nature genetics

Article

https://doi.org/10.1038/s41588-025-02128-y

The landscape of N^6-methyladenosine in localized primary prostate cancer

Received: 10 February 2023

Accepted: 13 February 2025

Published online: 24 March 2025

Check for updates

Xin Xu[1,24], Helen Zhu[1,2,3,24], Rupert Hugh-White[4,5,6,7,24], Julie Livingstone[4,5,6,7,24], Stefan Eng[4,5,6,7], Nicole Zeltser[4,5,6,7], Yujuan Wang[1], Kinga Pajdzik[8,9], Sujun Chen[1,2,21], Kathleen E. Houlahan[2,3,4,5,6,7], Wenqin Luo[1,10], Shun Liu[8,9], Xi Xu[1], Minzhi Sheng[2,11], Wang Yuan Guo[1], Jaron Arbet[4,5,6,7], Yuxi Song[4,5,6,7], Miranda Wang[1], Yong Zeng[1], Shiyan Wang[1,22], Guanghui Zhu[1,2,21], Tingxiao Gao[1,2], Wei Chen[1,23], Xinpei Ci[1], Wenjie Xu[12], Kexin Xu[13], Michele Orain[14], Valerie Picard[14], Helene Hovington[14], Alain Bergeron[14], Louis Lacombe[14], Bernard Têtu[14], Yves Fradet[14], Mathieu Lupien[1,2], Gong-Hong Wei[12,15], Marianne Koritzinsky[1,2], Robert G. Bristow[16,17], Neil E. Fleshner[1], Xue Wu[18], Yang Shao[18,19], Chuan He[8,9], Alejandro Berlin[1], Theodorus van der Kwast[1], Hon Leong[2,11], Paul C. Boutros[2,3,4,5,6,7,20] ✉ & Housheng Hansen He[1,2] ✉

N^6-methyladenosine (m^6A), the most abundant internal RNA modification in humans, regulates most aspects of RNA processing. Prostate cancer is characterized by widespread transcriptomic dysregulation; therefore, we characterized the m^6A landscape of 162 localized prostate tumors with matched DNA, RNA and protein profiling. m^6A abundance varied dramatically across tumors, with global patterns emerging via complex germline–somatic cooperative regulation. Individual germline polymorphisms regulated m^6A abundance, cooperating with somatic mutation of cancer driver genes and m^6A regulators. The resulting complex patterns were associated with prognostic clinical features and established the biomarker potential of global and locus-specific m^6A patterns. Tumor hypoxia dysregulates m^6A profiles, bridging prior genomic and proteomic observations. Specific m^6A sites, such as those in *VCAN*, drive disease aggression, associating with poor outcomes, tumor growth and metastasis. m^6A dysregulation is thus associated with key events in the natural history of prostate cancer: germline risk, microenvironmental dysregulation, somatic mutation and metastasis.

Cancer is a disease of the genome, driven by somatic mutations[1]. The cancer genome is actuated by the cellular machinery of the central dogma, leading to transcriptomic and proteomic dysregulation and eventual clinical presentation. The proteomic and clinical features of cancer are not fully explained by the genome and the transcriptome, suggestive of widespread post-transcriptional and post-translational regulation[2,3]. Chemical modification of RNA is a leading candidate to explain these discrepancies[4].

m^6A is the most abundant internal chemical modification of RNA in humans. This epitranscriptomic mark regulates almost every facet of the RNA life cycle, including secondary structure formation, splicing, nuclear export, translation and degradation[5]. At least 28 enzymes read, write or erase m^6A, facilitating dynamic transcript regulation[6]. Three lines of evidence suggest that m^6A drives tumor initiation and progression[4]. First, global m^6A abundance is altered relative to that of normal tissue in a cancer type-specific manner[7,8]. Second, mutations

A full list of affiliations appears at the end of the paper. ✉e-mail: pboutros@mednet.ucla.edu; hansenhe@uhnresearch.ca

in m^{6}A enzymes are common in many cancer types, and their perturbation influences tumor phenotypes in multiple model systems[9–13]. Third, specific m^{6}A sites in oncogenes and tumor suppressors influence protein abundance via transcript stability and translation rates[11,14,15].

Here, we mapped m^{6}A modifications transcriptome wide in 162 intermediate-risk localized prostate tumors with extensive prior molecular, microenvironmental and clinical characterization[2,16–21]. m^{6}A methylation defined clinically distinct tumor subgroups and influenced specific signaling pathways, with cross-talk between somatic mutational and epitranscriptomic profiles. m^{6}A sites interacted with clinically relevant germline genotypes, with 1,350 m^{6}A quantitative trait loci (m^{6}A-QTLs) identified. Similarly, both global and specific m^{6}A information reflected underlying tumor hypoxia, which is a strong prognostic factor. Consistent with this, m^{6}A peaks predicted multiple clinical phenotypes, including patient survival. Finally, we demonstrate that specific, targetable m^{6}A modifications functionally drive tumor growth and progression.

Results

The m^{6}A landscape of localized prostate cancer

We quantified the epitranscriptomes of 162 primary localized prostate tumors using methylated RNA immunoprecipitation sequencing (meRIP-seq)[22–24]. All patients were clinically managed for National Comprehensive Cancer Network intermediate-risk disease via radical prostatectomy with curative intent[25]. Clinical annotation included pretreatment prostate-specific antigen (PSA) abundance, tumor size and extent (T category), surgical International Society of Urological Pathology (ISUP) Grade Group, biochemical relapse, subhistologies and age at diagnosis (Fig. 1a and Supplementary Table 1). Tumors were treatment naive, and samples were taken from index lesions as assessed by two uropathologists. Median follow-up was 6.72 years. Samples were profiled with multiple additional assays including tumor–reference whole-genome sequencing[16–18,21], DNA methylation and histone H3 lysine 27 (H3K27) acetylation (H3K27ac) profiling[16,18,20], RNA-seq[19] and proteomics[2] (Fig. 1a and Extended Data Fig. 1a).

For each sample, two libraries were created and sequenced: an immunoprecipitation (IP) library generated by RNA IP with an anti-m^{6}A antibody and an input library generated from the total RNA pool. After alignment and quantitation (Extended Data Fig. 1b), stringent quality control led to the exclusion of 14 samples (Extended Data Fig. 1c). In the final 148-patient cohort, input library results were strongly positively correlated with prior deep transcriptome sequencing (median Spearman's $\rho = 0.87$; Extended Data Fig. 1d)[19]. Germline variants identified from input libraries validated sample identity (Extended Data Fig. 1e)[18], and metrics from the ENCODE Consortium and other prior m^{6}A studies were within expected ranges (Extended Data Fig. 1f)[26].

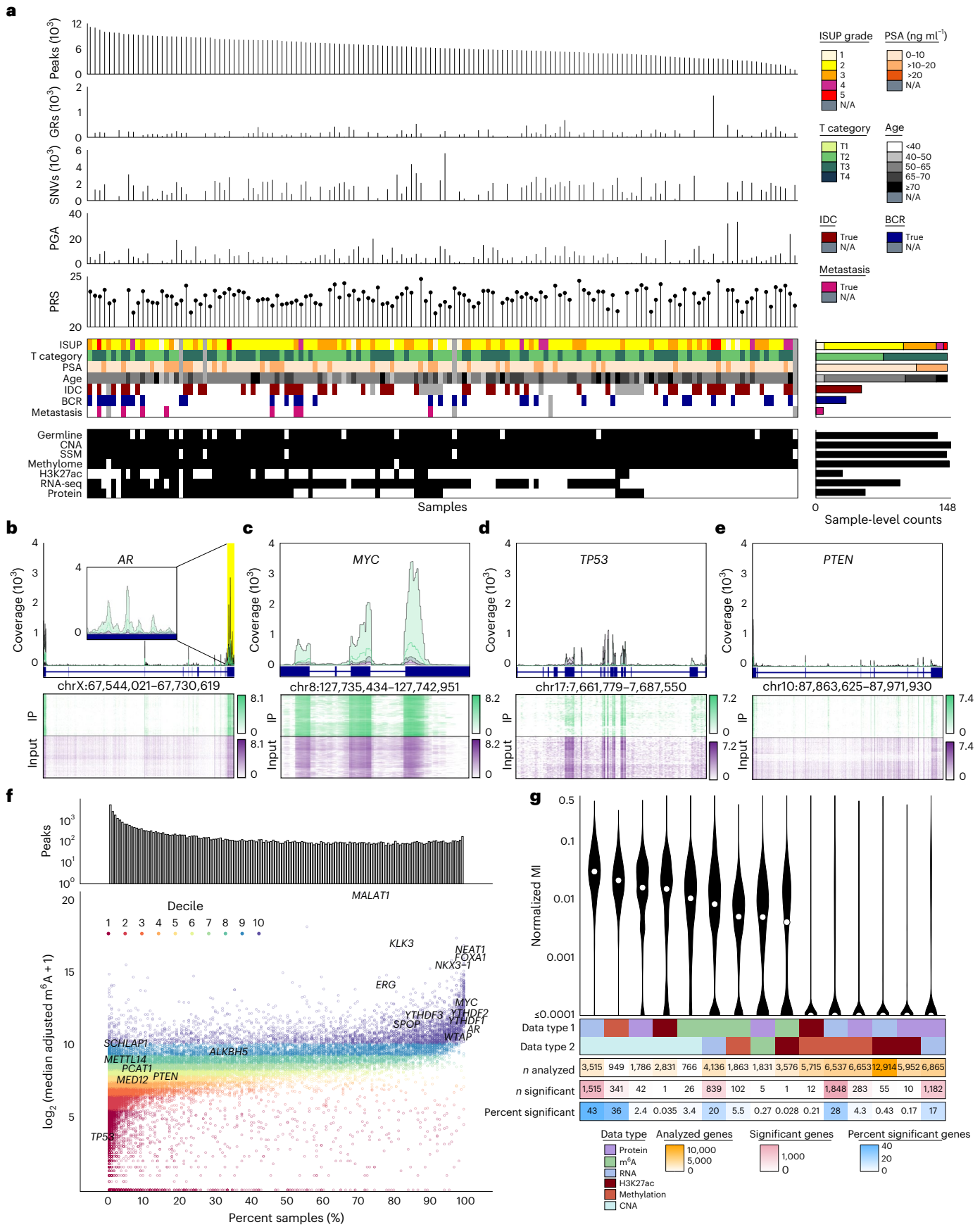
m^{6}A peaks were identified with both a highly specific algorithm (MeTPeak) and a highly sensitive one (exomePeak; Extended Data Fig. 2a)[27,28]. The number of peaks detected was modestly correlated with IP library size (Extended Data Fig. 2b), but read downsampling showed that libraries reached or approached saturation (Extended Data Fig. 2c,d)[29]. The vast majority of identified peaks overlapped known peaks (Extended Data Fig. 2e, inner panel)[26], and the canonical RRACH motif was enriched in peaks identified in every sample[30]. Downstream analyses used high-specificity MeTPeak results (Fig. 1a and Extended Data Fig. 2e, outer panel).

To rationalize peaks cohort wide, we created an algorithm to integrate sample-level peaks into cohort-level joint peaks called HistogramZoo[31]. Briefly, peaks for each sample were aggregated onto transcript backbones, forming a coverage histogram for each transcript. These histograms were then segmented to create a set of joint peaks[32–35]. m^{6}A methylation was quantitated for each sample on each peak as the number of IP reads, normalized for library size across the cohort[36] and then adjusted for transcript abundance using matched input libraries[37]. In the full 148-patient cohort, HistogramZoo (Methods) identified 32,051 high-confidence m^{6}A peaks across 9,571 genes; these were specifically enriched near stop codons (Extended Data Fig. 2f). The median tumor harbored 7,611 ± 2,922 m^{6}A peaks (s.d.). Several key prostate cancer oncogenes such as *MYC*, *AR*, *MALAT1* and *FOXA1* had frequently methylated m^{6}A peaks (Fig. 1b,c and Extended Data Fig. 2g,h), while tumor suppressors like *TP53*, *PTEN* and *RB1* were very infrequently methylated (Fig. 1d,e and Extended Data Fig. 2i). One notable exception to the trend is *NKX3-1*, which was frequently methylated across samples (Extended Data Fig. 2j). About 16% of peaks were identified in a single sample (5,146 peaks), and ~20% were identified in at least half (6,467 peaks), with only 0.5% (167 peaks) detected in every sample (Fig. 1f).

To ascertain how m^{6}A influences other aspects of the central dogma, we used mutual information, which can capture complex associations without assumptions of linearity or monotonicity[38]. For every gene, mutual information was calculated between each pair of molecular characteristics (Extended Data Fig. 1a). The number of genes with significant mutual information varied widely between different types of data (Fig. 1g and Supplementary Table 2). m^{6}A was associated with RNA abundance for ~20% of analyzed genes (839 genes), with less frequent associations with copy number alterations (CNAs) (26 genes), DNA methylation (102 genes) and H3K27ac (one gene) (Fig. 1g). Genes with very low RNA abundance displayed fewer samples with a peak (Extended Data Fig. 2k), likely due to limitations in RNA signature detection. Similarly, a significant negative correlation was detected between m^{6}A and RNA abundance (Extended Data Fig. 2l). The extent to which RNA explained protein abundance was weakly associated with mean gene-level m^{6}A abundance but not with the number of samples for which the m^{6}A peak was detected (Extended Data Fig. 2m,n). This provides minor support for previously reported regulatory roles of m^{6}A in modulating RNA translation into protein[23,39].

Fig. 1 | The m^{6}A landscape of localized prostate cancer. a, Top to bottom, bar plots. The number of MeTPeak peaks across samples ordered from greatest to smallest. The number of genomic rearrangements (GRs) identified in each sample. The number of single-nucleotide variants (SNVs) identified in each sample. The PGA[40] as a proxy of the total CNAs for each sample. The Conti PRS[50] calculated for each sample. Top heatmap shows the associated clinical covariates including ISUP Grade Group, anatomical T category, PSA, age, IDC/CA, biochemical relapse (BCR) and metastasis. Bottom heatmap, complementary molecular profiling data collected for each sample including germline polymorphisms in samples of European ancestry[18], somatic CNAs and simple somatic mutations[17], DNA methylome[18], H3K27ac[20], ultra-deep RNA-seq[19] and proteomics[2]. SSM, simple somatic mutation; N/A, not available. **b**–**e**, Exemplar plots for *AR* (the inset magnifies the region highlighted in yellow; **b**), *MYC* (**c**), *TP53* (**d**) and *PTEN* (**e**). In the top polygon plots, the median IP (green) and input (purple) coverage (reads per kilobase per million mapped reads-normalized bigWig files) is represented using lines, while background colors represent the range of IP and input coverage across samples. Exons identified in GENCODE version 34 are annotated below in dark blue. Bottom heatmaps represent the distribution of IP and input coverage (log1p transformed) across samples (*y* axis), and darker colors correspond to greater read coverage. Samples are clustered using Euclidean distance and Ward's minimum variance method with squared distances. chr, chromosome. **f**, Top bar plot, the number of peaks uniquely found in a given number of samples. Bottom scatterplot, the median adjusted m^{6}A abundance of a joint peak (*y* axis) versus the number of samples in which the peak is identified (*x* axis). Colors indicate the deciles of the adjusted m^{6}A abundance. Multiple joint peaks can be identified in a single gene, but the most prevalent and abundant peaks for a given gene are labeled where applicable. **g**, Distribution of the normalized mutual information (MI) for each data type pair. For visualization, values < 10^{-4} are shown as 10^{-4}. Top heatmap indicates data type pair. Subsequent heatmaps show the number of genes with data available for each data type pair, numbers of genes for which mutual information is significant ($Q < 0.1$) and percentage of significant genes.

https://doi.org/10.1038/s41588-025-02128-y



https://doi.org/10.1038/s41588-025-02128-y

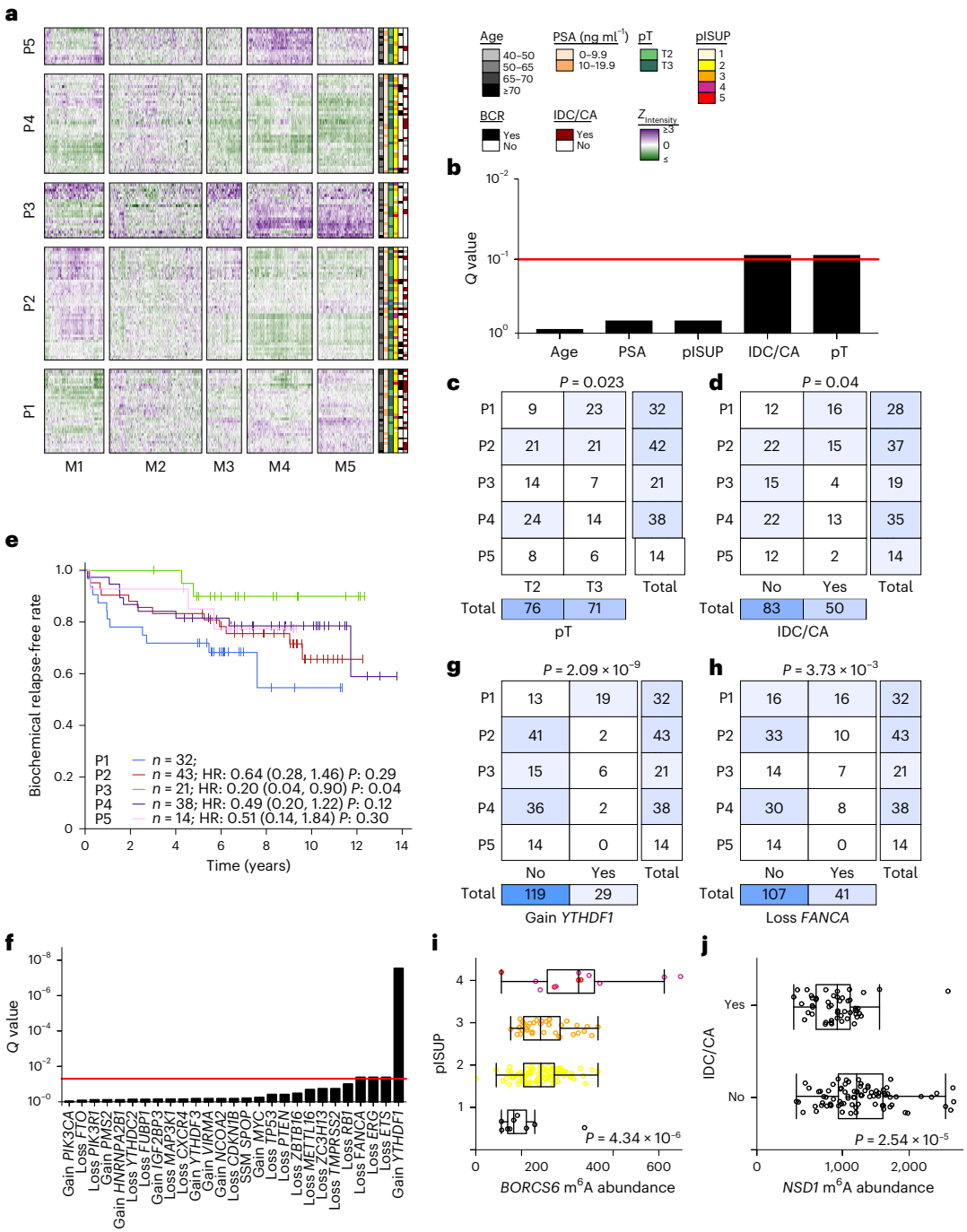


Fig. 2 | Molecular subtyping and clinical correlates of m⁶A. a, Clustering of the top quartile of interquartile range (IQR)-ranked m⁶A peaks (n = 5,203) identified five patient subtypes and five m⁶A subtypes. Clinical covariates are shown to the right of the heatmap. Heatmap coloring indicates z-score-scaled peak intensities. **b**, Association between m⁶A patient subtypes and clinical features. Q values are from Pearson's χ^2 test for categorical variables and a one-way ANOVA for continuous variables. **c**,**d**, Overlap of m⁶A patient clusters with pathologic tumor extent (pT) (**c**) and IDC/CA status (**d**). Patient clusters are indicated in rows, and clinical features are shown in columns. Row and column totals are depicted in the right and bottom heatmaps. Independence of clusters and clinical variables was assessed via Pearson's χ^2 test (P value shown). The relative intensity of blue shading indicates the size of the group. **e**, Biochemical relapse rate across the five m⁶A patient subtypes. A Cox proportional hazards model was fit with P1 as the baseline group. Hazard ratios (HR) and P values are shown with confidence intervals in parentheses. **f**–**h**, Association between m⁶A patient subtypes and mutations in m⁶A regulators. **f**, Bar plot shows Q values from Pearson's χ^2 test. ETS, E26 transformation-specific (ETS) transcription factor family members. **g**, Overlap of m⁶A patient clusters with presence of copy number gain of *YTHDF1*. **h**, Overlap of m⁶A patient clusters with presence of copy number gain of *FANCA*. Patient clusters are indicated in rows, and mutations are shown in columns. Row and column totals are depicted in the right and bottom heatmaps. Independence of clusters and mutations was assessed via Pearson's χ^2 test (P value shown). The relative intensity of blue shading indicates the size of the group. **i**, *BORCS6* m⁶A peak abundance varies by pISUP Grade Group. P value from one-way ANOVA is shown (n = 148). Box plots represent the median (center line) and upper and lower quartiles (box limits), and whiskers extend to the minimum and maximum values within 1.5× the IQR. **j**, *NSD1* m⁶A peak abundance is greater with presence of IDC/CA. P value from two-sided U-test is shown; samples where IDC status was missing have been removed (n = 133). Box plots represent the median (center line) and upper and lower quartiles (box limits), and whiskers extend to the minimum and maximum values within 1.5× the IQR.

