/10.1016/j.tig.2008.01.008

5 Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol* **20**, 590-607 (2019). https://doi.org/10.1038/s41580-019-0159-6

6 Zhao, S. G. *et al.* The DNA methylation landscape of advanced prostate cancer. *Nat Genet* **52**, 778-789 (2020). https://doi.org/10.1038/s41588-020-0648-8

7 Robinson, D. *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215-1228 (2015). https://doi.org/10.1016/j.cell.2015.05.001

8 Quigley, D. A. *et al.* Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. *Cell* **174**, 758-769.e759 (2018). https://doi.org/10.1016/j.cell.2018.06.039

9 Beltran, H. *et al.* Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat Med* **22**, 298-305 (2016). https://doi.org/10.1038/nm.4045

10 Nouruzi, S. *et al.* ASCL1 activates neuronal stem cell-like lineage programming through remodeling of the chromatin landscape in prostate cancer. *Nat Commun* **13**, 2282 (2022). https://doi.org/10.1038/s41467-022-29963-5

11 Hua, J. T. *et al.* Risk SNP-Mediated Promoter-Enhancer Switching Drives Prostate Cancer through lncRNA PCAT19. *Cell* **174**, 564-575.e518 (2018). https://doi.org/10.1016/j.cell.2018.06.014

12 Aggarwal, R. *et al.* Clinical and Genomic Characterization of Treatment-Emergent Small-Cell Neuroendocrine Prostate Cancer: A Multi-institutional Prospective Study. *J Clin Oncol* **36**, 2492-2503 (2018). https://doi.org/10.1200/JCO.2017.77.6880

13 Aggarwal, R. R. *et al.* Whole-Genome and Transcriptional Analysis of Treatment-Emergent Small-Cell Neuroendocrine Prostate Cancer Demonstrates Intraclass Heterogeneity. *Mol Cancer Res* **17**, 1235-1240 (2019). https://doi.org/10.1158/1541-7786.Mcr-18-1101

14 Pinskaya, M. *et al.* Reference-free transcriptome exploration reveals novel RNAs for prostate cancer diagnosis. *Life Sci Alliance* **2** (2019). https://doi.org/10.26508/lsa.201900449
15 Chen, S. *et al.* Widespread and Functional RNA Circularization in Localized Prostate Cancer. *Cell* **176**, 831-843.e822 (2019). https://doi.org/10.1016/j.cell.2019.01.025
16 Pomerantz, M. M. *et al.* Prostate cancer reactivates developmental epigenomic programs during metastatic progression. *Nat Genet* **52**, 790-799 (2020). https://doi.org/10.1038/s41588-020-0664-8
17 Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Res* **48**, D682-d688 (2020). https://doi.org/10.1093/nar/gkz966
18 Jinesh, G. G. & Kamat, A. M. RalBP1 and p19-VHL play an oncogenic role, and p30-VHL plays a tumor suppressor role during the blebbishield emergency program. *Cell Death Discov* **3**, 17023 (2017). https://doi.org/10.1038/cddiscovery.2017.23
19 Dai, C., Heemers, H. & Sharifi, N. Androgen Signaling in Prostate Cancer. *Cold Spring Harb Perspect Med* **7** (2017). https://doi.org/10.1101/cshperspect.a030452
20 Jin, H. J., Zhao, J. C., Wu, L., Kim, J. & Yu, J. Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. *Nat Commun* **5**, 3972 (2014). https://doi.org/10.1038/ncomms4972
21 Puig, R. R., Boddie, P., Khan, A., Castro-Mondragon, J. A. & Mathelier, A. UniBind: maps of high-confidence direct TF-DNA interactions across nine species. *BMC Genomics* **22**, 482 (2021). https://doi.org/10.1186/s12864-021-07760-6
22 Chen, W. S. *et al.* Genomic Drivers of Poor Prognosis and Enzalutamide Resistance in Metastatic Castration-resistant Prostate Cancer. *European urology* **76**, 562-571 (2019). https://doi.org/10.1016/j.eururo.2019.03.020
23 Qiu, X. *et al.* MYC drives aggressive prostate cancer by disrupting transcriptional pause release at androgen receptor targets. *Nat Commun* **13**, 2559 (2022). https://doi.org/10.1038/s41467-022-30257-z
24 Tran, M. G. B. *et al.* Independence of HIF1a and androgen signaling pathways in prostate cancer. *BMC Cancer* **20**, 469 (2020). https://doi.org/10.1186/s12885-020-06890-6
25 Lin, C. Y. *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151**, 56-67 (2012). https://doi.org/10.1016/j.cell.2012.08.026
26 Barfeld, S. J. *et al.* c-Myc Antagonises the Transcriptional Activity of the Androgen Receptor in Prostate Cancer Affecting Key Gene Networks. *EBioMedicine* **18**, 83-93 (2017). https://doi.org/10.1016/j.ebiom.2017.04.006
27 Wang, J. *et al.* EZH2 noncanonically binds cMyc and p300 through a cryptic transactivation domain to mediate gene activation and promote oncogenesis. *Nat Cell Biol* **24**, 384-399 (2022). https://doi.org/10.1038/s41556-022-00850-x
28 Kim, J. *et al.* Polycomb- and Methylation-Independent Roles of EZH2 as a Transcription Activator. *Cell Rep* **25**, 2808-2820.e2804 (2018). https://doi.org/10.1016/j.celrep.2018.11.035
29 Serresi, M. *et al.* Functional antagonism of chromatin modulators regulates epithelial-mesenchymal transition. *Sci Adv* **7** (2021). https://doi.org/10.1126/sciadv.abd7974

30 van Leenders, G. J. *et al.* Polycomb-group oncogenes EZH2, BMI1, and RING1 are overexpressed in prostate cancer with adverse pathologic and clinical features. *Eur Urol* **52**, 455-463 (2007). https://doi.org/10.1016/j.eururo.2006.11.020

31 Guo, B. H. *et al.* Bmi-1 promotes invasion and metastasis, and its elevated expression is correlated with an advanced stage of breast cancer. *Mol Cancer* **10**, 10 (2011). https://doi.org/10.1186/1476-4598-10-10

32 Lundberg, A. *et al.* The genomic and epigenomic landscape of double-negative metastatic prostate cancer. *Cancer research* (2023). https://doi.org/10.1158/0008-5472.Can-23-0593

33 Lei, J. & Howard, M. J. Targeted deletion of Hand2 in enteric neural precursor cells affects its functions in neurogenesis, neurotransmitter specification and gangliogenesis, causing functional aganglionosis. *Development* **138**, 4789-4800 (2011). https://doi.org/10.1242/dev.060053

34 Hendershot, T. J. *et al.* Conditional deletion of Hand2 reveals critical functions in neurogenesis and cell type-specific gene expression for development of neural crest-derived noradrenergic sympathetic ganglion neurons. *Dev Biol* **319**, 179-191 (2008). https://doi.org/10.1016/j.ydbio.2008.03.036

35 Mitchell, P. J., Timmons, P. M., Hébert, J. M., Rigby, P. W. & Tjian, R. Transcription factor AP-2 is expressed in neural crest cell lineages during mouse embryogenesis. *Genes Dev* **5**, 105-119 (1991). https://doi.org/10.1101/gad.5.1.105

36 Sjöström, M. *et al.* The 5-Hydroxymethylcytosine Landscape of Prostate Cancer. *Cancer Res*, Of1-of15 (2022). https://doi.org/10.1158/0008-5472.can-22-1123

37 Shukla, S. K. *et al.* MUC1 and HIF-1alpha Signaling Crosstalk Induces Anabolic Glucose Metabolism to Impart Gemcitabine Resistance to Pancreatic Cancer. *Cancer Cell* **32**, 71-87.e77 (2017). https://doi.org/10.1016/j.ccell.2017.06.004

38 Lau, H. H., Ng, N. H. J., Loo, L. S. W., Jasmen, J. B. & Teo, A. K. K. The molecular functions of hepatocyte nuclear factors - In and beyond the liver. *J Hepatol* **68**, 1033-1048 (2018). https://doi.org/10.1016/j.jhep.2017.11.026

39 Bonham, K., Ritchie, S. A., Dehm, S. M., Snyder, K. & Boyd, F. M. An alternative, human SRC promoter and its regulation by hepatic nuclear factor-1alpha. *J Biol Chem* **275**, 37604-37611 (2000). https://doi.org/10.1074/jbc.M004882200

40 Ci, X. *et al.* Heterochromatin Protein 1α Mediates Development and Aggressiveness of Neuroendocrine Prostate Cancer. *Cancer Res* **78**, 2691-2704 (2018). https://doi.org/10.1158/0008-5472.can-17-3677

41 Qamra, A. *et al.* Epigenomic Promoter Alterations Amplify Gene Isoform and Immunogenic Diversity in Gastric Adenocarcinoma. *Cancer Discov* **7**, 630-651 (2017). https://doi.org/10.1158/2159-8290.cd-16-1022

42 Zhang, M., Moreno-Rodriguez, T. & Quigley, D. A. in *Cancer Discov* Vol. 12 2017-2019 (©2022 American Association for Cancer Research., 2022).

