



# The DNA methylation landscape of advanced prostate cancer

Shuang G. Zhao<sup>1,2,3,4,5,50</sup>, William S. Chen<sup>3,4,6,50</sup>, Haolong Li<sup>3,4,50</sup>, Adam Foye<sup>4,7</sup>, Meng Zhang<sup>3,4</sup>, Martin Sjöström<sup>6</sup>, Rahul Aggarwal<sup>4,7</sup>, Denise Playdle<sup>4,7</sup>, Arnold Liao<sup>8</sup>, Joshi J. Alumkal<sup>9,10</sup>, Rajdeep Das<sup>10</sup>, Jonathan Chou<sup>3,4,7</sup>, Junjie T. Hua<sup>11</sup>, Travis J. Barnard<sup>3,4</sup>, Adina M. Bailey<sup>4,7</sup>, Eric D. Chow<sup>12,13</sup>, Marc D. Perry<sup>10</sup>, Ha X. Dang<sup>14,15,16</sup>, Rendong Yang<sup>10</sup>, Ruhollah Moussavi-Baygi<sup>3,4</sup>, Li Zhang<sup>4,7</sup>, Mohammed Alshalalfa<sup>3</sup>, S. Laura Chang<sup>3</sup>, Kathleen E. Houlahan<sup>10</sup>, Yu-Jia Shiah<sup>18</sup>, Tomasz M. Beer<sup>9,21</sup>, George Thomas<sup>9,22</sup>, Kim N. Chi<sup>23,24</sup>, Martin Gleave<sup>23</sup>, Amina Zoubeidi<sup>23</sup>, Robert E. Reiter<sup>25</sup>, Matthew B. Rettig<sup>25,26</sup>, Owen Witte<sup>27</sup>, M. Yvonne Kim<sup>28</sup>, Lawrence Fong<sup>4,7</sup>, Daniel E. Spratt<sup>1</sup>, Todd M. Morgan<sup>10</sup>, Rohit Bose<sup>10</sup>, Franklin W. Huang<sup>4,7</sup>, Hui Li<sup>3,4</sup>, Lisa Chesner<sup>3,4</sup>, Tanushree Shenoy<sup>4,7</sup>, Hani Goodarzi<sup>12,30</sup>, Irfan A. Asangani<sup>32</sup>, Shahneen Sandhu<sup>33</sup>, Joshua M. Lang<sup>34</sup>, Nupam P. Mahajan<sup>16,35</sup>, Primo N. Lara<sup>36,37</sup>, Christopher P. Evans<sup>37,38</sup>, Phillip Febbo<sup>8</sup>, Serafim Batzoglou<sup>8</sup>, Karen E. Knudsen<sup>39</sup>, Housheng H. He<sup>10</sup>, Jiaoti Huang<sup>40</sup>, Wilbert Zwart<sup>41</sup>, Joseph F. Costello<sup>28</sup>, Jianhua Luo<sup>42</sup>, Scott A. Tomlins<sup>5</sup>, Alexander W. Wyatt<sup>23</sup>, Scott M. Dehm<sup>10</sup>, Alan Ashworth<sup>4,7</sup>, Luke A. Gilbert<sup>10</sup>, Paul C. Boutros<sup>10</sup>, Kyle Farh<sup>10</sup>, Arul M. Chinnaiyan<sup>10</sup>, Christopher A. Maher<sup>10</sup>, Eric J. Small<sup>10</sup>, David A. Quigley<sup>10</sup>, and Felix Y. Feng<sup>10</sup>✉

**Although DNA methylation is a key regulator of gene expression, the comprehensive methylation landscape of metastatic cancer has never been defined. Through whole-genome bisulfite sequencing paired with deep whole-genome and transcriptome sequencing of 100 castration-resistant prostate metastases, we discovered alterations affecting driver genes that were detectable only with integrated whole-genome approaches. Notably, we observed that 22% of tumors exhibited a novel epigenetic subtype associated with hypermethylation and somatic mutations in *TET2*, *DNMT3B*, *IDH1* and *BRAF*. We also identified intergenic regions where methylation is associated with RNA expression of the oncogenic driver genes *AR*, *MYC* and *ERG*. Finally, we showed that differential methylation during progression preferentially occurs at somatic mutational hotspots and putative regulatory regions. This study is a large integrated study of whole-genome, whole-methylome and whole-transcriptome sequencing in metastatic cancer that provides a comprehensive overview of the important regulatory role of methylation in metastatic castration-resistant prostate cancer.**

DNA methylation of cytosine residues is a pervasive epigenomic mechanism of gene regulation<sup>1,2</sup>. DNA methyltransferases add a methyl group to the 5' carbon of cytosine nucleotides adjacent to guanines (CpG dinucleotides), creating 5mC nucleotides<sup>3</sup>. Most CpG dinucleotides are methylated, with the exception of hypomethylated regions (HMRs) enriched for CpGs termed islands, shores ( $\pm 2\text{ kb}$  around islands) and shelves ( $\pm 2\text{ kb}$  around shores<sup>4</sup>). These regions frequently mark gene regulatory loci such as promoters or enhancers<sup>5</sup>. Aberrant methylation has been implicated in oncogenesis, and differences in methylation patterns between tumors and benign tissues have been reported in many tumor types<sup>6</sup>. Cancer cells are frequently less methylated at CpGs than are normal cells, although hypermethylation at tumor CpG islands has also been reported<sup>1,2</sup>.

Several studies have compared DNA methylation patterns between primary prostate cancer (PCa) and benign prostate

tissue, and between subtypes of primary PCa<sup>5,7–15</sup>. Metastatic castration-resistant prostate cancer (mCRPC) is the lethal form of the disease. Although the genomic and transcriptomic landscape of mCRPC has been well characterized<sup>16–19</sup>, the complete epigenetic landscape remains largely unknown. Prior studies of mCRPC assessed a small percentage of the genome, primarily focused on promoter regions<sup>20,21</sup>. Many important regulatory regions are outside the profiled areas, and whole-genome bisulfite sequencing (WGBS) is required to systematically study the entire genome at single-base-level resolution. At the time of this analysis, WGBS has only been applied to a few relatively small cancer cohorts<sup>5,11,22–31</sup>. Moreover, WGBS has rarely been integrated with other genome-wide sequencing approaches such as whole-genome sequencing (WGS) and whole-transcriptome RNA-seq<sup>23,28,30</sup>. Herein, we describe a WGBS study in a metastatic cancer integrated with matched deep WGS and RNA-seq in the same samples.

A full list of affiliations appears at the end of the paper.

## Results

A prospective multi-institution Institutional Review Board-approved study (NCT02432001) obtained fresh-frozen core biopsies of metastases from 100 patients with mCRPC as described previously<sup>17</sup>. WGBS was performed on 100 biopsy samples and on 10 matched benign tissue samples, and a mean aligned sequencing depth of 46 $\times$  and 33 $\times$ , respectively, was obtained (Supplementary Table 1 and Supplementary Fig. 1a). Bone, lymph node and liver biopsies were represented in these benign-adjacent samples, which exhibited distinct methylation patterns from the tumor samples (Supplementary Fig. 1b). We integrated the methylation data with WGS (average tumor coverage 109 $\times$ , benign-adjacent coverage 38 $\times$ ) and whole-transcriptome RNA-seq (average 114 M reads per sample) performed on these same tumors<sup>17</sup>. The median tumor purity by histological assessment was 70%. Coverage of 10 $\times$  was achieved in 96–99% of mappable CpGs across our samples (Supplementary Table 1, excluding the Y chromosome, which is frequently lost in mCRPC), and coverage of 10 $\times$  in 95% of samples was achieved in 87% of mappable CpGs. Sample identity and tumor content were confirmed by the observed high concordance between copy number estimates derived from WGS and WGBS sequencing depth (Supplementary Fig. 1c). Analysis also incorporated previously published WGBS of primary PCa and benign prostate samples<sup>32</sup>, chromatin immunoprecipitation sequencing (ChIP-seq) performed on metastatic and primary PCa samples<sup>33–38</sup>, and chromatin interaction analysis paired-end tag sequencing (ChIA-PET) performed on the VCaP cell line<sup>39</sup>.

**Novel CpG methylation subtype of metastatic castration-resistant prostate cancer.** The total number of HMRs ranged from 24,388 to 85,474 per sample (Fig. 1a and Supplementary Table 1). HMR methylation levels were a median of 43% lower than the same locus in samples lacking an HMR. Most inter-sample variation was outside promoters and CpG islands, shores or shelves, and manifested in gene bodies and regulatory regions, such as transcription-factor-binding sites (TFBS; for example, those for AR, ERG, FOXA1 and HOXB13), enhancer sites (marked by H3K27ac ChIP-seq peaks) and repressed regions (marked by an H3K27me3 ChIP-seq signal) (Fig. 1a). Tumors with more HMRs had significantly higher genome copy number alteration frequencies (Spearman's  $\rho=0.42$  (0.23–0.59),  $P=1.5 \times 10^{-5}$ ), as observed previously<sup>40</sup>. HMR frequency was not associated with mutation or structural variant frequency (Fig. 1a).

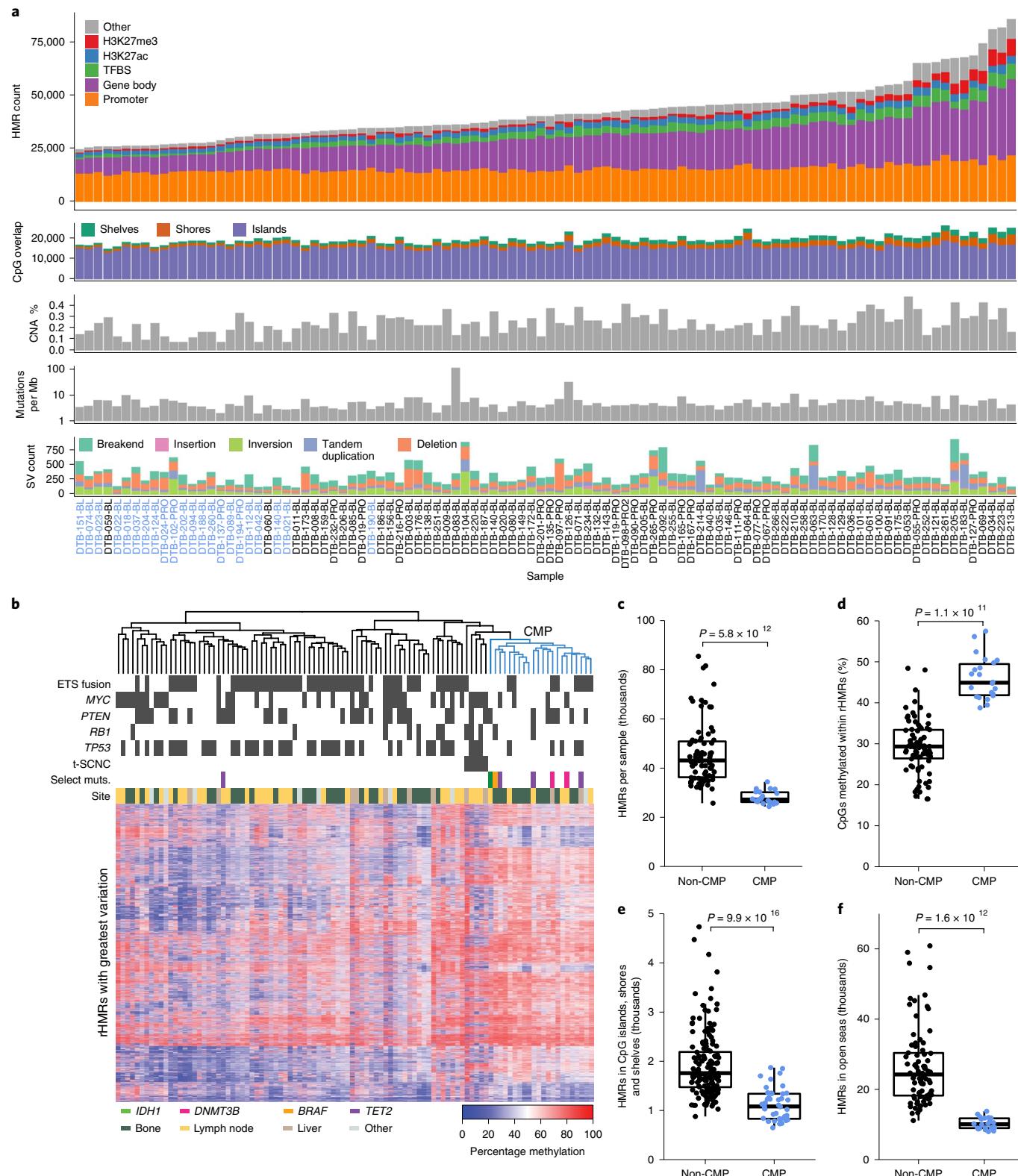
DNA methylation has been best characterized at the CpG islands that are present in the promoter regions of genes<sup>41–43</sup>. However, 74% of the 97,747 recurrent HMRs (rHMRs) (present in  $\geq 5\%$  of samples) were outside CpG islands, shores or shelves (Supplementary Table 2). We reasoned that recurrent intergenic HMRs would be associated with regulatory loci. Indeed, 88% of recurrent HMR sites overlapped putative regulatory regions (Fig. 1a). Unsupervised hierarchical clustering of rHMRs identified subgroups of tumors with distinct patterns of methylation (Fig. 1b). One cluster consisted of tumors that were previously identified as treatment-emergent small-cell neuroendocrine cancer<sup>44</sup> (t-SCNC), which is characterized by decreased AR signaling, elevated expression of neuroendocrine markers<sup>20,44,45</sup> and a distinct methylation profile<sup>20</sup>. We also identified a novel subtype of mCRPC (Fig. 1b) with significantly higher methylation levels at rHMRs than at all other clusters ( $P$  values  $<0.05$ , Wilcoxon signed-rank test; Supplementary Fig. 2a,b) and fewer HMRs (Fig. 1c,d). These tumors harbored fewer HMRs both at CpG islands, shores and shelves ( $P=9.9 \times 10^{-16}$ , Wilcoxon signed-rank test; Fig. 1e) and in CpG open seas (that is, the regions outside CpG islands, shores and shelves)<sup>4</sup> ( $P=1.6 \times 10^{-12}$ , Wilcoxon signed-rank test; Fig. 1f), and were designated a CpG methylator phenotype (CMP). Bootstrap resampling analysis of the cluster composition indicated that it was stable (Jaccard index 0.81)<sup>46</sup>. CMP tumors less frequently harbored ETS family-associated gene