https://doi.org/10.1038/s41588-025-02128-y

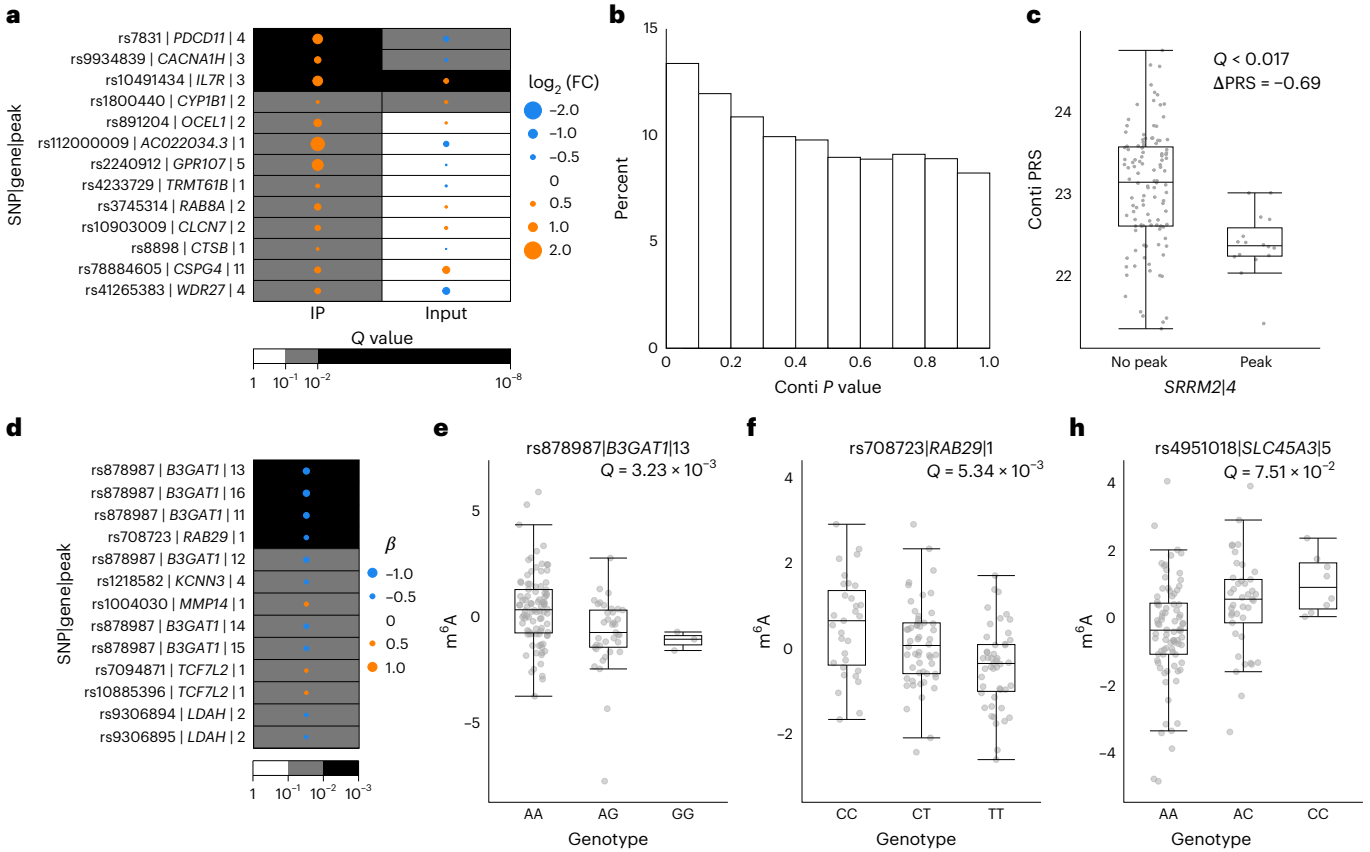


Fig. 3 | Germline correlates of m^{6}A in primary prostate cancer. a, Allelic imbalance of reads at SNPs in annotated m^{6}A sites in heterozygous samples in either the IP or input library. Allelic imbalance is evaluated using a paired *t*-test at a statistical threshold of $Q < 0.1$. Effect size ($\log_2$ (fold change (FC))) is calculated with respect to the A allele. Effect size and direction are represented by the size and color of the disks, respectively. Statistical significance is represented by the background color. **b**, Distribution of *P* values in associations of m^{6}A peak status with Conti PRS (*t*-test)[50]. **c**, m^{6}A methylation of *SRRM2* (peak 4) is associated with a lower Conti PRS (*t*-test). Box plots represent the median (center line) and upper and lower quartiles (box limits), and whiskers extend to the minimum and maximum values within 1.5× the IQR. $n = 133$. **d**, Significant prostate cancer risk SNP local m^{6}A-QTLs identified using a linear additive model. Effect size (β) and direction are represented by the size and color of the disks, respectively. Statistical significance is represented by the background color. **e**,**f**, Two local risk SNP quantitative trait loci are depicted: the associations of genotype with m^{6}A methylation for *B3GAT1* (rs878987 | peak 13) (**e**) and *RAB29* (rs708723 | peak 1) (**f**), respectively. Box plots represent the median (center line) and upper and lower quartiles (box limits), and whiskers extend to the minimum and maximum values within 1.5× the IQR. $n = 133$. **g**, Manhattan plot of genome-wide m^{6}A-QTL analysis. Results from a linear additive model implemented in Matrix eQTL[83]. Points representing the tag SNPs of m^{6}A-QTLs where the gene has been annotated in the Cancer Gene Census[56] or by Armenia et al.[57], Fraser et al.[16] and Quigley et al.[58] are indicated with black squares, while the SNPs of significant ($Q < 0.1$) m^{6}A-QTLs where the SNP has been annotated in RMVar[43] or by Cotter et al.[44] are indicated with black diamonds. Labeled black disks selectively identify top hits that fall into neither of the former categories. SLC9A3R2 (NHERF2). **h**, A genome-wide significant m^{6}A-QTL: SNP rs4951018 with *SLC45A3* m^{6}A abundance. Box plots represent the median (center line) and upper and lower quartiles (box limits), and whiskers extend to the minimum and maximum values within 1.5× the IQR. $n = 133$.

In addition to these 148 prostate tumor samples, we profiled benign tissue from seven individuals to allow tumor–normal comparisons (Supplementary Fig. 1 and Supplementary Tables 3 and 4). Furthermore, m^{6}A-selective allyl chemical labeling and sequencing was applied to eight samples, facilitating single-nucleotide-resolution m^{6}A profiling (Supplementary Fig. 2 and Supplementary Table 5). The analysis of these data is presented in the Supplementary Note.

m^{6}A subtypes of localized prostate cancer

To identify global trends in m^{6}A variation across patients, consensus clustering was performed on peak-level m^{6}A abundance. We identified five patient subtypes (Supplementary Fig. 3a; P1 through P5) and five m^{6}A subtypes (Supplementary Fig. 3b,c; M1 through M5; Fig. 2a and Supplementary Tables 6–8). m^{6}A-derived patient subtypes were associated with multiple clinico-molecular features (Fig. 2b), including tumor size and extent (pathologic T category (pT); Fig. 2c), presence of the aggressive intraductal carcinoma and cribriform architecture subhistologies (IDC/CA; Fig. 2d), genomic instability assessed as the proportion of the genome with a CNA (percent genome altered (PGA)[40]; Supplementary Fig. 3d) and relapse rate after surgery (Fig. 2e; unadjusted *P* values shown) but not with tumor purity (Fig. 2a and Supplementary Fig. 3e).

m^{6}A subtypes were related to known messenger RNA (mRNA) and CNA subtypes (Supplementary Fig. 3f,g), which themselves are associated with genomic instability and patient outcome[16,19]. The m^{6}A subtypes were characterized by patterns of somatic CNAs in m^{6}A writers, readers and erasers (Fig. 2f) including gain of *YTHDF1* (Fig. 2g) and loss of *FANCA* (Fig. 2h). Thus, there are multiple m^{6}A subtypes with distinct clinical presentations, which partially overlap previous DNA and RNA subtypes.

Clinical correlates of locus-specific m^{6}A

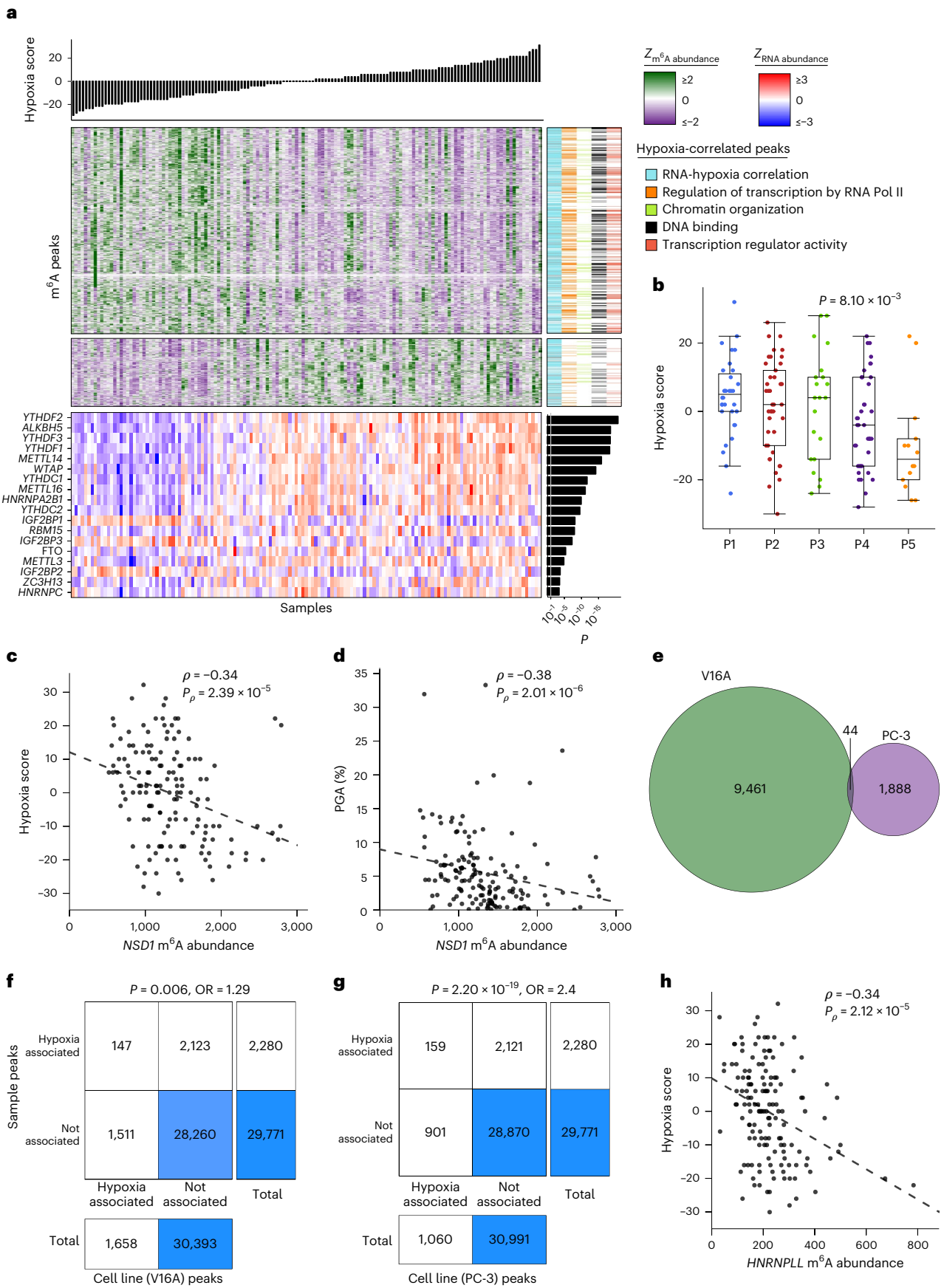
Several clinical features of prostate cancer influence patient management, including age at diagnosis, pathological tumor extent (pT), pretreatment serum PSA abundance, pathological ISUP (pISUP) Grade Group and aggressive intraductal carcinoma subhistology (IDC/CA). For each clinical indicator, we identified individual m^{6}A peaks associated with their status ($Q < 0.1$; Extended Data Fig. 3a). Six m^{6}A peaks were associated with age (Spearman's rank correlation, $Q < 0.1$; Extended Data Fig. 3b and Supplementary Table 9), eight peaks were associated with grade (one-way ANOVA, $Q < 0.1$; Fig. 2i), and 13 were associated with IDC/CA (two-sided Mann–Whitney *U*-test, $Q < 0.1$). These included *NSD1*, which enhances *AR* transactivation (Fig. 2j)[41]. The number of m^{6}A peaks in a tumor was elevated in IDC/CA-positive disease (Extended Data Fig. 3c,d), and loss of the tumor suppressor genes *FANCA* and *TP53* was more common in patients with more peaks (two-sided Mann–Whitney *U*-test, $P < 0.05$; Extended Data Fig. 3e). Thus, locus-specific m^{6}A patterns reflect clinical features of prostate cancer.

Heritable tumor-specific m^{6}A regulation

Given that prostate cancer is the most heritable solid cancer[42], we next considered whether germline genetics influence m^{6}A methylation. First, we examined the potential *cis* regulatory effects of germline polymorphisms on m^{6}A. Across all patients, around six million unique SNPs were identified[18]. Of these, 4,755 were located at the specific residue of a previously annotated m^{6}A site[43,44]. We identified 35 germline polymorphisms that (1) overlapped an m^{6}A peak identified in this cohort, (2) occurred at an m^{6}A annotated site and (3) had an A allele on the transcribed strand (Extended Data Fig. 4a). A alleles in active m^{6}A sites were present at much higher frequencies than non-A alleles (median_{A}, 0.83; range_{A}, 0.11–0.95; $\text{median}_{\text{non-A}}$, 0.45; $\text{range}_{\text{non-A}}$, 0.053–0.95) (Extended Data Fig. 4b). A lower A allele frequency was observed at inactive m^{6}A sites: that is, annotated sites lacking an m^{6}A peak in localized prostate cancer (Extended Data Fig. 4b). This imbalance suggests that non-A alleles are associated with reduced m^{6}A abundance. A quantitative analysis identified many SNPs associated with allelic imbalance of IP and/or input reads at the m^{6}A site (Fig. 3a; paired *t*-test, $Q < 0.1$). These were often correlated with clinical features. For example, a SNP in *GPR107*

Fig. 4 | Tumor hypoxia is associated with widespread m^{6}A modulation. **a**, Hypoxia correlates with m^{6}A peak abundance. Top bar plot indicates sample hypoxia score. Left–middle heatmaps show the corresponding *z*-scored m^{6}A peak abundance per sample for peaks with a hypoxia correlation (Spearman's correlation, $Q < 0.05$, 1,184 peaks). Peaks are clustered by abundance values, and the two largest resulting clusters are displayed in separate heatmaps. Right–middle heatmaps display m^{6}A peak annotations. 'RNA–hypoxia correlation' indicates whether RNA abundance for the corresponding genes is also correlated with hypoxia. Hypoxia-correlated peaks are enriched for several biological pathways, and subsequent columns indicate whether each corresponding gene is tagged with each of the enriched pathway terms. Bottom heatmap shows *z*-scored RNA abundance for m^{6}A writers, readers and erasers. Right bar plot shows *P* values for Spearman's correlation between RNA abundance and hypoxia. Pol, polymerase. **b**, Patient m^{6}A subtypes are associated with varying hypoxia levels. *P* value from one-way ANOVA. Box plots represent the median (center line) and upper and lower quartiles (box limits), and whiskers extend to the minimum and maximum values within 1.5× the IQR. $n = 146$. **c**,**d**, Associations between m^{6}A abundance for an example peak on *NSD1*. Hypoxia ($n = 146$) (**c**), PGA ($n = 148$) (**d**). Plots annotated with results from Spearman's correlation. **e**, Overlap between hypoxia-specific peaks in V16A and PC-3 cells. **f**,**g**, Patient hypoxia-associated peaks significantly overlap with hypoxia-specific peaks in prostate cancer cell lines. Contingency tables showing results from two-sided Fisher's exact test for association between patient hypoxia-associated peaks and V16A (**f**) and PC-3 (**g**) hypoxia-specific peaks. The relative intensity of blue shading indicates the size of the group. OR, odds ratio. **h**, Correlation between tumor m^{6}A abundance and hypoxia for a peak region in *HNRNPLL*. This peak shows a corroboratory hypoxia association in both V16A and PC-3 cell lines. Plots annotated with results from Spearman's correlation ($n = 146$).

https://doi.org/10.1038/s41588-025-02128-y



was associated with biochemical recurrence (Cox proportional hazards model, hazard ratio = 13.8; confidence interval, 3.2–60.3; $Q = 6.3 \times 10^{-3}$; ExaLT $P = 2.0 \times 10^{-2}$; Extended Data Fig. 4c). rs2240912 is neither a hit in prior GWAS studies[45–48] nor in linkage disequilibrium with a GWAS SNP from those studies, likely due to the low minor allele frequency. Validation of its molecular effects and clinical consequences will be needed to better estimate its associations with disease aggression.

Next, to understand potential epitranscriptomic alterations characteristic of prostate cancer incidence, we evaluated the association of m^{6}A peaks with two polygenic risk scores (PRSs) for incidence[49,50] and one for hazard (PHS290)[51] (Fig. 3b, Extended Data Fig. 4d–f and Supplementary Table 10). The number of m^{6}A peaks across samples was weakly associated with polygenic risk (Fig. 1a and Extended Data Fig. 4g). The *P*-value distribution suggested that many subthreshold associations of m^{6}A peaks with the incidence PRS remain to be elucidated in larger cohorts (Fig. 3b), and an m^{6}A site in *SRRM2* was associated with genetic risk (Fig. 3c and Supplementary Table 11).

We then related the 272 individual risk SNPs to m^{6}A peaks within 10 kbp and to more distal m^{6}A peaks identified using three-dimensional spatial genomic data[52] (Extended Data Fig. 4h,i). We identified 13 local prostate risk SNP m^{6}A tags, including six associated with peaks in *B3GAT1* and one with a peak in *RAB29* (Fig. 3d–f and Supplementary Table 12).

To generalize this observation, we identified m^{6}A-QTLs genome wide, comparing each m^{6}A peak to all SNPs within 10 kbp. This yielded 14,775 significant m^{6}A-QTLs representing 1,350 unique peaks ($Q < 0.1$; Fig. 3g and Supplementary Table 13). Eleven percent (151 of 1,350) of the m^{6}A-QTL SNPs overlap with the peaks they are associated with (Extended Data Fig. 4j), in concordance with previous studies showing that most of the m^{6}A-QTLs are not located within m^{6}A peaks[53,54]. Among these SNPs, 1% (14 of 1,350) overlap with methylated 'A' sites in the peaks (Supplementary Table 13). For SNPs that are located in regulatory regions, applying publicly available[55] and in-house H3K27ac high-throughput chromosome conformation capture with chromatin immunoprecipitation (HiChIP) data, we found evidence of physical interactions between the m^{6}A-QTL SNPs and the associated peaks in 4% (60 of 1,350) of SNP–peak pairs (Supplementary Table 13). These m^{6}A-QTLs affected 1,017 genes, including 60 cancer or prostate cancer driver genes[16,56–59]. Of the 1,350 m^{6}A-QTLs, 101 were associated with RNA abundance changes of the corresponding gene, nine with protein abundance changes and four with both ($Q < 0.1$; Extended Data Fig. 4k). For example, rs4951018 was associated with changes in m^{6}A, RNA and protein abundances of *SLC45A3*, a component of the common prostate cancer fusion gene *SLC45A3-ELK4* (Fig. 3h and Extended Data Fig. 4l–n). Three independent m^{6}A-QTL SNPs (rs143089027, rs57557217 and rs76338659) were associated with pISUP Grade Group (Pearson's χ^2 test, $Q < 0.1$; Extended Data Fig. 4o). In sum, these data reveal broad germline regulation of m^{6}A sites, including on cancer driver genes.

Broad m^{6}A dysregulation by tumor hypoxia

Hypoxia is an adverse prognostic feature in localized prostate cancer associated with grade, IDC/CA and disease relapse[17]. We identified 2,280 hypoxia-associated m^{6}A peaks (Spearman's rank correlation, $Q < 0.05$; Fig. 4a and Supplementary Table 14). Hypoxia-associated peaks tended to be less abundant in hypoxic tumors and preferentially occurred on genes involved in gene expression regulation (Fig. 4a). The RNA abundance of m^{6}A regulators was also correlated with hypoxia (Spearman's rank correlation, $P < 0.1$; Fig. 4a). Further highlighting the widespread nature of the hypoxia–m^{6}A relation, m^{6}A patient subtypes (Fig. 2a) were hypoxia associated, with the mutationally quiet P5 cluster being the most normoxic (one-way ANOVA, $P = 8.1 \times 10^{-3}$; Fig. 4b). Eight peaks showed additional mutational or clinical associations, including a peak located on *NSD1* (Figs. 2j and 4c,d and Supplementary Table 9).

To validate these hypoxia-associated m^{6}A peaks, we exposed two prostate cancer cell lines (V16A and PC-3) to 0.2% O_2 for 24 h and profiled m^{6}A using the same approach. Both cell lines had many hypoxia-specific peaks (Fig. 4e and Supplementary Table 15), with these peaks significantly overlapping the hypoxia-responsive m^{6}A peaks in primary patients (Fig. 4f,g). Nine peaks showed a consistent hypoxia correlation across both cell lines in addition to patient samples, including *HNRNPLL* (Fig. 4h). Thus hypoxia correlates with m^{6}A profiles in patients and model systems.

