nature communications

Article

https://doi.org/10.1038/s41467-025-64213-4

Chromatin remodeling restrains oncogenic functions in prostate cancer

Received: 7 October 2024

Accepted: 11 September 2025

Published online: 16 October 2025

Check for updates

Valentina Rosti[1,2,11] ✉, Giovanni Lembo[3,11], Cristiano Petrini[3,11], Francesca Gorini[2], Roberto Quadri[2], Chiara Cordiglieri[2], Margherita Mutarelli[4], Elisa Salviato[3], Elisabetta Casari[5,6], Emanuele Di Patrizio Soldateschi[1,2], Emanuele Montanari[7,8], Giancarlo Albo[7,8], Francesco Ripa[7], Alessandra Fasciani[2], Mariacristina Crosti[2], Elisa De Lorenzis[7,8], Marco Maggioni[5], Valentina Vaira[5,6], Maria Vivo[9], Francesco Ferrari[3,10,12] ✉ & Chiara Lanzuolo[1,2,12] ✉

Primary prostate cancer presents with multifocal lesions and unpredictable clinical behavior, posing significant challenges for effective prognosis. To address this, we investigate the epigenomic landscape of prostate tumor biopsies from 25 treatment-naïve male patients by analyzing chromatin compartmentalization patterns. Our analysis reveals two distinct molecular subtypes: one with a Low Degree of Decompartmentalization (LDD) and another with a High Degree of Decompartmentalization (HDD). Here we show that the HDD subgroup exhibits extensive chromatin reorganization associated with diminished oncogenic potential. This subtype shows repression of molecular pathways involved in extracellular matrix remodeling and cellular plasticity. From this distinction, we derive an 18-gene transcriptional signature capable of differentiating HDD from LDD cases. Importantly, we validate the prognostic relevance of this signature in multiple independent cohorts totaling over 900 patients. Our findings suggest that epigenetic-derived signature at the time of diagnostic biopsy can offer a powerful tool for risk stratification in prostate cancer.

Prostate cancer (PCa) is the second most common tumor type among males and accounts for 10% of cancer-related deaths[1]. While the majority of PCa cases present as indolent and slow-growing, about 20% of prostate tumors progress to metastatic and lethal forms. Currently, PCa diagnosis relies on histological evaluation of multiple biopsies, usually conducted following elevated prostate-specific antigen (PSA) levels or abnormal digital rectal examination (DRE)[2]. Despite the use of tumor grading systems, the clinical variability and frequent multifocality of PCa complicate the accurate prediction of cancer outcomes, hindering treatment guidance[3]. Consequently, many patients with silent PCa are over-treated[4]. Although the identification of driver coding mutations in solid tumors has significantly advanced the field

[1]ITB-CNR, Institute of Biomedical Technologies, National Research Council, Segrate, Italy. [2]INGM Istituto Nazionale Genetica Molecolare "Romeo ed Enrica Invernizzi", Milan, Italy. [3]IFOM-ETS, The AIRC Institute of Molecular Oncology, Milan, Italy. [4]ISASI-CNR, Institute of Applied Science and Intelligent Systems, National Research Council, Pozzuoli, Italy. [5]Division of Pathology, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. [6]Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy. [7]Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy. [8]Department of Urology, IRCCS Foundation Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. [9]Department of Chemistry and Biology "A. Zambelli", University of Salerno, Fisciano, Italy. [10]IGM-CNR, Institute of Molecular Genetics "Luigi Luca Cavalli-Sforza", National Research Council, Pavia, Italy. [11]These authors contributed equally: Valentina Rosti, Giovanni Lembo, Cristiano Petrini. [12]These authors jointly supervised this work: Francesco Ferrari, Chiara Lanzuolo. ✉e-mail: rosti@ingm.org; francesco.ferrari@cnr.it; chiara.lanzuolo@cnr.it

of clinical oncology[5], the translation of PCa genetic biomarkers into clinical practice has only marginally improved prognosis accuracy[6]. This limitation underscores the need for different biomarkers that can more precisely guide clinical decision-making in PCa management.

Emerging evidence suggest that epigenetic information, which regulates transcriptional plasticity, plays a critical role in tumor evolution[7]. Epigenetic dysfunctions have been shown to initiate carcinogenesis even in the absence of driver mutations[8], offering a different perspective on cancer research[9].