fusions ( $P=0.03$ , OR = 0.31 (0.10–0.90), Fisher's exact test) or *TP53* bi-allelic inactivation ( $P=0.02$ , OR = 0.26 (0.07–0.81), Fisher's exact test) (Fig. 1b). The CMP subtype was not significantly associated with the anatomic site of the biopsy. A *t*-distributed stochastic neighbor embedding plot incorporating all recurrently hypomethylated sites, benign prostate and primary prostate tumor samples demonstrated that CMP tumors, benign prostate tumors and t-SCNC tumors formed separate clusters (Supplementary Fig. 2c).

Several CMP tumors harbored mutually exclusive mutations in *TET2*, *IDH1* and *BRAF* (Fig. 1b and Supplementary Table 3). Mutations in these genes have been associated with increased CpG methylation in other tumor types<sup>32,47,48</sup>. Two additional CMP tumors harbored somatic mutations in the DNA methyltransferase gene *DNMT3B* (Supplementary Fig. 3a). CMP tumors were enriched for mutations in *TET2*, *IDH1*, *BRAF* and *DNMT3B* compared to non-CMP tumors ( $P=8 \times 10^{-5}$ , OR = 34.1 (3.4–1622.9), Fisher's exact test; Supplementary Table 3). To assess the potential for misattribution of somatic mutations to mutations introduced through clonal hematopoiesis<sup>49</sup>, we confirmed the absence of these mutations in peripheral blood germline DNA using both WGS and Sanger sequencing. *TET2* mutations are frequent in hematological malignancies, with missense mutations frequently clustered in the catalytic DSBH domain of *TET2* near the metal-binding sites at residues 1382 and 1884 (refs. <sup>50,51</sup>). Three of the four *TET2* mutations we observed (encoding p.His1380Leu, p.Tyr1421His and p.Arg1808Thr alterations) occurred in or near these hotspot regions (Supplementary Fig. 3b). The fourth mutation, encoding p.Thr1499Arg, occurred in the single *TET2*-mutated sample that did not cluster in the CMP subtype. Computational prediction of mutation consequences by FATHMM<sup>52</sup> predicted p.His1380Leu, p.Tyr1421His and p.Arg1808Thr to be deleterious and p.Thr1499Arg to be benign (Supplementary Fig. 3b). The *TET2* mutation that encodes p.His1380Leu has previously been reported in hematopoietic and lymphoid malignancies (COSMIC identifier COSM4170052)<sup>53,54</sup>.

In line with prior observations in tumors that harbor hypermethylation phenotypes<sup>32</sup>, not all CMP tumors harbored a somatic alteration in a gene that is known to affect methylation biology. No somatic mutations were observed in any *DNMT* or *TET* family genes other than *DNMT3B* and *TET2*. A ranked list of somatic associations with CMP is noted in Supplementary Table 4. Tumor purity was not associated with distinct methylation patterns within the CMP or non-CMP group (Supplementary Fig. 4). CMP status was independently associated with HMR number in CpG islands, shores and shelves, and CpG open seas after adjusting for tumor purity ( $P=0.008$  and  $P=2.19 \times 10^{-11}$ , respectively, linear model).

**Regional analysis of methylation.** Long-range epigenetic activation and repression is a phenomenon in which large regions containing multiple genes are concomitantly activated or repressed in PCa due to concordant epigenetic changes, such as histone modification or DNA methylation<sup>55,56</sup>. We identified 14 candidate long-range interactions (Supplementary Table 5), two of which (7p15.2 and 16q13) overlapped with previously identified long-range epigenetically silenced domains<sup>55</sup>. Partially methylated domains (PMDs) are genomic regions with incomplete loss of methylation<sup>57</sup>. There was modest correlation between PMD frequency and HMR frequency (Spearman's  $\rho=0.24$  (0.04–0.42),  $P=0.02$ ). Whereas the fraction of the genome harboring PMDs (21% to 61%) was not significantly different between benign prostate, primary PCa and mCRPC samples (Supplementary Fig. 5a), methylation levels within PMDs were lower in primary PCa and mCRPC in comparison to benign prostate tissue (Supplementary Fig. 5a). The genome PMD fraction was not significantly correlated with tumor purity ( $P=0.68$ ), total number of mutations ( $P=0.30$ ) or percentage copy number altered in mCRPC ( $P=0.26$ ). Regions of the genome harboring PMDs had



**Fig. 1 | CpG methylator phenotype.** **a**, Sample-level summary of HMR frequency and somatic alterations in 100 independent mCRPC samples. Bar plots show HMR counts within genomic features (HMR count), counts of HMRs overlapping with CpG islands, shores and shelves (CpG overlap), percentage of the genome with DNA copy number alterations (CNA %), somatic mutations per megabase (mutations per Mb) and counts of structural variants (SV count). CMP samples are labeled in blue. **b**, Hierarchical clustering of the top 10% of the most variable rHMRs in 100 mCRPC samples. Blue dendrogram denotes CMP samples. **c**, HMR count (in thousands) per sample in non-CMP ( $n=78$ ) and CMP ( $n=22$ ) samples. **d**, Percentage of CpGs methylated at loci harboring rHMRs in non-CMP ( $n=78$ ) and CMP ( $n=22$ ) samples, plotted and assessed as in **c**. **e**, rHMRs located in CpG islands, shores and shelves (count per sample in thousands), plotted and assessed as in **c**. **f**, rHMRs located in open seas (count per sample in thousands), plotted and assessed as in **c**. Box plots show the median, first and third quartiles, and outliers are shown if outside the  $1.5 \times$  interquartile range. Significance was assessed with a two-sided Wilcoxon signed-rank test.

an increased mutation burden and were less likely to include exons of genes compared to regions of the genome not covered by a PMD (Supplementary Fig. 5b,c), as was previously observed in breast cancer<sup>58</sup>. Whereas the fraction of the genome covered by PMDs was not associated with CMP status, the level of PMD methylation was significantly higher in the CMP subtype than in the non-CMP samples ( $P=0.03$ , Wilcoxon signed-rank test; Supplementary Fig. 5d).

We next identified DNA methylation valleys (DMVs), broad regions of hypo-methylation<sup>59,60</sup> that are associated with either the activating histone mark H3K4me3 or the repressive histone mark H3K27me3 (ref. <sup>60</sup>). The number of DMVs in mCRPC samples varied from a few hundred to over 20,000 (Fig. 2a). H3K27me3-associated DMVs tend to be dynamically methylated, and the polycomb complex has been shown to play a key role in maintaining the repressive and self-interacting state of DMVs<sup>61</sup>. DMVs in tumors with low DMV frequencies were more frequently associated with H3K4me3, but tumors with many DMVs coincided with a nearly equal proportion of H3K4me3 and H3K27me3 marks.

Up to 20% of patients with mCRPC develop treatment-induced small cell neuroendocrine carcinoma (t-SCNC)<sup>20,44,45,62,63</sup>. Tumors of the type t-SCNC harbored distinct genome-wide methylation patterns (Fig. 1b), as previously reported by a study that used enhanced reduced-representation bisulfite sequencing<sup>20</sup>. Genome-wide assessment of differential methylation demonstrated that the *AR* locus was the most differentially hypomethylated locus in t-SCNC (Fig. 2b and Supplementary Fig. 6). Methylation levels in this region predicted t-SCNC status independently from copy number ( $P=0.01$ , logistic regression). These data are compatible with a model in which epigenetic alterations drive t-SCNC<sup>64</sup>, and suggest a role for methylation at the *AR* locus in this phenotype.

**Differential prostate cancer gene promoter methylation.** Genes with higher expression had more frequent promoter hypomethylation and gene-body hypermethylation (Supplementary Fig. 7a), as previously observed<sup>23,65–67</sup>. Negative correlation of CpG methylation and gene expression peaked at the gene promoter, and positive correlation peaked in the gene body (Supplementary Fig. 7b,c), which is also consistent with previous observations<sup>24</sup>. We identified rHMRs correlated with expression of genes within 10 kb and termed these HMRs ‘expression-associated hypomethylated regions’ (eHMRs). Negatively correlated eHMRs (70% of the total) were predominantly located at the transcription start site (TSS) (Supplementary Fig. 7d). The strongest positive correlations (30% of the total) fell at the 3' end of the gene body (Supplementary Fig. 7e), which is consistent with prior studies<sup>24,68</sup>. We expanded our analysis to test for associations in candidate enhancer regions and HMRs identified in a 1-Mb window around the TSS. Candidate enhancers were identified by the presence of H3K27ac peaks in primary prostate tumors. At a 5% false discovery rate (FDR), 10,412 genes harbored at least one significant association with a candidate enhancer region, and 11,928 genes harbored at least one significant association with an HMR. When both locus types were combined, 71,163 associations were significant overall (reported in Supplementary Table 6). The association between methylation levels and expression tended to be stronger in regions that were physically close to the TSS (Supplementary Fig. 8).

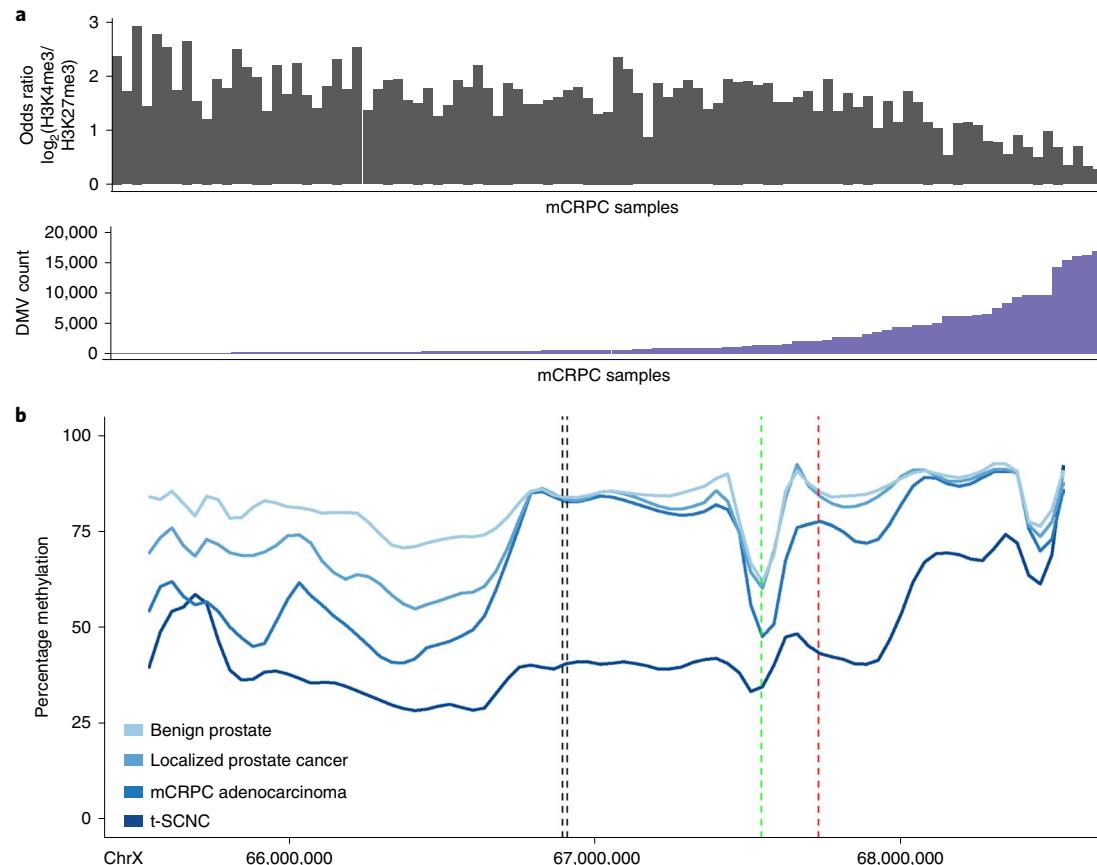
We found that key androgen-response genes demonstrated promoter hypomethylation in mCRPC compared to benign prostate samples, including *AR*, *KLK3* (which encodes prostate-specific antigen), *NKX3-1*, *FOLH1* (which encodes prostate-specific membrane antigen), *SCHLAP1* and *PIK3CA* (Supplementary Fig. 9). We did not observe promoter hypermethylation of tumor suppressors such as *TP53* or *RB1* in mCRPC tumors compared to benign prostate samples. However, numerous genes that were previously reported to be hypermethylated in PCa (for example, *GSTP1*)<sup>69</sup> were differentially methylated in mCRPC compared to benign prostate (Supplementary Fig. 9).