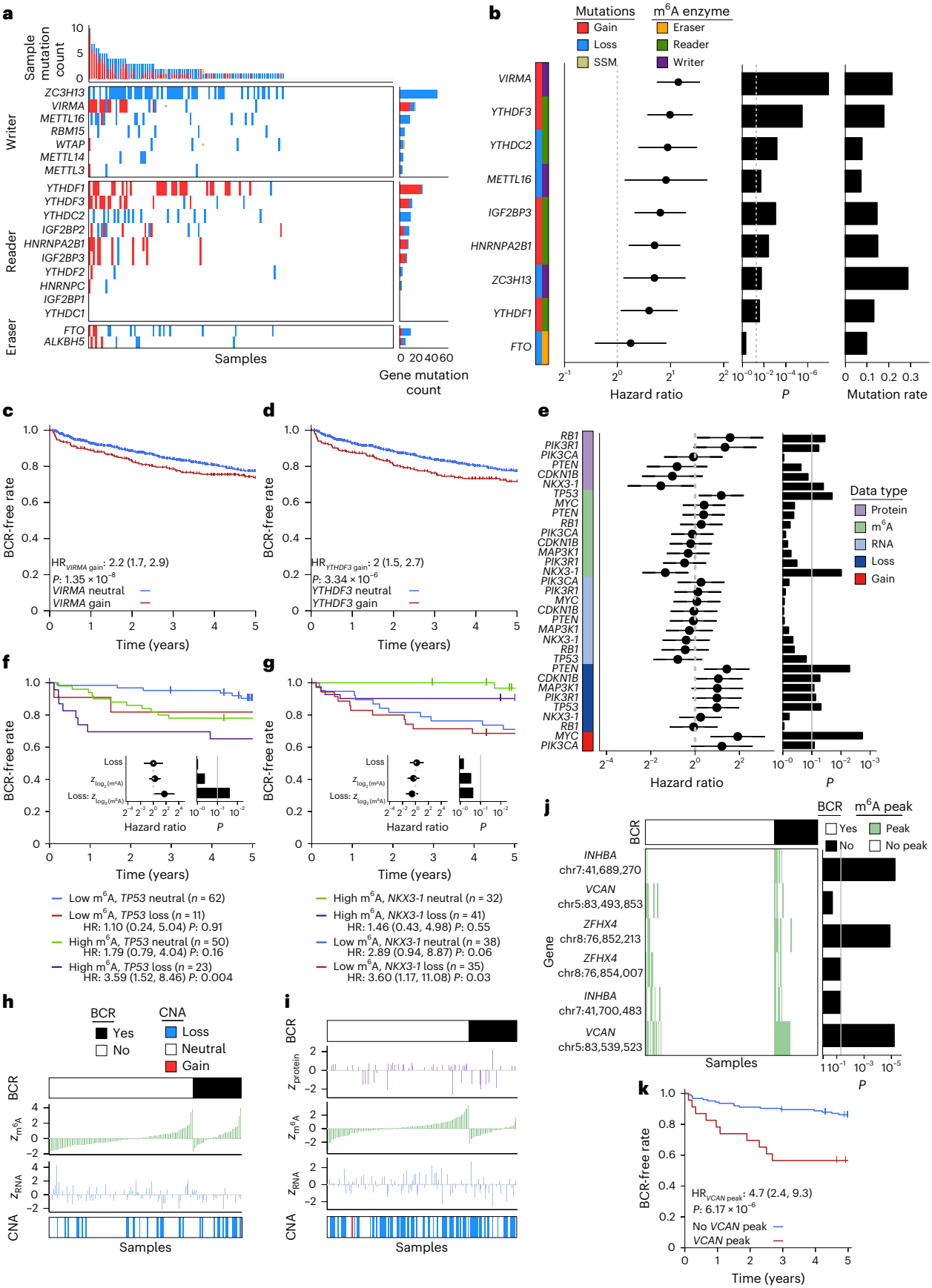
m^{6}A is a biomarker of prostate cancer patient outcome

We next sought to evaluate the biomarker potential of m^{6}A. The enzymes that read, write or erase m^{6}A are frequently altered by somatic mutations, with a majority of tumors having a mutation affecting an m^{6}A enzyme (Fig. 5a and Supplementary Table 16). Copy number loss of these genes was accompanied by reduced RNA and protein abundance in several instances (Extended Data Fig. 5a), and these modulations were associated with pISUP Grade Group for loss of *METTL16* and gain of *VIRMA*, *HNRNPA2B1* and *YTHDF3* (Extended Data Fig. 5a). A majority (seven of nine) of these associations were replicated in the Cancer Genome Atlas (TCGA) Prostate Adenocarcinoma (PRAD) cohort ($N = 421$) (Supplementary Table 17). Similarly, RNA abundance of *HNRNPA2B1* and *YTHDC2* was also associated with ISUP Grade Group (Extended Data Fig. 5b).

To further investigate the clinical importance of these m^{6}A regulator mutational profiles, we analyzed six independent cohorts comprising 1,239 patients[16,17,59–62]. CNAs were significantly predictive of disease relapse (Fig. 5b), with the strongest effect sizes being observed for gain of the m^{6}A regulators encoded by *VIRMA* and *YTHDF3* (Fig. 5b–d and

Fig. 5 | Quantifying the biomarker potential of m^{6}A sites. a, m^{6}A regulatory genes are frequently mutated. m^{6}A enzyme-encoding genes are shown on the *y* axis, and patient samples are on the *x* axis. Top bar plot represents the number of mutations per sample, while the right bar plot shows the number of mutations in each m^{6}A regulatory gene across samples. Central heatmaps display the mutations in each enzyme for each sample. **b**–**d**, Meta-analysis of six independent patient cohorts identifies mutations in m^{6}A regulatory genes as predictive of patient biochemical recurrence. **b**, Left heatmap depicts m^{6}A regulatory gene, mutation type and category of m^{6}A regulatory action. Forest plot shows hazard ratios and 95% confidence intervals of each mutation on biochemical recurrence, and left bar plot indicates corresponding *P* values from a Cox model. The far right bar plot displays average sample mutation rate across the patient cohorts (protein, $n = 54$, m^{6}A, $n = 148$; RNA, $n = 92$; gain or loss, $n = 146$). **c**,**d**, Gain of *VIRMA* (**c**) and gain of *YTHDF3* (**d**) predict decreased biochemical recurrence-free rate. *VIRMA* gain, $n = 22$; loss, 191. *YTHDF3* gain, $n = 23$; loss, 186. *P* values displayed are from Cox proportional hazards models. **e**–**i**, m^{6}A provides complementary prognostic information in profiling prostate cancer driver genes. **e**, Left heatmap shows driver genes and molecular data type analyzed. Forest plot shows hazard ratio and 95% confidence intervals for biochemical recurrence, and the right bar plot indicates corresponding *P* values from a Cox model. **f**,**g**, Influence of m^{6}A and copy loss on biochemical recurrence-free rate for tumor suppressors *TP53* (**f**) and *NKX3-1* (**g**). Kaplan–Meier survival curves show the results from median dichotomizing m^{6}A abundance to create four patient groups, while inset forest plots show results from treating m^{6}A as a continuous variable. *P* values displayed are from Cox proportional hazards models. **h**,**i**, Relative m^{6}A abundance by biochemical recurrence status for *TP53* (**h**) and *NKX3-1* (**i**). Plots show copy number, *z*-scored RNA, m^{6}A and protein abundance, and biochemical recurrence status at censoring time. Patients are depicted in columns and are ordered by biochemical recurrence status followed by m^{6}A abundance. **j**,**k**, m^{6}A peak status predicts biochemical recurrence for specific peak sites on *INHBA*, *VCAN* and *ZFHX4*. **j**, Patients are represented in columns. Top heatmap displays biochemical recurrence status at censoring time. Middle heatmap shows patient peak status (peak sites depicted in row labels). Right bar plot shows *P* values from a Cox model. **k**, Presence of *a VCAN* m^{6}A peak increases risk of biochemical recurrence. *VCAN* peak, $n = 23$; no peak, 125. *P* values displayed are from Cox proportional hazards models.

https://doi.org/10.1038/s41588-025-02128-y



Supplementary Table 18). Furthermore, CNAs in prognostic m^{6}A regulatory genes co-occurred with spatially distant prostate cancer driver events, providing evidence of positive selection (Extended Data Fig. 5c). Of these significantly co-occurring driver–m^{6}A CNA events, 52% (29 of 56) replicated in the TCGA-PRAD cohort (Supplementary Table 19).

Mutations in m^{6}A enzymes may determine patient outcomes through direct modification of site-specific m^{6}A. We investigated whether these mutations, in addition to canonical prostate cancer driver events, were associated with differential methylation of m^{6}A peaks. Of the 6,467 m^{6}A peaks identified in at least 50% of samples, over half (3,432 peaks) were differentially methylated in relation to a prostate cancer driver or m^{6}A regulator, with clustering by effect size resulting in five peak clusters and five mutation clusters (MC; Extended Data Fig. 5d and Supplementary Table 20). A subset of m^{6}A clustering reflects known mutational co-occurrence, such as for *YTHDF1* gain and *FANCA* loss (MC4, Extended Data Fig. 5c,d). In many cases, unrelated somatic mutations yielded similar m^{6}A profiles, suggesting convergent epitranscriptomic dysregulation (for example, *ZC3H13* and *SPOP* (MC5)). Co-regulated peaks were enriched for specific biological functions. MC1 preferentially dysregulates several pathways relating to cellular organization, while MC4 was associated with regulation of nucleic acid metabolism and RNA methylation (Extended Data Fig. 5e and Supplementary Table 21). Thus, in addition to predicting patient outcome, the somatic mutation landscape influences the epitranscriptome in a driver- and regulator-specific manner.

We next examined the clinical importance of locus-specific m^{6}A on known prostate cancer drivers. A comprehensive survival analysis of canonical prostate cancer driver genes across available biomolecular data types was performed (Fig. 5e and Supplementary Table 22). Consistent with previous findings, CNAs were predictive of biochemical recurrence for a subset of driver events (*MYC*, *PIK3R1* and *PTEN*[16,63]). For tumor suppressors *TP53* and *NKX3-1*, gene-level m^{6}A abundance provided significant prognostic information (Fig. 5e). Median dichotomization of m^{6}A revealed a trend toward increasing risk of biochemical recurrence upon *TP53* loss in the context of higher m^{6}A abundance. Further investigation identified a significant interaction effect, with a hazard ratio of 3.35 per standard deviation increase in $\log_2$ (m^{6}A) abundance in patients with *TP53* loss (Fig. 5f,i). Finally, we expanded the biomarker evaluation transcriptome wide by considering the 20,334 m^{6}A peaks identified in at least six patients. Strong associations with biochemical recurrence were detected for peaks on the *INHBA*, *VCAN* and *ZFHX4* transcripts (Fig. 5j, Extended Data Fig. 5f and Supplementary Tables 9 and 23). Additional peaks present on these transcripts showed a similar trend for association with biochemical recurrence but with nonsignificant effect sizes (Fig. 5j). The presence of each peak was associated with worse patient outcome (Fig. 5k and Extended Data Fig. 5g,h). In summary, m^{6}A provides prognostic value in patients with prostate cancer via both epitranscriptomic and somatic mutational information.

m^{6}A modification of *VCAN* associates with prostate cancer progression

To gain deeper insights into the intricate mechanisms underlying the relationship between m^{6}A peaks and tumorigenesis, we conducted a more detailed exploration. The RNA abundances of *VCAN*, *INHBA* and *ZFHX4* were significantly correlated in two large patient cohorts (Extended Data Fig. 6a). *VCAN* m^{6}A peaks were observed in ~15% of tumors, while *INHBA* and *ZFHX4* peaks were observed in 5%. A second *VCAN* peak correlated well with tumor hypoxia (Supplementary Table 14), prompting further investigation.

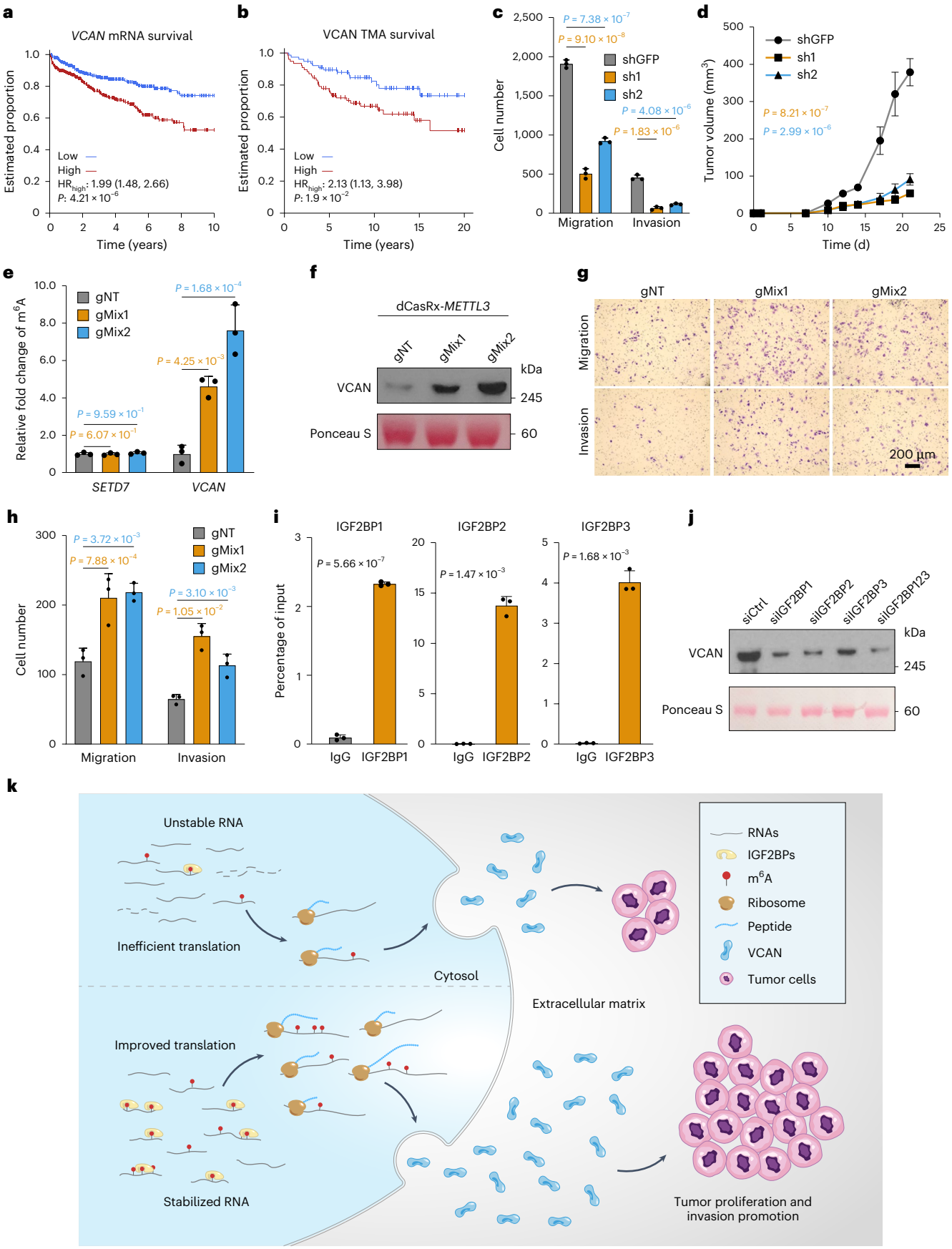
VCAN encodes a large chondroitin sulfate proteoglycan called versican (VCAN). It is a key component of the extracellular matrix[64]; the secretion of VCAN by fibroblasts promotes prostate cancer invasion[65]. The mRNA abundance of *VCAN* is significantly higher in tumors with m^{6}A peaks than in those without (Extended Data Fig. 6b). The abundance of VCAN protein is also higher in tumors with m^{6}A peaks (Extended Data Fig. 6c). Among the five prostate cancer cell lines tested, PC-3 had the highest level of *VCAN* m^{6}A (Extended Data Fig. 6d), comparable to patient samples. High abundance of *VCAN* mRNA was significantly associated with a worse outcome in five cohorts comprising 981 independent patients (Fig. 6a and Extended Data Fig. 6e).

VCAN protein abundance showed a trend but was not significantly associated with patient survival, possibly due to the relatively small sample size (Supplementary Fig. 4a). To address this limitation, we performed immunohistochemistry staining of VCAN in a tissue microarray with a larger sample size ($n = 154$). The staining results revealed the presence of VCAN in both tumor cells and stromal cells, albeit in modest amounts, with its primary localization observed in the extracellular matrix (Supplementary Fig. 4b). Importantly, patients with higher levels of VCAN had significantly poorer survival outcomes than those with lower levels (Fig. 6b and Supplementary Table 24). Furthermore, we observed a significant increase in the total amount of VCAN protein in tumors compared to normal tissues (Supplementary Fig. 4c). Additionally, VCAN abundance was notably elevated in high-risk groups compared to low- and intermediate-risk groups (Supplementary Fig. 4d). These findings highlight the prognostic value of VCAN in prostate cancer.

Knockdown of *VCAN* using short hairpin RNA (shRNA) or small interfering RNA (siRNA) significantly reduced cell proliferation, migration and invasion in PC-3 cells (Fig. 6c, Supplementary Fig. 4e and

Fig. 6 | *VCAN* drives proliferation and progression in vitro and in vivo. **a**, Kaplan–Meier survival curves showing biochemical recurrence stratified by *VCAN* mRNA abundance across five cohorts (Gerhauser, Ross-Adams, Taylor, TCGA-PRAD and International Cancer Genome Consortium (ICGC)-PRAD-CA (CA, Canada)), grouped by median abundance of *VCAN* mRNA. A Cox proportional hazards model was used to compare the hazard of biochemical recurrence between the groups. **b**, Kaplan–Meier survival curves showing biochemical recurrence stratified by VCAN average optical density from the tissue microarrays (TMA), grouped by median VCAN average optical density. A Cox proportional hazards model was used to compare the hazard of biochemical recurrence between the groups. **c**, Quantification of a migration and invasion assay after shRNA-mediated knockdown of *VCAN* in PC-3 cells. Data are represented as mean ± s.d. of three biological replicates. *P* values were calculated using one-way ANOVA with Dunnett's multiple-comparisons test. **d**, Tumor growth of xenografts derived from injected PC-3 cells infected with shRNAs targeting *VCAN* or green fluorescent protein (GFP) as the control. Data are shown as mean ± s.e.m. of five biological replicates. *P* values are from one-way ANOVA with Dunnett's multiple-comparisons test. **e**, Writing efficiency was assessed by m^{6}A meRIP-quantitative PCR (qPCR) in PC-3 cells. *SETD7* was introduced here as a negative control. Data are represented as mean ± s.d. of three biological replicates. *P* values are calculated using one-way ANOVA with Dunnett's multiple-comparisons test. g, guide RNA; gNT, non-targeting guide RNA. **f**, VCAN protein abundance after dCasRx-*METTL3*-based m^{6}A writing in PC-3 cells was detected by western blot. Ponceau S stain serves as the loading control. Numbers on the right indicate the positions of molecular mass (kDa) standards. **g**,**h**, Migration and invasion of PC-3 cells after writing by the dCasRx-*METTL3*-based RNA-editing system guided by *VCAN* mRNA-targeting guide RNA mixes. Representative images (**g**) and quantification (**h**, mean ± s.d. of three biological replicates; one-way ANOVA with Dunnett's multiple-comparisons test) are shown. Scale bar, 200 µm. **i**, RIP–qPCR showing the physical association of *VCAN* mRNA with IGF2BPs in PC-3 cells. Data are represented as mean ± s.d. of three biological replicates. *P* values were calculated using an unpaired two-sided Student's *t*-test. IgG, immunoglobulin G. **j**, Western blot showing VCAN protein abundance change after knocking down each IGF2BP protein (siIGF2BP1, siIGF2BP2 and siIGF2BP3) or all three IGF2BPs (siIGF2BP123) in PC-3 cells. Ponceau S stain serves as the loading control. Numbers on the right indicate the positions of molecular mass (kDa) standards. Ctrl, control. **k**, Graphical representation of *VCAN* regulation by m^{6}A.

https://doi.org/10.1038/s41588-025-02128-y



Extended Data Fig. 7a–g). *VCAN* knockdown with shRNA significantly reduced xenograft tumor growth in vivo (Fig. 6d and Extended Data Fig. 7h). Cell extravasation, which is a key step during cancer metastasis, was reduced after *VCAN* suppression in vivo in a chick embryo model (Extended Data Fig. 7i and Supplementary Videos 1–3). RNA-seq of PC-3 mouse xenografts with and without shRNA-mediated knockdown of *VCAN* identified 2,050 differentially expressed genes (Extended Data Fig. 7j), with downregulated genes enriched in cell adhesion, consistent with decelerating cell growth and motility (Extended Data Fig. 7k). In sum, in vitro, in vivo and clinical data suggest that *VCAN* drives aggressive prostate cancer.

VCAN-mediated cross-talk in prostate cancer progression

VCAN is recognized as a pivotal constituent in prostate stroma, secreted by fibroblasts, and exerts substantial influence on tumor initiation and progression[65,66]. To explore cross-talk between tumor and stromal cells in VCAN-mediated prostate cancer progression, we used the human prostate fibroblast cell line WPMY-1, which expresses an abundance of VCAN comparable to PC-3 cells (Extended Data Fig. 8a). Knocking down *VCAN* in WPMY-1 cells significantly reduced proliferation, underscoring its pivotal role in facilitating cellular proliferation (Extended Data Fig. 8b). Co-culturing PC-3 and WPMY-1 cells enhanced PC-3 motility in a dose-dependent manner with increasing WPMY-1 cell numbers (Extended Data Fig. 8c,d). Furthermore, co-culturing PC-3 cells with WPMY-1 cells with either *VCAN* or control knockdown (Extended Data Fig. 8e,f) showed that WPMY-1 could rescue the phenotype caused by *VCAN* knockdown (Extended Data Fig. 8g,h), suggesting that the phenotype was VCAN dependent. These findings highlight the complex interplay between tumor and stromal cells mediated by VCAN in prostate cancer progression.

m^{6}A on *VCAN* drives prostate cancer aggression

Given that tumors with m^{6}A peaks have higher *VCAN* mRNA abundance, we hypothesize that the m^{6}A modification stabilizes *VCAN* mRNA. We initially observed that *VCAN* mRNA abundance decreased with *METTL3* knockdown and increased with *ALKBH5* and *FTO* knockdown (Extended Data Fig. 9a–c). To confirm that *VCAN* m^{6}A modifications directly influence *VCAN* gene expression, we used catalytically dead CasRx (dCasRx)/*METTL3*-based programmable site-specific base editing on *VCAN* transcripts. meRIP–qPCR showed that, guided by the *VCAN*-targeting guide RNAs, the modification level on *VCAN* was significantly increased, while *SETD7* as a negative control was unchanged (Fig. 6e). After m^{6}A writing, *VCAN* mRNA and protein abundance was significantly increased (Fig. 6f and Extended Data Fig. 9d), consistent with perturbation of m^{6}A regulators. In contrast to knockdown of *VCAN*, writing additional m^{6}A modification enhanced proliferation and migration of PC-3 cells (Fig. 6g,h and Extended Data Fig. 9e). These data suggest that specific m^{6}A modifications on *VCAN* stabilize its mRNA and promote subsequent translation to functional protein.