43 Linder, S. *et al.* Drug-Induced Epigenomic Plasticity Reprograms Circadian Rhythm Regulation to Drive Prostate Cancer toward Androgen Independence. *Cancer Discov* **12**, 2074-2097 (2022). https://doi.org/10.1158/2159-8290.cd-21-0576

44 Labbé, D. P. *et al.* TOP2A and EZH2 Provide Early Detection of an Aggressive Prostate Cancer Subgroup. *Clin Cancer Res* **23**, 7072-7083 (2017). https://doi.org/10.1158/1078-0432.ccr-17-0413
45 Xu, K. *et al.* EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. *Science* **338**, 1465-1469 (2012). https://doi.org/10.1126/science.1227604
46 Anwar, T., Gonzalez, M. E. & Kleer, C. G. Noncanonical Functions of the Polycomb Group Protein EZH2 in Breast Cancer. *Am J Pathol* **191**, 774-783 (2021). https://doi.org/10.1016/j.ajpath.2021.01.013
47 Beltran, H. *et al.* The Role of Lineage Plasticity in Prostate Cancer Therapy Resistance. *Clin Cancer Res* **25**, 6916-6924 (2019). https://doi.org/10.1158/1078-0432.ccr-19-1423
48 Ferguson, A. M. & Rubin, M. A. Lineage plasticity in prostate cancer: Looking beyond intrinsic alterations. *Cancer Lett* **548**, 215901 (2022). https://doi.org/10.1016/j.canlet.2022.215901
49 Kulis, M., Queirós, A. C., Beekman, R. & Martín-Subero, J. I. Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. *Biochim Biophys Acta* **1829**, 1161-1174 (2013). https://doi.org/10.1016/j.bbagrm.2013.08.001
50 Kim, J. H. *et al.* Deep sequencing reveals distinct patterns of DNA methylation in prostate cancer. *Genome Res* **21**, 1028-1041 (2011). https://doi.org/10.1101/gr.119347.110

METHODS

RNA sequencing and processing

Human studies were approved and overseen by the UCSF Institutional Review Board. All individuals provided written informed consent to obtain fresh tumor biopsies and to perform comprehensive molecular profiling of tumor and germline samples. Patients donated samples and were not compensated.

For the WCDT cohort, RNA was extracted from between 2 and 8 25um sections (50-200um total) of frozen tissue using the Qiagen UCP RNeasy Micro kit. Total RNA libraries were generated using the Kapa Hyperprep Total RNAseq kit, with Qiagen FastSelect rRNA depletion. The libraries were sequenced on the Illumina NovaSeq platform at paired end 150bp. Fastq files were first trimmed to remove adapter sequences using Cutadapt[51] and then aligned to the human reference genome build hg38 using the splice-aware aligner STAR[52] in 2-pass mode. The splice junction output (_SJ.out.tab) from STAR was used for promoter activity estimation described in the section below.

Raw fastq files for RNA-seq data from the PAIR cohort, CPCG cohort, and shFOXA1 and Ctrl LNCaP cells were downloaded from GSE115414, EGAD00001004424, and GSE119757, respectively. These RNA-seq data were processed using the same pipeline as our in-house RNA-seq data. We assessed positional biases in the RNA-seq datasets (WCDT, PAIR and CPCG) using the biasPlot function from R package "EDASeq"[53], and did not observe notable differences between datasets (**Extended Data Figure Extended Data Figure 2D**). To assess the effect of differing sequencing depths on our analysis, we computationally down-sampled the RNA-seq data in all three cohorts to 80M reads per sample, which was the lowest depth of all samples. We then re-calculated promoter activities using the down-sampled dataset, assessed promoter activity for each category of genes, and performed differential promoter

usage analysis between disease stages. As expected, we identified fewer genes that switched from single-promoter active to multiple-promoter activity in the down-sampled dataset, but the conclusions remained the same: both genes with upregulated expression in localized PCa or mCRPC (compared to benign) and known oncogenes were more likely to switch from a single promoter active in benign tissue to multiple promoters active in tumors compared to all expressed genes (**Extended Data Figure Extended Data Figure 2E**). Reassuringly, in the down-sampled dataset we observed a similar separation of the samples by disease type in the principal component analysis (**Extended Data Figure 2F**).

Data exclusion statement

One localized prostate cancer RNA-seq sample from the CPCG cohort (CPCG0416) was excluded from the analysis due to extremely low unique mapping rate (2.21%) using our processing pipeline.

Estimation of promoter activity

Promoter annotation

Promoters were defined as the region around the first 5' TSSs (transcription start sites, defined as the start of the first annotated exon) of overlapping first exons[3,54]. To create a comprehensive gene model, we combined GENCODE 32[55] and NCBI RefSeq[56]. GENCODE 32 was used as the main gene model. Additional transcripts in RefSeq that have distinct inner exons from GENCODE 32 (excluding first and last exons to avoid redundancy due to noisy annotation of the gene starts and ends) were added. Promoters were then annotated from the customized gene model using the promoterAnnotation function from the R package proActiv[3].

Promoter region definition

To identify transcription factor binding enrichment near TSS sites, we employed the 400bp promoter region definition recommended by the widely used transcription factor motif discovery tool Homer "-size -300, 100"[54]. This choice minimized the overlap between different promoters of the same gene. Using the 400bp definition, 15% of promoters from multi-promoter genes (7,765/49,976) overlapped with another promoter in the same gene; a wider search space (e.g. 1,500bp, -1000, 500) would produce overlaps in 41.4% of genes."

Promoter activity estimation

Absolute and relative promoter activities were calculated using the proActiv function from the R package proActiv[3]. Briefly, the absolute activity of a promoter is the total count of the junction reads aligning to the set of first introns belonging to the transcripts associated with that promoter. The absolute activities were normalized by DESeq2 size factors[57]. The relative activity of a promoter is the proportion of the absolute activity of that promoter divided by the sum of the absolute activity of all promoters of that gene. Major promoters are defined as the promoters with the highest activity for a gene.

Internal promoter correction

An internal promoter is defined as a promoter where the first exon overlaps with internal exons in other isoforms (*P2* in **Extended Data Figure 1B**). The junction reads mapping to these exons cannot be unambiguously assigned to the internal promoter as they could also be from the transcription of other isoforms that use the exons internally. Although the internal promoters were excluded in prior pan-cancer analysis[3], they developed a split read ratios method to correct for them. The ratios of the first exons' donor sites junction reads against their acceptor sites junction reads were used as the corrected promoter activity (**Extended Data Figure 1B).** To assess the effectiveness of the correction, we performed correlation tests between the

CAGE (cap analysis of gene expression) tag reads and the uncorrected and corrected promoter activity calculated from the RNA-seq data from matching samples from FANTOM5[58]. The split read ratios method corrected activities showed an improved correlation with the CAGE signal (median of Spearman's Rho 0.43 vs 0.35, Student's test p value = 2.4×10^{-25}, **Extended Data Figure 1C**).

However, the ratios are not reflective of the actual abundance of transcription and not comparable to non-internal promoters. We developed a split read subtractions method which uses subtractions instead of ratios for correction (**Extended Data Figure 1B**). The correction excludes the transcription of other isoforms from that initiated from the internal promoters and preserves the actual expression levels. The correlation with the CAGE signal was further improved (median of Spearman's Rho 0.51 vs 0.35, Student's test p value = 7.71×10^{-55}, **Extended Data Figure 1C**), and even to a similar extent as the non-internal promoters (Spearman's Rho 0.52, **Extended Data Figure 1C**).

To be conservative, we restricted candidate internal promoters to those with additional evidence: either the promoter was annotated as canonical in the Ensembl database[17], or it overlapped with H3K4me3 ChIP peaks in benign prostate or localized PCa[16], or it was located in recurrently hypomethylated regions in mCRPC[6]. A larger proportion of highly confident promoters were not internal, but the correlation with CAGE signal for internal promoters demonstrated the value of including these promoters after correction (**Extended Data Figure 1D**).

There is an edge case where a unique prior intron cannot be found for the isoform containing the internal exon (**Extended Data Figure 1B**, TX_B). In this case, SJ_a and SJ_b would be the same junction, and thus the split read subtractions method [SJ_c-$SJ_{a|b}$] underestimates the activity of internal promoter *P2* by incorrectly subtracting transcriptional activity from the SJ_a containing isoform (**Extended Data Figure 1B**, TX_A). 9.1% (822 / 9,015) of the highly confident internal promoters fall into this scenario and were annotated in **Supplementary Table 1**.

Differential promoter usage analysis

Identification of differentially used promoters between two conditions

Both absolute and relative promoter activities were considered as differential alternative promoter identification criteria. A differential exon usage (DEU) analysis was performed on the promoter activities by using the R package DEXSeq[59] to identify differentially used alternative promoters between conditions within the mCRPC cohort (MYC expression high vs low, t-SCNC vs adeno, and GI signature high vs low). The DEXSeq analysis measured changes in the relative usage of promoters under different conditions, where the relative usage of a promoter is defined as $\frac{\text{number of transcripts from the gene that start with this promoter}}{\text{number of all transcripts from the gene}}$. A promoter was considered differentially used if the adjusted *p* value was < 0.05.

For comparisons between localized PCa vs benign, mCRPC vs benign and mCRPC vs localized, an additional linear model was fit using the getAlternativePromoters function from the R package proActiv[3], and a more stringent threshold was used to identify the strongest differentially active promoters in tumors. The disease stage was the only covariate used for both DEXSeq and proActiv based analyses. Cohort/data source was not a covariate because all mCRPC samples were from the WCDT cohort, while all benign prostate samples were from the PAIR cohort. Promoters meeting the following criteria were considered significantly upregulated (or downregulated) in tumors: DEXseq adjusted *p* value < 0.05, log2fc > 1 (or < -1); proActiv linear model: adjusted *p* value for both absolute and relative promoter activity estimates < 0.05, log2fc of absolute activity > 1 (or < -1), difference in relative activity > 0 (or < 0).

Identification of differentially used promoters in individual samples

The absolute and relative activity in each individual sample was compared to the median values of a cohort. Promoters were considered to have differential usage if the log2fc of the absolute activity over the median value was greater than 1 and the difference in the relative activity from the median value was greater than 10%.

Differential gene expression analysis

Gene counts were calculated using the featureCounts function from the R package subread[60]. Differential gene expression analysis was carried out on the raw counts using the R package DESeq2[57]. Genes with adjusted *p* value < 0.01 and log2 fold change > 1 (or < -1) were considered differentially upregulated (or downregulated).