Many genes with PCa-specific expression lack PCa-specific DNA-sequence alterations. To test the model that methylation influences disease-specific expression of PCa-specific genes, we performed an unbiased analysis that compared eHMR correlation strength in all genes to their expression variability. PCa-specific genes had stronger associations with methylation than did other genes (Fig. 3a), even after adjusting for gene size, average expression and variation in expression ( $P<2 \times 10^{-16}$ , Wilcoxon signed-rank test). Many genes whose expression was most strongly independently linked to methylation were associated with PCa, or exclusively expressed in PCa, including *TMEFF2* (ref. <sup>70</sup>) ( $P=4.1 \times 10^{-13}$ ,  $F=28.2$ , d.f.=3, analysis of variance (ANOVA)), *SPON2* (ref. <sup>71</sup>) ( $P=6.6 \times 10^{-19}$ ,  $F=25.4$ , d.f.=7, ANOVA), *TDRD1* (ref. <sup>72</sup>) ( $P=3.3 \times 10^{-29}$ ,  $F=78.2$ , d.f.=4, ANOVA), *SLC45A3* (ref. <sup>73</sup>) ( $P=9.2 \times 10^{-23}$ ,  $F=51.0$ , d.f.=4, ANOVA), and the long non-coding RNAs *SCHLAP1* (ref. <sup>74</sup>) ( $P=1.4 \times 10^{-22}$ ,  $F=88.3$ , d.f.=2, ANOVA) and *PCAT14* (ref. <sup>75</sup>) ( $P=7.4 \times 10^{-20}$ ,  $F=132.7$ , d.f.=1, ANOVA) (Fig. 3b).

**Novel intergenic regulatory regions of the AR gene.** DNA methylation may operate in tandem with other somatic DNA alterations that influence gene expression. Gene expression was significantly associated with local DNA copy number alterations, mutations or structural variants in 15,014 of 51,708 genes (29%), and with local methylation in 10,118 genes (19.5%). Of the 10,118 genes in which expression was associated with methylation, 4,735 had associations with both methylation and DNA alterations, and 5,383 genes were only associated with methylation. Methylation improved the fit of a model for gene expression beyond DNA alterations alone for 16.4% of all genes and 26.3% of housekeeping genes<sup>76</sup> (FDR  $\leq 0.05$ , ANOVA). The top enriched Molecular Signatures Database (MSigDb) hallmark pathway<sup>77,78</sup> for genes with improved fit was ‘androgen response’; methylation significantly improved the model fit in 73.7% of transcripts in the pathway (FDR = 0.0002 versus housekeeping genes<sup>76</sup>, OR = 2.06 (1.49–2.85), Fisher’s exact test; Fig. 4a). Key *AR*-associated genes that showed correlation with methylation independent of DNA alterations included *KLK3* ( $P=4.0 \times 10^{-15}$ ,  $F=86.8$ , d.f.=1, ANOVA), *NKX3-1* ( $P=2.4 \times 10^{-8}$ ,  $F=36.9$ , d.f.=1, ANOVA) and *FOLH1* ( $P=7.7 \times 10^{-16}$ ,  $F=36.5$ , d.f.=3, ANOVA; Fig. 4b). This finding supports the role of methylation in androgen pathway activity in mCRPC.

We and others have previously identified a distal *AR* enhancer region in which DNA copy number amplifications are associated with elevated *AR* expression<sup>17,18,33</sup>. We identified multiple eHMRs near *AR*, including adjacent to the *AR* promoter, the previously identified *AR* enhancer and additional loci upstream and downstream of *AR* (Fig. 4c). Whereas the *AR* promoter was hypomethylated in all tissues evaluated, other eHMRs were identified only in mCRPC samples and not in benign-adjacent tissue, benign prostate or primary PCa samples. Five of the seven eHMRs colocalized with H3K27ac (a mark of enhancer activity), and/or with the binding sites for HOXB13, FOXA1, AR, or ERG. Furthermore, AR and ERG ChIA-PET data indicated that long-range chromatin interactions exist between many of these loci, which supports the potential for physical interactions between these loci (Fig. 4c). In a linear model that predicts *AR* expression based on the number of hypomethylated eHMRs, *AR* expression was positively associated with the number of hypomethylated eHMR loci ( $P=3.7 \times 10^{-5}$ , linear model).

The *AR* gene body and/or the upstream enhancer were amplified in a total of 81% of mCRPC samples. The number of amplified eHMR loci was positively associated with *AR* expression ( $P=3.8 \times 10^{-8}$ , linear model), which is consistent with the hypothesis that these eHMR loci are *AR* regulatory regions (Supplementary Table 7). These data are compatible with a model in which selective pressure of androgen deprivation therapy favors broad amplifications that span multiple enhancers to drive *AR* expression in mCRPC.



**Fig. 2 | DNA methylation valleys.** **a**, Upper: Sample-level  $\log_2(\text{OR})$  calculated from the number of DMVs that overlap H3K4me3 vs. H3K27me3 sites. Lower values favor H3K27me3 and higher values favor H3K4me3. Lower: Sample-level count of DMVs, order matching upper panel. **b**, Mean percentage of methylation across the AR locus for benign prostate ( $n=4$ ), localized PCa ( $n=5$ ), mCRPC adenocarcinoma ( $n=95$ ) and t-SCNC samples ( $n=5$ ). Vertical black dashed lines show the location of the previously identified AR enhancer<sup>17</sup>. The vertical green and red dashed lines show the TSS and transcriptional terminator of the androgen receptor, respectively.

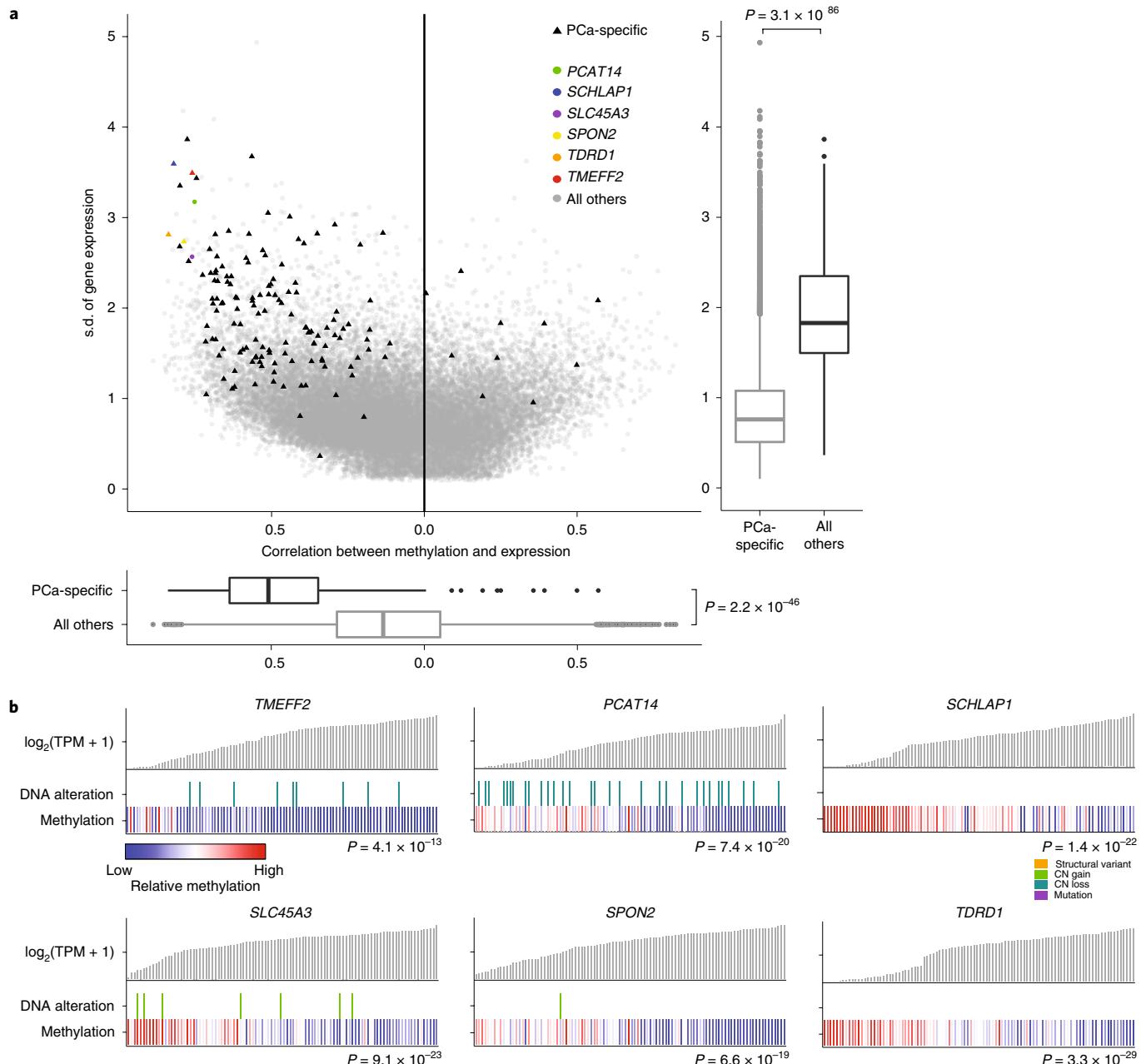
Hypomethylation in non-t-SCNC mCRPC samples was focal, and correlation between hypomethylation and copy number amplification was not present at genomic loci immediately adjacent to the focal eHMRs (Supplementary Table 7). This analysis identified focal genomic loci that may represent novel intergenic regulatory regions of AR that are potentially important in the development of androgen deprivation therapy-resistance<sup>17,18</sup>.

**Methylation associated with TMPRSS2-ERG and MYC expression.** Approximately half of prostate cancers are defined by over-expression of the oncogenic transcription factor encoded by ERG. ERG expression is negligible in PCa unless it is activated by gene fusions that bypass the ERG promoter<sup>79</sup>. The predominant 5' ERG fusion partner is the AR-regulated gene TMPRSS2, and the fusion brings the TMPRSS2 promoter into proximity with the ERG gene body, transforming ERG into an AR-driven gene<sup>79</sup>. ERG expression levels vary widely within TMPRSS2-ERG fusion-positive tumors, and a linear model that predicts ERG expression from AR expression and mutation status provided a poor fit ( $P=0.49$ ,  $F=0.72$ , d.f.=38, ANOVA; Fig. 5b). We reasoned that methylation in the promoter or upstream region of TMPRSS2 could influence ERG expression when the fusion was present. We identified rHMRs upstream of TMPRSS2 that colocalized with the TFBS of HOXB13, FOXA1, AR or ERG (Fig. 5a). Hypomethylation frequencies of these loci were similar in both the fusion-positive and fusion-negative samples. However, methylation at these loci was negatively associated with ERG expression in only the fusion-positive samples, which is consistent with

a model in which TFBS methylation modulates expression of the downstream fusion gene<sup>80–82</sup> (Supplementary Fig. 10). Prediction of ERG expression was significantly improved by the addition of methylation at all rHMRs upstream of TMPRSS2, only in fusion-positive tumors ( $P=0.0002$ ,  $F=5.1$ , d.f.=16, for fusion-positive versus  $P=0.76$ ,  $F=0.72$ , d.f.=16, for fusion-negative samples, ANOVA; Fig. 5b). These data suggest that methylation at regulatory regions upstream of TMPRSS2 contribute to this subtype.