To elucidate the regulatory mechanisms of m^{6}A modification on *VCAN* expression, we investigated the involvement of m^{6}A readers in regulating *VCAN* mRNA or protein abundance. YTHDF1 was initially considered a potential regulator of *VCAN* expression. However, analysis of publicly available YTHDF1 RNA immunoprecipitation sequencing (RIP-seq) data, including in PC-3 cells, revealed that *VCAN* is not a target of YTHDF1 (refs. 67–70). Furthermore, knockdown of YTH domain readers (YTHDF1–YTHDF3) in PC-3 cells did not affect VCAN protein abundance (Extended Data Fig. 9f–j). However, despite the lack of involvement of YTH domain-containing readers in *VCAN* regulation, previous studies have reported *VCAN* as a high-confidence target shared by the m^{6}A readers IGF2BP1–IGF2BP3, which stabilize target mRNA and promote translation[71]. We tested whether the IGF2BP proteins also bind *VCAN* mRNA in PC-3 cells by the RIP assay. Significant enrichment of *VCAN* was detected in all three pulldowns (Fig. 6i and Extended Data Fig. 9k). Knockdown of these three readers also reduced *VCAN* mRNA and protein abundance (Fig. 6j and Extended Data Fig. 9l,m). Consistent with *VCAN* knockdown, cell proliferation and invasion were also significantly reduced upon suppression of these three readers (Extended Data Fig. 10a–c). Ribosome profiling revealed increased *VCAN* abundance in the polysome-free fraction upon IGF2BP knockdown (Extended Data Fig. 10d,e). A rescue assay introducing m^{6}A modification onto *VCAN* mRNA under the background of *IGF2BP2* knockdown restored VCAN expression and cell phenotypes (Extended Data Fig. 10f). These data support a mechanism for m^{6}A modifications to increase *VCAN* mRNA abundance by stabilizing and promoting its translation through IGF2BP proteins, enhancing prostate cancer cell aggression (Fig. 6k).

Discussion

Post-transcriptional modifications such as m^{6}A facilitate tight regulation of gene expression and are essential for normal tissue physiology[67,72–74]. Germline and somatic genetics directly influence site-specific m^{6}A levels on thousands of transcripts, while m^{6}A levels influence downstream RNA and protein abundance. Our data provide an initial functional characterization of m^{6}A peaks and highlight the complexity of these regulatory relationships and the need for additional functional analyses.

Hundreds of studies have sought to identify genomic or transcriptomic features of prostate cancer or its microenvironment that can predict variable aggressiveness of diseases[2,16–19,21,57–60,63,75–81]. Our data support the hypothesis that the epitranscriptome is clinically predictive: it is influenced by germline, somatic and microenvironmental features (for example, hypoxia) and shows correlation with clinical behavior. It appears to synergize with other biomarkers and may be particularly useful with long-read, single-molecular sequencing technologies that simultaneously quantify abundance, structure and the epitranscriptome[82]. m^{6}A may not only predict clinical behavior but directly drive aggressive phenotypes, such as proliferation, migration and invasion.

Our study demonstrates the role of IGF2BPs in recognizing m^{6}A modifications on *VCAN*. Nevertheless, we acknowledge the possibility of other m^{6}A readers and their potential collaboration in regulating *VCAN* mRNA or protein abundance. Future investigations in different cell lines and with additional m^{6}A readers are warranted to comprehensively elucidate the regulatory landscape of *VCAN*.

While our study represents the most extensive m^{6}A profiling endeavor in prostate cancer to the best of our knowledge, the use of a single discovery cohort (n = 148) limits the generalizability of our findings. Future studies with larger cohorts incorporating diverse ancestries are warranted to establish the broader clinical importance of m^{6}A dysregulation in prostate cancer progression.

Our results propose m^{6}A modification as a central player in the life history of cancer, influencing its initiation by intermediating genetic factors, being dysregulated by somatic mutations and ultimately predicting and driving aggressive clinical and molecular cancer biology.

Online content

Any methods, additional references, Nature Portfolio 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/s41588-025-02128-y.

References

1. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
2. Sinha, A. et al. The proteogenomic landscape of curable prostate cancer. *Cancer Cell* **35**, 414–427 (2019).
3. Liu, J. et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* **173**, 400–416 (2018).

4. Huang, H., Weng, H. & Chen, J. m^{6}A modification in coding and non-coding RNAs: roles and therapeutic implications in cancer. *Cancer Cell* **37**, 270–288 (2020).
5. Zhao, B. S., Roundtree, I. A. & He, C. Post-transcriptional gene regulation by mRNA modifications. *Nat. Rev. Mol. Cell Biol.* **18**, 31–42 (2017).
6. Shi, H., Wei, J. & He, C. Where, when, and how: context-dependent functions of RNA methylation writers, readers, and erasers. *Mol. Cell* **74**, 640–650 (2019).
7. Wang, Q. et al. METTL3-mediated m^{6}A modification of *HDGF* mRNA promotes gastric cancer progression and has prognostic significance. *Gut* **69**, 1193–1205 (2020).
8. Gu, C. et al. Mettl14 inhibits bladder TIC self-renewal and bladder tumorigenesis through N^6-methyladenosine of Notch1. *Mol. Cancer* **18**, 168 (2019).
9. Panneerdoss, S. et al. Cross-talk among writers, readers, and erasers of m^{6}A regulates cancer growth and progression. *Sci. Adv.* **4**, eaar8263 (2018).
10. Lin, S., Choe, J., Du, P., Triboulet, R. & Gregory, R. I. The m^{6}A methyltransferase METTL3 promotes translation in human cancer cells. *Mol. Cell* **62**, 335–345 (2016).
11. Choe, J. et al. mRNA circularization by METTL3–eIF3h enhances translation and promotes oncogenesis. *Nature* **561**, 556–560 (2018).
12. Zhang, S. et al. m^{6}A demethylase ALKBH5 maintains tumorigenicity of glioblastoma stem-like cells by sustaining FOXM1 expression and cell proliferation program. *Cancer Cell* **31**, 591–606 (2017).
13. Vu, L. P. et al. The N^6-methyladenosine (m^{6}A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nat. Med.* **23**, 1369–1376 (2017).
14. Chang, G. et al. YTHDF3 induces the translation of m^{6}A-enriched gene transcripts to promote breast cancer brain metastasis. *Cancer Cell* **38**, 857–871 (2020).
15. Su, R. et al. *R*-2HG exhibits anti-tumor activity by targeting FTO/m^{6}A/MYC/CEBPA signaling. *Cell* **172**, 90–105 (2018).
16. Fraser, M. et al. Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* **541**, 359–364 (2017).
17. Bhandari, V. et al. Molecular landmarks of tumor hypoxia across cancer types. *Nat. Genet.* **51**, 308–318 (2019).
18. Houlahan, K. E. et al. Genome-wide germline correlates of the epigenetic landscape of prostate cancer. *Nat. Med.* **25**, 1615–1626 (2019).
19. Chen, S. et al. Widespread and functional RNA circularization in localized prostate cancer. *Cell* **176**, 831–843 (2019).
20. 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).
21. Espiritu, S. M. G. et al. The evolutionary landscape of localized prostate cancers drives clinical aggression. *Cell* **173**, 1003–1013 (2018).
22. Zeng, Y. et al. Refined RIP-seq protocol for epitranscriptome analysis with low input materials. *PLoS Biol.* **16**, e2006092 (2018).
23. Meyer, K. D. et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).
24. Dominissini, D. et al. Topology of the human and mouse m^{6}A RNA methylomes revealed by m^{6}A-seq. *Nature* **485**, 201–206 (2012).
25. Schaeffer, E. M. et al. Prostate Cancer, Version 4.2023, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Canc. Netw.* **21**, 1067–1096 (2023).
26. Liu, S., Zhu, A., He, C. & Chen, M. REPIC: a database for exploring the N^6-methyladenosine methylome. *Genome Biol.* **21**, 100 (2020).
27. Meng, J., Cui, X., Rao, M. K., Chen, Y. & Huang, Y. Exome-based analysis for RNA epigenome sequencing data. *Bioinformatics* **29**, 1565–1567 (2013).
28. Cui, X., Meng, J., Zhang, S., Chen, Y. & Huang, Y. A novel algorithm for calling mRNA m^{6}A peaks by modeling biological variances in meRIP-seq data. *Bioinformatics* **32**, i378–i385 (2016).
29. Robinson, D. G. & Storey, J. D. subSeq: determining appropriate sequencing depth through efficient read subsampling. *Bioinformatics* **30**, 3424–3426 (2014).
30. 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).
31. Zhu, H., Eng, S., Boutros, P. & He, H. H. HistogramZoo. *Zenodo* https://doi.org/10.5281/ZENODO.14713036 (2025).
32. Delon, J., Desolneux, A., Lisani, J.-L. & Petro, A. B. A nonparametric approach for histogram segmentation. *IEEE Trans. Image Process.* **16**, 253–261 (2007).
33. Delon, J., Desolneux, A., Lisani, J. L. & Petro, A. B. Color image segmentation using acceptable histogram segmentation. In *Pattern Recognition and Image Analysis* 239–246 (Springer, 2005).
34. Balaguer, A. B. P. *Analytical Methods for the Study of Color in Digital Images*. PhD thesis, Universitat de les Illes Balears (2006).
35. Lisani, J.-L. & Petro, A. B. Automatic 1D histogram segmentation and application to the computation of color palettes. *Image Process. Line* **11**, 76–104 (2021).
36. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
37. Zhang, Z. et al. RADAR: differential analysis of MeRIP-seq data with a random effect model. *Genome Biol.* **20**, 294 (2019).
38. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 623–656 (1948).
39. Mao, Y. et al. m^{6}A in mRNA coding regions promotes translation via the RNA helicase-containing YTHDC2. *Nat. Commun.* **10**, 5332 (2019).
40. Lalonde, E. et al. Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. *Lancet Oncol.* **15**, 1521–1532 (2014).
41. Heemers, H. V. & Tindall, D. J. Androgen receptor (AR) coregulators: a diversity of functions converging on and regulating the AR transcriptional complex. *Endocr. Rev.* **28**, 778–808 (2007).
42. Lichtenstein, P. et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* **343**, 78–85 (2000).
43. Luo, X. et al. RMVar: an updated database of functional variants involved in RNA modifications. *Nucleic Acids Res.* **49**, D1405–D1412 (2021).
44. Cotter, K. A. et al. Mapping of m^{6}A and its regulatory targets in prostate cancer reveals a METTL3-low induction of therapy resistance. *Mol. Cancer Res.* **19**, 1398–1411 (2021).
45. Al Olama, A. A. et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat. Genet.* **46**, 1103–1109 (2014).
46. Wang, A. et al. Characterizing prostate cancer risk through multi-ancestry genome-wide discovery of 187 novel risk variants. *Nat. Genet.* **55**, 2065–2074 (2023).
47. Xu, J. et al. Inherited genetic variant predisposes to aggressive but not indolent prostate cancer. *Proc. Natl Acad. Sci. USA* **107**, 2136–2140 (2010).
48. He, Y. et al. The prostate cancer susceptibility variant rs2735839 near *KLK3* gene is associated with aggressive prostate cancer and can stratify Gleason score 7 patients. *Clin. Cancer Res.* **20**, 5133–5139 (2014).
49. Schumacher, F. R. et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat. Genet.* **50**, 928–936 (2018).

50. Conti, D. V. et al. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nat. Genet.* **53**, 65–75 (2021).
51. Huynh-Le, M.-P. et al. Prostate cancer risk stratification improvement across multiple ancestries with new polygenic hazard score. *Prostate Cancer Prostatic Dis.* **25**, 755–761 (2022).
52. Yuan, J. et al. Prostate cancer transcriptomic regulation by the interplay of germline risk alleles, somatic mutations, and 3D genomic architecture. *Cancer Discov.* **12**, 2838–2855 (2022).
53. Xiong, X. et al. Genetic drivers of m^{6}A methylation in human brain, lung, heart and muscle. *Nat. Genet.* **53**, 1156–1165 (2021).
54. Zhang, Z. et al. Genetic analyses support the contribution of mRNA N^6-methyladenosine (m^{6}A) modification to human disease heritability. *Nat. Genet.* **52**, 939–949 (2020).
55. Wei, Z. et al. MYC reshapes CTCF-mediated chromatin architecture in prostate cancer. *Nat. Commun.* **14**, 1787 (2023).
56. Futreal, P. A. et al. A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
57. Armenia, J. et al. The long tail of oncogenic drivers in prostate cancer. *Nat. Genet.* **50**, 645–651 (2018).
58. Quigley, D. A. et al. Genomic hallmarks and structural variation in metastatic prostate cancer. *Cell* **174**, 758–769 (2018).
59. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* **163**, 1011–1025 (2015).
60. Taylor, B. S. et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11–22 (2010).
61. Hieronymus, H. et al. Copy number alteration burden predicts prostate cancer relapse. *Proc. Natl Acad. Sci. USA* **111**, 11139–11144 (2014).
62. Ross-Adams, H. et al. Integration of copy number and transcriptomics provides risk stratification in prostate cancer: a discovery and validation cohort study. *eBioMedicine* **2**, 1133–1144 (2015).
63. Weischenfeldt, J. et al. Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer. *Cancer Cell* **23**, 159–170 (2013).
64. Schmalfeldt, M., Dours-Zimmermann, M. T., Winterhalter, K. H. & Zimmermann, D. R. Versican V2 is a major extracellular matrix component of the mature bovine brain. *J. Biol. Chem.* **273**, 15758–15764 (1998).
65. Read, J. T. et al. Androgen receptor regulation of the versican gene through an androgen response element in the proximal promoter. *J. Biol. Chem.* **282**, 31954–31963 (2007).
66. Wu, Y. J., La Pierre, D. P., Wu, J., Yee, A. J. & Yang, B. B. The interaction of versican with its binding partners. *Cell Res.* **15**, 483–494 (2005).
67. Wang, X. et al. N^6-methyladenosine modulates messenger RNA translation efficiency. *Cell* **161**, 1388–1399 (2015).
68. Liu, T. et al. The m^{6}A reader YTHDF1 promotes ovarian cancer progression via augmenting EIF3C translation. *Nucleic Acids Res.* **48**, 3816–3831 (2020).
69. Wang, S. et al. N^6-methyladenosine reader YTHDF1 promotes ARHGEF2 translation and RhoA signaling in colorectal cancer. *Gastroenterology* **162**, 1183–1196 (2022).
70. Li, P. et al. ELK1-mediated YTHDF1 drives prostate cancer progression by facilitating the translation of Polo-like kinase 1 in an m^{6}A dependent manner. *Int. J. Biol. Sci.* **18**, 6145–6162 (2022).
71. Huang, H. et al. Recognition of RNA *N*-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat. Cell Biol.* **20**, 285–295 (2018).
72. Wang, X. et al. N^6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117–120 (2014).
73. Batista, P. J. et al. m^{6}A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **15**, 707–719 (2014).
74. Geula, S. et al. m^{6}A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science* **347**, 1002–1006 (2015).
75. Wedge, D. C. et al. Sequencing of prostate cancers identifies new cancer genes, routes of progression and drug targets. *Nat. Genet.* **50**, 682–692 (2018).
76. Baca, S. C. et al. Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666–677 (2013).
77. Barbieri, C. E. et al. Exome sequencing identifies recurrent *SPOP*, *FOXA1* and *MED12* mutations in prostate cancer. *Nat. Genet.* **44**, 685–689 (2012).
78. Gerhauser, C. et al. Molecular evolution of early-onset prostate cancer identifies molecular risk markers and clinical trajectories. *Cancer Cell* **34**, 996–1011 (2018).
79. Houlahan, K. E. et al. A polygenic two-hit hypothesis for prostate cancer. *J. Natl. Cancer Inst.* **115**, 468–472 (2023).
80. Erho, N. et al. Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. *PLoS ONE* **8**, e66855 (2013).
81. Jairath, N. K. et al. A systematic review of the evidence for the decipher genomic classifier in prostate cancer. *Eur. Urol.* **79**, 374–383 (2021).
82. Liu, H. et al. Accurate detection of m^{6}A RNA modifications in native RNA sequences. *Nat. Commun.* **10**, 4079 (2019).
83. Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012).

[1]Princess Margaret Cancer Center, University Health Network, Toronto, Ontario, Canada. [2]Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada. [3]Vector Institute, Toronto, Ontario, Canada. [4]Department of Urology, University of California, Los Angeles, Los Angeles, CA, USA. [5]Institute for Precision Health, University of California, Los Angeles, Los Angeles, CA, USA. [6]Jonsson Comprehensive Cancer Centre, University of California, Los Angeles, Los Angeles, CA, USA. [7]Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA, USA. [8]Department of Chemistry, the University of Chicago, Chicago, IL, USA. [9]Howard Hughes Medical Institute, the University of Chicago, Chicago, IL, USA. [10]Department of Urology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China. [11]Sunnybrook Research Institute, Toronto, Ontario, Canada. [12]MOE Key Laboratory of Metabolism and Molecular Medicine and Department of Biochemistry and Molecular Biology of School of Basic Medical Sciences and Fudan University Shanghai Cancer Center, Shanghai Medical College of Fudan University, Shanghai, China. [13]Department of Molecular Medicine, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA. [14]Research Centre of CHU de Québec-Université Laval, Québec City, Quebec, Canada. [15]State Key Laboratory of Common Mechanism Research for Major Disease, Suzhou Institute of Systems Medicine, Chinese Academy of Medical Sciences and Peking Union Medical College, Suzhou, China. [16]Division of Cancer Sciences, University of Manchester, Manchester, UK. [17]Christie NHS Trust and CRUK Manchester Institute and Cancer Centre, Manchester, UK. [18]Geneseeq Research Institute, Geneseeq Technology lnc., Toronto, Ontario, Canada. [19]School of Public Health, Nanjing Medical University, Nanjing, China. [20]Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario, Canada. [21]Present address: West China School of Public Health, West China Fourth Hospital and State Key Laboratory of Biotherapy, Sichuan University, Chengdu, China. [22]Present address: Institute of Precision Medicine, First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. [23]Present address: Department of Respiratory and Critical Care Medicine, the Affiliated Huaian No.1 People's Hospital of Nanjing Medical University, Huai'an, China. [24]These authors contributed equally: Xin Xu, Helen Zhu, Rupert Hugh-White, Julie Livingstone.
✉e-mail: pboutros@mednet.ucla.edu; hansenhe@uhnresearch.ca

Methods

Ethics statement and patient consent

All samples obtained in this study complied with the relevant ethical regulations approved by the institutional ethics committee and Research Ethics Board at the University Health Network (UHN). All procedures were performed in accordance with the International Guidelines for the Use of Animals and approved by the Animal Care Committee at the UHN. Informed consent was obtained at clinical follow-up and was consistent with local Research Ethics Board and ICGC guidelines.

Cohort development

Patient selection, tumor sectioning and sample processing followed established methods[16]. Tumor tissues had been collected following UHN Research Ethics Board and CHU de Québec-Université Laval-approved study protocols (UHN 06-0822-CE, UHN 11-0024-CE, CHUQc-UL 2012-913:H12-03-192). Patients were treated via radical prostatectomy, with biochemical recurrence after radical prostatectomy defined as two consecutive measurements of PSA > 0.2 ng ml^{-1} or the administration of salvage radiotherapy. All patients were classified as intermediate risk by National Comprehensive Cancer Network standards. ISUP score and tumor cellularity were evaluated by two genitourinary pathologists (T.v.d.K. and B.T.) using scanned hematoxylin- and eosin-stained slides from surgical samples[16]. This categorization yielded a distribution of nine patients in G1, 87 in G2, 36 in G3, eight in G4 and four in G5, in which G1–G5 classification determined prostate cancer-specific mortality. Tissue areas with >70% tumor cellularity were selected for macrodissection and multiomic data generation[16]. Cellularity was validated in silico using qpure version 1.1 with OncoScan array data[84].

m^{6}A meRIP and RNA sequencing

The m^{6}A meRIP assay was performed as described in the 'Refined RIP-seq' protocol[22]. Briefly, 4.5 ng of *Escherichia coli* K-12 total RNA was added to 1 µg of total RNA for each sample as a spike-in control. The RNA samples were fragmented into ~180-nucleotide-long pieces at 70 °C with RNA Fragmentation Buffer. The reaction was stopped with EDTA, followed by ethanol precipitation. Ten percent of the fragmented RNA was saved as input. Protein A magnetic beads (30 µl, Thermo Fisher, 10002D) and 30 µl protein G magnetic beads (Thermo Fisher, 10004D) were mixed and washed twice using 1 ml IP buffer (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.1% IGEPAL CA-630) and resuspended in 500 µl IP buffer. Next, 5 µg anti-m^{6}A antibody (Millipore, ABE572) was added to the beads, and samples were incubated at 4 °C for at least 6 h with constant rotation. The beads were then washed twice with 1 ml IP buffer and resuspended in 500 µl IP mixture containing 100 µl of 5× IP buffer, fragmented RNA and 5 µl RNasin Plus RNase Inhibitor (Promega, N2611) and then incubated at 4 °C for another 2 h. The beads were then washed twice with cold IP buffer, twice with cold low-salt buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 0.1% IGEPAL CA-630) and twice with cold high-salt IP buffer (10 mM Tris-HCl, pH 7.5, 500 mM NaCl, 0.1% IGEPAL CA-630) for 10 min each at 4 °C. After the washing steps, RNA enriched on the beads was eluted from the beads with 200 µl RLT buffer supplied in the RNeasy Mini Kit (Qiagen, 74106) and then cleaned up with an RNeasy MinElute spin column. The anti-m^{6}A antibody-enriched RNA was eluted with 14 µl ultrapure water.

MeRIP-seq libraries were prepared by using the SMARTer Stranded Total RNA-Seq Kit version 2 (Pico Input Mammalian, Takara, 634413) according to the manufacturer's protocol, which incorporates an rRNA depletion step. This involves the targeted cleavage of rRNA by ZapR in the presence of the mammalian-specific R-Probes. Libraries for siRNA-mediated knockdown of *VCAN* in PC-3 cells were prepared by using the TruSeq Stranded mRNA kit (Illumina, 20020594). The libraries were sequenced on Illumina HiSeq X Ten with paired-end 150-bp read length to a depth of at least 30 million reads per sample. Samples were mapped, and counts were quantified using STAR (version 2.7.2), and differentially expressed genes were identified using DESeq2 (version 1.26.0)[36]. $P_{adjusted} < 0.05$ and $|FC| > 1.5$ were used to define differentially expressed genes. GO enrichment analysis for significantly upregulated or downregulated genes after knocking down *VCAN* was performed using the R package clusterProfiler (version 3.14.3)[85]. Benjamini–Hochberg multiple-testing adjustment was used (Q value < 0.1).