The three-dimensional (3D) genome architecture, which is hierarchically organized on multiple, highly regulated structural levels in the nucleus, has emerged as crucial for regulating cellular programs[9,10]. Notably, the physical separation between euchromatin, the more accessible and transcriptionally active part of the chromatin, and heterochromatin, which is compacted and gene-poor, is a hallmark of healthy cells. Morphological changes in nuclear organization, referred to as nuclear atypia, still represent a gold standard for diagnosis and staging of various cancers, including prostate cancer[11]. A recent study on metastatic prostate cancer patients reported multilevel alterations in chromatin structures associated with poor survival outcomes[12]. However, this work was performed on advanced PCa stages, thus providing limited information for earlier diagnostic samples. The same consideration applies to previous transcriptional studies performed on surgically removed prostates[13–15].

In this work, we comprehensively characterize chromatin compartment organization in PCa using our recently developed 4f-SAMMY-seq[16]. We analyze fresh needle biopsies from PCa patients at their initial clinical presentation and identify two subgroups based on the degree of large-scale chromatin remodeling. We derive an 18-gene expression signature that reflects this epigenomic landscape and asses its predictive power across multiple patient cohorts, including an independent validation set of patients' biopsies, overall encompassing more than 900 individuals. Of note, our findings reveal that extensive chromatin remodeling is associated with reduced tumor malignancy. This signature demonstrates significant potential as a tool for risk stratification in PCa and could be readily implemented in clinical practice.

Results

The prostate cancer biopsies cohort

To characterize chromatin architecture in prostate cancer patients at the time of diagnosis, we collected fresh needle biopsies from chemonaïve patients undergoing the standard diagnostic procedure. Each patient was punctured to collect 14 needle biopsy cores for histopathological examination (diagnostic biopsies) and 1 to be processed for our research project (research biopsy – Fig. 1a). Each research biopsy was divided into two parts: one-third was used for histopathological and immuno-histological analyses. At the same time, the rest of the tissue sample was enzymatically digested to obtain a cell suspension further split for epigenomic (4f-SAMMY-seq), transcriptomic (RNA-seq) and cellular composition (flow-cytometry) analyses.

Our cohort includes seven non-neoplastic controls (CTR), *i.e.* patients that turned out to be negative for cancer, and ten primary prostate cancer patients (PCa) with histologically confirmed tumor and Gleason Score (GS) between 3 + 3 and 5 + 5 (Fig. 1b). We also report in Fig. 1b the percentage of diagnostic biopsy cores positive for cancer cells (PPC), their distribution in a prostate gland diagram (DPC), and the puncture site of the research biopsy used for this study. All the research biopsies of prostate cancer patients were obtained from punctures adjacent to one or more positive diagnostic biopsy cores (Fig. 1b). In addition to the Gleason Score assigned to the patient in the clinical report (GS-P), we recorded the Gleason Score assigned to the closest diagnostic biopsy (GS-CDB) and the score assigned to the research dedicated biopsy (GS-RDB, Supplementary Fig. S1a). We noted that the histopathological examination on research biopsies did not confirm the presence of tumor cells in four out of ten cases (Supplementary Fig. S1a). However, it must be noted that the Gleason Score assignment was done on a third of the tissue biopsy and that all research biopsy samples express diagnostic PCa tumoral markers like PCA3, HPN and GOLM1[17–19] (Supplementary Fig. S1b). Therefore, we retained all ten PCa samples in our experimental analyses.

Our cohort of PCa patients is relatively homogeneous in terms of age (median 72 years; interquartile range, abbreviated as IQR, 62–89 years) and serum PSA level (median 16 ng/mL, IQR: 7.1–45.5) but heterogeneous in terms of clinical features as percentage of cancer positive biopsy cores (median: 21.4%, IQR: 21.4–39.2), clinical stage (T1c: 60%, T2a-T2b: 20%, T3: 20%) and index lesions on Magnetic Resonance Imaging (MRI) (40% with Prostate Imaging Reporting and Data System (PI-RADS) ≥ 3 and 60% with PI-RADS 1/2) (Table 1). Survival outcomes are not available due to the long clinical course of primary PCa and the short follow-up time since the biopsies were collected (median time from biopsy less than four years).

We also examined tissue samples of our cohort by confocal microscopy. Non-neoplastic controls and PCa patients' cells showed similar nuclear areas but a significant difference in nuclear circularity, with PCa patients' cells showing nuclei deformations (Supplementary Fig. S1c, d). Staining of the nuclear lamina with the Lamin A/C antibody highlighted a slight statistically significant depletion of the signal in the nuclear interior, corresponding to a parallel yet not significant decrease in the Hoechst staining (Supplementary Fig. S1e). Although descriptive, these data suggest a cancer-specific remodeling of the nuclear and genome organization in PCa cells.

When dissociating the needle biopsy tissue samples (size around 2 cm in length, Supplementary Fig. S2a, see Methods), we achieved a viable cell yield in the range of 30,000-80,000 cells per sample (Supplementary Fig. S2b).