The oncogene MYC is amplified in 38% of our mCRPC samples<sup>17</sup>. MYC gene copy number amplification was modestly correlated with MYC expression ( $P=0.002$ , Spearman's  $\rho=0.31$  (0.11–0.49)). Distal enhancers in the downstream gene PVT1 have been reported to regulate MYC via physical DNA–DNA interactions<sup>83</sup>. DNA interactions between PVT1 and MYC were present in the VCaP ChIA-PET data (Fig. 5c). We observed rHMRs in the MYC promoter and in PVT1, which are associated with MYC expression (Fig. 5c). These rHMRs improved the fit of a model that predicts MYC expression over one that uses MYC amplification alone ( $P=0.001$ ,  $F=3.2$ , d.f.=11, ANOVA; Fig. 5d). Enhancer methylation has been shown to modulate enhancer activity, which provides a plausible explanation of this observation<sup>27,84</sup>. Altogether, these findings support the model that methylation may affect the activity of key PCa drivers.

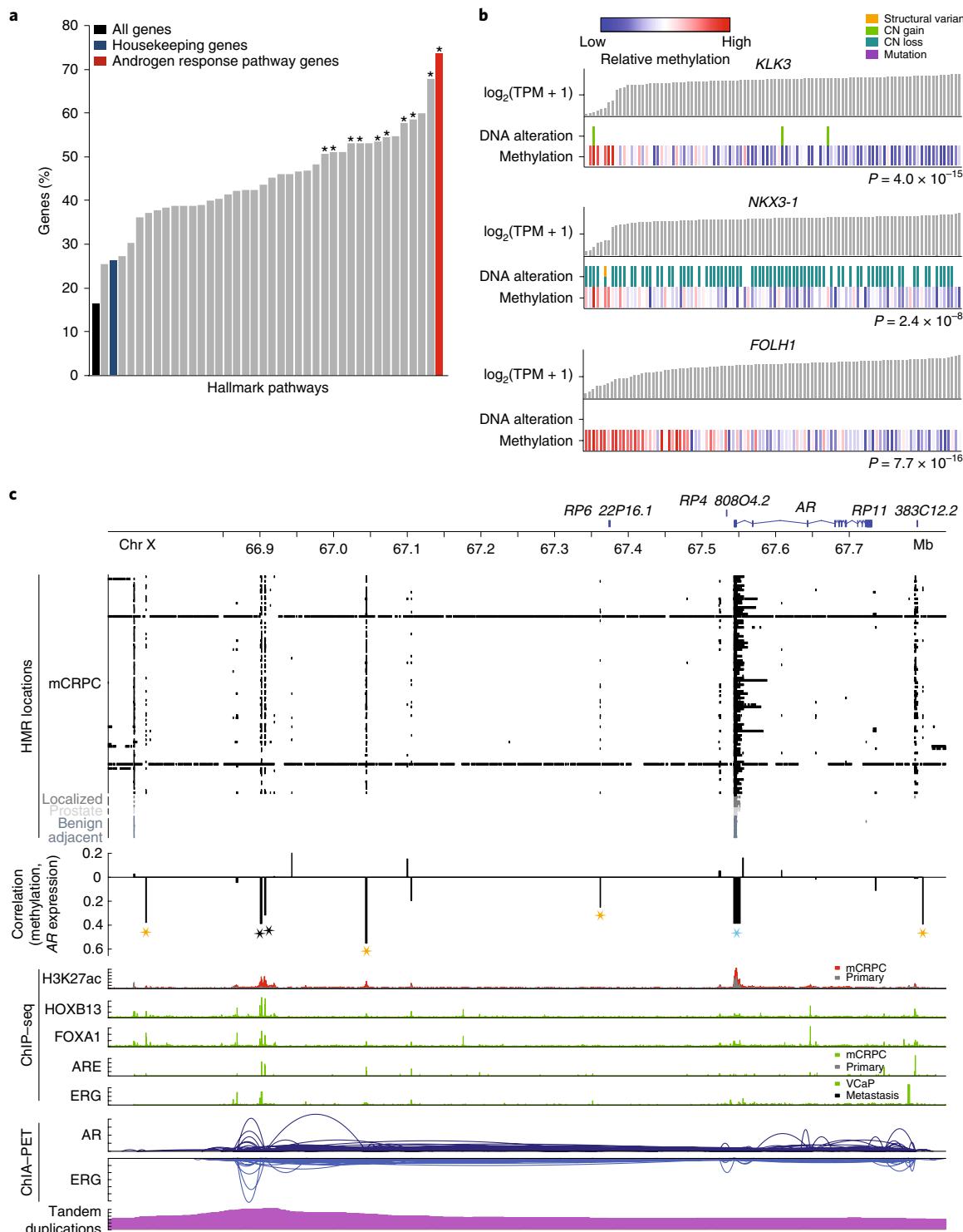
**Methylation and prostate cancer progression.** We used publicly available WGBS data on benign prostate and localized PCa samples<sup>11</sup> to identify differentially methylated regions (DMRs) when comparing benign prostate versus primary PCa and primary



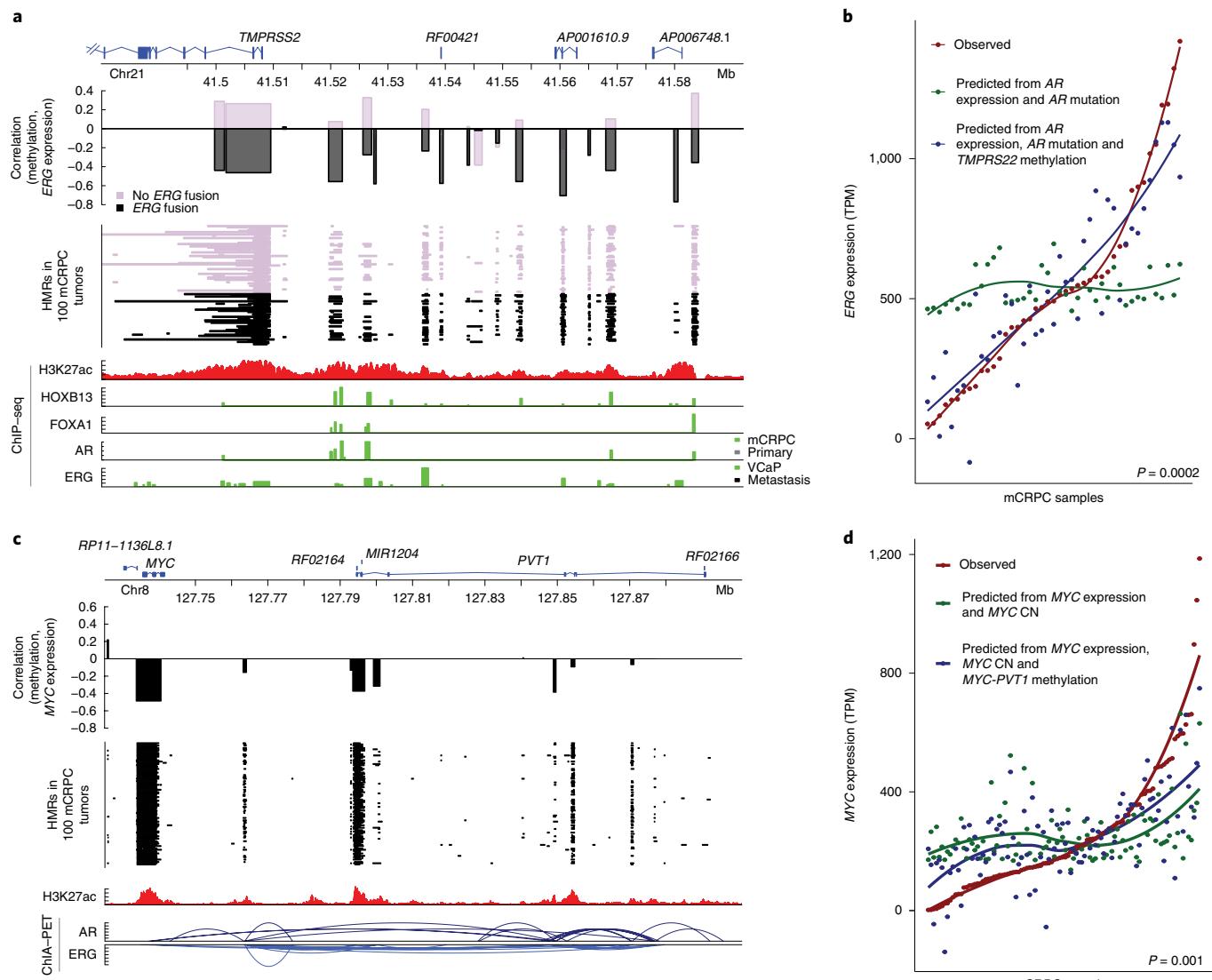
**Fig. 3 | Methylation associated with prostate-cancer-specific genes.** **a**, Variability in gene expression levels versus the correlation between gene expression and methylation. Expression variability was calculated as the s.d. of  $\log_2(\text{TPM} + 1)$ , and the correlation was calculated at the most significant promoter or gene-body eHMR for each gene. The y-axis box plot shows gene expression variability for prostate-cancer-specific genes vs. all other genes. The x-axis box plot shows the correlation of methylation with gene expression of prostate-cancer-specific genes vs. all other genes. Significance was assessed with a two-sided Wilcoxon signed-rank test;  $n=169$  vs. 51502 for prostate-cancer-specific genes vs. all other genes, respectively. Box plots show the median, first and third quartiles, and outliers are shown if outside the 1.5 $\times$  interquartile range. **b**, Sample-level gene expression levels compared to the presence of DNA alterations and methylation at the most significant promoter or gene-body eHMR. Alterations predicted to be activating (for example, those in *SLC45A3*, *SPON2*, *TDRD1* and *SCHLAP1*) or inactivating (for example, those in *TMEFF2* and *PCAT14*) are shown<sup>17</sup>. Significance of methylation levels was assessed by ANOVA, in which a model that predicts gene expression from DNA alterations alone was compared to a second model that had methylation as an added factor;  $n=100$  independent mCRPC samples. CN, copy number; TPM, transcripts per million.

PCa versus mCRPC (Fig. 6a). Primary PCa was predominantly less methylated than benign prostate (97% of 113,622 DMRs; Supplementary Table 8). In addition, mCRPC samples were predominantly less methylated than primary PCa (96% of 508,313 DMRs; Supplementary Table 9). Of the DMRs from benign versus primary PCa, 55% overlapped with the DMRs from primary PCa versus mCRPC.

Global hypomethylation in cancer may contribute to genomic instability<sup>85–87</sup>. When we compared DMRs between benign prostate and mCRPC (Supplementary Table 10) samples with the locations of mCRPC somatic mutations, we found that regions with more differential hypomethylation in mCRPC had an elevated somatic mutation rate in mCRPC (in 1-Mb windows, Spearman's  $\rho=-0.70$  (0.68 to  $-0.72$ ),  $P<2 \times 10^{-16}$ ; Fig. 6b). The mutation rate was 58.5% higher



**Fig. 4 | Methylation association with the androgen-response pathway.** **a**, Percentage of genes in MSigDB hallmark pathways for which methylation predicted expression independently from DNA alterations in a linear model. An asterisk indicates significant enrichment (two-sided FDR  $\leq 0.05$ ) relative to the set of all housekeeping genes. Significance was assessed with a two-sided Fisher's exact test.  $n=100$  independent mCRPC samples. **b**, Sample-level gene expression levels compared to the presence of DNA alterations and methylation at the most significant promoter or gene-body eHMR. Alterations that are predicted to be activating (*KLK3*, *FOLH1*) or inactivating (*NKX3-1*) are shown<sup>17</sup>. Significance of methylation levels was assessed by ANOVA, in which a model that predicts gene expression from DNA alterations alone was compared to a second model that had methylation as an added factor;  $n=100$  independent mCRPC samples. **c**, HMRs, correlation between methylation at loci that harbor rhMRs and AR expression, ChIP-seq peaks (H3K27ac<sup>33</sup>, AR<sup>36</sup>, ERG<sup>38</sup>, FOXA1 (ref. <sup>37</sup>), HOXB13 (ref. <sup>37</sup>)) and ChIA-PET interactions (AR and ERG)<sup>39</sup> at the AR locus. Stars denote HMRs at which methylation was associated with AR expression (eHMRs), colored black for previously reported AR upstream enhancers, blue for the AR promoter and gold for new putative AR regulatory regions. Significance was assessed with a two-sided Spearman's correlation test;  $n=100$  independent mCRPC samples. 'Primary' in the ChIP-seq tracks indicates localized primary PCa.