Pathway enrichment analysis

The significance of overlapping between the selected genes (such as significantly upregulated genes called in the differential gene expression analysis, and genes with differentially used APs) and annotated gene sets was performed using the hypergeometric test implemented in the enricher function from the R package clusterProfiler[61]. The HALLMARK gene sets[62], Wiki Pathways[63], and the TOMLINS_PROSTATE_CANCER_UP pathway[64] were obtained from the molecular signatures database (MSigDB)[62]. Prostate cancer specific genes constructed using the pan-cancer TCGA expression data as previously described[6] were used. The GO_BP gene sets were included in the analysis for genes with upregulated APs in t-SCNC vs adenocarcinoma and in GI signature high vs low samples. All genes listed in the database of these pathways were used as the background genes.

Transcription factor binding enrichment analysis

The enrichment of transcription factor (TF) binding in the selected promoter regions was assessed by the UniBind Enrichment Analysis (https://unibind.uio.no/enrichment/)[21]. The UniBind enrichment tool predicts which sets of TFBSs from the UniBind database are enriched in a set of given genomic regions, in our case, the -300bp/+100bp region of the transcription start sites. Enrichment computations are preformed using the LOLA tool[65], which uses Fisher's exact tests with false discovery rate correction to assess the significance of overlap in each pairwise comparison. We used the *Enrichment with a background* type of analysis with the upregulated APs as the query set, and all high confident promoters with a non-NA p-value in corresponding DEXSeq output as the background set. The *Robust* collection of 4,166 ChIP-seq datasets of 268 *Homo Sapiens* TFs were used as the reference database. For the analysis of upregulated alternative promoters in localized PCa vs benign and in mCRPC vs benign, 31 TFs from the 386 ChIP-seq datasets in prostate tissues/cell lines were used.

Methylation analysis

Differential methylation analysis was performed using the DSS R package v2.26.0[66] with smoothing set to true, and otherwise using default parameters as described in our previous study[6]. Recurrent hypomethylated regions (rHMRs) were identified with the MethylSeekR[67] tool as described in our previous study[6]: rHMRs were defined by running a 100-bp sliding window across the genome and identifying contiguous regions where MethylSeekR called an HMR in ≥5% of mCRPC samples. For the correlation between promoter methylation and gene expression analysis, the -1000/+500bp regions around the overlapping first TSSs were used to calculate DNA methylation levels.

Statistics & Reproducibility

No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those reported in previous publications[3]. This study was a retrospective analysis of tumor samples, and randomization was not applicable, and no groups were allocated so blinding was not relevant. Transcriptomic sequence data were not presumed to be normally distributed and were analyzed using standard methods implemented in the DESeq and DEXseq package.

All statistical analyses were conducted using the R statistical software version 4.0.3[68]. All correlation analyses were performed using Spearman's method. Differences between groups were assessed by Student's t-tests. Fisher's exact tests were applied to determine enrichment of transcription factor binding in distinct promoter categories. All tests were two-sided unless otherwise noted, and $p < 0.05$ was considered statistically significant. Results were corrected for multiple testing using the Benjamini-Hochberg method, when applicable. All measurements were taken from distinct individual samples.

METHODS ONLY REFERENCES

51 Martin, M. (**EMBnet.journal**, 2011).

52 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**, 15-21 (2013). https://doi.org/10.1093/bioinformatics/bts635

53 Risso, D., Schwartz, K., Sherlock, G. & Dudoit, S. GC-content normalization for RNA-Seq data. *BMC Bioinformatics* **12**, 480 (2011). https://doi.org/10.1186/1471-2105-12-480

54 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010). https://doi.org/10.1016/j.molcel.2010.05.004

55 Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47**, D766-d773 (2019). https://doi.org/10.1093/nar/gky955

56 O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733-745 (2016). https://doi.org/10.1093/nar/gkv1189

57 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550 (2014). https://doi.org/10.1186/s13059-014-0550-8

58 Lizio, M. *et al.* Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol* **16**, 22 (2015). https://doi.org/10.1186/s13059-014-0560-6

59 Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res* **22**, 2008-2017 (2012). https://doi.org/10.1101/gr.133744.111

60 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)* **30**, 923-930 (2014). https://doi.org/10.1093/bioinformatics/btt656

61 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284-287 (2012). https://doi.org/10.1089/omi.2011.0118

62 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015). https://doi.org/10.1016/j.cels.2015.12.004

63 Pico, A. R. *et al.* WikiPathways: pathway editing for the people. *PLoS Biol* **6**, e184 (2008). https://doi.org/10.1371/journal.pbio.0060184

64 Tomlins, S. A. *et al.* Integrative molecular concept modeling of prostate cancer progression. *Nat Genet* **39**, 41-51 (2007). https://doi.org/10.1038/ng1935
65 Sheffield, N. C. & Bock, C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* **32**, 587-589 (2016). https://doi.org/10.1093/bioinformatics/btv612
66 Wu, H. *et al.* Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic acids research* **43**, e141 (2015). https://doi.org/10.1093/nar/gkv715
67 Burger, L., Gaidatzis, D., Schubeler, D. & Stadler, M. B. Identification of active regulatory regions from DNA methylation data. *Nucleic acids research* **41**, e155 (2013). https://doi.org/10.1093/nar/gkt599
68 R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*, **https://www.R-project.org/** (2021).

Stand Up 2 Cancer (SU2C) / Prostate Cancer Foundation (PCF) West Coast Prostate Cancer Dream Team

Rahul Aggarwal[1,2]; Joshi J. Alumkal[3]; Felix Y. Feng[1,2,4,5]; Adam Foye[2,5]; David A. Quigley[2,4,6]; Eric J. Small[1,2]

[1]Division of Hematology and Oncology, Department of Medicine, University of California San Francisco, San Francisco, California, United States; [2]Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, California, United States; [3]Department of Medicine, University of Michigan, Ann Arbor, Michigan, United States; [4]Department of Urology, University of California San Francisco, San Francisco, California, United States; [5]Department of Radiation Oncology, University of California San Francisco, San Francisco, California, United States; [6]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, United States

Figure 1. Activation and upregulation of alternative promoters are associated with increased expression of disease related genes during prostate cancer progression.

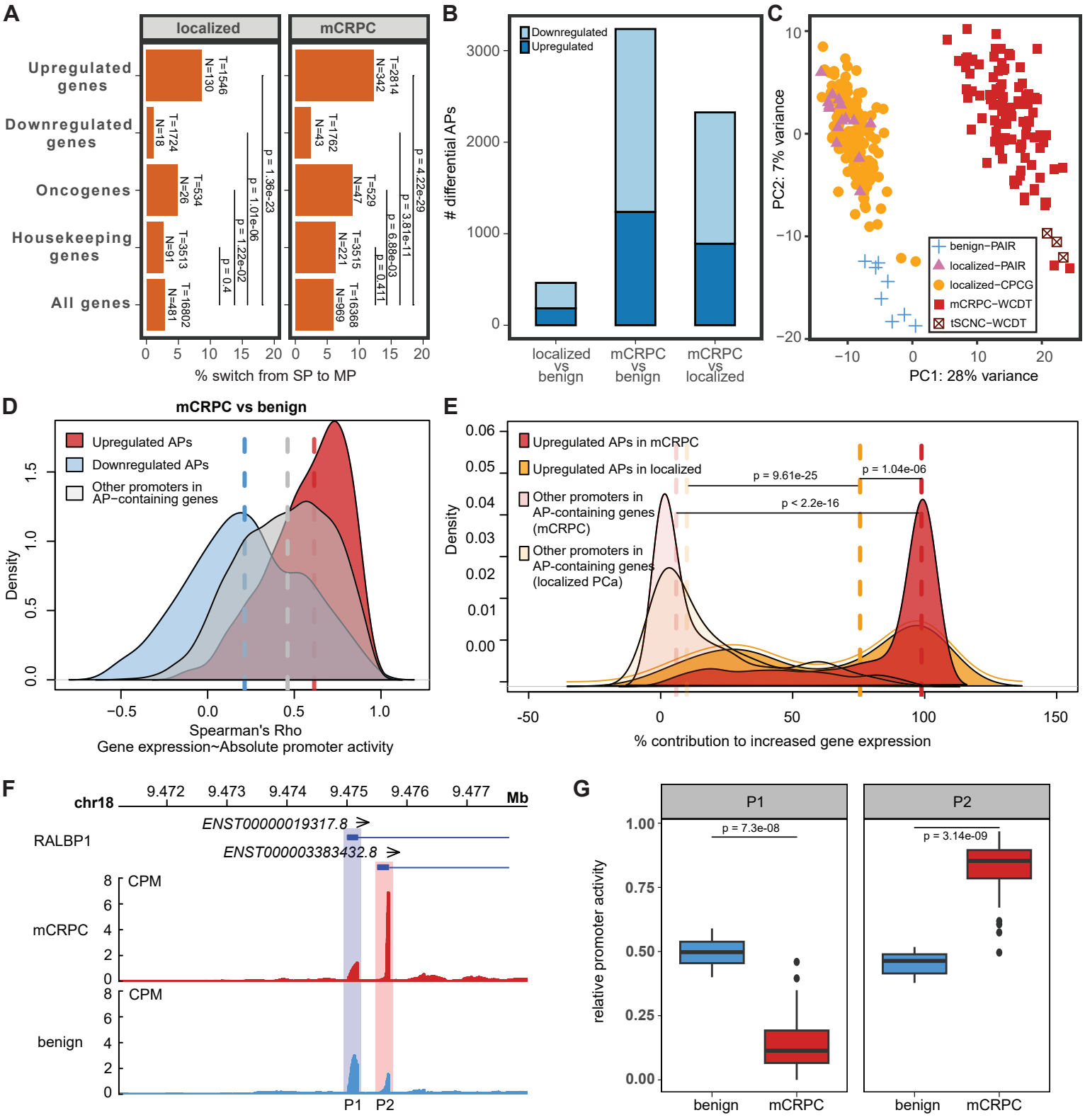


Figure 2. FOXA1 binding and androgen signaling are associated with alternative promoter usage in PCa