As a further control, we examined the cell type composition in a set of 13 prostate biopsies by flow cytometry (Supplementary Fig. S2c–f). We didn't detect significant differences in the relative proportion of epithelial (EpCAM/CD326 + , CD45-), leukocyte (EpCAM/CD326-, CD45 +) or stromal cells (EpCAM/CD326-, CD45-) between non-neoplastic controls and PCa samples (Supplementary Fig. S2g). All samples showed a major fraction of stromal cells (average 68,8% controls and 70,9% PCa patients), while we found an average of 18,5% and 15,2% of epithelial cells in control and PCa biopsies, respectively.

Epigenomic state mirrors prostate cancer heterogeneity

We applied 4f-SAMMY-seq to examine genome architecture in prostate biopsies[16]. This method relies on the sequential isolation and high-throughput sequencing of distinct chromatin fractions separated according to their accessibility and solubility properties, which are expected to correlate with chromatin epigenetic and transcriptional status (Supplementary Fig. S3a). The more soluble fractions (S2S, S2L) are associated with gene activation markers, and the less soluble fractions (S3, S4) are enriched for constitutive heterochromatin.

To confirm the applicability of this technique on fresh prostate biopsies, we first sequenced 4f-SAMMY fractions of non-neoplastic control samples. We examined the DNA sonication and short fragments (S2S) selection to obtain high-quality libraries (Supplementary Fig. S3b). We verified that we obtained sequencing reads with comparable quality across samples (Supplementary Fig. S3c). The sequencing reads distribution profiles for individual fractions from CTR samples show a high level of concordance (Supplementary Fig. S3d), with the highest correlation values for S2L fractions (median: 0.94).

To study euchromatic and heterochromatic domains, we defined the solubility profile as the ratio of high-throughput sequencing reads distribution in S2L over S3 fractions (see Methods), where higher and lower values correspond to regions enriched for open and closed