**Fig. 5 | Methylation association with *TMPRSS2-ERG* and *MYC*. a**, HMRs, correlation between methylation in loci harboring rHMRs and *ERG* expression, and ChIP-seq peaks (H3K27ac<sup>34</sup>, AR<sup>36</sup>, *ERG*<sup>38</sup>, FOXA1 (ref. <sup>37</sup>) and HOXB13 (ref. <sup>37</sup>)) at the *TMPRSS2* locus. Significance was assessed with two-sided Spearman's correlation;  $n=100$  independent mCRPC samples. *TMPRSS2* isoform 204 was not shown as its TSS was  $\sim 20$  kb upstream of the other 5 protein coding isoforms. **b**, Observed *ERG* expression in *TMPRSS2-ERG* fusion-positive mCRPC and *ERG* expression predicted in those tumors using two linear models: one that includes *AR* expression and *AR* mutations and another that includes *AR* expression, *AR* mutations and methylation at the *TMPRSS2* promoter and upstream locus. Significance was assessed by a two-sided ANOVA ( $n=41$  independent fusion-positive samples). **c**, HMRs, correlation between methylation in rHMRs and *MYC* expression, ChIP-seq peaks (H3K27ac<sup>34</sup>) and ChIA-PET interactions (AR and *ERG*)<sup>39</sup> at the *MYC-PVT1* locus. Significance was assessed with two-sided Spearman's correlation;  $n=100$  independent mCRPC samples. 'Primary' in the ChIP-seq tracks indicates localized primary PCa. **d**, Observed *MYC* expression and *MYC* expression predicted in those tumors using two linear models: one that includes *MYC* copy number alone and another that includes *MYC* copy number and methylation at the *MYC-PVT1* locus. Significance was assessed by a two-sided ANOVA ( $n=100$  independent mCRPC samples).

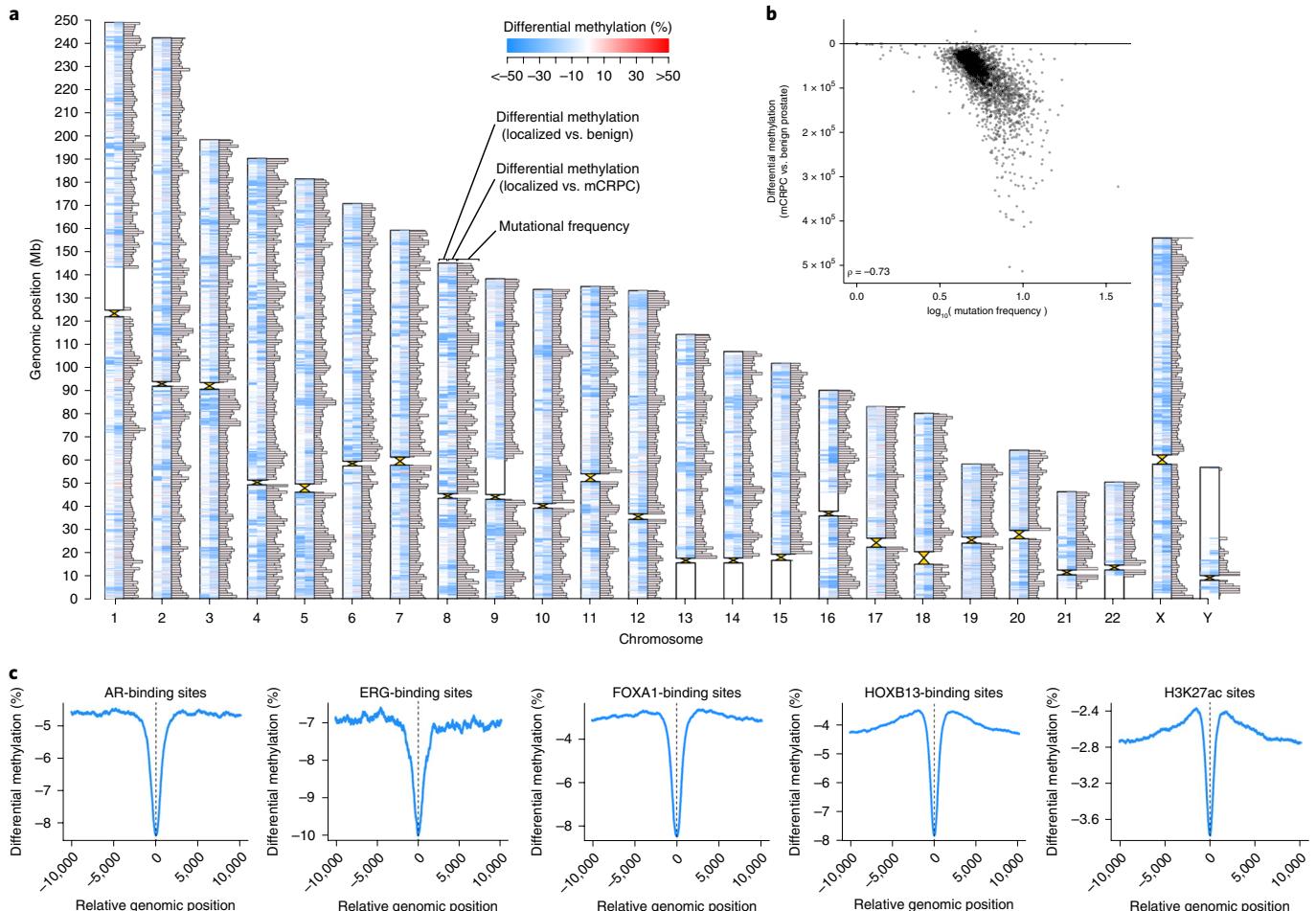
within a DMR than outside a DMR (6.77 versus 4.28 mutations per Mb), which suggests that certain regions of the genome are more frequently somatically altered by both mutation and methylation. Finally, we tested whether differential methylation occurs preferentially in regulatory regions across the genome. When we examined putative regulatory regions (marked by AR, ERG, FOXA1, HOXB13 and H3K27ac ChIP-seq) for differential methylation, HMRs in mCRPC, in comparison to those in benign prostate tissue, were enriched at these sites compared to the surrounding genome (Fig. 6c).

## Discussion

Here, we present global analysis of methylation in mCRPC with WGBS on 100 tumor samples and 10 matched benign-adjacent

metastatic samples, integrated with matched deep WGS and RNA-seq of the same samples. These data identified a novel epigenetic subtype of mCRPC, new intergenic regulatory regions of AR, and the interplay between somatic and epigenetic alterations in the regulation of AR, ERG, MYC and other important PCa drivers. We also demonstrated global methylome changes that distinguish benign prostate, primary PCa and mCRPC. We found that somatic mutations and putative regulatory regions are frequently located in regions that are differentially hypomethylated.

Although genomic and transcriptomic subtypes of PCa have been described<sup>12,16,18–20,88</sup>, we have identified a new epigenetic CMP subtype of mCRPC that is characterized by hypermethylation both within and outside CpG islands, shores and shelves. We propose



**Fig. 6 | Genome-wide analysis of differential methylation.** **a**, DMRs and mutation frequency in mCRPC. Ideogram shows, for each chromosome, from left to right: DMRs comparing primary PCa ( $n=5$ ) to benign prostate ( $n=4$ ), DMRs comparing mCRPC (adenocarcinoma,  $n=95$ ) to primary PCa ( $n=5$ ), and mutational frequency in 1-Mb windows in the mCRPC samples (excluding two hypermutated samples<sup>17</sup>). Maximum bar height in mutation frequency represents an average mutational frequency of  $\geq 10$  mutations per Mb per sample. **b**, Differential methylation (comparing mCRPC adenocarcinoma to benign prostate) compared to mutational frequency (excluding 2 hypermutated samples<sup>17</sup>);  $n=98$ . Each point represents a fixed 1-Mb window of the genome and all points collectively represent all 1-Mb windows across the genome excluding centromeres and telomeres. **c**, Average differential methylation values across all sites identified from publicly available ChIP-seq data (AR<sup>36</sup>, ERG<sup>38</sup>, FOXA1 (ref. <sup>37</sup>), HOXB13 (ref. <sup>37</sup>) and H3K27ac<sup>35</sup>). For each ChIP-seq peak, a 20 kb window centered on the midpoint of the peak ( $x=0$ ) was assessed for differential methylation between mCRPC adenocarcinoma vs. benign prostate samples.

that this phenomenon is analogous to the CpG island methylator phenotype (CIMP) that has been described in other tumor types. The mCRPC CIMP subtype was enriched for mutations in *TET2*, *BRAF* and *IDH1*, which have been associated with the CIMP subtype in other cancer types<sup>32</sup>. *IDH1* mutations were associated with CpG island hypermethylation in The Cancer Genome Atlas primary PCa data<sup>12</sup>. The present study cannot determine whether any mutations that we observed could drive methylation changes. Previous experimental studies of *TET2* and *DNMT3B* mutations have demonstrated that their impact may vary by tissue type and genomic region<sup>89–94</sup>, and phenotypic studies will be required to elucidate the mechanistic basis of the CIMP phenotype. There are potential therapeutic implications of the mCRPC CIMP subtype, as methylation inhibitors such as 5-azacytidine and 5-aza-2-deoxycytidine are FDA-approved anti-neoplastic drugs. In vitro data, as well as clinical data, suggest that hypermethylated tumors may preferentially benefit from these treatments<sup>95,96</sup>.

Our results highlight the importance of cancer-associated hypomethylation in overexpression of oncogenic drivers in mCRPC. The androgen receptor is the dominant driver and therapeutic target in

PCa. Recent studies have characterized amplifications of the *AR* gene body and an enhancer upstream of *AR*<sup>17,18,33</sup>. We found that intergenic eHMRs in these regions at putative *AR* enhancers were associated with *AR* expression in mCRPC. Many of these putative enhancers overlap transcription-factor-binding sites<sup>80–82,97</sup>. These enhancers were distant from the *AR* gene body, but this region demonstrated complex DNA looping, which may bring these loci into proximity with the *AR* promoter. The *MYC-PVT1* interaction is another example of the interplay between long-range *cis*-enhancers and methylation<sup>83</sup>. Distal enhancers are known to activate oncogenes across cancers<sup>27,84</sup>, and these data emphasize the complex interactions between methylation, transcription factors, DNA alterations and the three-dimensional structure of the genome in the pathogenesis of mCRPC.

Comparisons between methylation in mCRPC and primary PCa were limited by the small number of primary PCa samples on which WGBS has been performed<sup>5,11</sup>. Future work that integrates WGS, WGBS and RNA-seq data in large cohorts of primary PCa samples would enable a more robust analysis of the way in which DNA methylation changes during progression to advanced disease, and

would better capture the molecular heterogeneity of primary PCa. Integrated sequencing on additional mCRPC cohorts would allow us to understand the impact of rare alterations on methylation. Furthermore, combining WGS, WGBS and RNA-seq with additional complementary sequencing approaches that measure protein–DNA binding or chromatin structure (for example, ChIP-seq and ChIA-PET) on the same tumors would allow direct observation of how these processes work together to regulate gene expression.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-020-0648-8>.