Processing meRIP-seq data and quality control

Reads were trimmed using cutadapt (version 2.8) using adaptor sequences[86]. Samples were mapped, and reads were quantitated at the gene level using STAR (version 2.7.2) using two-pass mapping and default parameters for hg38 using annotations from gencode.v34.chr_patch_hapl_scaff[87,88]. Using the PCAWG pipeline as a reference, different parameters were chosen for the parameters outFilterMultimapNmax, the threshold for multiple mapping reads, and outFilterMismatchNoveReadLmax, the threshold for an acceptable number of mismatches, to determine the robustness of mapped reads[89]. The quality of each sample was determined using FastQC (version 0.11.8)[90], RSeQC (version 3.0.1)[91], Picard (version 2.21.7) and the ENCODE QC guidelines as described in ref. 26 and the output metrics of STAR. Using 26 QC metrics, the sum of negative z scores was computed, and 14 samples failing a number of QC metrics were excluded (Extended Data Fig. 1b). Input reads were counted using RSEM (version 1.3.2)[92]. Concordant with past observation, a high percentage (median: IP, 84.32%; input, 47.86% and range: IP, (0.51, 0.97); input, (0.28, 0.94)) of duplicated reads was observed[26], potentially exacerbated by the use of low-input starting biological material[93], but, as RNA-seq is prone to natural read duplicates, the duplicated reads were retained in subsequent peak calling and m^{6}A quantitation. Germline variants were called using the GATK pipeline[94–96] and compared to variants called by Houlahan et al.[18] using vcfeval (rtg-tools, version 3.10.1) to verify sample identity[97,98]. TPMs were compared between the 92 overlapping samples from this study and previous ultra-deep bulk RNA-seq[19].

Peak calling, m^{6}A quantitation and motif identification

Peak calling employed MeTPeak (version 1.1) and exomePeak (version 2.16.0) with default parameters[27,28]. As MeTPeak is specialized for single-end sequencing, forward reads were extracted and used as input. Subsampling was conducted with Picard DownsampleSam (version 2.21.7) and three seeds (314,159, 271,828, 1,618) at 50%, 25%, 10% and 1%. Motifs were called for each sample using HOMER (version 4.1.1)[30].

Joint peaks were called using HistogramZoo (version 1.0.1). Briefly, the coverage of gene exons by sample-level peak calls using MeTPeak was computed on a gene-based coordinate system with the removal of introns to generate a cross-cohort pileup. The pileup was binned into 1-bp bins as a gene-level histogram. Using a fine-to-coarse segmentation algorithm, a histogram segmentation was computed (hyperparameter $\varepsilon = 0.005$) and regions of low entropy were filtered[32–35]. Histograms were fit with probability distributions by optimizing against goodness-of-fit metrics (mean squared error, Kolmogorov–Smirnov statistic, area of intersection, Jaccard index, χ^2 statistic) using the DEoptim R package to characterize distribution shape and skew[99]. The best-fitting distribution was selected using majority voting; ties were broken by the best fit using the Jaccard index.

Uniquely mapped reads were quantitated on the joint peaks using featureCounts from subread (version 2.0.0) using the parameter –largestOverlap to assign reads to the peak with the largest overlap. Uniquely mapped reads were quantitated on genes using RSEM (version 1.3.2). The two datasets were used to estimate peak-level and gene-level m^{6}A abundance, respectively, using the following post-processing steps based on the protocol described by Zhang et al.[37]. Specifically, for the 1% of peaks or genes with the greatest counts, the mean fold enrichment of peak-level or gene-level IP counts over the corresponding gene-level input counts was calculated. These enrichment values were then used in library size factor calculation (RLE) with the DESeq2 (version 1.26.0)[36]

'estimateSizeFactorsForMatrix' function. After normalization, IP reads in peaks or genes were adjusted by the normalized gene-level read counts of the input libraries and rounded to yield whole numbers[37,54].

For m^{6}A peaks, the resulting peak-by-sample count matrix was either (1) used as direct input or (2) 'soft-masked'. The identification of a peak in a sample is a function of applying a statistical threshold on observed IP and input read counts, with the intention of determining a binary methylation status for a sample. A peak may therefore have nonzero counts in a sample for which the peak-calling threshold was not met. Masking thus addressed the goal of filtering spurious reads observed in 'peak-less' samples. Soft-masking was defined as setting the peak counts to zero after input adjustment in samples where the peak was not called.

Soft-masked peak-level m^{6}A data were used for differential driver methylation analysis, while the unmasked peak-by-count matrix was used in subtype discovery, clinical associations, nominating m^{6}A-QTLs, hypoxia associations and biomarker discovery. Gene-level m^{6}A data were used in biomarker discovery for prostate cancer drivers. The binary status of whether a particular peak was observed in a given sample was used for mutual information analysis.

Identifying m^{6}A-QTLs

m^{6}A-QTLs were identified using a linear additive model implemented in Matrix eQTL (version 2.3) and validated using QTLtools (version 1.2)[83,100,101]. The 'soft-masked' peaks were filtered to remove peaks with zero reads in greater than 20% of samples (30,303 of 32,051 peaks remaining). Normalized, adjusted counts were then transformed using rank-inverse normal transformation[102]. The first seven principal components of the genotype matrix were identified using PCA with PLINK at default pruning for linkage disequilibrium[103], and, in combination with age, were used as covariates. SNPs within 10-kbp windows to the peaks were tested for local m^{6}A-QTLs, while SNPs identified as linked to genes by Yuan et al.[52] were tested for spatial m^{6}A-QTLs, and the top peak–SNP associations in the set of significant associations ($Q < 0.1$) were nominated as candidate m^{6}A-QTLs. eQTLs and pQTLs were identified using the same transformation and covariates for significant hits.

m^{6}A sites were downloaded from the RMVar database (hg38, accessed 21 August 2021) and filtered for high-confidence single-base pair m^{6}A characterization methods in human cell lines. This dataset represented 132,428 single-base pair sites across 70 studies, 335 tissues, cell lines and conditions and 1,056 samples and were captured using one of six experimental techniques (miCLIP-seq, DART-seq, m^{6}ACE-seq, m^{6}A-REF-seq, PA-m^{6}A-seq and m^{6}A-label-seq). Sites were filtered for those found in lncRNAs, protein-coding genes and pseudogenes. Specifically, sites were required to be localized to exons, UTRs, stop codons and coding sequences. Next, sites were filtered for those that overlapped the 5,647 regions found in at least two cancer cell lines using meRIP-seq. Finally, sites were filtered for being identified in at least two studies, identified by at least two experimental techniques and identified in at least two tissues, resulting in a dataset of 2,883 m^{6}A sites.

HiChIP loop calls from Wei et al.[55] were obtained and used to determine whether m^{6}A-QTL SNPs and their associated peaks were located in anchored loops.

The GenomicRanges (version 1.54.1) R package was used to intersect the genomic coordinates and calculate the distance between m^{6}A-QTL SNPs and their associated m^{6}A peaks.

m^{6}A hypoxia analysis

Hypoxia scores were calculated using microarray-derived RNA abundance values in a previous study[17]. To maximize sample overlap, the corresponding gene-level mRNA abundance values[16] were used when calculating correlation between m^{6}A regulatory gene RNA abundance and sample hypoxia. m^{6}A peaks that were identified in every sample ($n = 20,811$) were assessed for correlation with patient hypoxia score via Spearman's rank correlation. Pathway enrichment analysis of the resulting genes with significant peaks was performed using the gprofiler2 R package (version 0.2.0)[104] against the Gene Ontology Biological Process domain and the Reactome database, with the analysis being limited to pathway terms containing a maximum of 1,500 genes.

Prostate cancer drivers and m^{6}A enzymes

Somatic mutations were selected based on prior work[17]. These data were obtained from the ICGC Data Portal under project 'PRAD-CA' (https://dcc.icgc.org/projects/PRAD-CA). Specifically, CNA calls from the Affymetrix OncoScan array were used. CNAs were annotated with genes of interest, and, if conflicting gains or losses overlapped a given gene, the CNA type with the largest overlap was selected. Single-nucleotide variants used were those called via SomaticSniper version 1.0.5.0 and indels as called by the 'PCAWG Consensus INDEL caller'. Only simple somatic mutation events with the following predicted consequence type were used: 'missense_variant', 'stop_gained', 'stop_lost', 'splice_donor_variant', 'splice_acceptor_variant'. The set of m^{6}A regulatory enzymes was curated from ref. 105. The set of prostate cancer driver CNAs was curated from Rheinbay et al.[106]; a representative Cancer Gene Census gene was selected from every peak identified in prostate cancer through the Genomic Identification of Significant Targets in Cancer project that was annotated with at least one Cancer Gene Census gene[106]. The set of prostate cancer driver simple somatic mutations was curated from the TCGA-PRAD[107]. Structural variants resulting in ERG or ETV fusion events were called via identification of genomic breakpoints inferred through whole-genome sequencing or OncoScan array[16] or the presence of *TMPRSS2-ERG* fusion transcripts inferred through RNA-seq. For statistical analyses, only genes mutated in >5% of samples were included, yielding 18 drivers and ten m^{6}A enzymes. Mutation co-occurrence was assessed using a hypergeometric test.

Mutation differential methylation analysis

For differential methylation analysis, initial filtering of peaks was performed using a threshold of >0 normalized, 'soft-masked', IP-adjusted read counts in ≥50% of samples (74 of 144 samples with both meRIP-seq and mutation status data), resulting in a total of 6,467 peaks. To reduce the influence of outliers, the top and bottom 2.5% of samples for each peak were trimmed. For each mutation, differences in normalized IP-adjusted read counts for each peak were assessed between mutation-positive and mutation-negative samples using a Poisson random effect model as implemented in the diffIP function of the RADAR R package (version 0.2.4)[37]. Significance was called at $Q < 0.1$. Pathway enrichment analysis of significant genes was performed using ActivePathways (version 1.1.0)[108] against the Gene Ontology and Reactome databases (data obtained from g:profiler release/2022-02-14). Terms were limited to between 5 and 500 genes. The threshold for merged P values and for pathway enrichment was set to 0.1. P values were merged using Brown's method, and multiple-testing correction was conducted using Holm's method.

Biomarker discovery using univariate survival analysis

CNAs in nine m^{6}A regulatory genes were tested for association with time until biochemical recurrence. For each gene, a Cox proportional hazards model was used to compare the hazard of biochemical recurrence between the more common alteration event (that is, copy number gain or loss) versus neutral copy number. The model was fit using the survival R package (version 3.3.1)[109] with Wald test P values. When modeling a given gene, the Cox model was fit separately to six different cohorts: Taylor et al. ($n = 154$)[60], Hieronymus et al. ($n = 104$)[61], Ross-Adams et al. 1 ($n = 105$)[62], Ross-Adams et al. 2 ($n = 76$)[62], ICGC-PRAD-CA ($n = 379$)[16,17] and TCGA-PRAD ($n = 421$)[59]. A random effect meta-analysis was used to combine the six cohort-specific log hazard ratios into an overall hazard ratio via the meta R package[110] (version 6.5.0).

Taylor et al., Ross-Adams et al. 1 and Ross-Adams et al. 2 CNA data were obtained from https://github.com/crukci-bioinformatics/camcAPP/tree/master/data. Taylor et al. biochemical recurrence data were obtained from https://www.cbioportal.org/study/clinicalData?id=prad_mskcc. Hieronymus et al. data were obtained from https://www.cbioportal.org/study/summary?id=prad_mskcc_2014, TCGA-PRAD data were downloaded using the TCGAbiolinks R package (version 2.25.3)[111], and ICGC-PRAD-CA data are a superset of the CNA data described in the current study and were obtained as described above.

Survival analysis of prostate cancer drivers was performed similarly. m^{6}A was considered as a continuous variable using gene-level normalized, IP-adjusted counts before median dichotomization. Other continuous variables (RNA and protein abundance) were also median dichotomized, and CNAs were analyzed as previously detailed.

For peak-level survival analysis, peaks were initially filtered to remove those identified in less than six samples (n = 20,334). A two-stage approach was then used to identify survival associations. First, Cox proportional hazards models were fitted with the R package survival (version 3.2-13) for each peak using presence or absence of the peak as a discrete variable and the Wald test for nonzero coefficients[109]. The proportional hazards assumption was tested for each peak. Next, to correct for the comparison of potentially imbalanced populations, ExaLT was used to validate P values using an approximation factor of 10 and a P-value threshold of 0.01 (ref. 112).

For survival analysis of *VCAN* RNA abundance, the assumptions of the Cox model were tested using the cox.zph function in the R survival package (version 2.41-3) with 0.1 as a cutoff. The Cox proportional hazards model was fit using patient groups dichotomized by the median RNA abundance, with biochemical recurrence as the endpoint. The analysis used bulk RNA-seq data from several studies, including TCGA-PRAD[59], Taylor et al.[60], Gerhauser et al.[78], Ross-Adams et al.[62] and ICGC-PRAD-CA[19]. TCGA-PRAD RNA-seq data were downloaded using the R package TCGA2STAT (version 1.2)[113].

m^{6}A datasets

Reference m^{6}A sites were downloaded from RMVar (29 August 2021)[43], and additional m^{6}A sites from the prostate epithelial cell line RWPE and the cancer cell line LNCaP were downloaded from Cotter et al.[44]. For the latter, m^{6}A sites were identified as the nearest adenine within 3 bp downstream of the cross-linking mutation by strand.

Additional datasets

Data from previous publications were used in their original state[2,16–18,20] with the exception of genotyping data as described above and RNA-seq data from ref. 19, which were remapped to hg38 using annotations from gencode.v34.chr_patch_hapl_scaff, using the RNA-seq pipeline described above. Quality control of the input library, mutual information analysis, eQTLs, enzyme associations with RNA abundance, comparison of m^{6}A subtypes to RNA subtypes and survival analysis of target genes based on transcript abundance were conducted using RNA-seq data[19] due to greater sequencing depth. Additional RNA-seq and/or meRIP-seq datasets were processed using identical pipelines.

Statistics and reproducibility

Data analysis was conducted using R (version 3.6.1 to version 4.1.0). Unless indicated otherwise, multiple-hypothesis testing was conducted for each statistical analysis using the method outlined by Benjamini and Hochberg[114], commonly referred to as the false discovery rate or Q value (denoted here as Q). In vitro and in vivo data were expressed as mean ± s.d. and mean ± s.e.m., respectively. Two-tailed unpaired Student's t-test and one-way ANOVA were performed to identify significant differences between groups in our experiments. Data were confirmed to meet the assumptions of normality and equal variance for the statistical tests used. Details of the test method used for statistical analysis are specified in the relevant figure legends and the Methods. All values were considered significantly different at $P < 0.05$. Western blot experiments were performed at least twice with similar results. Data visualization was conducted using BoutrosLab.plotting.general (version 6.0.3)[115], ggseqlogo (version 0.1)[116] and Guitar (version 2.4.0)[117]. No statistical method was used to predetermine sample size. The meRIP-seq data of 14 samples were excluded from the analyses due to stringent quality control. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

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

Data availability

m^{6}A meRIP-seq raw sequencing data are available at the EGA under accession EGAS00001006925. Other previously published molecular data characterizing the ICGC-PRAD-CA cohort analyzed in this study are available at the EGA under accession EGAS00001000900. PC-3 shRNA knockdown *VCAN* RNA-seq data were deposited in the Gene Expression Omnibus under the accession number GSE220432. Previously published mRNA microarray data are available at the Gene Expression Omnibus under accession code GSE84043. Source data are provided with this paper.

Code availability

HistogramZoo is available at https://github.com/uclahs-cds/public-R-HistogramZoo ref. 31. Scripts for data analysis and plotting are available at https://github.com/uclahs-cds/prostate-cancer-m6a-paper ref. 118.

References

84. Song, S. et al. qpure: a tool to estimate tumor cellularity from genome-wide single-nucleotide polymorphism profiles. *PLoS ONE* **7**, e45835 (2012).
85. 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).
86. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
87. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
88. Harrow, J. et al. GENCODE: the reference human genome annotation for the ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
89. PCAWG Transcriptome Core Group et al. Genomic basis for RNA alterations in cancer. *Nature* **578**, 129–136 (2020).
90. Andrews, S. et al. FastQC. *GitHub* https://github.com/s-andrews/FastQC (2023).
91. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
92. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
93. Parekh, S., Ziegenhain, C., Vieth, B., Enard, W. & Hellmann, I. The impact of amplification on differential expression analyses by RNA-seq. *Sci. Rep.* **6**, 25533 (2016).
94. Van der Auwera, G. A. & O'Connor, B. D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (O'Reilly Media, 2020).
95. Poplin, R. et al. Scaling accurate genetic variant discovery to tens of thousands of samples. Preprint at *bioRxiv* https://doi.org/10.1101/201178 (2017).
96. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

97. Krusche, P. et al. Best practices for benchmarking germline small-variant calls in human genomes. *Nat. Biotechnol.* **37**, 555–560 (2019).
98. Cleary, J. G. et al. Comparing variant call files for performance benchmarking of next-generation sequencing variant calling pipelines. Preprint at *bioRxiv* https://doi.org/10.1101/023754 (2015).
99. Mullen, K. M., Ardia, D., Gil, D. L., Windover, D. & Cline, J. DEoptim: an R package for global optimization by differential evolution. *J. Stat. Softw.* **40**, 1–26 (2011).
100. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).
101. Delaneau, O. et al. A complete tool set for molecular QTL discovery and analysis. *Nat. Commun.* **8**, 15452 (2017).
102. McCaw, Z. R., Lane, J. M., Saxena, R., Redline, S. & Lin, X. Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics* **76**, 1262–1272 (2020).
103. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
104. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2—an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res.* **9**, ELIXIR-709 (2020).
105. Jiang, X. et al. The role of m^6A modification in the biological functions and diseases. *Signal Transduct. Target. Ther.* **6**, 74 (2021).
106. Rheinbay, E. et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature* **578**, 102–111 (2020).
107. Bailey, M. H. et al. Comprehensive characterization of cancer driver genes and mutations. *Cell* **174**, 1034–1035 (2018).
108. Paczkowska, M. et al. Integrative pathway enrichment analysis of multivariate omics data. *Nat. Commun.* **11**, 735 (2020).
109. Therneau, T. M. & Mayo Foundation. *A Package for Survival Analysis in S* sites.google.com/site/statsr4us/docs/SurvivalanalysisinR_package.pdf (1999).
110. Schwarzer, G. et al. meta: an R package for meta-analysis. *R News* **7**, 40–45 (2007).
111. Colaprico, A. et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).
112. Vandin, F., Papoutsaki, A., Raphael, B. J. & Upfal, E. Accurate computation of survival statistics in genome-wide studies. *PLoS Comput. Biol.* **11**, e1004071 (2015).
113. Wan, Y.-W., Allen, G. I. & Liu, Z. TCGA2STAT: simple TCGA data access for integrated statistical analysis in R. *Bioinformatics* **32**, 952–954 (2016).
114. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
115. P'ng, C. et al. BPG: seamless, automated and interactive visualization of scientific data. *BMC Bioinformatics* **20**, 42 (2019).
116. Wagih, O. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* **33**, 3645–3647 (2017).
117. Cui, X. et al. Guitar: an R/Bioconductor package for gene annotation guided transcriptomic analysis of RNA-related genomic features. *BioMed Res. Int.* **2016**, 8367534 (2016).
118. Xu, X. et al. The landscape of N^6-methyladenosine in localized primary prostate cancer. *Zenodo* https://doi.org/10.5281/ZENODO.14655534 (2025).

Acknowledgements

We thank all members of the Boutros and He laboratory for insightful suggestions, all patients who made this study possible and X. Xiong for sending supporting GTEx data. The results here are in part based on data generated by the TCGA Research Network: https://www.cancer.gov/tcga. This work was supported by the Princess Margaret Cancer Foundation (886012001223 to H.H.H.), the Canadian Cancer Society (TAG2018-2061), CIHR operating grants (142246, 152863, 152864 and 159567 to H.H.H.) and a Terry Fox New Frontiers Program Project Grant (PPG19-1090 to H.H.H.). H.H.H. holds the Joey and Toby Tanenbaum Brazilian Ball Chair in Prostate Cancer and the Tier 1 Canada Research Chair in RNA Medicine. The data presented in this paper were partially funded by the Terry Fox Research Institute's Marathon of Hope Cancer Centres Network. Xin Xu was supported by the Prostate Cancer Foundation Young Investigator Award (21YOUN06). H.Z. was supported by a CIHR doctoral award. K.E.H. was supported by a CIHR Vanier Fellowship. N.Z. was supported by the NIH through awards T32HG002536 and F31CA281168. S.C. was supported by the Prostate Cancer Foundation Young Investigator Award (21YOUN06) and a CIHR fellowship (181755). P.C.B. was supported by NIH grants P30CA016042, U2CCA271894, U24CA248265, U01CA214194 and R01CA272678. This work was supported by a Prostate Cancer Foundation Special Challenge Award to P.C.B. (award ID 20CHAS01) made possible by the generosity of L. Ruvo. This study is also supported by the Medical Research Data Center at Shanghai Medical College of Fudan University. The funders had no role in the study design, data collection and analysis, decision to publish or manuscript preparation.

Author contributions

Designed studies: Xin Xu, H.Z., R.H.-W., J.L., P.C.B. and H.H.H. Sample and clinical data acquisition: M.O., V.P., H.H., A. Bergeron, L.L., B.T., Y.F., R.G.B., N.E.F., A. Berlin and T.v.d.K. Performed experiments: Xin Xu, Y.W., K.P., W.L., Xi Xu, M.S., W.Y.G., M.W., S.W., W.C., X.C., K.X., M.L., M.K., C.H. and H.L. Data analysis: H.Z., R.H.-W., J.L., Xin Xu, S.E., N.Z., S.C., K.E.H., S.L., J.A., Y. Song, Y.Z., G.Z., T.G., W.X., G.-H.W., X.W. and Y. Shao. Manuscript first draft: Xin Xu, H.Z., R.H.-W., J.L., P.C.B. and H.H.H. Revised and approved manuscript: all authors.

Competing interests

For a portion of the time during the preparation of this paper, H.Z. was an employee at Deep Genomics. All contributions to the design, analysis and interpretation of results of this project were completed outside of the term of employment. P.C.B. sits on the scientific advisory boards of Sage Bionetworks, Intersect Diagnostics and BioSymetrics. Y. Shao and X.W. are shareholders and/or employees of Geneseeq Technology. The other authors declare no competing interests.