https://doi.org/10.1038/s41467-025-64213-4

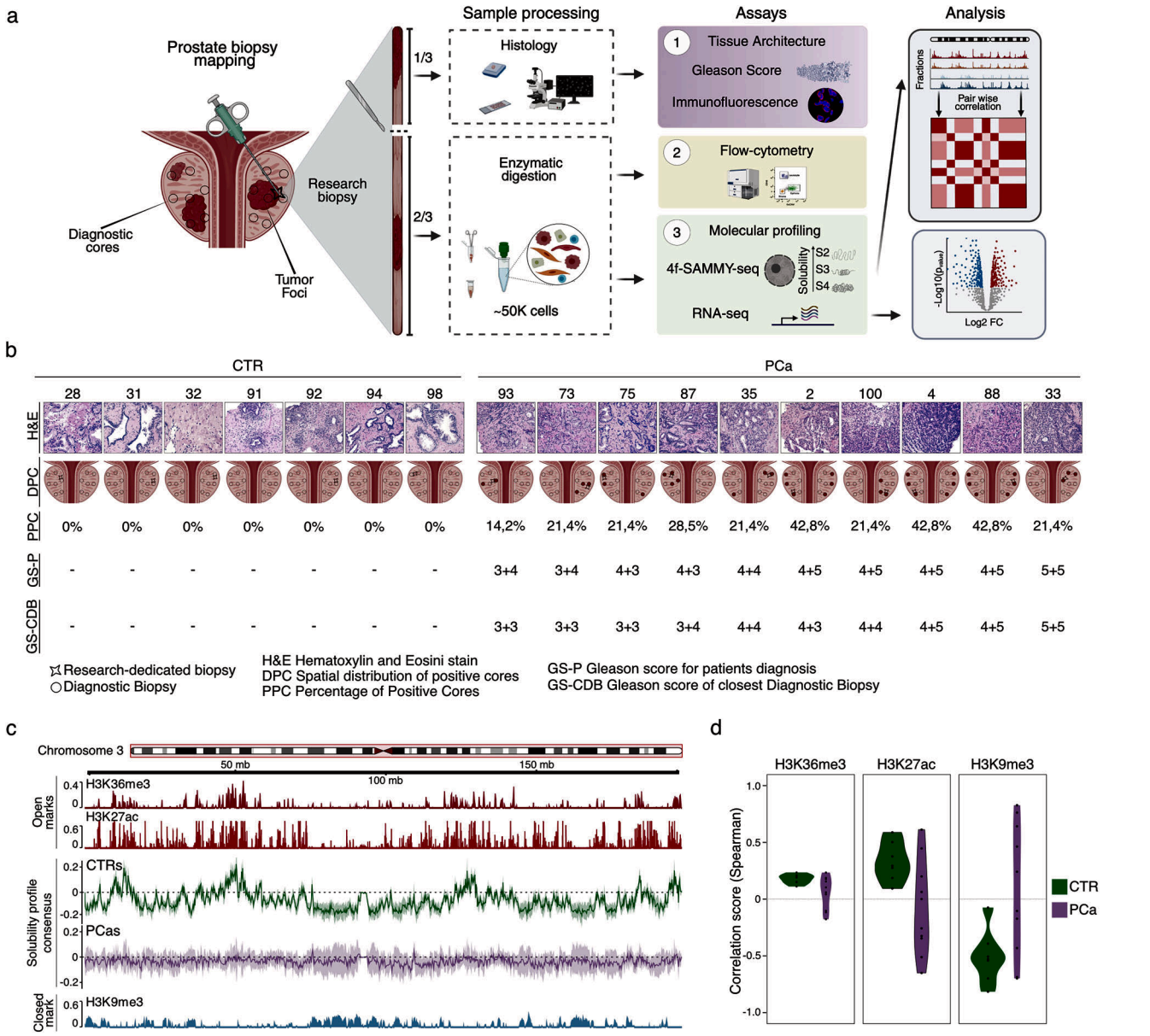


Fig. 1 | Chromatin architecture alterations in prostate cancer patients. **a** Overview of the experimental design. Fresh biopsies were dissected into 2 portions: one third was flash frozen and embedded in OCT for histopathological and immunofluorescence analyses (1), while the rest was enzymatically digested into single cell suspensions that were further divided for flow cytometry-based phenotype assay (2), 4f-SAMMY-seq and RNA-seq (3). **b** Study cohort biopsies: representative images of Hematoxylin and Eosin staining (H&E) at magnification x20 prostate biopsies of non-neoplastic patients (CTR) and prostate cancer patients (PCa). In the schematic prostate diagram, the circles mark sites of diagnostic biopsies ($n = 14$ for each patient), the star indicates the research biopsy position. Positive cores (black circles) and Percentage of Positive Cores (PPC) are also reported. GS-P indicates diagnostic Gleason Score assigned to each patient, while GS-CDB indicates the Gleason Score assigned to the biopsy adjacent to the research biopsy. **c**. Epigenome profiles: distribution along chromosome 3 of ChIP-seq enrichment profiles for open (H3K27ac and H3K36me3) and closed (H3K9me3) chromatin marks, along with 4f-SAMMY-seq consensus solubility profiles of 7 CTR (green) and of 10 PCa (purple) samples. The solid line shows the mean, and the shadowed areas marks the confidence interval: +/−1.96 standard error in the solubility profiles. **d** Violin plots for genome-wide Spearman correlation values (y-axis) for 4f-SAMMY-seq solubility profiles of CTR (green, $n = 7$) and PCa (purple, $n = 10$) samples with respect to ChIP-seq enrichment profiles in normal prostate gland for three distinct histone marks (H3K36me3, H3K27ac, H3K9me3). Each data point in the violin plots shows the correlation value of individual patients computed with genomic profiles binned at 150 kb. The schematic illustration in panels **a** and **b** were created using BioRender.com.

chromatin, respectively. We compared the consensus solubility profiles across all CTR and PCa specimens with epigenomic marks in normal prostate tissue for post-translational histone modifications (histone marks) associated with active (H3K27ac, H3K36me3) and inactive (H3K9me3) chromatin states[20] (Fig. 1c). We observed a high similarity of solubility profiles among CTR samples and a clear concordance with the location of euchromatin marks (H3K27ac and H3K36me3), along with a negative correlation with the heterochromatin mark profile (H3K9me3). These observations were also confirmed in a genome-wide quantification of correlations for individual samples (Fig. 1d).

Instead, in PCa biopsies we observed that the solubility profile is more heterogeneous and its correlation with histone marks is more variable with respect to CTR samples (Fig. 1c, d). Visual inspection of individual solubility profiles showed a patient-specific degree of chromatin alterations (Supplementary Fig. S3e). Moreover, we noticed

https://doi.org/10.1038/s41467-025-64213-4

Table 1 | Clinicopathological characteristics of the study participants undergoing prostate biopsy