Received: 11 January 2020; Accepted: 20 May 2020;

Published online: 13 July 2020

## References

- Jones, P. A. & Baylin, S. B. The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* **3**, 415–428 (2002).
- Feinberg, A. P., Koldobskiy, M. A. & Gondor, A. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat. Rev. Genet.* **17**, 284–299 (2016).
- Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257 (1999).
- Rechache, N. S. et al. DNA methylation profiling identifies global methylation differences and markers of adrenocortical tumors. *J. Clin. Endocrinol. Metab.* **97**, E1004–E1013 (2012).
- Skvortsova, K. et al. DNA hypermethylation encroachment at CpG island borders in cancer is predisposed by H3K4 monomethylation patterns. *Cancer Cell* **35**, 297–314.e8 (2019).
- Saghafinia, S., Mina, M., Riggio, N., Hanahan, D. & Ciriello, G. Pan-cancer landscape of aberrant DNA methylation across human tumors. *Cell Rep.* **25**, 1066–1080.e8 (2018).
- Kobayashi, Y. et al. DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer. *Genome Res.* **21**, 1017–1027 (2011).
- Kim, J. H. et al. Deep sequencing reveals distinct patterns of DNA methylation in prostate cancer. *Genome Res.* **21**, 1028–1041 (2011).
- Maruyama, R. et al. Aberrant promoter methylation profile of prostate cancers and its relationship to clinicopathological features. *Clin. Cancer Res.* **8**, 514–519 (2002).
- Bhasin, J. M. et al. Methylome-wide sequencing detects DNA hypermethylation distinguishing indolent from aggressive prostate cancer. *Cell Rep.* **13**, 2135–2146 (2015).
- Yu, Y. P. et al. Whole-genome methylation sequencing reveals distinct impact of differential methylation on gene transcription in prostate cancer. *Am. J. Pathol.* **183**, 1960–1970 (2013).
- Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* **163**, 1011–1025 (2015).
- Borno, S. T. et al. Genome-wide DNA methylation events in TMPRSS2-ERG fusion-negative prostate cancers implicate an EZH2-dependent mechanism with miR-26a hypermethylation. *Cancer Discov.* **2**, 1024–1035 (2012).
- Gerhauser, C. et al. Molecular evolution of early-onset prostate cancer identifies molecular risk markers and clinical trajectories. *Cancer Cell* **34**, 996–1011.e8 (2018).
- Yegnasubramanian, S. et al. Hypermethylation of CpG islands in primary and metastatic human prostate cancer. *Cancer Res.* **64**, 1975–1986 (2004).
- Robinson, D. et al. Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015).
- Quigley, D. A. et al. Genomic hallmarks and structural variation in metastatic prostate. *Cell* **174**, 758–769.e9 (2018).
- Viswanathan, S. R. et al. Structural alterations driving castration-resistant prostate cancer revealed by linked-read genome sequencing. *Cell* **174**, 433–447.e19 (2018).
- Fraser, M. et al. Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* **541**, 359–364 (2017).
- Beltran, H. et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat. Med.* **22**, 298–305 (2016).
- Aryee, M. J. et al. DNA methylation alterations exhibit intra-individual stability and inter-individual heterogeneity in prostate cancer metastases. *Sci. Transl. Med.* **5**, 169ra10 (2013).
- Hama, N. et al. Epigenetic landscape influences the liver cancer genome architecture. *Nat. Commun.* **9**, 1643 (2018).
- Kretzmer, H. et al. DNA methylome analysis in Burkitt and follicular lymphomas identifies differentially methylated regions linked to somatic mutation and transcriptional control. *Nat. Genet.* **47**, 1316–1325 (2015).
- Kulis, M. et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat. Genet.* **44**, 1236–1242 (2012).
- Berman, B. P. et al. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat. Genet.* **44**, 40–46 (2011).
- Klughammer, J. et al. The DNA methylation landscape of glioblastoma disease progression shows extensive heterogeneity in time and space. *Nat. Med.* **24**, 1611–1624 (2018).
- Queiros, A. C. et al. Decoding the DNA methylome of mantle cell lymphoma in the light of the entire B cell lineage. *Cancer Cell* **30**, 806–821 (2016).
- Hovestadt, V. et al. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510**, 537–541 (2014).
- McDonald, O. G. et al. Epigenomic reprogramming during pancreatic cancer progression links anabolic glucose metabolism to distant metastasis. *Nat. Genet.* **49**, 367–376 (2017).
- Chun, H. E. et al. Genome-wide profiles of extra-cranial malignant rhabdoid tumors reveal heterogeneity and dysregulated developmental pathways. *Cancer Cell* **29**, 394–406 (2016).
- Spencer, D. H. et al. CpG island hypermethylation mediated by DNMT3A is a consequence of AML progression. *Cell* **168**, 801–816.e13 (2017).
- Hughes, L. A. et al. The CpG island methylator phenotype: what's in a name? *Cancer Res.* **73**, 5858–5868 (2013).
- Takeda, D. Y. et al. A somatically acquired enhancer of the androgen receptor is a noncoding driver in advanced prostate cancer. *Cancer Cell* **174**, 422–432.e13 (2018).
- Kron, K. J. et al. TMPRSS2-ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer. *Nat. Genet.* **49**, 1336–1345 (2017).
- Stellico, S. et al. Integrative epigenetic taxonomy of primary prostate cancer. *Nat. Commun.* **9**, 4900 (2018).
- Sharma, N. L. et al. The androgen receptor induces a distinct transcriptional program in castration-resistant prostate cancer in man. *Cancer Cell* **23**, 35–47 (2013).
- Pomerantz, M. M. et al. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat. Genet.* **47**, 1346–1351 (2015).
- Yu, J. et al. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell* **17**, 443–454 (2010).
- Zhang, Z. et al. An AR-ERG transcriptional signature defined by long-range chromatin interactomes in prostate cancer cells. *Genome Res.* **29**, 223–235 (2019).
- Sun, W. et al. The association between copy number aberration, DNA methylation and gene expression in tumor samples. *Nucleic Acids Res.* **46**, 3009–3018 (2018).
- Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. *J. Mol. Biol.* **196**, 261–282 (1987).
- Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. USA* **103**, 1412–1417 (2006).
- Ioshikhes, I. P. & Zhang, M. Q. Large-scale human promoter mapping using CpG islands. *Nat. Genet.* **26**, 61–63 (2000).
- 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).
- 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).
- Hennig, C. Cluster-wise assessment of cluster stability. *Comput. Stat. Data Anal.* **52**, 258–271 (2007).
- Weisenberger, D. J. et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat. Genet.* **38**, 787–793 (2006).
- Figueroa, M. E. et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567 (2010).
- Buscarlet, M. et al. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood* **130**, 753–762 (2017).

50. Delhommeau, F. et al. Mutation in *TET2* in myeloid cancers. *N. Engl. J. Med.* **360**, 2289–2301 (2009).
51. Langemeijer, S. M. et al. Acquired mutations in *TET2* are common in myelodysplastic syndromes. *Nat. Genet.* **41**, 838–842 (2009).
52. Dong, C. et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* **24**, 2125–2137 (2015).
53. Wong, T. N. et al. Cellular stressors contribute to the expansion of hematopoietic clones of varying leukemic potential. *Nat. Commun.* **9**, 455 (2018).
54. Odejide, O. et al. A targeted mutational landscape of angioimmunoblastic T-cell lymphoma. *Blood* **123**, 1293–1296 (2014).
55. Coolen, M. W. et al. Consolidation of the cancer genome into domains of repressive chromatin by long-range epigenetic silencing (LRES) reduces transcriptional plasticity. *Nat. Cell Biol.* **12**, 235–246 (2010).
56. Bert, S. A. et al. Regional activation of the cancer genome by long-range epigenetic remodeling. *Cancer Cell* **23**, 9–22 (2013).
57. Zhou, W. et al. DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat. Genet.* **50**, 591–602 (2018).
58. Brinkman, A. B. et al. Partially methylated domains are hypervariable in breast cancer and fuel widespread CpG island hypermethylation. *Nat. Commun.* **10**, 1749 (2019).
59. Xie, W. et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
60. Jeong, M. et al. Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat. Genet.* **46**, 17–23 (2014).
61. Li, Y. et al. Genome-wide analyses reveal a role of Polycomb in promoting hypomethylation of DNA methylation valleys. *Genome Biol.* **19**, 18 (2018).
62. Kumar, A. et al. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. *Nat. Med.* **22**, 369–378 (2016).
63. Mu, P. et al. SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. *Science* **355**, 84–88 (2017).
64. Beltran, H. et al. The role of lineage plasticity in prostate cancer therapy resistance. *Clin. Cancer Res.* **25**, 6916–6924 (2019).
65. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492 (2012).
66. Pacis, A. et al. Gene activation precedes DNA demethylation in response to infection in human dendritic cells. *Proc. Natl Acad. Sci. USA* **116**, 6938–6943 (2019).
67. Teschendorff, A. E. et al. DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nat. Commun.* **7**, 10478 (2016).
68. Yang, X. et al. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* **26**, 577–590 (2014).
69. Massie, C. E., Mills, I. G. & Lynch, A. G. The importance of DNA methylation in prostate cancer development. *J. Steroid Biochem. Mol. Biol.* **166**, 1–15 (2017).
70. Gery, S., Sawyers, C. L., Agus, D. B., Said, J. W. & Koeffler, H. P. *TMEFF2* is an androgen-regulated gene exhibiting antiproliferative effects in prostate cancer cells. *Oncogene* **21**, 4739–4746 (2002).
71. Qian, X. et al. Spondin-2 (SPON2), a more prostate-cancer-specific diagnostic biomarker. *PLoS One* **7**, e37225 (2012).
72. Boormans, J. L. et al. Identification of *TDRD1* as a direct target gene of *ERG* in primary prostate cancer. *Int. J. Cancer* **133**, 335–345 (2013).
73. Xu, J. et al. Identification and characterization of prostein, a novel prostate-specific protein. *Cancer Res.* **61**, 1563–1568 (2001).
74. Prensner, J. R. et al. RNA biomarkers associated with metastatic progression in prostate cancer: a multi-institutional high-throughput analysis of SChLAP1. *Lancet Oncol.* **15**, 1469–1480 (2014).
75. White, N. M. et al. Multi-institutional analysis shows that low *PCAT-14* expression associates with poor outcomes in prostate cancer. *Eur. Urol.* **71**, 257–266 (2017).
76. Eisenberg, E. & Levanon, E. Y. Human housekeeping genes, revisited. *Trends Genet.* **29**, 569–574 (2013).
77. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
78. Zhao, S. G. et al. The immune landscape of prostate cancer and nomination of PD-L2 as a potential therapeutic target. *J. Natl. Cancer Inst.* **111**, 301–310 (2019).
79. Tomlins, S. A. et al. Role of the TMPRSS2-*ERG* gene fusion in prostate cancer. *Neoplasia* **10**, 177–188 (2008).
80. Domcke, S. et al. Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* **528**, 575–579 (2015).
81. Feldmann, A. et al. Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet.* **9**, e1003994 (2013).
82. Yin, Y. et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, eaaj2239 (2017).
83. Cho, S. W. et al. Promoter of lncRNA gene *PVT1* is tumor-suppressor DNA boundary element. *Cell* **173**, 1398–1412.e22 (2018).
84. Bell, R. E. et al. Enhancer methylation dynamics contribute to cancer plasticity and patient mortality. *Genome Res.* **26**, 601–611 (2016).
85. Yegnasubramanian, S. et al. DNA hypomethylation arises later in prostate cancer progression than CpG island hypermethylation and contributes to metastatic tumor heterogeneity. *Cancer Res.* **68**, 8954–8967 (2008).
86. Kulis, M. & Esteller, M. DNA methylation and cancer. *Adv. Genet.* **70**, 27–56 (2010).
87. Chen, R. Z., Pettersson, U., Beard, C., Jackson-Grusby, L. & Jaenisch, R. DNA hypomethylation leads to elevated mutation rates. *Nature* **395**, 89–93 (1998).
88. Zhao, S. G. et al. Associations of luminal and basal subtyping of prostate cancer with prognosis and response to androgen deprivation therapy. *JAMA Oncol.* **3**, 1663–1672 (2017).
89. Yamazaki, J. et al. *TET2* mutations affect non-CpG island DNA methylation at enhancers and transcription factor-binding sites in chronic myelomonocytic leukemia. *Cancer Res.* **75**, 2833–2843 (2015).
90. Yamazaki, J. et al. Effects of *TET2* mutations on DNA methylation in chronic myelomonocytic leukemia. *Epigenetics* **7**, 201–207 (2012).
91. Ko, M. et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant *TET2*. *Nature* **468**, 839–843 (2010).
92. Rasmussen, K. D. et al. Loss of *TET2* in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis. *Genes Dev.* **29**, 910–922 (2015).
93. Liao, J. et al. Targeted disruption of *DNMT1*, *DNMT3A* and *DNMT3B* in human embryonic stem cells. *Nat. Genet.* **47**, 469–478 (2015).
94. Duymich, C. E., Charlet, J., Yang, X., Jones, P. A. & Liang, G. *DNMT3B* isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. *Nat. Commun.* **7**, 11453 (2016).
95. Itzykson, R. et al. Impact of *TET2* mutations on response rate to azacitidine in myelodysplastic syndromes and low blast count acute myeloid leukemias. *Leukemia* **25**, 1147–1152 (2011).
96. Bejar, R. et al. *TET2* mutations predict response to hypomethylating agents in myelodysplastic syndrome patients. *Blood* **124**, 2705–2712 (2014).
97. Brocks, D. et al. Intratumor DNA methylation heterogeneity reflects clonal evolution in aggressive prostate cancer. *Cell Rep.* **8**, 798–806 (2014).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

<sup>1</sup>Department of Radiation Oncology, University of Michigan, Ann Arbor, MI, USA. <sup>2</sup>Rogel Cancer Center, University of Michigan, Ann Arbor, MI, USA.

<sup>3</sup>Department of Radiation Oncology, University of California San Francisco, San Francisco, CA, USA. <sup>4</sup>Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA, USA. <sup>5</sup>Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, MI, USA. <sup>6</sup>Yale School of Medicine, New Haven, CT, USA. <sup>7</sup>Division of Hematology and Oncology, Department of Medicine, University of California San Francisco, San Francisco, CA, USA. <sup>8</sup>Illumina Inc., San Diego, CA, USA. <sup>9</sup>Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA.