Additional information

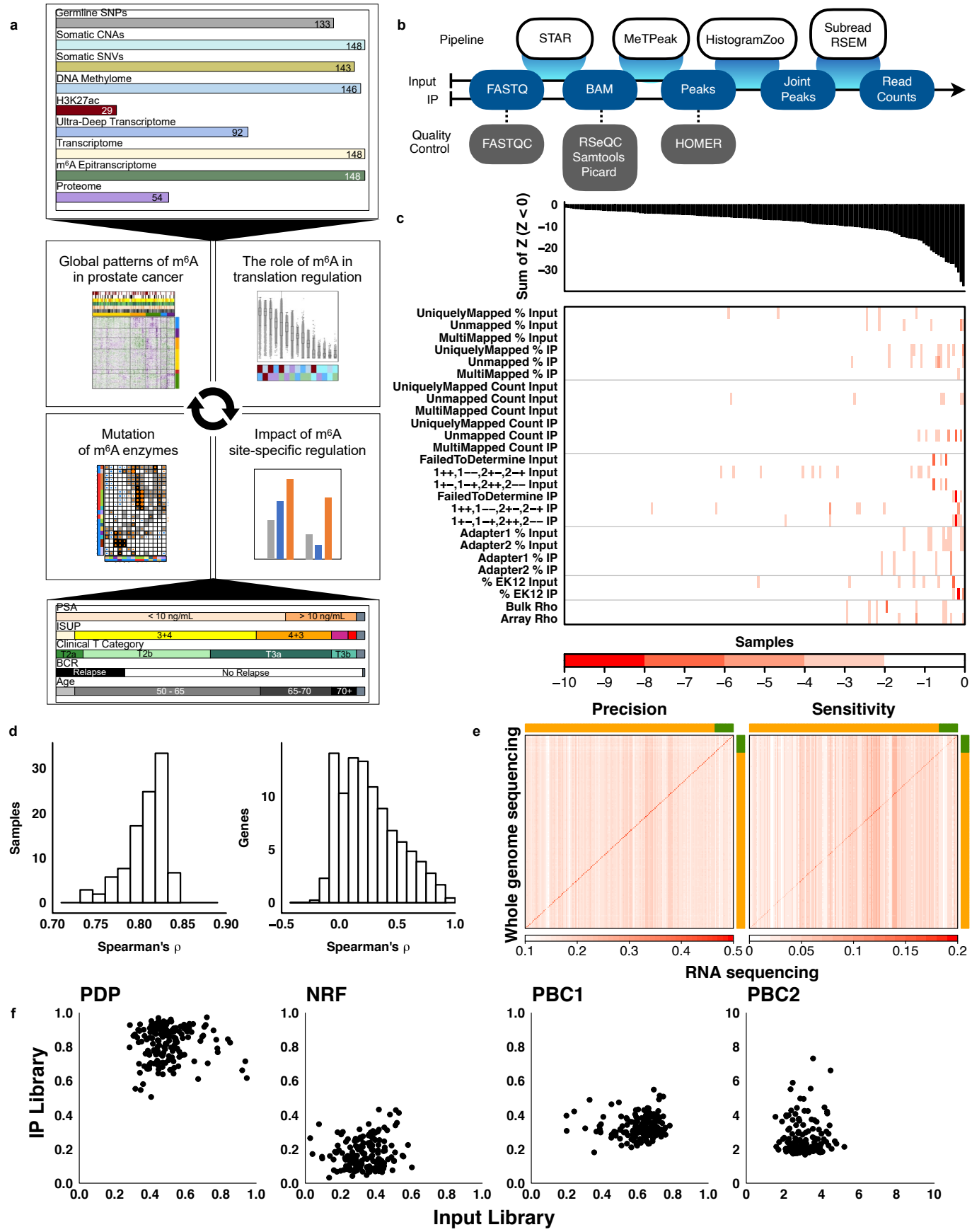
Extended data is available for this paper at https://doi.org/10.1038/s41588-025-02128-y.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-025-02128-y.

Correspondence and requests for materials should be addressed to Paul C. Boutros or Housheng Hansen He.

Peer review information *Nature Genetics* thanks Jianjun Chen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41588-025-02128-y

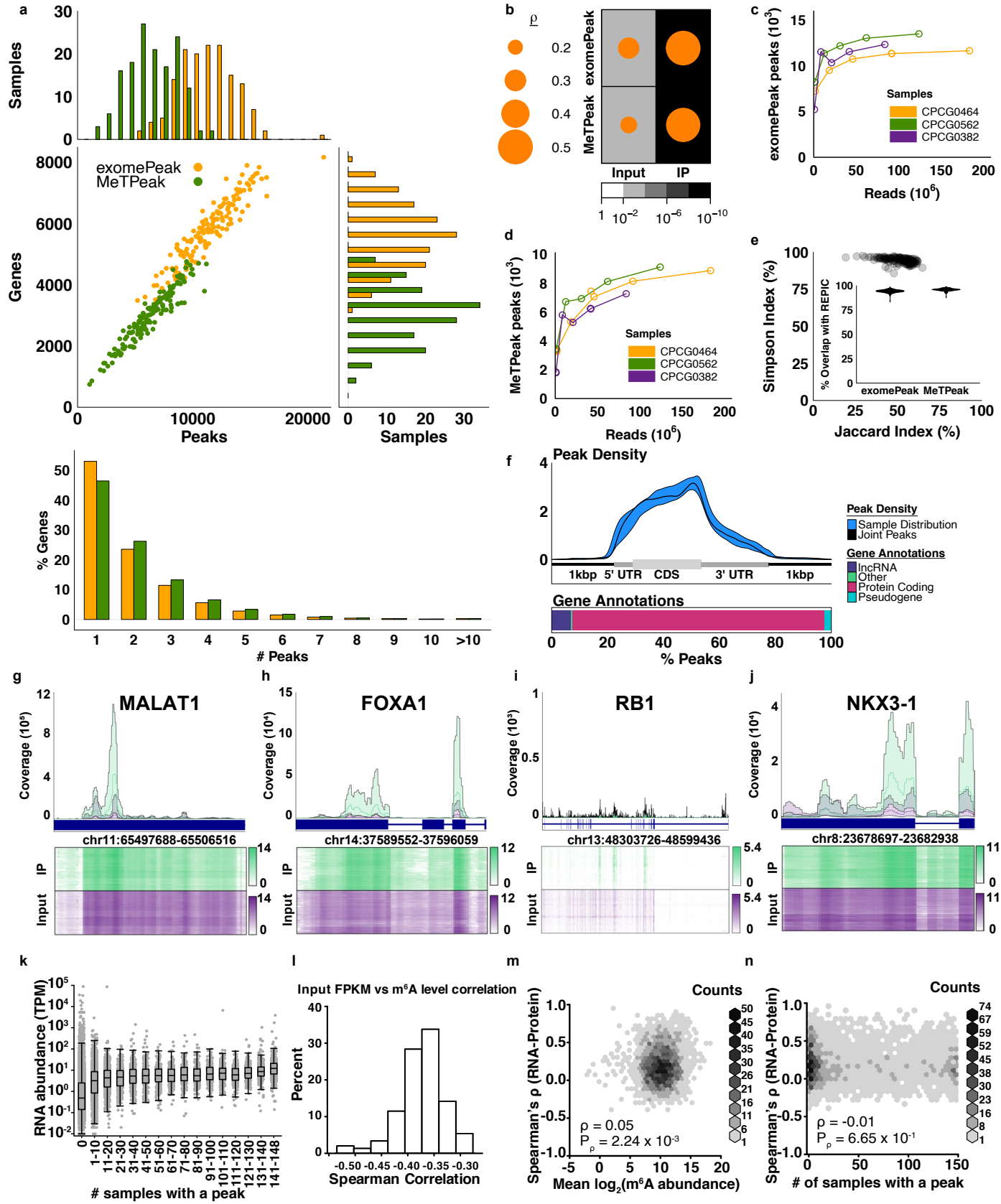


Extended Data Fig. 1 | See next page for caption.

https://doi.org/10.1038/s41588-025-02128-y

Extended Data Fig. 1 | Data Analysis & Quality Control. a) The study workflow: we investigate the role of the post-transcriptional RNA modification m^{6}A in primary prostate tumors by examining global m^{6}A patterns, germline origins, clinical associations, microenvironmental associations and functional impact. This builds on previous molecular profiling studies of this primary prostate tumor cohort including 133 genotypes[18], 148 CNA and 143 somatic mutation profiles[17], 146 DNA methylation[18], 29 H3K27ac ChIP-seq[20], 92 ultra-deep RNA-sequencing[19] and 54 proteomic profiles[2]. **b)** Data processing pipeline for meRIP-sequencing. **c)** Quality control metrics derived from the output of STAR, RSeQC, cutadapt and correlation of the Input library with previous transcriptome profiling datasets[16,19]. Top barplot: the sum of negative Z-scores for each sample. Bottom heatmap: metric by sample matrix where colors denote the magnitude of negative Z-scores. **d)** Left panel: sample-level correlation of transcript abundance in 92 samples common between Input libraries of this study and ultra-deep bulk RNA-seq from Chen et al.[19] (Spearman's ρ). Right panel: gene-level correlation of transcript abundance in the same datasets (Spearman's ρ)[19]. **e)** Validation of sample identity by comparing germline variants identified in the Input libraries and Houlahan et al.[18]. Both axes are ordered identically, first, by ancestry (covariate: orange represents European ancestry, green represents non-European ancestry) and then, by sample ID. Heatmap color represents precision (left) and sensitivity (right). Precision is calculated as the intersection of RNA-seq and WGS variants divided by all RNA-seq variants. Sensitivity is calculated as the intersection of RNA-seq and WGS variants divided by all WGS variants. **f)** Comparisons of the PCR Duplicate Proportion (PDP), the Non-Redundant Fraction (NRF) and the PCR Bottlenecking Coefficients 1 and 2 (PBC1/2) for each sample. The metric values of the IP library (y-axis) are compared to the metric values of the Input library (x-axis).

https://doi.org/10.1038/s41588-025-02128-y

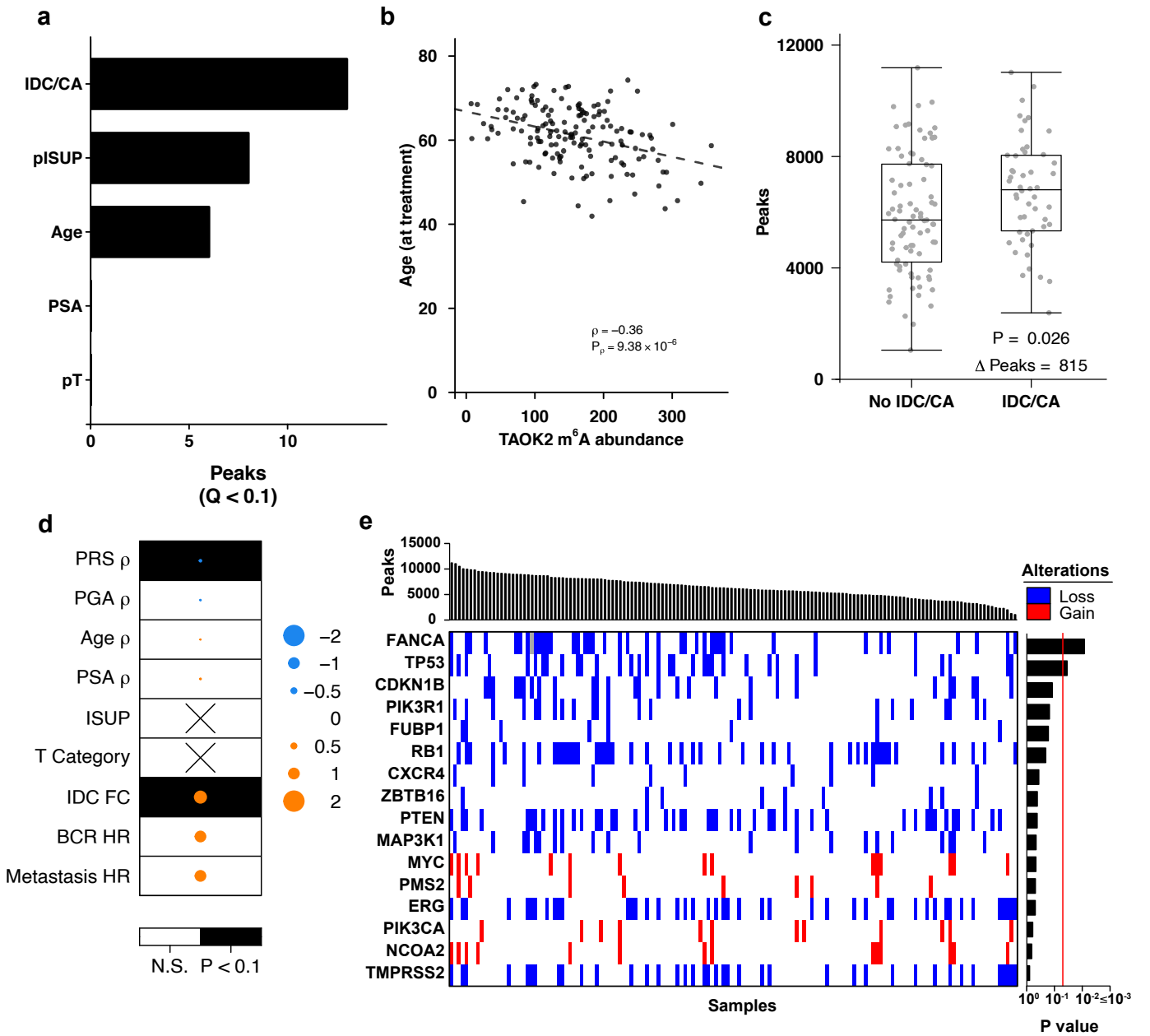


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s41588-025-02128-y

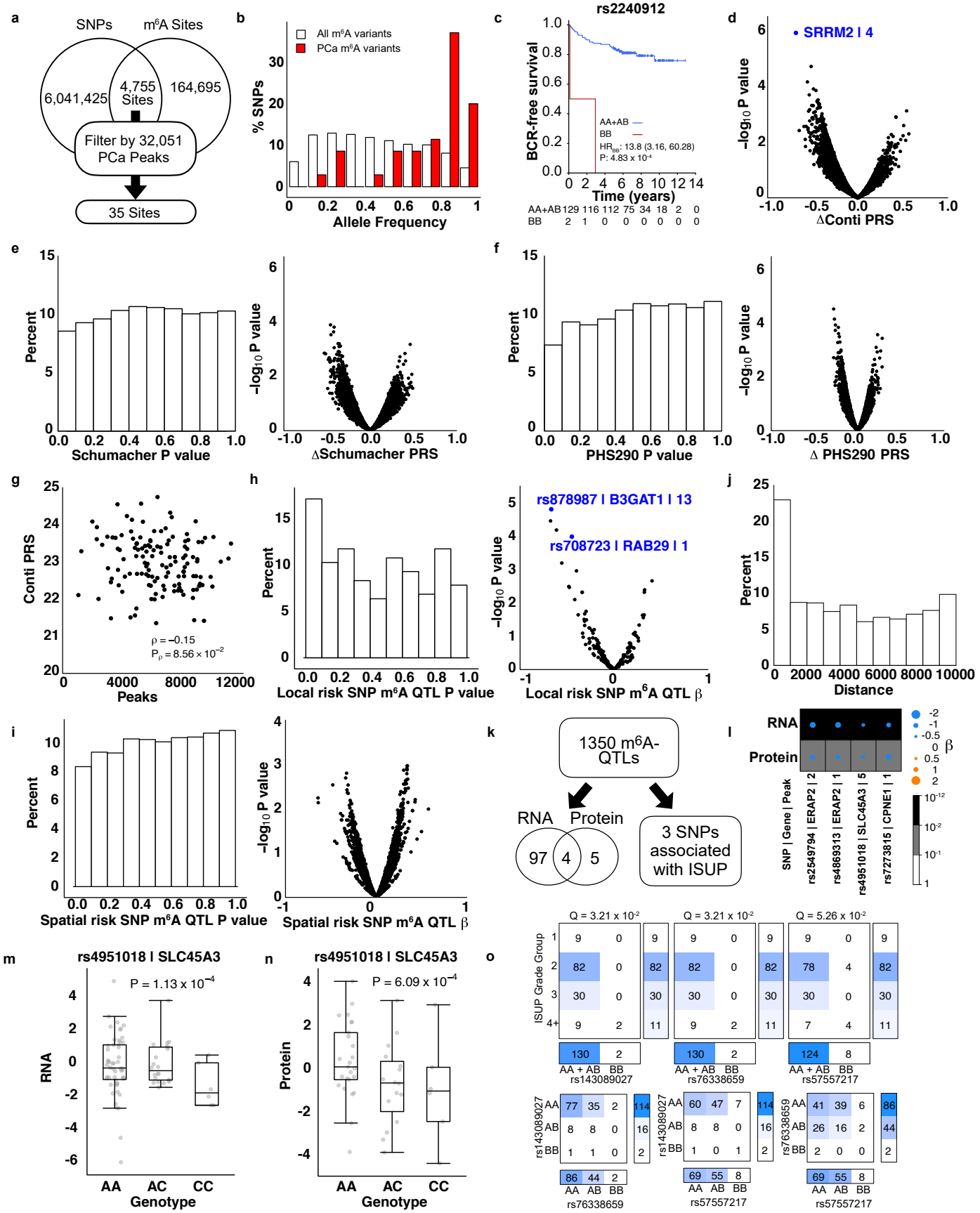
Extended Data Fig. 2 | Peak Calling & Central Dogma Integration. a) The distribution of peaks identified by exomePeak and MeTPeak. Each dot in the scatterplot represents a sample, histograms show the distribution of m^{6}A methylated genes and peaks across samples. Bottom histogram shows the distribution of the number of peaks identified per gene. **b)** Spearman's correlation of the number of peaks and the library size. The size of the disks corresponds to the magnitude of Spearman's ρ and the background color indicate the P value (two-tailed test). **c-d)** The number of peaks called using three IP libraries when randomly subsampled to 50%, 25% 10% and 1% of the original reads for exomePeak (**c**) and MeTPeak (**d**). **e)** Comparing the coverage of MeTPeak and exomePeak peaks using the Simpson Index (intersection/smallest set, median = 94.80%) and the Jaccard Index (intersection/union, median = 53.01%) for each sample. Inner panel: Overlap by genomic coverage with previously identified peaks in the REPIC database[26]. **f)** Distribution of peak density across a metagene as called by MeTPeak. Line shows the density of the peaks identified by HistogramZoo while the blue background shows the variation of peak density across individual samples. Bottom plot shows the distribution of peaks annotated across transcriptomic elements. **g-j)** Exemplar plots for *MALAT1* (**g**), *FOXA1* (**h**), *RB1* (**i**) and *NKX3-1* (**j**). The median IP (green) and Input (purple) coverage are represented using lines while background colors represent the range of IP and Input coverage across samples. Exons are annotated below in dark blue. Bottom heatmaps represent the distribution of IP and Input coverage (log1p-transformed) across samples. **k)** Relationship between the number of samples with an m^{6}A peak at each gene and the corresponding gene-level RNA abundance. Box plots represent the median (center line) and upper and lower quartiles (box limits), and whiskers extend to the minimum and maximum values within 1.5x the IQR. **l)** Distribution of spearman's correlation between m^{6}A level and input abundance. **m-n)** Relationship between gene-level RNA and protein abundance correlation and (**m**) gene-level m^{6}A abundance and (**n**) the number of samples with an m^{6}A peak at each gene. Spearman's correlation and corresponding P value displayed (n = 3,233).

https://doi.org/10.1038/s41588-025-02128-y



Extended Data Fig. 3 | Clinical Correlates of m^{6}A. a) m^{6}A peak abundance correlates with patient clinical features. **b)** *TAOK2* m^{6}A peak abundance is correlated with age. Spearman's correlation and corresponding P value displayed. **c)** Presence of IDC/CA co-occurs with greater patient peak number. P value is from a Student's t-test (n = 133). Box plots represent the median (center line) and upper and lower quartiles (box limits), and whiskers extend to the minimum and maximum values within 1.5x the IQR. **d)** Association between patient peak number and patient clinical features. Dot size represents the indicated effect size (where applicable) and background color indicates P value. Spearman's rank correlation was employed for PGA, age, and PSA, Mann Whitney U test was used for IDC status and pT, one-way ANOVA was used for analysis of ISUP Grade Group, and a Cox Proportional Hazards model was applied to analyse association between median-dichotomized sample peak number and biochemical recurrence. **e)** Association between patient peak number and mutations in prostate cancer driver genes. Samples are ranked by the number of peaks identified in descending order. Barplot indicates the statistical significance of each association, assessed *via* U-test.