Patient	Histology (GS-P)	Age at time of biopsy (years)	PSA level (ng/ml)	Positive cores (%)	Clinical stage	PI-RADS[49]	D'Amico risk classification[50]
CTR 28	Neg	62	3.7	-	-	-	-
CTR 31	Neg	65	7.18	-	-	-	-
CTR 32	Neg	67	3.73	-	-	-	-
CTR 91	Neg	67	8.15	-	-	-	-
CTR 92	Neg	78	4.6	-	-	-	-
CTR 94	Neg	72	8.9	-	-	-	-
CTR 98	Neg	72	20.9	-	-	-	-
PCa 93	3 + 4	73	8.9	14.2	T1c	2	1 (Intermediate risk)
PCa 73	3 + 4	65	19	21.4	T1c	5	2 (Intermediate risk)
PCa 75	4 + 3	82	13	21.4	T1c	2	2 (Intermediate risk)
PCa 87	4 + 3	72	6.5	28.5	T1c	2	1 (Intermediate risk)
Pca 35	4 + 4	71	4.88	21.4	T1c	4	4 (High risk)
PCa 2	4 + 5	89	51	42.8	T2b	2	9 (High risk)
PCa 100	4 + 5	87	29	21.4	T2a	5	8 (High risk)
PCa 4	4 + 5	74	487	42.8	T3a	4	12 (High risk)
PCa 88	4 + 5	72	186	42.8	T3a	2	12 (High risk)
PCa 33	5 + 5	78	4	21.4	T1c	5	8 (High risk)
PCa 14	3 + 3	67	7.6	50	T1c	2	0 (Low risk)
PCa 38	3 + 4	73	4.1	28.5	T2a	-	1 (Intermediate risk)
PCa 105	3 + 4	78	6.84	35.7	T2a	4	1 (Intermediate risk)
PCa 22	4 + 3	78	7.07	21.4	T1c	-	1 (Intermediate risk)
PCa 102	4 + 3	60	3.8	28.5	T1c	4	1 (Intermediate risk)
PCa 108	4 + 4	79	4.25	78.6	T1c	-	4 (High risk)
PCa 104	4 + 5	74	6.5	50	T2c	5	8 (High risk)
PCa 106	4 + 5	77	4.16	42.8	T3a	5	8 (High risk)

GS-P Gleason Score for patients' diagnosis.
PSA Prostate Specific Antigen.
PI-RADS Prostate Imaging Reporting and Data System.

clear differences among PCa samples, with five out of ten samples losing the regular chromatin compartmentalization, as also highlighted by their correlation with histone marks (Fig. 1d). These differences are remarkably not attributable to the cell composition of the samples (Supplementary Fig. S2g).

Thus, 4f-SAMMY-seq data describe a conserved genome architecture among control biopsies with a separation between euchromatin and heterochromatin. On the other hand, diagnostic early-stage PCa biopsies show a patient-specific chromatin solubility dysregulation that mirrors PCa's highly heterogeneous nature.

Compartment analysis identifies two epigenetic PCa subtypes

To further analyze the chromatin organization, we used the sequencing data of all the 4f-SAMMY-seq fractions to identify active and inactive compartments, named "A" and "B", respectively, as per the convention adopted for Hi-C derived compartments[16] (Fig. 2a, b, see Methods). However, unlike Hi-C, the chromatin compartments inferred from 4f-SAMMY-seq are defined based on their chromatin accessibility and solubility properties[16]. CTR samples showed a consistent pattern across all seven individuals with 27% of the entire genome having conserved active A compartment and 39% having conserved inactive B compartment across all CTRs (Fig. 2b, c). Instead, PCa samples showed patient-specific chromatin compartment alterations (Fig. 2b, c). We quantified the genomic regions with a different compartment assignment in each PCa sample with respect to the consensus among CTR samples. We labelled these compartment switches as "A to B" or "B to A" for the regions that in CTR samples were A or B, respectively. Unsupervised clustering based on compartment similarity identified two subgroups of cancer patients (Fig. 2d) that, considering the amount of compartment alterations with respect to CTR samples, we named LDD (Low Degree of Decompartmentalization; average 5% over the entire genome for "A to B" and 5% for "B to A" compartment switches) and HDD (High Degree of Decompartmentalization; average 14% over the entire genome "A to B" and 24% "B to A"). We also noted a consistent intra-group similarity for LDD samples with 0.78 mean Jaccard Index (JI) (variance 0.001). On the other hand, HDD samples show more heterogeneity in chromatin compartmentalization (JI 0.42 mean and 0.012 variance) (Fig. 2d).

In line with patient-specific genome remodeling, we did not find regions with compartment changes concordant across all LDD and HDD tumoral samples compared to CTR consensus. However, we found recurrent "B to A" regions, indicating higher chromatin accessibility compared to controls, concordant across all HDD patients and covering 2% of the entire genome (538 genes) (Supplementary Fig. S4a). Gene ontology analysis highlighted an enrichment of genes associated with Protein Deubiquitination and Regulation of Programmed Cell Death (such as *FAS*) (Supplementary Fig. S4b). Instead, the HDD-specific "A to B" compartment switch regions, which correspond to reduced chromatin accessibility compared to controls, (0.6% of the genome – 59 genes) included genes involved in PI3-kinase activities (such as *PIK3R5* and *PIK3R6*) and lipid catabolic processes (such as *PNLIP* and *STS*), both of which are known drivers of prostate cancer progression[21–23].