<sup>10</sup>Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR, USA. <sup>11</sup>Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada. <sup>12</sup>Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, CA, USA.

<sup>13</sup>Center for Advanced Technology, University of California San Francisco, San Francisco, CA, USA. <sup>14</sup>McDonnell Genome Institute, Washington University, St. Louis, MO, USA. <sup>15</sup>Department of Internal Medicine, Washington University, St. Louis, MO, USA. <sup>16</sup>Siteman Cancer Center, Washington University, St. Louis, MO, USA. <sup>17</sup>The Hormel Institute, University of Minnesota, Austin, MN, USA. <sup>18</sup>Ontario Institute for Cancer Research, Toronto, Ontario, Canada.

<sup>19</sup>Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada. <sup>20</sup>Department of Human Genetics, Institute for Precision Health, UCLA, Los Angeles, CA, USA. <sup>21</sup>Division of Hematology/Medical Oncology, Department of Medicine, Oregon Health & Science University, Portland, OR, USA. <sup>22</sup>Department of Pathology, Oregon Health & Science University, Portland, OR, USA. <sup>23</sup>Vancouver Prostate Centre, Department of Urologic Sciences, University of British Columbia, Vancouver, British Columbia, Canada. <sup>24</sup>British Columbia Cancer Agency, Vancouver Centre, Vancouver,

British Columbia, Canada. <sup>25</sup>Jonsson Comprehensive Cancer Center, Departments of Medicine and Urology, University of California Los Angeles, Los Angeles, CA, USA. <sup>26</sup>Department of Medicine, VA Greater Los Angeles Healthcare System, Los Angeles, CA, USA. <sup>27</sup>Department of Microbiology, Immunology, and Molecular Genetics at the David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. <sup>28</sup>Department of Neurological Surgery, University of California San Francisco, San Francisco, CA, USA. <sup>29</sup>Department of Urology, University of Michigan, Ann Arbor, MI, USA. <sup>30</sup>Department of Urology, University of California San Francisco, San Francisco, CA, USA. <sup>31</sup>Department of Anatomy, University of California San Francisco, San Francisco, CA, USA. <sup>32</sup>Department of Cancer Biology, University of Pennsylvania, Philadelphia, PA, USA. <sup>33</sup>Peter MacCallum Cancer Centre, University of Melbourne, Melbourne, Victoria, Australia. <sup>34</sup>Department of Medicine, University of Wisconsin, Madison, WI, USA. <sup>35</sup>Department of Surgery, Washington University, St. Louis, MO, USA. <sup>36</sup>Division of Hematology Oncology, Department of Internal Medicine, University of California Davis, Sacramento, CA, USA. <sup>37</sup>Comprehensive Cancer Center, University of California Davis, Sacramento, CA, USA. <sup>38</sup>Department of Urologic Surgery, University of California Davis, Sacramento, CA, USA. <sup>39</sup>Department of Cancer Biology, Thomas Jefferson University, Philadelphia, PA, USA. <sup>40</sup>Department of Pathology, Duke University, Durham, NC, USA. <sup>41</sup>Netherlands Cancer Institute, Oncode Institute, Amsterdam, the Netherlands. <sup>42</sup>Department of Pathology, University of Pittsburgh, Pittsburgh, PA, USA. <sup>43</sup>Masonic Cancer Center, University of Minnesota, Minneapolis, MN, USA. <sup>44</sup>Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA. <sup>45</sup>Department of Pathology, University of Michigan, Ann Arbor, MI, USA. <sup>46</sup>Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. <sup>47</sup>Howard Hughes Medical Institute, University of Michigan Medical School, Ann Arbor, MI, USA. <sup>48</sup>Department of Biomedical Engineering, Washington University, St. Louis, MO, USA. <sup>49</sup>Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, USA. <sup>50</sup>These authors contributed equally: Shuang G. Zhao, William S. Chen and Haolong Li. <sup>51</sup>These authors jointly supervised the work: Arul M. Chinnaiyan, Christopher A. Maher, Eric J. Small, David A. Quigley and Felix Y. Feng. ✉e-mail: [Felix.Feng@ucsf.edu](mailto:Felix.Feng@ucsf.edu)

## Methods

**Biopsy samples.** Fresh-frozen image-guided mCRPC biopsy samples were obtained as described previously<sup>17</sup>. Benign-adjacent metastatic biopsies were identified for a subset of patients following centralized pathology review. DNA extraction was performed as described previously<sup>17</sup>. WGBS libraries were prepared from 250 ng of genomic DNA with 0.5% unmethylated  $\lambda$  phage DNA (Promega) spiked in to measure bisulfite conversion efficiency. Bisulfite conversion efficiency was >99.5% in all samples, as measured by  $\lambda$  phage DNA spike-in. Samples were fragmented using a Covaris M220 focused-ultrasonicator to an average size of 500 bp. Bisulfite conversion was performed using the EZ DNA methylation gold kit (Zymo Research). Library preparation was performed using Accel-NGS Methyl-Seq (Swift BioSciences). Library quality was monitored by using a 2100 Bioanalyzer (Agilent). Sequencing was performed at the University of California San Francisco Center for Advanced Technology sequencing core. Paired-end reads of 151 bp were sequenced on the Illumina Novaseq 6000 system.

**Data processing.** Alignment, trimming and methylation calling was performed using the Illumina Basespace platform. Ten bases were trimmed off the 5' end of every read according to the Bismark User Guide recommendations for the library kit used. Quality trimming was performed according to default recommendations of the Illumina MethylSeq application v2.0.0 (trim bases at the 5'-end with a quality score of less than 30; trim bases at the 3'-end with a quality score of less than 30; trim the 3'-end of reads with a quality score of less than 15; trim the 3'-end of reads using a sliding-window approach with a window length of 4). Alignment to the GRCh38.p12 genome assembly, deduplication and base-level methylation calling were performed using Bismark v0.20.0 (ref. <sup>98</sup>) using the default parameters as recommended by the Bismark User Guide for the library kit. The '-paired-end' and '--no\_overlap' parameters were set. Bases with germline or somatic C>T or G>A mutations were excluded from analysis on a per-sample basis using the WGS germline and somatic results, as these specific mutations resulted in variants that are indistinguishable from bisulfite-converted reference bases by the sequencer. HMRs and PMDs were identified using MethylSeekR v1.22.0 (ref. <sup>99</sup>), with a UMR/LMR threshold of 30%, and otherwise using the default parameters. Only bases with a minimum coverage of five reads (the default MethylSeekR cutoff) were included for subsequent analysis. RNA-seq from laser-capture micro-dissected samples was aligned as described previously<sup>17</sup>, and the abundance was calculated using featureCounts with the default parameters<sup>100</sup>. Genes were defined using GENCODE release 28. Duplicate reads were ignored and junction counts were included. The transcripts per million (TPM) value was calculated for each gene to quantify expression<sup>17</sup>. WGS data were processed to call mutations, copy number alterations and structural variants, as described previously<sup>17</sup>. Tumor purity was assessed by histological evaluation, by analysis of DNA using Canvas<sup>101</sup> and in the RNA by ESTIMATE<sup>102</sup>. Purity estimates were all significantly intercorrelated (Spearman's *P* values were all <0.0001 for histological evaluation versus DNA, histological evaluation versus RNA and DNA versus RNA).

**Statistical methods.** Plotting and statistical tests were performed using R v3.4.4. All statistical tests performed in the manuscript were two-sided. Box plots were generated using the R ggplot2 function (center line = median; box limits = upper and lower quartiles; whiskers =  $1.5 \times$  interquartile range). Hierarchical clustering was performed using the Euclidean distance and the complete linkage method. A two-sided Wilcoxon signed-rank test was used to assess differences between two groups. Multiple testing correction was performed using the Benjamini–Hochberg method when applicable. Box plots show the median, first and third quartiles, and outliers are shown if outside the  $1.5 \times$  interquartile range. A reporting summary can be found in the attached Life Sciences Reporting Summary.

**Publicly available data.** WGBS for five primary prostate tumors and four matched benign-adjacent prostate samples (referred to as 'benign prostate' throughout the text to avoid confusion with the benign-adjacent metastatic biopsies) were obtained from the authors of ref. <sup>11</sup>. Quality trimming was performed as above, and alignment to the GRCh38.p12 assembly, deduplication and base-level methylation calls were performed using Bismark v0.20.0 as above<sup>98</sup>. The default Bismark parameters were again used, as well as the '-non\_directional' parameter needed for the specific library preparation protocol used on these samples. The '-paired-end' and '--no\_overlap' parameters were also set. MethylSeekR was called with identical parameters as above, except that a three-read minimum coverage<sup>100</sup> was applied due to lower sequencing depth.

Processed ChIP-seq data were obtained from the Gene Expression Omnibus. Raw data were not reprocessed. If raw density tracks were available in the form of BigWig files for plotting, these were used. Otherwise, the peaks were plotted. The peak calls from the original ChIP-seq studies were used without modification for all analyses that used peaks. H3K27ac data from mCRPC and primary PCa samples were obtained from GSE114385 (ref. <sup>33</sup>) (only available on chromosome X). Primary PCa H3K4me3 data were obtained from GSE96652 (ref. <sup>34</sup>). H3K4me3, H3K27ac and H3K27me3 primary PCa ChIP-seq data were obtained from GSE120738 (ref. <sup>35</sup>). Primary and metastatic PCa AR ChIP-seq data were obtained from GSE28219 (ref. <sup>36</sup>). Primary PCa FOXA1 and HOXB13 ChIP-seq data were obtained from GSE70079 (ref. <sup>37</sup>). Metastatic PCa and VCaP ERG ChIP-seq data

were obtained from GSE14097 (ref. <sup>38</sup>). Processed AR and ERG ChIP-PET data from VCaP were obtained from GSE54946 (ref. <sup>39</sup>). The ChIP-seq peaks and ChIP-PET interactions published in the original manuscripts were used, and coordinates were converted from the hg19 genome assembly to GRCh38 using the University of California Santa Cruz (UCSC) LiftOver tool.

**Recurrent hypomethylated regions.** HMRs were identified with the MethylSeekR tool<sup>99</sup>. rHMRs were defined by running a 100-bp sliding window across the genome and identifying contiguous regions where MethylSeekR called an HMR in  $\geq 5\%$  of mCRPC samples. For example, if on chr1, in the region from 10,000–10,099, 1 sample had an HMR; in the region from 10,100–10,199, 5 samples had an HMR; in the region from 10,200–10,299, 7 samples had an HMR; and in the region from 10,300–10,399, 2 samples had an HMR, the region from 10,100–10,299 would be marked as a recurrent HMR. Only focal HMRs ( $\leq 10$  kb) were used in this analysis. HMRs were assigned to the first group that they overlapped in the following order: promoter, gene body, publicly available PCa ChIP-seq for transcription factors (AR<sup>36</sup>, ERG<sup>38</sup>, FOXA1 (ref. <sup>37</sup>) and HOXB13 (ref. <sup>37</sup>)), H3K27ac<sup>35</sup> and H3K27me3 (ref. <sup>35</sup>).

**Definition of prostate-cancer-specific genes.** Prostate-cancer-specific genes were defined as those with elevated expression in primary PCa compared to all other tumor types and benign prostate<sup>103</sup>. We used The Cancer Genome Atlas pan-cancer fragments per kilobase of transcript per million mapped reads RNA-seq data<sup>104</sup> (downloaded via the UCSC Xena Browser<sup>105</sup>) to identify genes that were overexpressed in PCa compared to benign prostate tissue and compared these to all 32 other tumor or normal tissue types individually. Genes were deemed to be PCa-specific if all 33 comparisons had a one-sided Wilcoxon signed-rank test FDR of  $\leq 0.05$ , and a fold change of  $>2$ , when PCa samples were compared to non-PCa samples.

**Correlation analysis between methylation and gene expression.** All correlation analyses were performed using Spearman's correlation. Genes with RNA-seq expression values of  $<1$  TPM in all samples were excluded from such analyses, which resulted in a total of 51,708 genes that were retained for analysis. To estimate methylation levels and calculate eHMRs, the methylation levels of all CpGs in the rHMRs were first averaged in each sample, and then correlated with gene expression across samples. eHMRs were defined as rHMRs that were significantly associated with expression, using a threshold of FDR  $\leq 0.05$ . Whereas multiple eHMRs could exist for a single gene, a single eHMR with the smallest *P* value when correlating with gene expression was reported.