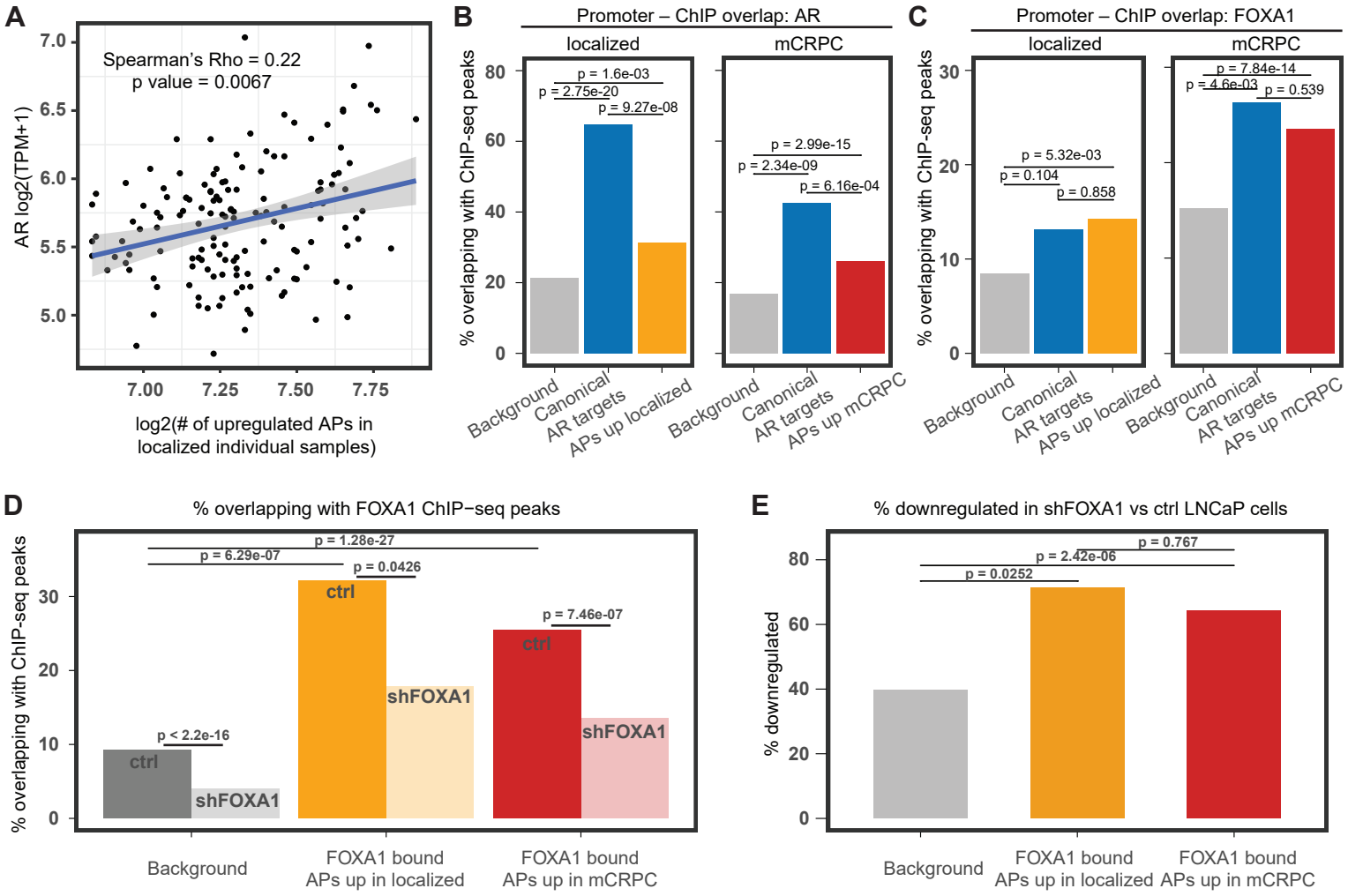


Figure 3. MYC is a potential driver of alternative promoter activation in mCRPC.

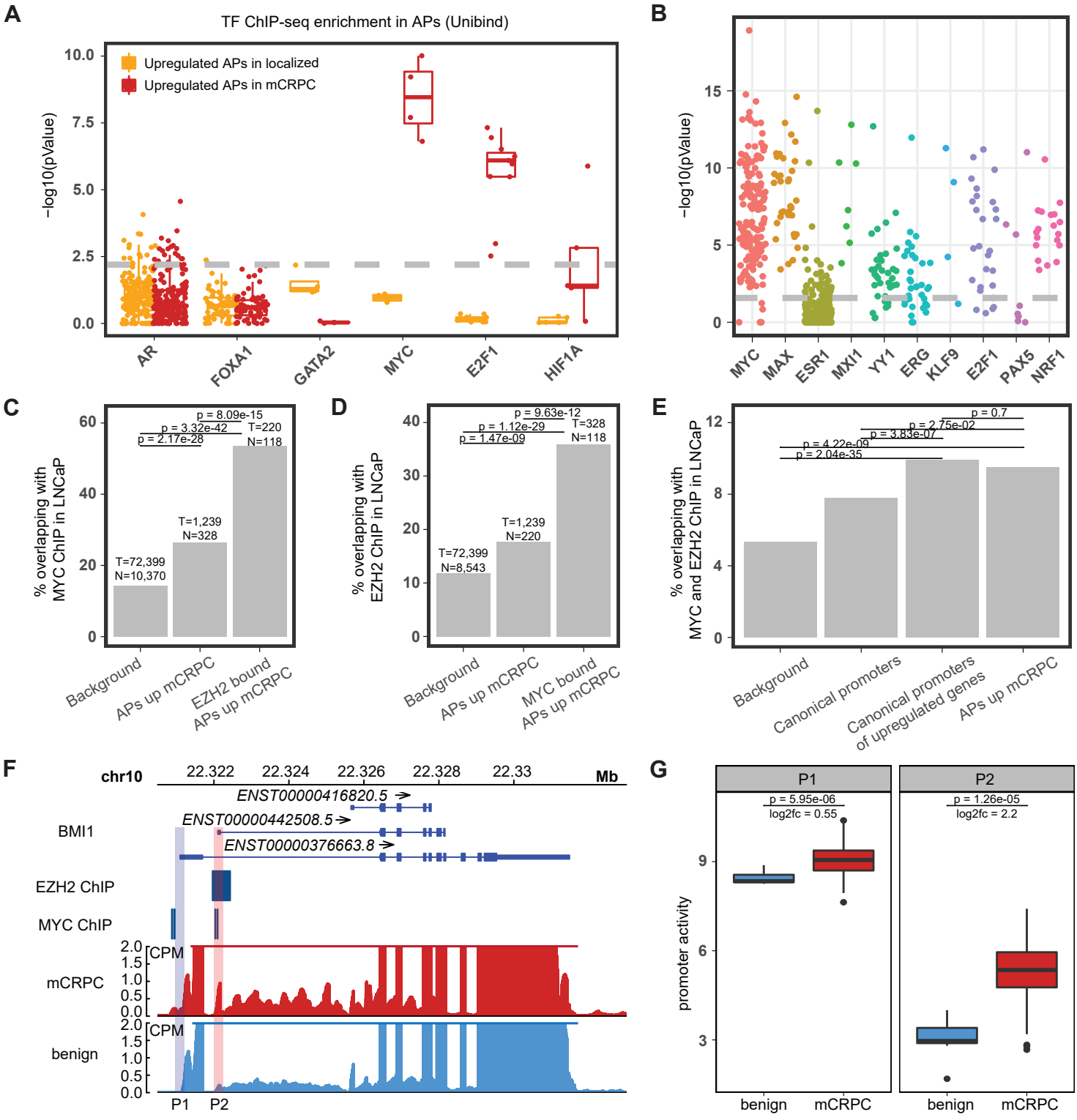


Figure 4. Alternative promoter usage reflects lineage plasticity in response to therapy.