https://doi.org/10.1038/s41588-025-02128-y

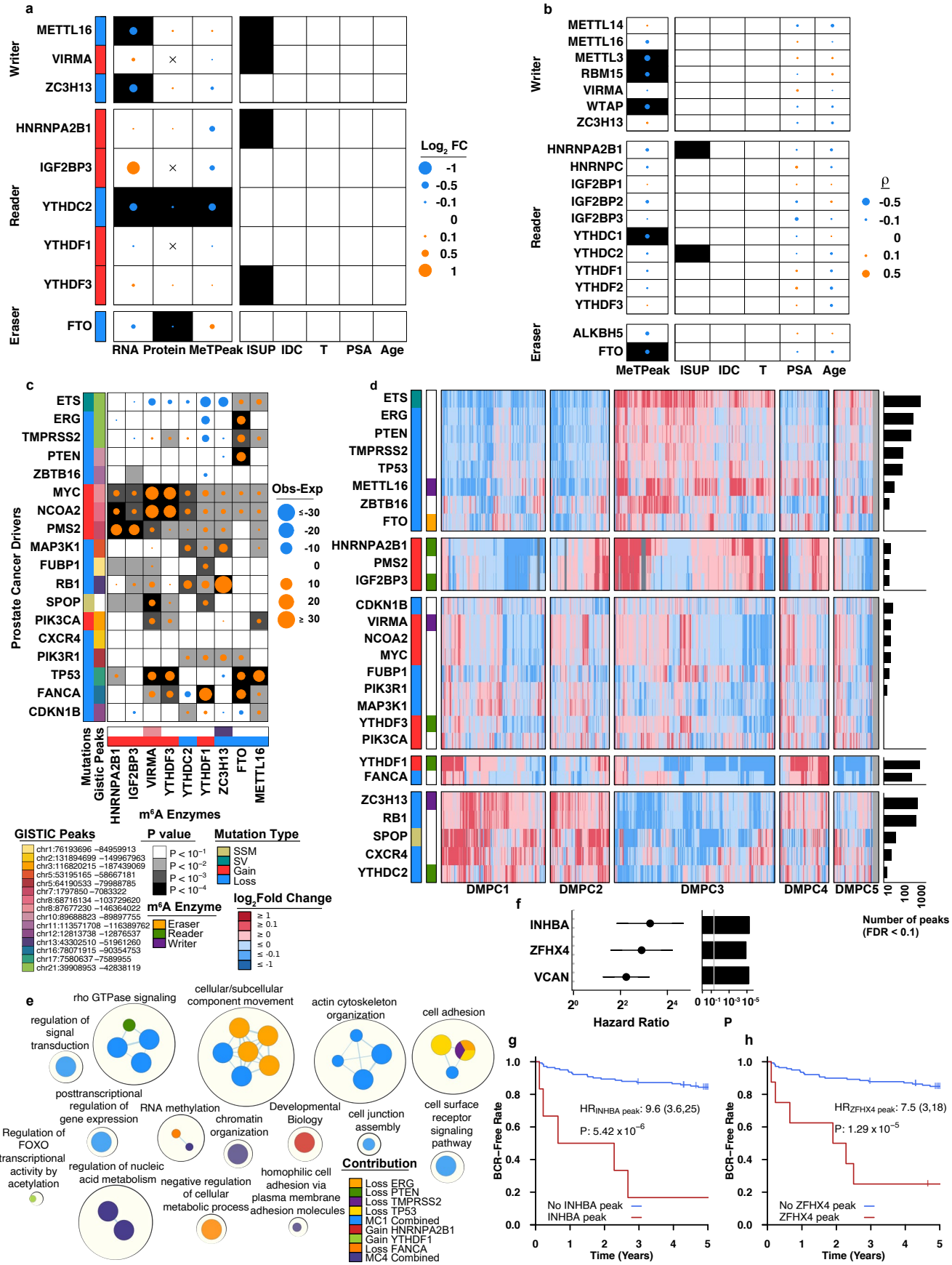


Extended Data Fig. 4 | See next page for caption.

https://doi.org/10.1038/s41588-025-02128-y

Extended Data Fig. 4 | Germline Correlates of m^{6}A. a) The overlap of m^{6}A sites identified in RMVar[42] and Cotter et al. 2021[43] with germline SNPs genotyped as described in Houlahan et al.[18]. **b)** The distribution of allele frequencies of the A for SNPs/m^{6}A sites in prostate cancer joint peaks (red) compared to the A for SNPs/m^{6}A sites outside of joint peaks (white). **c)** The SNP rs2240912 is associated with biochemical recurrence (Q < 0.1; Cox Proportional Hazards model). **d)** The difference in Conti PRS relative to m^{6}A peak status (PRS_{Peak} - $PRS_{No\ Peak}$) across peaks assessed using Student's T test (two-sided). **e-f)** Left: The distribution of P values in associations (two-sided T test) between m^{6}A peak status and Schumacher PRS (**e**) and PHS290 (**f**). Right: Volcano plots showing $\log_{10}$ P values *versus* effect sizes (Δ PRS). **g)** The association between Conti PRS *versus* the number of MeTPeak peaks per sample using Spearman's correlation analysis (two-sided). **h-i)** Left: distribution of P values in modelling associations between risk SNPs and local m^{6}A peak methylation (**h**) and spatially-associated m^{6}A peak methylation (**i**) Right: Volcano plots of P values *versus* effect sizes from additive linear models (β). Data from linear additive model. **j)** Histogram of the distance (bps) between the m^{6}A-QTL and associated m^{6}A peak. **k)** The identification of significant RNA, protein and clinical associations based on tag SNPs and gene pairs of m^{6}A-QTLs. **l)** RNA and protein associations identified using m^{6}A-QTL tag SNP-gene pairs. The disk sizes and color are indicative of the magnitude and direction of the β values of a linear additive model while the background color indicates the Q value. **m-n)** The associations of the SNP rs4951018 with *SLC45A3* RNA (**m**) and protein abundance (**n**). Box plots represent the median (center line) and upper and lower quartiles (box limits), and whiskers extend to the minimum and maximum values within 1.5x the IQR. P values from linear additive model (n = 133). **o)** Top: tables showing associations between ISUP-associated alleles and ISUP Grade Group. Q values are derived from the global X^2 test. Bottom: tables showing associations between pairs of alleles.

https://doi.org/10.1038/s41588-025-02128-y

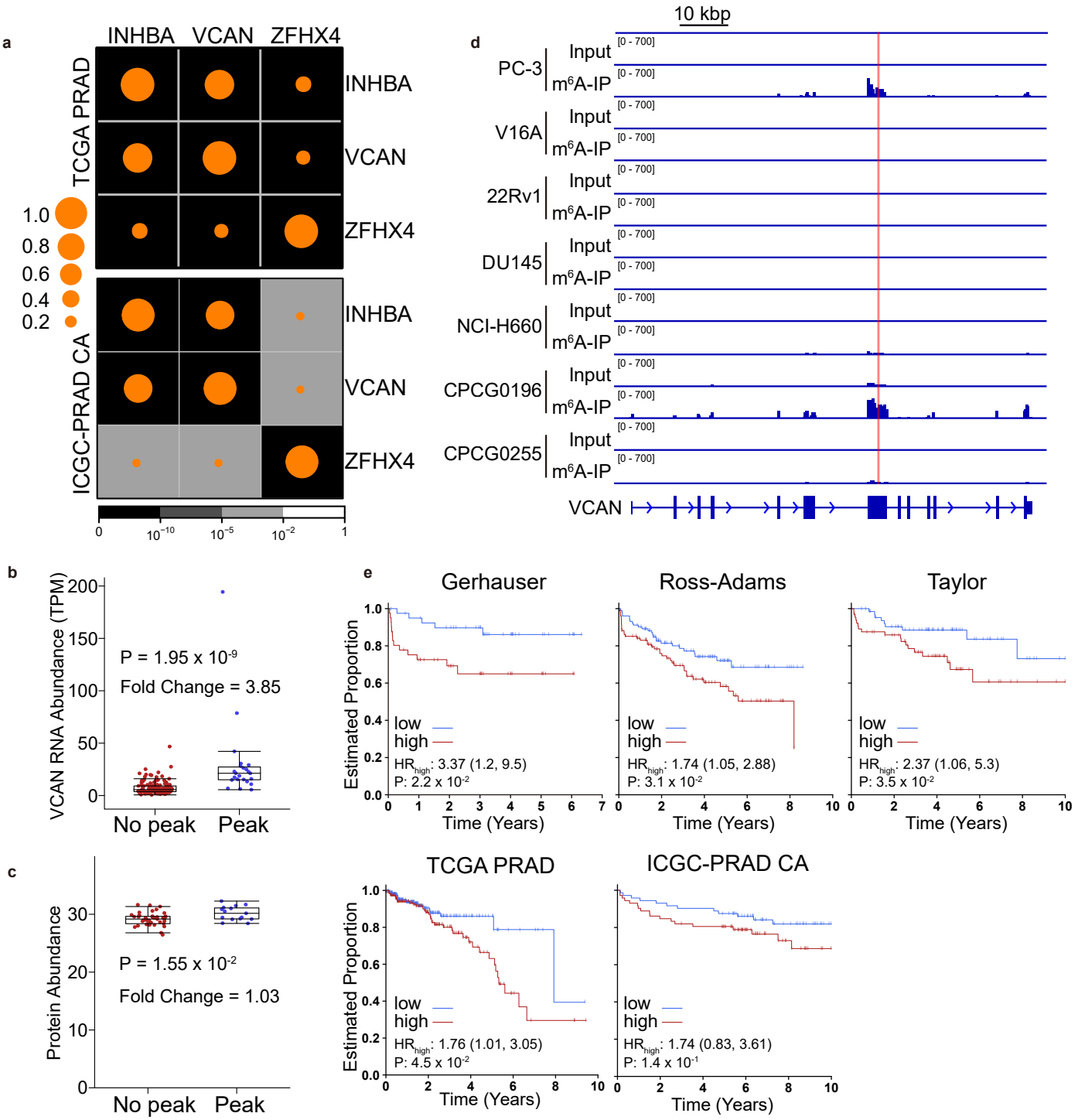


Extended Data Fig. 5 | See next page for caption.

https://doi.org/10.1038/s41588-025-02128-y

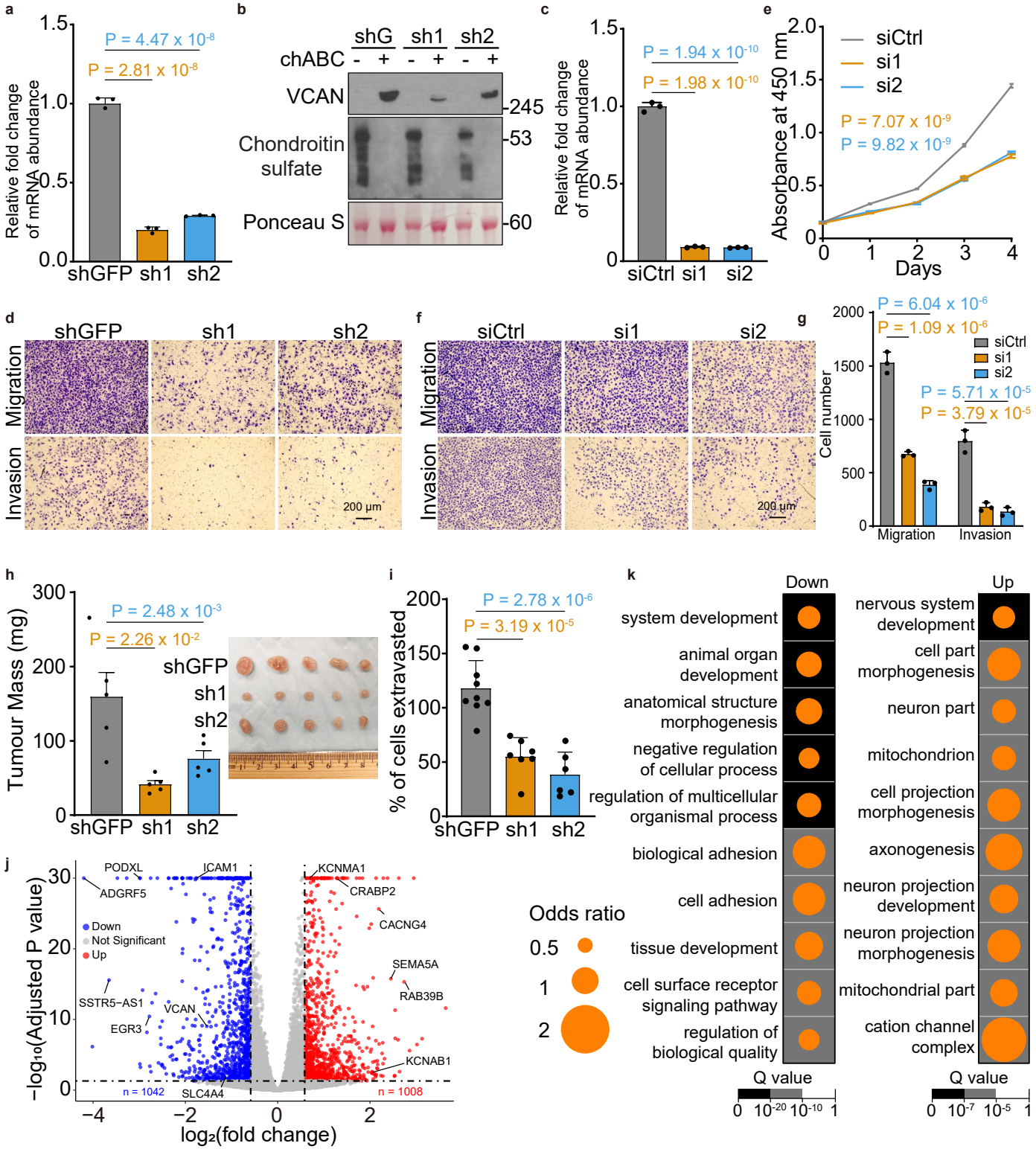
Extended Data Fig. 5 | Somatic Mutations Modulate the m^{6}A Landscape. **a-b)** Association of (**a**) copy number aberrations or (**b**) RNA abundance with 1) RNA or protein abundance of the corresponding gene, or total sample peak number (Student's t-test, two-sided), 2) clinical features: ISUP Grade Group, IDC status, pathological T category (Pearson's X^2 test, two-sided), PSA, and age (Student's t-test, two-sided). Black background indicates Q value < 0.1, whilst dot size represents (**a**) effect size or (**b**) Spearman's correlation. **c)** Association of mutations in m^{6}A regulatory genes and prostate cancer driver events. P values from a hypergeometric test (one-sided for over-representation) are represented by the color of the background, whilst the size of the dot represents the difference between observed and expected co-occurring mutations. Covariate bars indicate mutation type and, if applicable, GISTIC peak the mutation is located within. **d)** Consensus clustering of 3,432 differentially methylated m^{6}A peaks (k = 5) associated with driver mutations or mutations to m^{6}A regulators (k = 5). For each mutation, colors in the main heatmap indicate the $\log_2$ fold change in m^{6}A level, and right barplot the number of m^{6}A peaks significantly differentially m^{6}A methylated, between samples with and without the corresponding mutation. MC: mutation cluster; DMPC: differentially-methylated peak cluster. **e)** Pathway enrichment analysis performed on genes with differential m^{6}A abundance associated with mutations in m^{6}A enzymes or prostate cancer drivers. Each node represents an enriched pathway term (from GO:BP or REACTOME ontologies). Node colors represent mutations from (**d**), with some pathways only enriched through joint consideration of all genes associated with a whole mutation cluster (indicated through 'Contribution' key). **f)** Forest plot showing the hazard ratio and 95% confidence intervals for presence of an m^{6}A peak associated with biochemical recurrence, and left barplot indicates corresponding P values from a Cox Proportional Hazards model. **g-h**) Presence of a m^{6}A peak on **g)** *INHBA* or **h)** *ZFHX4*, increases risk of biochemical recurrence. *INHBA* peak samples n = 6, *ZFHX4* peak samples n = 9, total samples = 148. P values from a Cox Proportional Hazards model.

https://doi.org/10.1038/s41588-025-02128-y



Extended Data Fig. 6 | *VCAN* as a Candidate m6A Driver Event. a) Correlations in RNA abundance of *INHBA*, *VCAN*, and *ZFHX4* in the TCGA-PRAD and ICGC PRAD-CA cohorts. Dot size is Spearman's correlation; background shading is P value. **b)** Comparison of the *VCAN* mRNA abundance between samples with and without peaks by the Input RNA-seq data. Samples with m^{6}A peaks (n = 23) and without m^{6}A peaks (n = 124) were derived from independent biological samples. The P value was calculated with two-sided Mann-Whitney U test between samples with and without m^{6}A peaks on *VCAN* mRNA. The top and bottom of the box represent the 75th and 25th percentiles, respectively, with the line inside the box indicating the median. Whiskers extend to 1.5 times the interquartile range from the 25th and 75th percentiles. The fold change of peak mean TPM relative to no peak mean TPM is shown. **c)** Comparison of the VCAN abundance between samples with and without peaks by protein mass spectrometry data. Samples with m^{6}A peaks (n = 15) and without m^{6}A peaks (n = 39) were derived from independent biological samples. A two-sided Mann-Whitney U test was used to compare the difference between the two groups. The top and bottom of the box represent the 75th and 25th percentiles, respectively, with the line inside the box indicating the median. Whiskers extend to 1.5 times the interquartile range from the 25th and 75th percentiles. The fold change of peak mean value relative to no peak mean value is shown. **d)** m^{6}A peak signals of *VCAN* transcripts from meRIP-seq libraries of several prostate cancer cell lines and patient samples with (CPCG0196) or without (CPCG0255) *VCAN* m^{6}A peaks identified. The vertical red line indicates the location of the significant peak. **e)** Kaplan-Meier survival curves showing biochemical recurrence stratified by *VCAN* mRNA abundance in the Gerhauser, Ross-Adams, Taylor, TCGA-PRAD and ICGC-PRAD CA cohorts. A Cox Proportional Hazards model was used to compare the hazard of biochemical recurrence between the groups.

https://doi.org/10.1038/s41588-025-02128-y

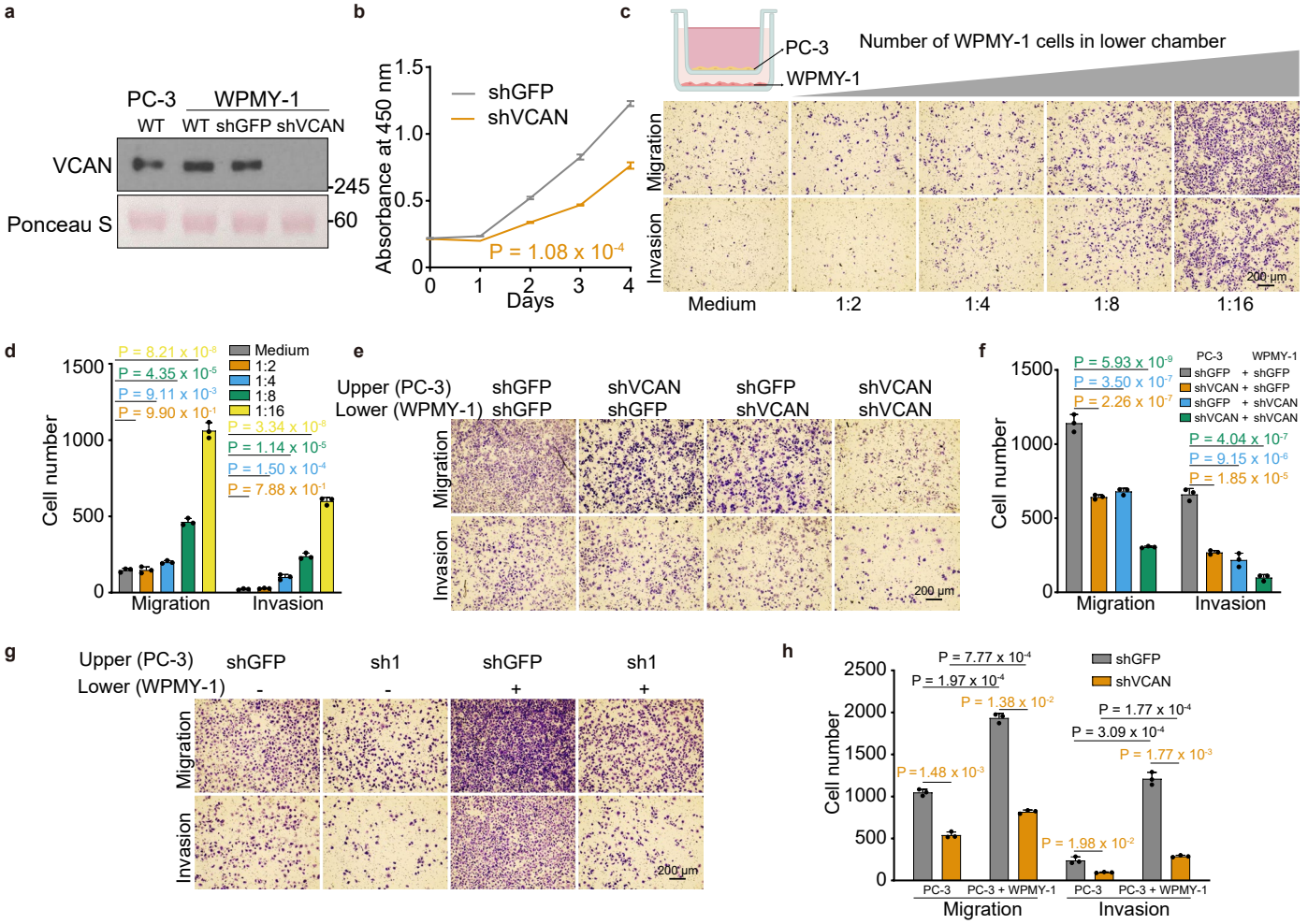


Extended Data Fig. 7 | See next page for caption.