### Methylation association with gene expression independent of DNA alterations.

In order to identify genes in which methylation was associated with gene expression independent of DNA alterations, we fit a linear model that predicts gene expression based on DNA-sequence alterations and all upstream/promoter or gene-body rHMRs collectively. Using ANOVA, we compared this model that included both DNA-sequence alterations and gene methylation to a linear model that included DNA alterations alone<sup>106</sup>. All recurrent promoter and gene-body HMRs were included (rather than only eHMRs) to avoid bias for regions known only to be associated with expression. Promoters were defined as  $\pm 1.5$  kb from the gene start site<sup>68</sup>. To assess which genomic pathways were most associated with methylation, we computed the number of genes in each MSigDB hallmark pathway v6.2 (refs. <sup>77,78</sup>) whose expression was associated with methylation independently of DNA alterations. Fisher's exact test was used to compare this statistically with the number of housekeeping genes<sup>76</sup> in which methylation added to DNA alterations.

**Differentially methylated regions.** Differential methylation was performed using the DSS R package v2.26.0 (ref. <sup>107</sup>) with smoothing set to true, and otherwise using default parameters. No minimum CpG read coverage was applied for this analysis, as DSS accounts for read depth when calling DMRs. To compute the correlation between DMRs and somatic mutational frequency, the differential methylation extent was computed in 1-Mb windows for the entire genome, defined as the sum of the DSS 'areaStat' within the 1-Mb window. Somatic mutational frequency was computed for the same 1-Mb windows and was averaged across all samples, excluding the two hypermutated samples<sup>17</sup>. Mutation and differential methylation calls that overlapped assembly gaps and centromeres (obtained from the UCSC genome browser) were excluded for this analysis. The correlation between differential methylation and mutational frequency in these windows was computed using Spearman's correlation. Differential methylation analysis at ChIP-seq loci was performed by first identifying published AR-, ERG-, FOXA1- and HOXB13-binding sites and H3K27ac sites as above. A 20-kb window centered on each TFBS was considered. Each base in a 20-kb window was represented as the degree of differential methylation if contained within a DMR (defined by DSS), or as 0 if not contained within a DMR. The per-base DSS values were averaged across all 20-kb windows to assess focal enrichment of differential methylation in or around each TFBS.

**DNA methylation valleys.** DMVs were defined as HMRs that are  $\geq 5$  kb in length<sup>59</sup>. To assess the balance between H3K4me3 and H3K27me3, for each sample, a  $2 \times 2$

table was constructed with the number of DMVs that overlapped an H3K4me3 site only, an H3K27me3 site only, both an H3K4me3 and an H3K27me3 site, or neither site. The odds ratio (OR) was then calculated and plotted.

**Partially methylated domains.** To globally assess the variability of PMDs in PCa, we defined PMDs for each sample using the MethylseekR tool with the same settings as when calling HMRs. The total length of PMDs for each sample was divided by the total genome length to calculate the proportion of the genome that contains PMDs. For each PMD called by MethylSeekR, the mean methylation level of all CpGs in that PMD was calculated and the mean methylation of all PMDs in each sample was calculated to obtain the mean PMD methylation value. GENCODE 28-annotated exons were merged to identify coding bases. To calculate the fraction of PMDs that overlapped with the coding genome, we divided the total number of coding bases inside PMDs by the total length of all PMDs in each sample. To calculate the fraction of the genome outside of PMDs that overlapped with the coding genome, we divided the total number of coding bases outside PMDs by the total length of the genome outside PMDs. This analysis was restricted to mCRPC samples. Mutational density inside and outside PMDs was calculated for each sample. The two previously identified hypermutated samples were excluded from this analysis<sup>17</sup>.

**Long-range epigenetic regulation.** To identify candidate long-range epigenetic regulated regions, we examined five-gene windows across the genome, where every gene was correlated with the nearest two genes upstream and downstream. We identified peaks with Spearman's correlation in this sliding window when the average correlation exceeded 0.3. Peaks needed to have at least five genes and peaks within two genes of each other were merged together. This same sliding-window approach was applied to CpG islands. Regions in which the gene expression and CpG island intercorrelated peaks overlapped with each other were identified, where the average correlation between expression and CpG island methylation exceeded 0.1 or -0.1.

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

## Data availability

WGBS, WGS and RNA-seq data are available at dbGAP ([phs001648](#)). All figures use these raw data. Processed ChIP-seq and CHIA-PET data were obtained from the Gene Expression Omnibus: [GSE114385](#); [GSE96652](#); [GSE120738](#); [GSE28219](#); [GSE70079](#); [GSE14097](#); [GSE54946](#).

## Code availability

All code used in the manuscript is available at [https://github.com/DavidQuigley/WCDT\\_WGBS](https://github.com/DavidQuigley/WCDT_WGBS).

## References

98. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
99. Burger, L., Gaidatzis, D., Schubeler, D. & Stadler, M. B. Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Res.* **41**, e155 (2013).
100. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
101. Roller, E., Ivakhno, S., Lee, S., Royce, T. & Tanner, S. Canvas: versatile and scalable detection of copy number variants. *Bioinformatics* **32**, 2375–2377 (2016).
102. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
103. Sanchez-Vega, F. et al. Oncogenic signaling pathways in The Cancer Genome Atlas. *Cell* **173**, 321–337.e10 (2018).
104. Bailey, M. H. et al. Comprehensive characterization of cancer driver genes and mutations. *Cell* **173**, 371–385.e18 (2018).
105. Goldman, M. et al. The UCSC Cancer Genomics Browser: update 2015. *Nucleic Acids Res.* **43**, D812–D817 (2015).
106. Faraway, J. *Linear Models with R* (Taylor & Francis, 2014).
107. Wu, H. et al. Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic Acids Res.* **43**, e141 (2015).

## Acknowledgements

We thank the patients who selflessly contributed samples to this study and without whom this research would not have been possible. We would also like to acknowledge the assistance of Steven Kronenberg and Barbara Panning. This research was supported by 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. S.G.Z., D.A.Q., Hui Li, R.A., J.T.H., R.B. and R.Y. were funded by Prostate Cancer Foundation Young Investigator Awards. 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. and A.A. were supported by National Institutes of Health (NIH)/National Cancer Institute (NCI) 1R01CA230516-01. F.Y.F. and N.M. were supported by NIH/NCI 1R01CA227025 and Prostate Cancer Foundation (PCF) 17CHAL06. F.Y.F. and A.M.C. were supported by NIH P50CA186786. A.M.C. is supported by NIH R35CA231996 and U01CA214170. D.A.Q. was funded by a BRCA Foundation Young Investigator Award. M.S. was supported by the Swedish Research Council (Vetenskapsrådet) with grant number 2018-00382 and the Swedish Society of Medicine (Svenska Läkaresällskapet). L.A.G. was supported by K99/R00 CA204602 and DP2 CA239597, as well as the Goldberg-Benioff Endowed Professorship in Prostate Cancer Translational Biology. P.C.B. was supported by the NIH/NCI under award number P30CA016042 and by an operating grant from the National Cancer Institute Early Detection Research Network (1U01CA214194-01).

## Author contributions

S.G.Z., W.S.C., E.J.S., D.A.Q. and F.Y.F. conceived and designed the study. S.G.Z., Haolong Li, A.F., R.A., D.P., J.J.A., R.D., T.J.B., A.M.B., E.C., T.M.B., G.T., K.N.C., M.G., A.Z., R.E.R., M.B.R., O.W., M.Y.K., P.N.L., C.P.E., P.F., S.B., J.H., J.F.C., J.L., A.W.W., K.F., E.J.S., D.A.Q. and F.Y.F. acquired the data. S.G.Z., W.S.C., Haolong Li, M.Z., M.S., R.A., A.L., R.D., J.C., J.T.H., M.P., H.X.D., R.Y., R.M.-B., L.Z., M.A., S.L.C., K.E.H., Y.J.S., M.Y.K., L.F., D.E.S., T.M.M., R.B., F.W.H., Hui Li, L.C., T.S., H.G., I.A.A., S.S., J.M.L., N.M., K.E.K., H.H.H., W.Z., S.A.T., A.W.W., S.M.D., A.A., L.A.G., P.C.B., A.M.C., C.A.M., E.J.S., D.A.Q. and F.Y.F. analysed and interpreted the data. All authors drafted the article or revised it critically for important intellectual content. All authors approved the final version of the manuscript.

## Competing interests

P.F., S.B., K.F. and A.L. are employees of Illumina Inc., which provided material support for this project. No other commercial entities contributed to or played a role in the writing of this article. J.M.L. holds equity in Salus Discovery, LLC. L.F. has funding from BMS, Abbvie, Janssen, Roche/Genentech and Merck. O.W. currently has consulting, equity and/or board relationships with Trethera Corporation, Kronos Biosciences, Sofie Biosciences, Breakthrough Properties, Vida Ventures, Nanmi Therapeutics and Allogene Therapeutics. The University of Michigan and Brigham and Women's Hospital have been issued patents on ETS gene fusions in prostate cancer, on which S.A.T. is a co-inventor. The diagnostic field of use was licensed to Hologic/Gen-Probe Inc., which has sublicensed rights to Roche/Ventana Medical Systems. S.A.T. has served as a consultant for and received honoraria from Janssen, AbbVie, Sanofi, Almac Diagnostics and Astellas/Medivation. S.A.T. has sponsored research agreements with Astellas/Medivation and GenomeDX. S.A.T. is a cofounder, previous consultant for and current employee of Strata Oncology. T.M.B. has research funding from Alliance Foundation Trials, Boehringer Ingelheim, Concept Therapeutics, Endocyte Inc., Janssen R&D, Medivation Inc./Astellas, oncoGenex, Sotio and Theraclone Sciences/Oncoreponse. T.M.B. has received consulting fees from AbbVie, AstraZeneca, Astellas Pharma, Bayer, Boehringer Ingelheim, Clovis Oncology, GlaxoSmithKline, Janssen Biotech, Janssen Japan, Merck and Pfizer. T.M.B. holds stock in Salaris Pharmaceuticals. M.R. reports consulting and Speakers' Bureau for Johnson & Johnson, research funding from Novartis, research support from Merck and Astellas/Medivation, and a provisional patent with UCLA on the development of small-molecule inhibitors of the androgen receptor N-terminal domain. 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. A.A. is a co-founder of Tango Therapeutics, Azkarra Therapeutics and Ovibio Corporation; is a consultant for SPARC, Bluestar, Prolynx, Earli, Cura, GenVivo and GSK; is a member of the SAB of Genentech and GLAdiator; receives grant/research support from SPARC and AstraZeneca; and holds patents on the use of PARP inhibitors held jointly with AstraZeneca, from which he has benefitted financially (and may do so in the future). F.Y.F. has consulted for Astellas, Bayer, BlueEarth Diagnostics, Celgene, Clovis, EMD Serono, Genentech, Janssen, Myovant, Ryovant and Sanofi, and is a co-founder and has an ownership stake in PFS Genomics. S.L.C. is in a leadership role at PFS Genomics. S.G.Z., S.L.C. and F.Y.F. have patent applications with Decipher Biosciences on molecular signatures in prostate cancer unrelated to this work. S.G.Z. and F.Y.F. have a patent application for a molecular signature in breast cancer unrelated to this work and licensed to PFS Genomics. S.G.Z. and F.Y.F. have patent applications with Celgene unrelated to this work.

## Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41588-020-0648-8>.

Correspondence and requests for materials should be addressed to F.Y.F.

Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).

Corresponding author(s): Felix Y. Feng

Last updated by author(s): Apr 28, 2020

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

No software was used for data collection

Data analysis

Illumina Basespace Platform - <https://basespace.illumina.com/>

Bismark 0.20.0

Illumina FastQ Toolkit 2.2.0

MethylSeekR 1.24.0

DSS 2.32.0

R 3.4.4

All custom code has been made available at [https://github.com/DavidQuigley/WCDT\\_WGBS](https://github.com/DavidQuigley/WCDT_WGBS)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

WGS and RNA-seq are available at dbGAP (phs001648) and WGBS are available under the same accession. All figures use these raw data. Processed ChIP-seq and ChIA-PET data were obtained from the Gene Expression Omnibus (GEO): GSE114385; GSE96652; GSE120738; GSE28219; GSE70079; GSE14097; GSE54946.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](http://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculations were performed, as the whole-genome bisulfite sequencing data were built upon an existing database with whole-genome sequencing and RNA-seq of a fixed size. This is 2.5X larger than largest published cohort with integrated sequencing of this kind in cancer, and captures all of the most common genetic subtypes of prostate cancer.
Data exclusions	No data excluded
Replication	Not applicable, one sample per patient was sequenced
Randomization	Not applicable
Blinding	Not applicable

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms		
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Men with metastatic castration-resistant prostate cancer
Recruitment	The patients were recruited at multiple different institutions as part of the Stand-Up-To-Cancer West Coast Dream Team consortium. One potential bias is that these patients had to be well enough to undergo a biopsy, and that they were seen at academic medical centers. However, since we are studying the genomics/transcriptomics/epigenomics of the tumors, rather than any clinical outcomes, we believe this would minimally impact our study.
Ethics oversight	UCSF Institutional Review Board

Note that full information on the approval of the study protocol must also be provided in the manuscript.