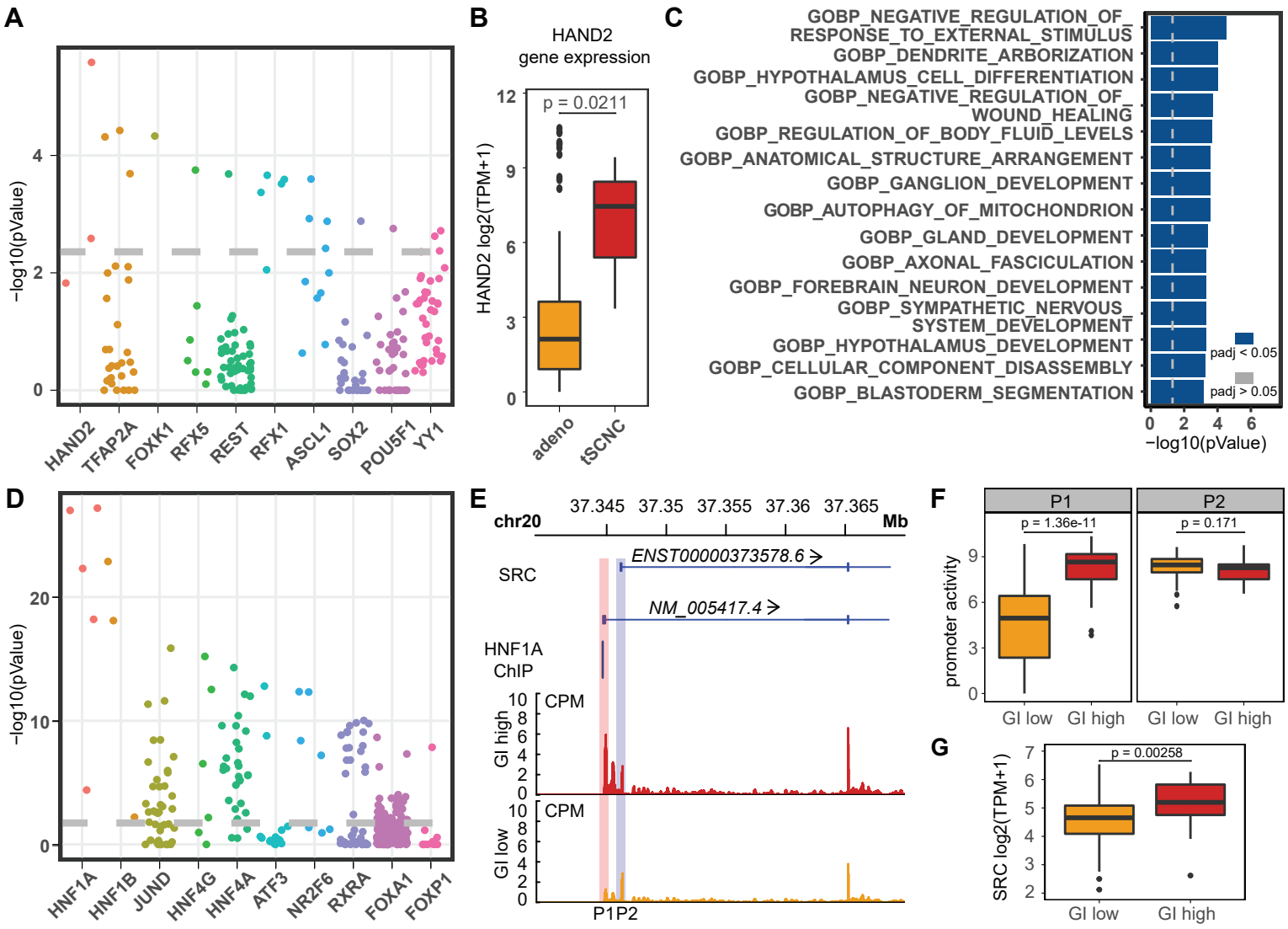
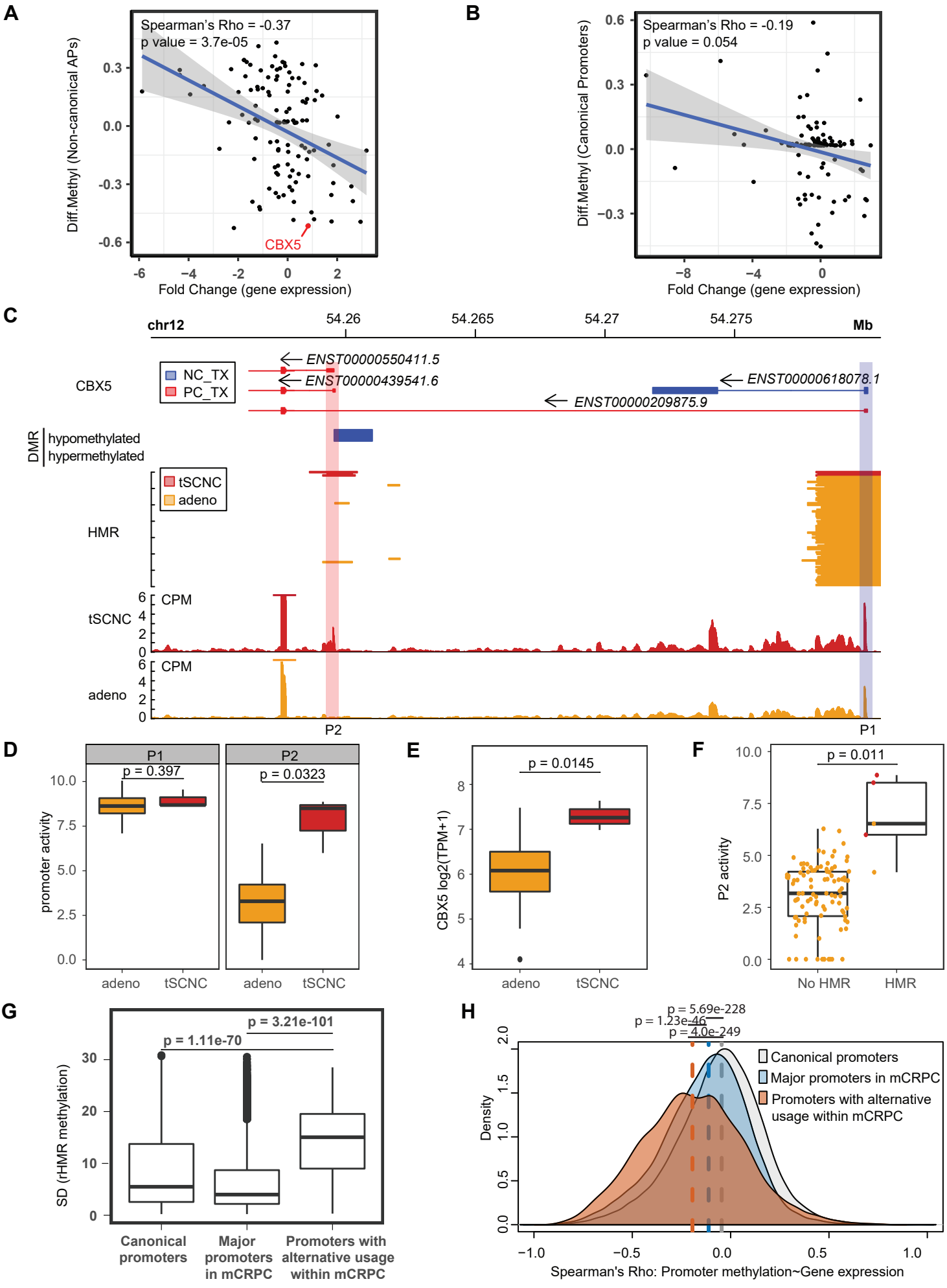
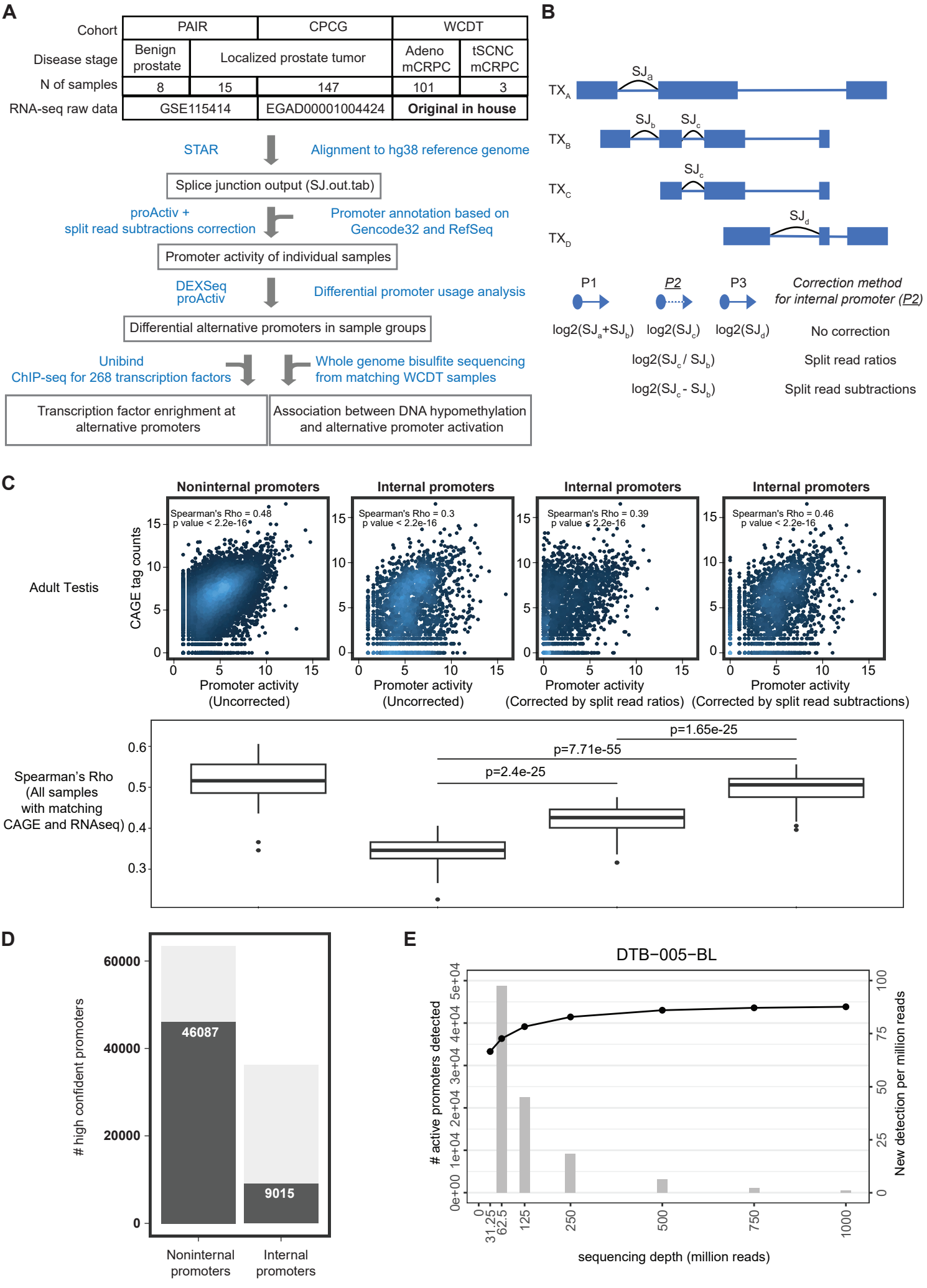


Figure 5. Activation of alternative promoters is associated with DNA hypomethylation.