It should be remarked that the LDD vs HDD groups stratification of PCa patients would not be immediately evident based on known clinically relevant pathophysiological features such as PSA level, Gleason Score, age at diagnosis and number of positive cores (Supplementary Fig. S4c). Furthermore, the analysis of the global

https://doi.org/10.1038/s41467-025-64213-4

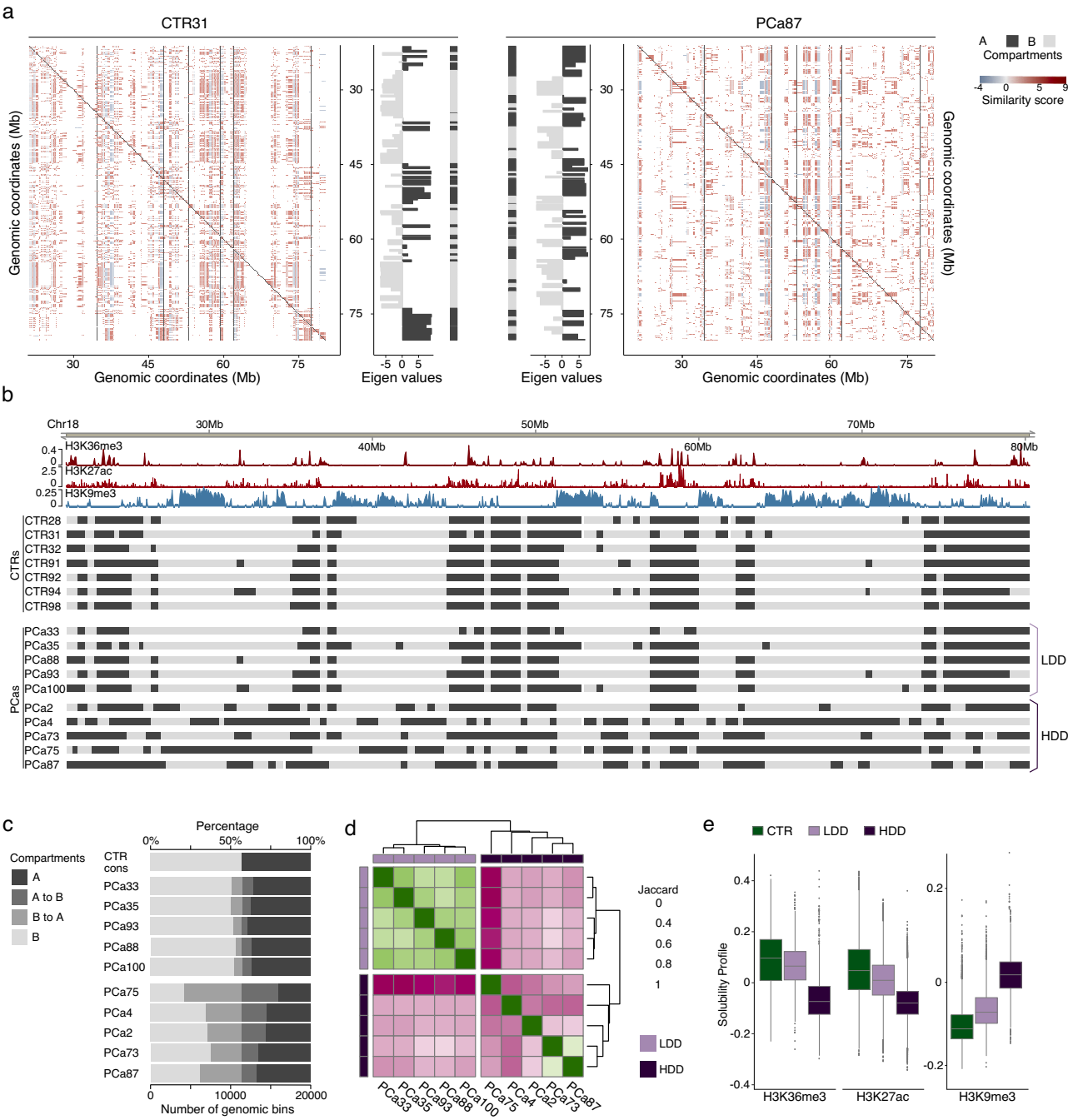


Fig. 2 | Prostate cancer subtypes identified by epigenomic profiles. a Pair-wise correlation matrices of reads distribution profiles for representative samples CTR31 (left) and PCa87 (right), computed at 150 kb genomic bins resolution on chromosome 18. On the side of each matrix the respective first eigenvector is colored to mark the position of active ("A" compartment, positive values) and inactive regions ("B" compartment, negative values). The two bars in the middle show the A and B compartments along the chromosome according to a color code. **b** Summary of chromatin compartments and chromatin marks on chromosome 18. From top to bottom: ChIP-seq enrichment profiles of H3K36me3 (dark red), H3K27ac (red) and H3K9me3 (light blue); A and B compartments segmentation for CTR and PCa patients (labels on the left), grouped in LDD and HDD subgroups (labels on the right). **c** Stacked barplot illustrating the count (number of genomic bins, bottom x-axis) and the relative distribution over the entire genome (percentage, top x-axis) of genomic regions with conserved compartment class ("A" - black or "B"- grey) and regions shifting compartment ("A to B" or "B to A" - different shades of grey) for each PCa sample. The discordant chromatin compartment classification is calculated with respect to the consensus of CTR (on the top). **d** Pair-wise Jaccard Index for compartments similarity among PCa samples. Light-purple (PCa33, PCa35, PCa93, PCa88 and PCa100) and dark purple (PCa75, PCa4, PCa2, PCa73, PCa87) clusters arranged by unsupervised hierarchical clustering, contain the LDD and HDD samples, respectively. **e**. Boxplot of solubility profiles average values (y-axis) for individual 150 kb genomic bins overlapping ChIP-seq enrichment peaks of open (H3K36me3, H3K27ac) and closed (H3K9me3) chromatin histone marks (labels on the x-axis). For each group of patients CTR ($n = 7$), LDD ($n = 5$) and HDD ($n = 5$) and for each genomic bin, we plot the mean 4f-SAMMY-seq solubility profile across the group. Each data point in the boxplot is one genomic bin enriched for H3K36me3 (n bins = 737), H3K27ac (n bins = 11536) or H3K9me3 (n bins = 3876), respectively. The horizontal lines mark the median, the boxes mark the interquartile range (IQR) and whiskers extend up to 1.5 times the IQR; individual data points beyond this range are plotted.