https://doi.org/10.1038/s41588-025-02128-y

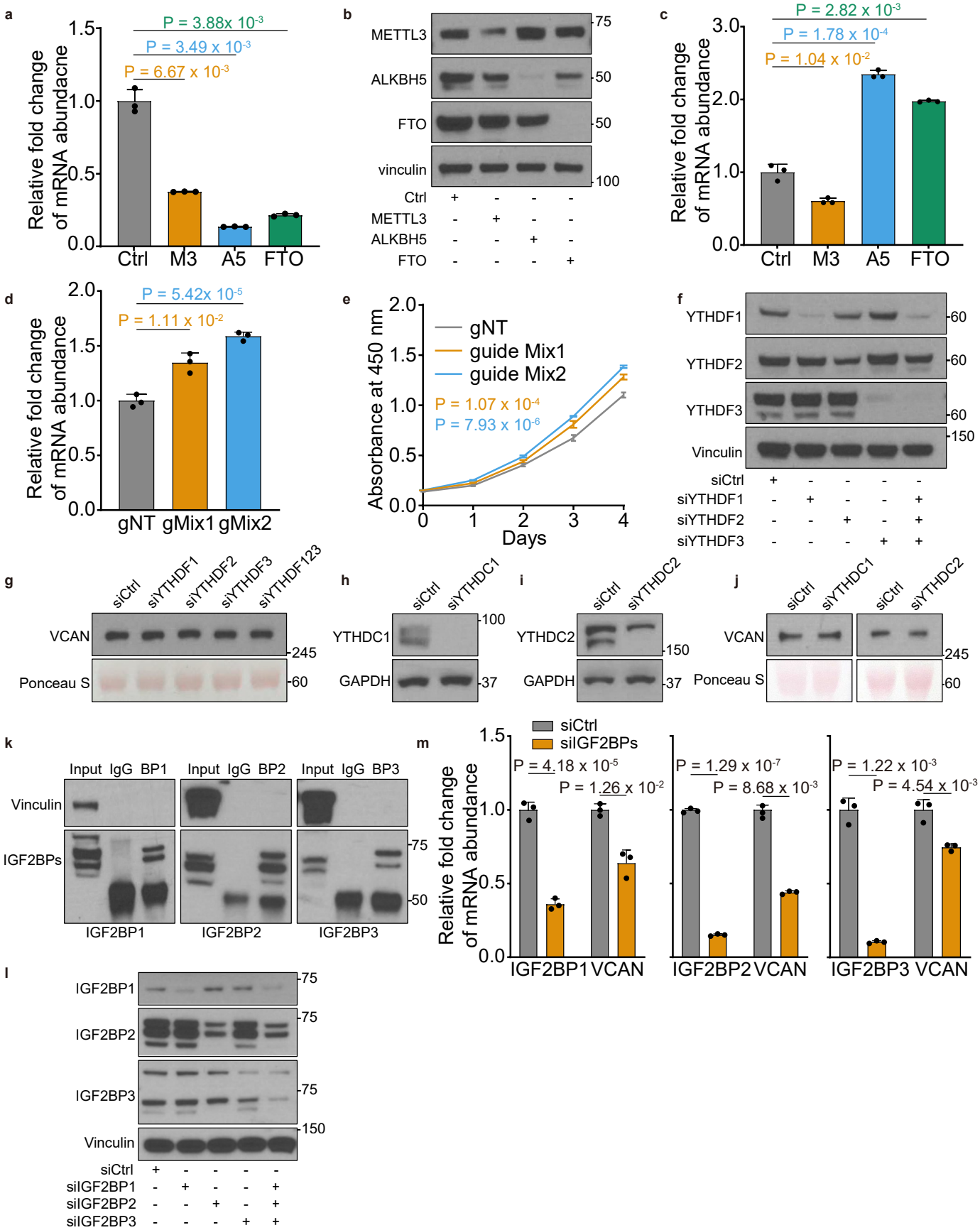
Extended Data Fig. 7 | *VCAN* Drives Prostate Cancer Progression. **a-c)** Knockdown efficiency of *VCAN* mRNA by shRNAs (**a**), protein abundance (**b**) and by siRNAs (**c**) in PC-3 cells. All other Western blots for VCAN in this study are treated by chABC by default. Ponceau S as the loading control. Numbers indicate the positions of molecular mass (kDa) standards. Data are represented as mean ± SD of three biological replicates. P values are calculated using one-way ANOVA with Dunnett's multiple comparisons test. **d)** The representative images of the migration and invasion assay of PC-3 cells after shRNA-mediated knockdown of *VCAN*. The experiment was conducted using three biological replicates for each condition. The scale bar represents 200 μm. **e)** The growth curve after siRNA-mediated knockdown of *VCAN* in PC-3. Data are represented as mean ± SD of three biological replicates. P values are calculated using one-way ANOVA with Dunnett's multiple comparisons test. **f-g)** The migration and invasion assay of PC-3 cells after siRNA-mediated knockdown of *VCAN*. The representative images (**f**), and the quantification (**g**). Data are represented as mean ± SD of three biological replicates. The scale bar represents 200 μm. **h-i)** The tumor weight (left) and the tumor size (right) of xenografts derived from injected PC-3 cells (**h**), and the extravasation ability of PC-3 (**i**) infected with shRNAs targeting *VCAN* or GFP as the control. Xenografts are shown as mean ± SEM of five biological replicates, and the extravasation data are presented as mean ± SD of biological replicates (shGFP, n = 9; *VCAN* sh1, n = 7; *VCAN* sh2, n = 6). P values are calculated using one-way ANOVA with Dunnett's multiple comparisons test. **j)** The volcano plot showed the significantly up/down-regulated genes (average FC > $\log_2(1.5)$, adjusted P value < 0.05, two-sided) between shGFP and shKD *VCAN* groups in PC-3 mouse xenografts. Adjusted P values were calculated using the Benjamini-Hochberg method. Vertical dashed lines indicate $-\log_2(1.5)$ and $\log_2(1.5)$, respectively, and the horizontal one marks adjusted P = 0.05. **k)** Dotmap representing top 10 enriched pathways after shKD *VCAN* in PC-3 mouse xenograft tumors. Dot size indicates the odds ratio and background color the Q values.

https://doi.org/10.1038/s41588-025-02128-y



Extended Data Fig. 8 | VCAN Regulates Tumor-Stromal Crosstalk. **a)** Knockdown efficiency of *VCAN* by shRNA in WPMY-1 cells was detected by Western blot. Ponceau S stain serves as the loading control, and molecular mass (kDa) standards are indicated on the right. Additionally, a sample from PC-3 cells is included for relative quantification, demonstrating comparable levels of VCAN extracted from the same volume of culture medium collected from an equal number of cells. The experiment was conducted using three biological replicates for each condition. **b)** The growth curve after shRNA-mediated knockdown of *VCAN* in WPMY-1. Data are represented as mean ± SD of three biological replicates. P value is calculated using an unpaired two-sided Student's T test. **c-f)** Migration and invasion analysis of indirect coculture of PC-3 and WPMY-1 cells. PC-3 cells (2×10^4) were seeded into the upper chamber containing serum-free medium, while WPMY-1 cells were seeded in the lower chamber containing 10% fetal bovine serum at varying ratios (PC-3 to WPMY-1) ranging from 1:2 to 1:16 as indicated (**c-d**). PC-3 cells (5×10^4) were seeded into the upper chamber containing serum-free medium, while WPMY-1 cells were seeded in the lower chamber containing 10% fetal bovine serum at a fixed ratio of 1:8 (PC-3 to WPMY-1) (**e-f**). Representative images (**c, e**) and quantification (**d**, **f**, mean ± SD of three biological replicates; one-way ANOVA with Dunnett's multiple comparisons test) are presented. The scale bar represents 200 µm. **g-h)** PC-3 cells with *VCAN* knockdown (shVCAN) or control (shGFP) were cultured either with or without WPMY-1 cells. In the coculture condition, WPMY-1 cells were seeded in the lower chamber at a ratio of 1:8 (PC-3 to WPMY-1) in medium containing 10% fetal bovine serum (FBS); in the monoculture condition, the lower chamber contained only medium with 10% FBS. Representative images (**g**) and quantification (**h**, mean ± SD of three biological replicates; two-sided Student's t-test) are presented. The treatments (knockdown) for each group are indicated in the figure. The scale bar represents 200 µm.

https://doi.org/10.1038/s41588-025-02128-y

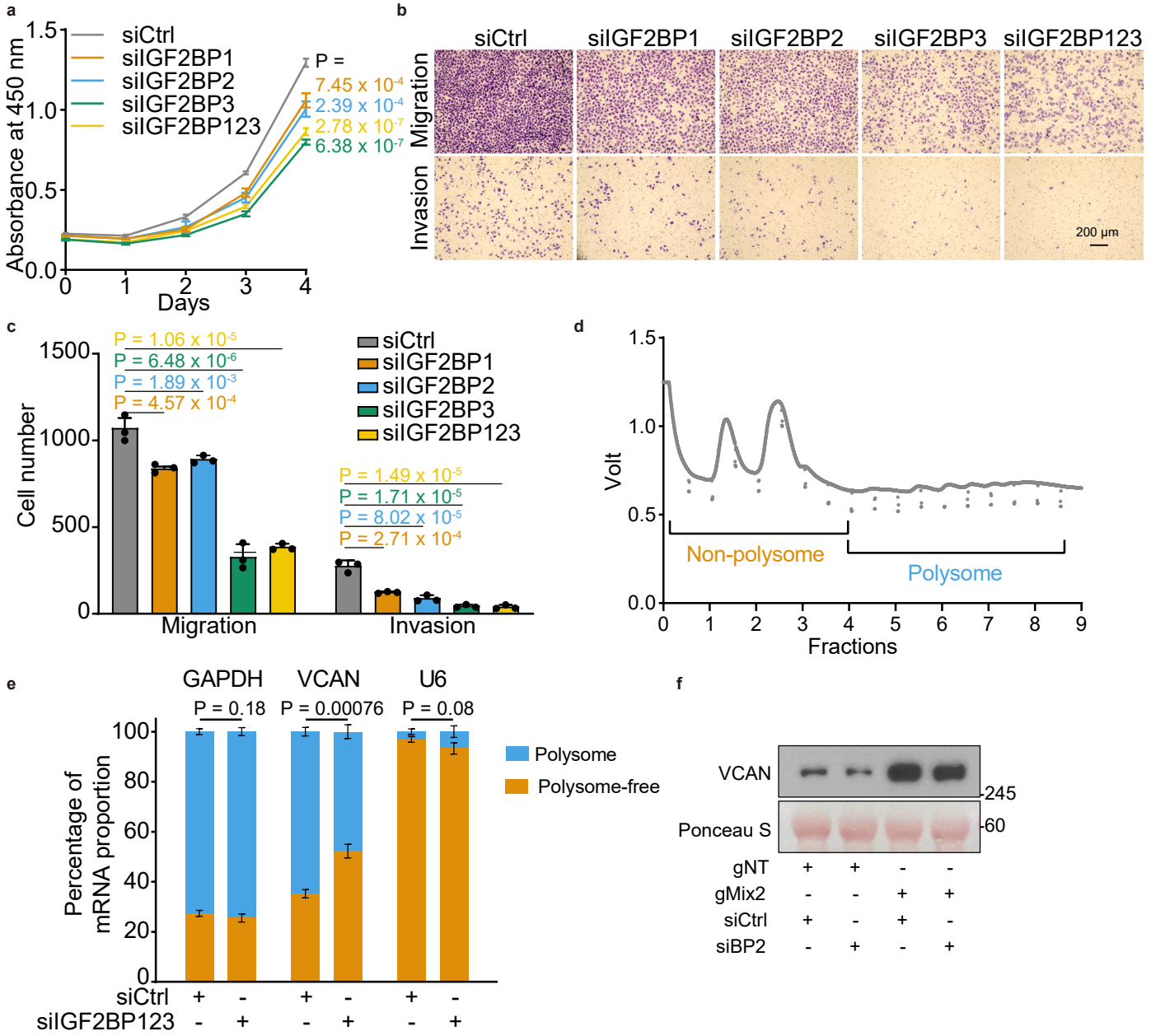


Extended Data Fig. 9 | See next page for caption.

https://doi.org/10.1038/s41588-025-02128-y

Extended Data Fig. 9 | m^{6}A Stabilizes *VCAN* mRNA and Promotes its Translation. a-b) Knockdown efficiencies of the m^{6}A regulators by shRNAs in PC-3 at RNA (**a**, Data are represented as mean ± SD of three biological replicates. P values are calculated using unpaired two-sided T test.) and protein level (**b**, Numbers indicate the positions of molecular mass (kDa) standards). **c)** Relative abundance of *VCAN* mRNA after knockdown m^{6}A regulators in PC-3 cells. Ctrl:shRNA targets GFP, M3:*METTL3*, A5:*ALKBH5*. Data are represented as mean ± SD of three biological replicates; P values, unpaired Student's T test. **d-e)** The *VCAN* mRNA abundance (**d**) and growth curve (**e**) of PC-3 cells after writing by *VCAN* mRNA-targeting guide RNA mixes. Data are represented as mean ± SD of three biological replicates. P values, one-way ANOVA with Dunnett's multiple comparisons test. **f-g)** Knockdown efficiencies of YTHDFs (**f**) and corresponding VCAN protein abundance change (**g**) in PC-3 cells infected with siRNAs against each YTHDF were assessed by western blot. Numbers indicate the positions of molecular mass (kDa) standards. **h-i)** Knockdown efficiencies of YTHDC1 (**h**) and YTHDC2 (**i**) in PC-3 cells infected with siRNAs against each YTHDC were assessed by western blot. Vinculin was used as a loading control. Numbers on the right indicate the positions of molecular mass (kDa) standards. **j)** Western blot showing VCAN protein abundance change after knocking down YTHDC1 and YTHDC2 in PC-3 cells. Ponceau S stain serves as the loading control. Numbers on the right indicate the positions of molecular mass (kDa) standards. **k)** RIP efficiencies of the IGF2BP proteins were evaluated by western blot. Vinculin was included as negative control of the IPs. Numbers on the right indicate the positions of molecular mass (kDa) standards. **l)** Knockdown efficiencies of IGF2BPs in PC-3 cells infected with siRNAs against each IGF2BP were assessed by western blot. Vinculin was used as a loading control. Numbers on the right indicate the positions of molecular mass (kDa) standards. **m)** Relative *VCAN* mRNA abundance after siRNA-mediated knockdown of IGF2BPs in PC-3 cells. Data are represented as mean ± SD of three biological replicates. P values are from unpaired T test.

https://doi.org/10.1038/s41588-025-02128-y



Extended Data Fig. 10 | m^{6}A Regulates *VCAN* Expression via IGF2BPs. **a)** Cell proliferation after siRNA-mediated knockdown of IGF2BPs in PC-3 cells. Data are represented as mean ± SD of three biological replicates. P values are calculated using one-way ANOVA with Dunnett's multiple comparisons test. **b-c)** Migration and invasion of PC-3 cells after siRNA-mediated knockdown of IGF2BPs in PC-3 cells. P values are calculated using one-way ANOVA with Dunnett's multiple comparisons test. The representative images (**b**), and the quantification (**c**). Data are represented as mean ± SD of three biological replicates. The scale bar represents 200 μm. **d)** Ribosome profiling analysis of sample fractions. The graph illustrates ribosome profiling, with the intensity of the signal (voltage) plotted on the y-axis and sample fractions collected on the x-axis. The fractions are grouped based on the presence of polysomes, with polysome and polysome-free fractions distinguished and labeled accordingly on the graph. **e)** Relative enrichment of *VCAN*, *GAPDH*, and *U6* in the polysome-free and polysome-associated fractions of PC-3 cells were measured by qPCR. *GAPDH* and *U6* were employed as controls. Data are represented as mean ± SD of three biological replicates. P values are calculated using an unpaired two-sided Student's T test. **f)** VCAN protein abundance after dCasRx-METTL3-based m^{6}A writing with or without knocking down of *IGF2BP2* was detected by western blot. Ponceau S stain serves as the loading control. Numbers on the right indicate the positions of molecular mass (kDa) standards.

nature portfolio

Corresponding author(s): Housheng Hansen He
Paul C. Boutros

Last updated by author(s): Oct 4, 2024

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies 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	qPCRs were run on StepOnePlus Real-Time PCR System (Applied Biosystems) or CFX96 Touch Real-Time PCR Detection System (Bio-Rad). The Western blots were exposed on Thermo Scientific CL-XPosure film. The average optical density of VCAN was subsequently analyzed utilizing HALO software (Indica Labs).
Data analysis	All data analysis details, including method and detailed parameters are described in Methods. Customized code for data analyzing and plotting will be available upon request. Software used: STAR (v2.7.2),FastQC (v0.11.8), RSeQC (v3.0.1), Picard (v2.21.7), DESeq2 (v1.26.0), MeTPeak (v1.1), exomePeak (v2.16.0), DownsampleSam (v2.21.7), HOMER (v4.1.1),HistogramZoo (v1.0.1), subread (v2.0.0), RSEM (v1.3.2), Entropy (v1.2.1), gprofiler2 R package (v0.2.0), ConsensusClusterPlus R package (v1.58.0), ANNOVAR (v2018-04-16), dbSNP database (v150), Plink (v1.9.0), R package Survival (v3.2-13), GATK ASEReadCounter (v4.1.8.1), clusterProfiler (v3.14.3), cutadapt (v2.8), Michigan Imputation Server (v1.7.3), SnpEff (v5.1), MatrixEQTL (v2.3), QTLtools (v1.2), ActivePathways (v1.1.0), survival R package (v3.3.1), R Survival package (v2.41-3), meta R package (v6.5.0), TCGAbiolinks R package (v2.25.3), R package TCGA2STAT (v1.2), BoutrosLab.plotting.general (v6.0.3), ggseqlogo (v0.1), and Guitar (v2.4.0).

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Bulk RNA-seq data of Gerhauser et al. and Taylor et al. were obtained from https://www.cbioportal.org/study/summary?id=prostate_dkfz_2018 and https://www.cbioportal.org/study/clinicalData?id=prad_mskcc, respectively. The Ross-Adams (GSE70770) and ICGC-PRAD CA (GSE113120) were downloaded from Gene Expression Omnibus. Raw sequencing data is available in EGA under accession EGAS00001006925. Previously published data are available in EGA under accession EGAS00001000900. The PC-3 siRNA knockdown VCAN RNA-seq data is deposited in the Gene Expression Omnibus (GEO) under the accession number GSE220432. Previously published mRNA microarray data are available at the Gene Expression Omnibus under accession code GSE84043.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

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

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

Life sciences study design

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

Sample size	MeRIP-seq was performed for 162 patient tumor and 10 benign RNA samples. m6A-SAC-seq was performed for four cell line and eight tumor samples. RNA-seq were performed with the sample size of three for each cell line at each condition. RT-qPCR were performed with biological replicates. 5-10 mice were used for each treatment arms. Sample sizes were not predetermined by any statistical methods. Generally, sample size was based on our prior experience and common standards in the field for detecting statistically significant differences between conditions, and also based on experimental feasibility, sample availability, and necessary to obtain definitive results.
Data exclusions	All data were used without exclusions.
Replication	RNA-seq were performed with the sample size of three for each cell line at each condition. RT-qPCR and western blots were performed with three biological replicates. 5 mice were used for each group.
Randomization	Randomization not deemed essential for this study because the fundamental findings were made using cell lines, since they are divided and expanded from a relatively homogeneous cell population. For xenograft, the mice were randomly assigned to receive subcutaneous injections of either control or treatment group cells.
Blinding	Blinding was not possible to our study, because of the time constraint and the limitation of the human resource for animal work.

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

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

anti-anti-VCAN: Abcam, ab270444;
anti-vinculin: CST, 13901S;
anti-IGF2BP1: proteintech, 22803-1-AP, or CST, 8482;
anti-IGF2BP2: proteintech, 11601-1-AP;
anti-IGF2BP3: proteintech, 14642-1-AP;
anti-Chondroitin Sulfate: Sigma, C8035;
anti-METTL3: proteintech, 15073-1-AP;
anti-ALKBH5: Sigma, HPA007196;
anti-FTO: proteintech, 27226-1-AP;
anti-YTHDF1: proteintech, 17479-1-AP;
anti-YTHDF2: proteintech, 24744-1-AP;
anti-YTHDF3: proteintech, 25537-1-AP;
anti-YTHDC1: proteintech, 29441-1-AP;
anti-YTHDC2: proteintech, 27779-1-AP;
anti-mouse IgG: CST, 7076S; Secondary antibody, 1:5,000 dilution
anti-rabbit IgG: CST, 7074S; Secondary antibody, 1:5,000 dilution

Validation

Validation statements and literature citations of each antibody can be found on the manufacturer's website and/or the CiteAb website:
anti-VCAN (https://www.abcam.com/products/primary-antibodies/versican-antibody-epr23374-21-ab270444.html)
anti-vinculin (https://www.cellsignal.com/products/primary-antibodies/vinculin-e1e9v-xp-rabbit-mab/13901)
anti-IMP1(IGF2BP1) (https://www.cellsignal.com/product/productDetail.jsp?productId=8482)
anti-IGF2BP1 (https://www.ptglab.com/products/IGF2BP1-Antibody-22803-1-AP.htm)
anti-IGF2BP2 (https://www.ptglab.com/products/IGF2BP2-Antibody-11601-1-AP.htm)
anti-IGF2BP3 (https://www.ptglab.com/products/IGF2BP3-Antibody-14642-1-AP.htm)
anti-Chondroitin Sulfate (https://www.sigmaaldrich.com/CA/en/product/sigma/c8035)
anti-METTL3 (https://www.ptgcn.com/products/METTL3-Antibody-15073-1-AP.htm)
anti-ALKBH5 (https://www.sigmaaldrich.com/CA/en/product/sigma/hpa007196)
anti-FTO (https://www.ptgcn.com/products/FTO-Antibody-27226-1-AP.htm)
anti-YTHDF1 (https://www.ptgcn.com/products/YTHDF1-Antibody-17479-1-AP.htm)
anti-YTHDF2 (https://www.ptgcn.com/products/YTHDF2-Antibody-24744-1-AP.htm)
anti-YTHDF3 (https://www.ptgcn.com/products/YTHDF3-Antibody-25537-1-AP.htm)
anti-YTHDC1 (https://www.ptgcn.com/products/YTHDC1-Antibody-29441-1-AP.htm)
anti-YTHDC2 (https://www.ptgcn.com/products/YTHDC2-Antibody-27779-1-AP.htm)
anti-mouse IgG (https://www.cellsignal.com/products/secondary-antibodies/anti-mouse-igg-hrp-linked-antibody/7076)
anti-rabbit IgG (https://www.cellsignal.com/products/secondary-antibodies/anti-mouse-igg-hrp-linked-antibody/7074)

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

HEK293T (Cat# CRL-3216), PC-3 (Cat# CRL-1435), 22Rv1 (Cat# CRL-2505), DU145 (Cat# HTB-81), NCI-H660 (Cat# 5813) and WPMY-1 (Cat# CRL-2854) were purchased from the American Type Culture Collection (ATCC). V16A cell line, originally established by Dr. Amina Zoubeidi's laboratory, was provided as a gift from her lab.

Authentication

HEK293T, PC-3, 22Rv1, DU145, NCI-H660, WPMY-1 and V16A cells have been authenticated by STR analysis, as declared by ATCC.

Mycoplasma contamination

HEK293T, PC-3, 22Rv1, DU145, NCI-H660, WPMY-1 and V16A cells have been tested negative for Mycoplasma contamination by the MycoAlert Mycoplasma Detection Kit (Cat. #LT07-118, Lonza).

Commonly misidentified lines
(See ICLAC register)

No

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Four to six-week old NOD/SCID and BALB/c mice were obtained from Princess Margaret Cancer Centre Animal Research Centre and housed under standard temperature at 21C-22C, humidity at 45%-60%, and timed lighting conditions at 12hr/12hr light/dark cycle mandated by the committee.
Wild animals	No wild animals were used in this study
Reporting on sex	Our finding only apply to male. The sex was not considered in study design because we are focusing on prostate cancer.
Field-collected samples	No field collected samples were used in the study
Ethics oversight	All procedures were performed in accordance with the International Guidelines for the Use of Animals and approved by the Animal Care Committee at UHN.

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

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A