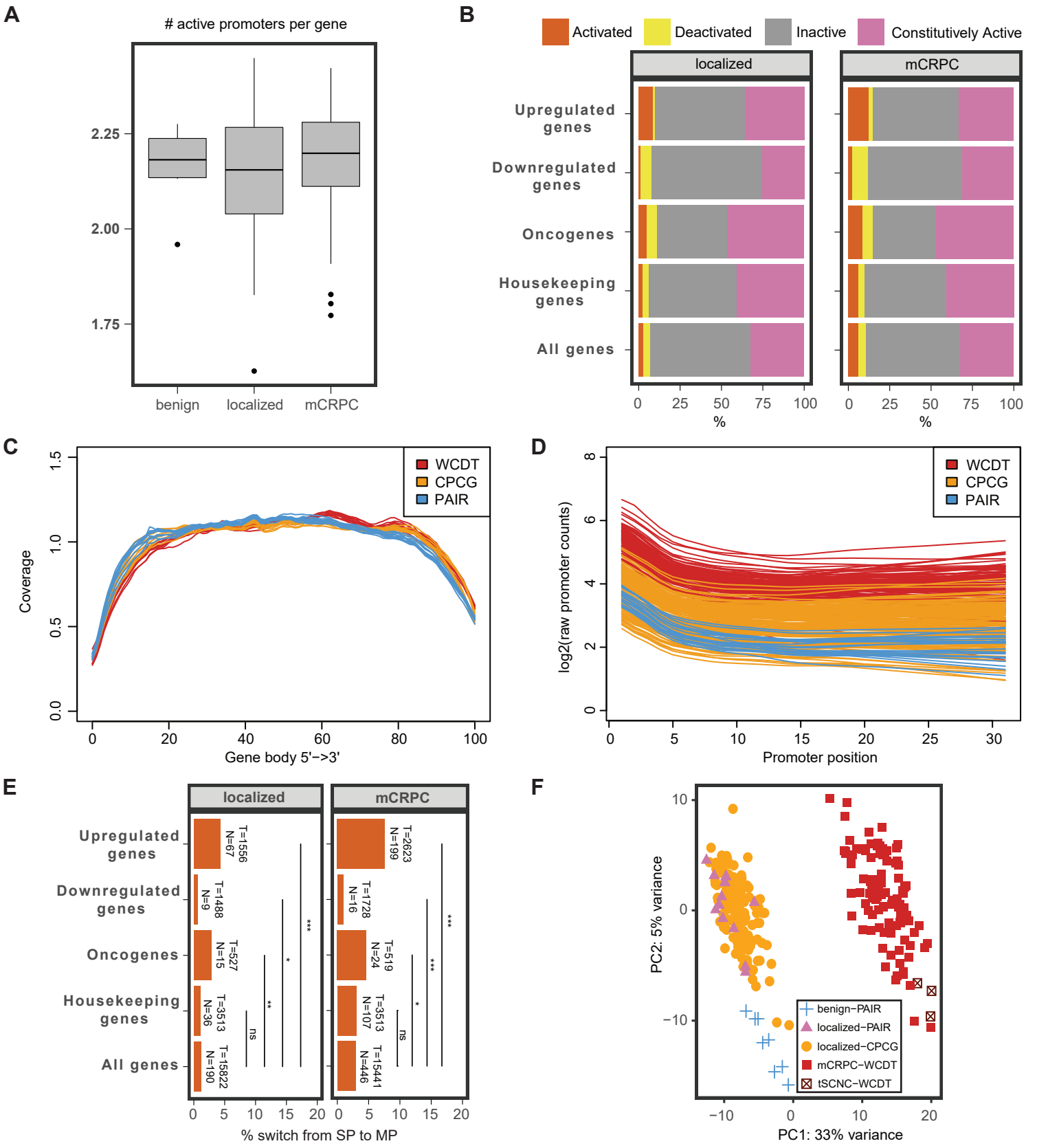


Supplementary Figure 1



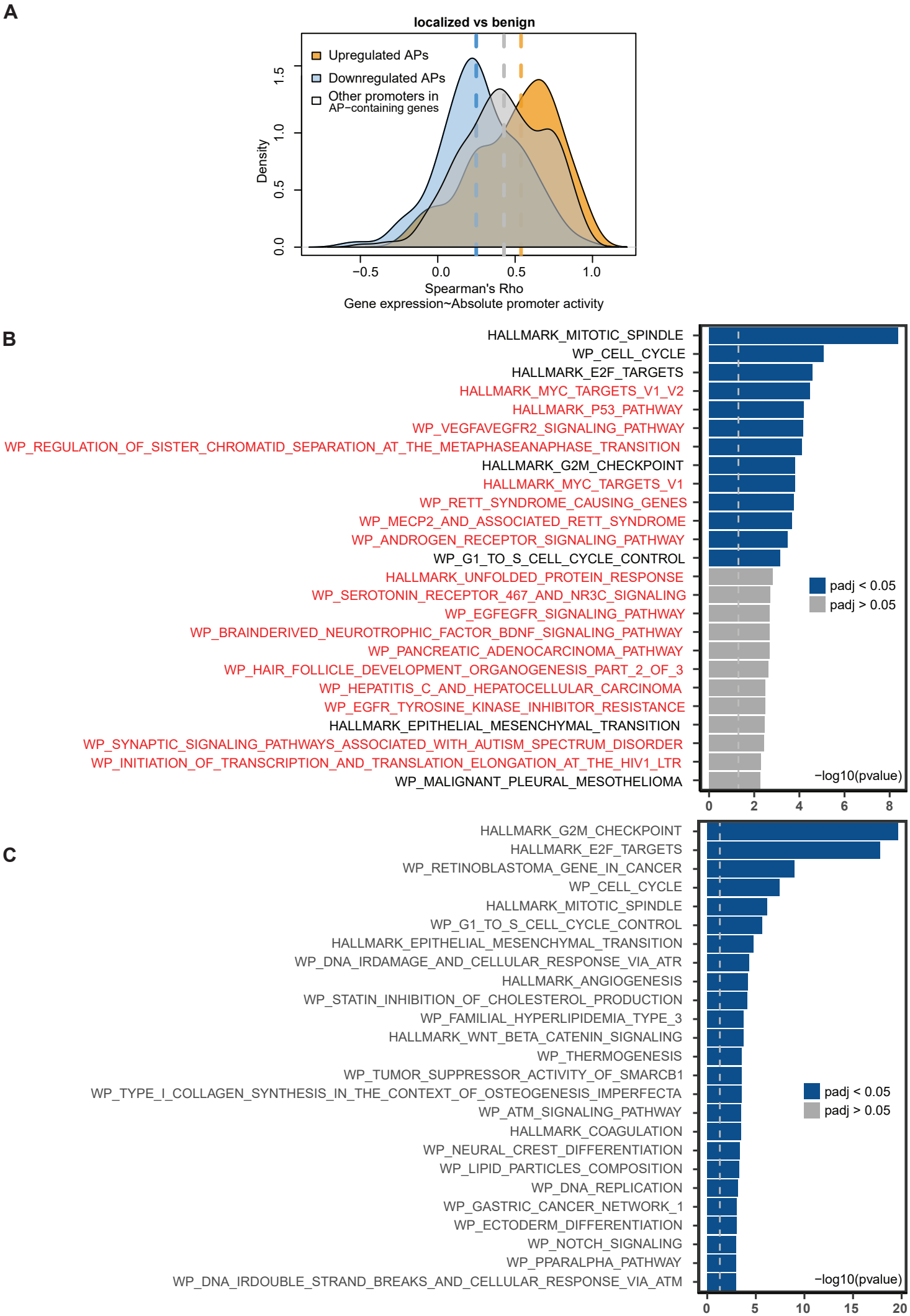
Supplementary Figure 1. Optimization for promoter activity estimation.
A) An overall schematic of the samples, data processing, and principle tools used for the analysis. PAIR: from the Henri Mondor institution, CPCG: Canadian Prostate Cancer Genome Network, WCDT: West Coast Dream Team, t-SCNC: treatment-emergent small cell neuroendocrine carcinoma.
B) An illustration of the promoter activity estimation methods. Solid boxes represent exons while the lines represent introns. The promoter (P1, P2, P3) are defined as the first 5' TSSs (transcription start sites) of overlapping first exons. The splice junction reads (SJ) from the overlapping first exons were summed and log2-normalized to represent the transcriptional activity of the promoters. The activity of the internal promoter P2 driven isoform C (TX_C) can be corrected by the split read ratios or split read subtractions method to exclude transcriptional activity from isoform B (TX_B) (see Methods for details). TX: transcript, SJ: splice jucntion.
C) Correlations between the CAGE (cap analysis of gene expression) tag reads and the promoter activity calculated using RNA-seq data of non-internal promoters without correction, internal promoters without correction, internal promoters corrected by the split read ratios method, and internal promoters corrected by the split read subtractions method. The matching CAGE and RNA-seq data from the same samples were from FANTOM5. Upper row: representative correlation plots showing one human adult testis sample. Lower row: a box plot showing Spearman's correlation coefficients for all 67 samples with matching CAGE and RNA-seq data.
D) Number of high confidence promoters (dark gray, see Methods for details) and non high confidence promoters (light gray) in the non-internal and internal promoters category.
E) A representative sample downsampled to 31.25 million (M), 62.5M, 125M, 250M, 500M, and 750M reads from 1000M. The bars show the number of active promoters detected at each read depth (left y axis). Lines connected by points show the number of new promoters detected per million reads, with values indicated on the right axis.