transcriptome on prostate samples did not separate so clearly the HDD and LDD subtypes (Supplementary Fig. S4d). We excluded the association of HDD and LDD subtypes with cell composition and immune cell infiltration, estimated by flow-cytometry (Supplementary Fig. S4e) and RNA-seq data deconvolution (see Methods, Supplementary Fig. S4f, S4g). Interestingly, upon separating HDD and LDD groups in the dataset of Supplementary Fig. S1d, we observed that nuclear circularity differs significantly from controls only in the LDD group (Supplementary Fig. S4h).

We further investigated epigenetic differences between LDD and HDD groups. We compared the histone mark enrichment peaks from normal prostate tissue with the 4f-SAMMY-seq solubility profiles of CTR, LDD and HDD samples. As expected, we found that the solubility profile of CTR samples is higher in ChIP-seq peaks for open histone marks (H3K27ac and H3K36me3) and lower in constitutive heterochromatin (H3K9me3 peaks) (Fig. 2e). The HDD group showed instead an inverted trend, thus confirming a large-scale chromatin architecture remodeling. Interestingly, the LDD subgroup, despite the limited alterations in chromatin compartmentalization (Fig. 2b, c), showed an intermediate solubility pattern between the other two groups, especially for H3K27ac and H3K9me3 domains.

HDD patients present an extensive chromatin reorganization

We next examined the functional effects of chromatin compartment reorganization by matched RNA-seq analysis. Comparing transcription profiles between CTR ($n = 7$) and PCa ($n = 10$) samples uncovered 66 differentially expressed genes (DEGs): 22 up-regulated and 44 down-regulated in PCa with respect to CTR (Fig. 3a). The limited number of DEGs is likely due to the heterogeneity of expression across tumor patients' biopsies.

Thus, we considered LDD and HDD subtypes separately. We identified only 30 DEGs in LDD comparison to CTR samples: 9 up- and 21 down-regulated. Instead, we found 162 up- and 410 down-regulated genes in HDD compared to CTR samples (Fig. 3a). This suggests that the higher degree of chromatin compartment remodeling is associated with a more extensive gene expression dysregulation. We found 5 DEGs up-regulated and 4 DEGs down-regulated in common among all comparisons of different PCa (Fig. 3b). Seven genes were already described in the literature as prostate cancer biomarkers (*HPN*, *BICD1*, *GOLM1*, *TARP*), and two down-regulated tumor suppressor genes (*SERPINB5* and *GATA6*)[19,24–28].