Supplementary Figure 2



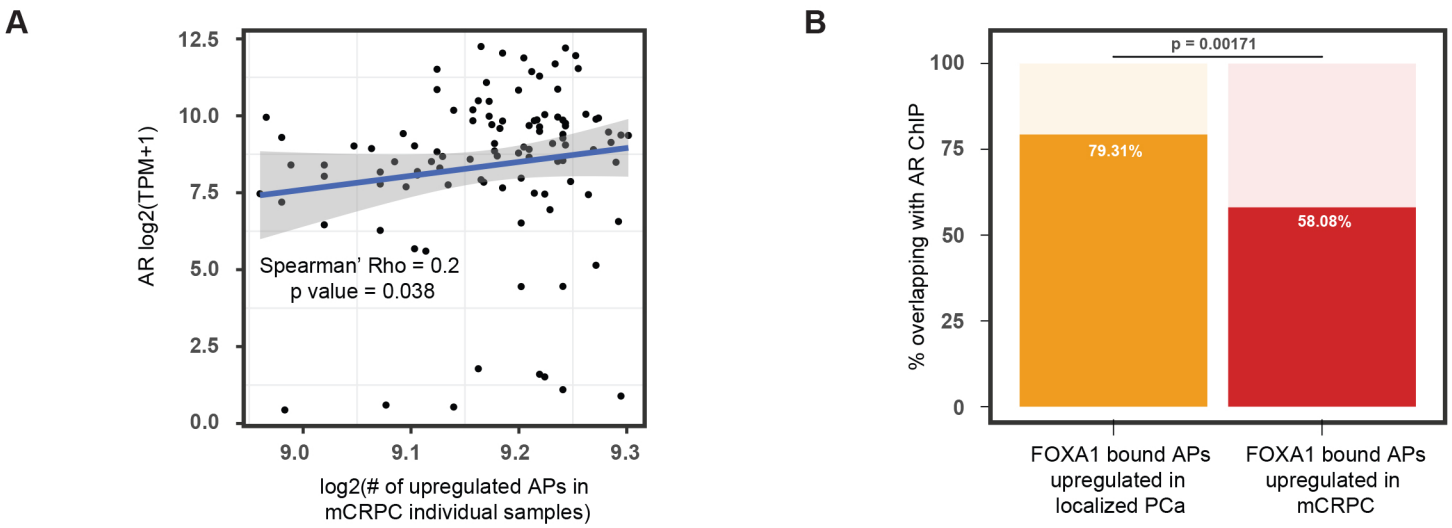
Supplementary Figure 2. Activation of additional promoters is associated with gene expression upregulation.
A) The number of active promoters normalized to the number of expressed genes for each individual sample grouped by disease stages. Genes with nonzero counts were considered as expressed.
B) Upregulated and downregulated genes were identified by differential gene expression analysis. Bar plot shows the percentage of genes in each category that switch between single-promoter active and multiple-promoter active in benign prostate and localized PCa (left) or mCRPC (right). Activated: switch from SP (single-promoter active) in benign to MP (multiple-promoter active) in tumors. Deactivated: switch from MP in benign to SP in tumors. Inactive: SP in both benign and tumors. Constitutively active: MP in both benign and tumors.
C) The RNA-seq coverage across gene body from 5' to 3' for ten random samples from each of the dataset (PAIR, CPCG, and WCDT) in our data collection.
D) The EDASeq bias plot of the positional biases in unnormalized promoter counts of all samples from the RNA-seq datasets (PAIR, CPCG, and WCDT) in our data collection.
E) The analysis of number of genes switching from single promoter active in benign prostate to multiple promoters active in localized (left) or mCRPC (right) using the RNA-seq dataset all down-sampled to 80M reads/sample. SP: single -promoter active, MP: multiple-promoter active. *p value < 0.05, **p value < 0.01, ***p value < 0.005 (Fisher's exact tests).
F. Principle component analysis of all samples of different disease stages from three cohorts using the down-sampled RNA-seq dataset.

Supplementary Figure 3



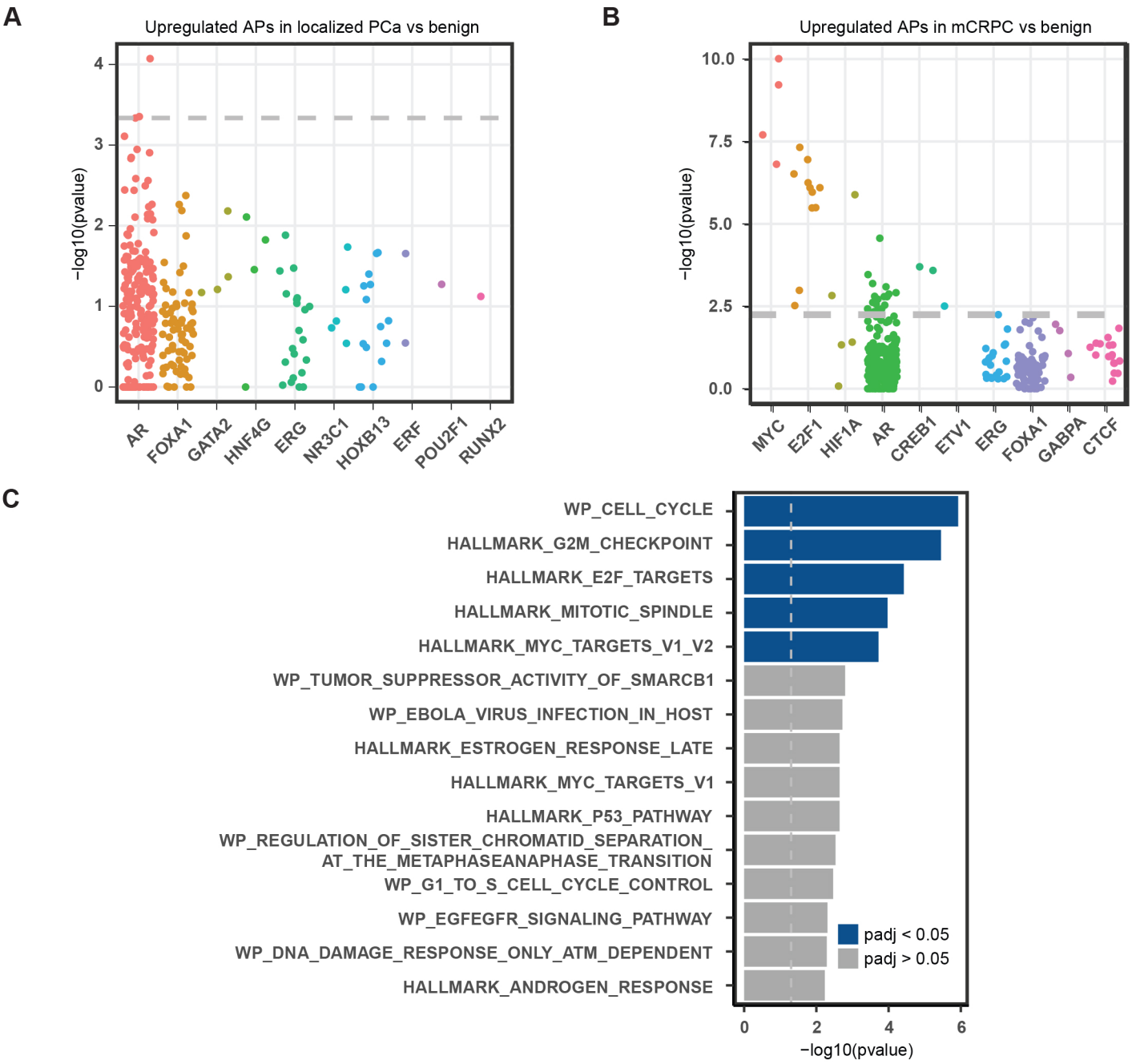
Supplementary Figure 3. Alternative promoter usage occurs in cancer related genes.
A) Density plot of the Spearman's correlation rho values between absolute promoter activity and corresponding gene expression for upregulated APs, downregulated APs and non-differential promoters in genes with differential APs in localized PCa vs benign prostate.
B) Pathway enrichment analysis of genes with upregulated APs in mCRPC vs benign. Highlighted in red are pathways enriched for the genes with upregulated APs but not in upregulated genes. Dashed line shows p value 0.05.
C) Pathway enrichment analysis result of genes upregulated in mCRPC vs benign prostate. Dashed line shows p value 0.05.

Supplementary Figure 4



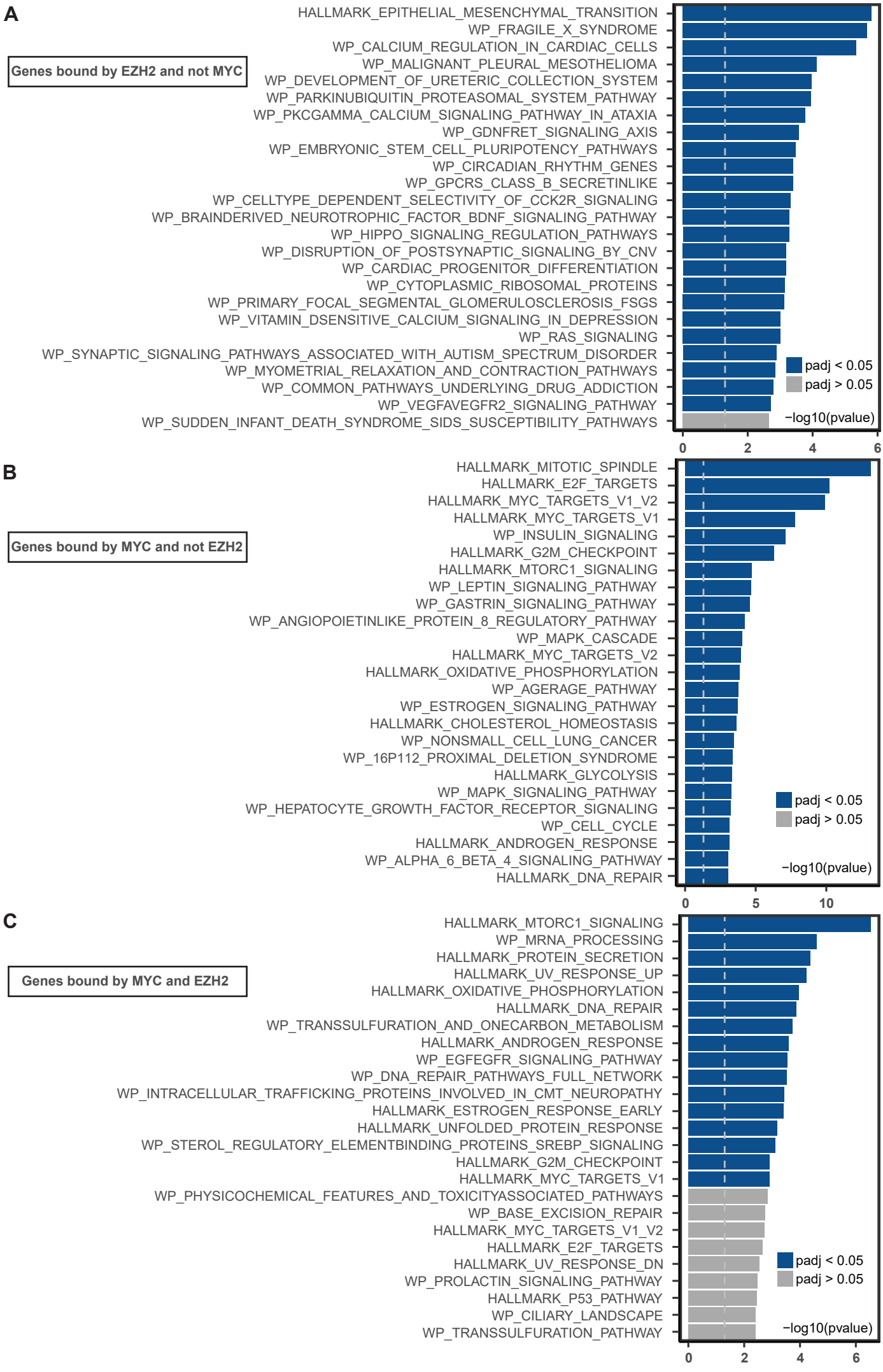
Supplementary Figure 4. Alternative promoter usage is associated with AR levels.
A) Correlation between the number of upregulated APs in individual mCRPC samples with AR expression levels.
B) The percentage of AR and FOXA1 co-binding in the FOXA1 bound upregulated APs in localized PCa and mCRPC (Fisher's exact test).

Supplementary Figure 5



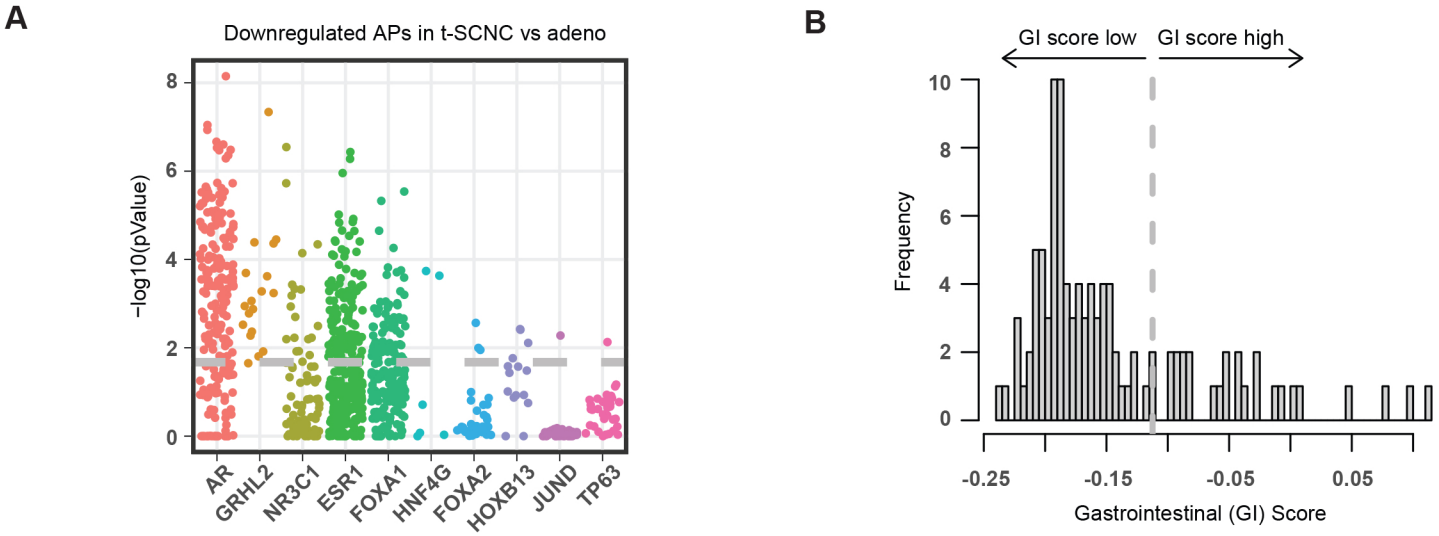
Supplementary Figure 5. Alternative promoter usage is associated with driver transcription factors.
A, B) Unibind results showing significance of overlap between transcription factor (TF) ChIP-seq peaks and upregulated APs in localized PCa (A) or mCRPC (B). Each dot represents one ChIP-seq dataset. TFs were ranked by the most significant ChIP-seq dataset. Dashed line: BH adjusted p value = 0.05.
C) Pathway enrichment analysis of genes with APs upregulated in mCRPC vs benign prostate and overlapping with MYC ChIP-seq peaks in LNCaP cells. Dashed line: p value 0.05.

Supplementary Figure 6



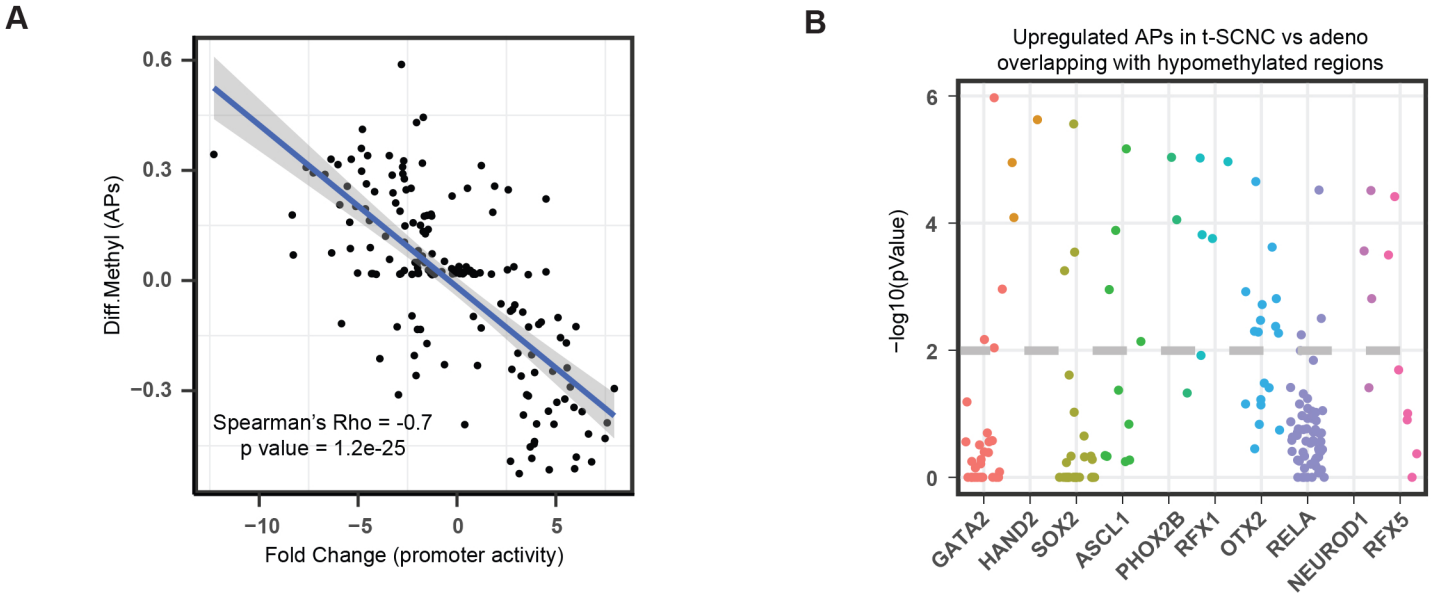
Supplementary Figure 6. Enriched pathways in genes whose promoters are bound by MYC, EZH2 or both.
Pathway enrichment analyses of genes with promoters overlapping with EZH2 LNCaP ChIP-seq peaks only (A), with MYC LNCaP ChIP-seq peaks only (B), and with both MYC and EZH2 LNCaP ChIP-seq peaks (C). Dashed line: p value 0.05.

Supplementary Figure 7



Supplementary Figure 7. Alternative promoter usage reflects lineage plasticity in mCRPC.
A) Unibind results showing significance of overlap between TF ChIP-seq peaks and downregulated APs in t-SCNC vs adenocarcinoma mCRPC. Each dot represents one ChIP-seq dataset. TFs were ranked by the most significant ChIP-seq dataset. Dashed line: BH adjusted p value = 0.05.
B) Histogram showing the distribution of gastrointestinal (GI) scores across mCRPC samples. Dashed line splits the fourth quartile vs others.

Supplementary Figure 8



Supplementary Figure 8. DNA methylation at alternative promoters is anticorrelated with their activity.
A) Correlation between the promoter activity fold change and methylation differences at differentially active APs between mCRPC t-SCNC and adenocarcinoma mCRPC.
B) Unibind results showing significance of overlap between TF ChIP-seq peaks and upregulated APs in mCRPC t-SCNC vs adenocarcinoma that overlapped with differentially hypomethylated regions in t-SCNC. Each dot represents one ChIP-seq dataset. TFs were ranked by the most significant ChIP-seq dataset. Dashed line: BH adjusted p value = 0.05.