To further dissect the molecular differences between LDD and HDD subtypes, we performed an enrichment analysis for Gene Ontology (GO) annotations on the list of DEGs in HDD vs CTR. We didn't perform this analysis on LDD vs CTR samples due to the small number of DEGs. Nevertheless, in LDD DEGs, we found the upregulation of two critical genes, *PDLIM5* and *CAMKK2*, involved in cell migration and invasion[29,30]. In the other comparison, we discovered that HDD up-regulated genes are enriched for the "negative regulation of cell motility" class. whereas down-regulated ones include GO classes related to tumor microenvironment remodeling (Fig. 3c), suggesting decreased functions usually associated with tumor progression. These findings were also confirmed through GeneSet Enrichment Analysis (GSEA)[31] (Supplementary Fig. S5a).

We then applied a complementary approach for the functional analysis of transcriptome profiles for both LDD and HDD comparison to CTR samples. We used PROGENy[32], a curated signaling pathway compendium derived from perturbation experiments. This compendium includes weights quantifying the response of each gene to pathway perturbations, thus allowing a weighted analysis of transcriptional response to specific signaling cues (see Methods, Supplementary Fig. S5b). Although Androgen pathway activity is enriched in both LDD and HDD subgroups, it is markedly stronger in HDD. Notably, we found a positive enrichment for TGFbeta activity in LDD, whereas it has negative enrichment in the HDD, indicating subgroup-specific regulatory patterns.

These results demonstrate that the LDD and HDD subgroups are different in their chromatin architecture and gene expression, thus representing previously undescribed PCa patient subtypes. The HDD chromatin reorganization is also reflected in more prominent chromatin remodeling and transcriptional profiles that counteracts malignant features. Then, we explored the possible connections between epigenetic remodeling and functions, analyzing compartment changes and transcription. Accordingly with chromatin compartmentalization, we found active transcription associated to the consensus A compartment across all our PCa samples (Supplementary Fig. S5c). However, by considering DEGs and the consensus of compartment switches (see methods), we observed that the majority of DEGs (97% for LDD; 71% for HDD) are located in domains not involved in a compartment transition (Fig. 3d). This suggests that, as extensively described in other models[33–35], changes in chromatin architecture are not enough to trigger transcriptional dysfunction, but they establish conditions that enable transcriptional deregulation. A higher proportion of DEGs (29%) were found in compartment switch regions for HDD. Within these genes, we identified a trend where up-regulated genes were associated with "B to A" switch regions, while down-regulated genes were associated with "A to B" compartment switch regions (Fig. 3e, f). Functional annotations analysis revealed that DEGs up-regulated in HDD and involved in "B to A" transitions are enriched for prostate cancer-specific terms, related to cellular identity, as the primary prostate cancer cell lines VCap, and signaling pathways regulated by the master transcription factors AR and FOXA1. This suggests that A compartment reorganization observed in HDD is associated with prostate cancer phenotype and plasticity. Conversely, DEGs downregulated in HDD and involved in "A to B" transitions are significantly enriched for classes related to extracellular matrix (ECM) remodeling and cell–matrix interactions (Fig. 3f), suggesting a dynamic crosstalk between tumor cells and the ECM. Transcription factor enrichment analysis identified the Polycomb protein SUZ12 (Fig. 3f), a subunit of the well-known Polycomb repressive Complex 2 (PRC2), as a key regulator of the identified genes, highlighting the role of epigenetic silencing in promoting the HDD transcriptional phenotype.

HDD status is associated to good prognosis

To identify a signature distinguishing HDD vs LDD subtypes, we examined the differentially expressed genes comparing HDD vs LDD transcriptomic profiles.

We identified 101 DEGs: 77 down- and 24 up-regulated genes in HDD with respect to LDD (Fig. 4a). Consistent with previous findings, we found differential expression of Epithelial-Mesenchymal Transition (EMT) genes (GSEA analysis – Fig. 4b and Supplementary Fig. S5d) as well as differential activity of Androgen and TGFbeta pathways (PROGENy analysis - Supplementary Fig. S5e).

Then, we asked if our signature of 101 genes could provide prognostic information on patients. Thus, we queried the Cancer Genome Atlas (TCGA) for the RNA-seq data of 499 prostate adenocarcinomas (PRAD)[36]. We defined a quantitative score named PCI (Prostate Compartmentalization Index) to summarize the HDD vs LDD transcriptional signature activity for each patient in this cohort (see Methods). We classified TCGA samples as HDD-like and LDD-like patients based on the PCI score, indicating a more prevalent HDD ($PCI > 0$) or LDD ($PCI < 0$) signature, respectively (Fig. 4c).

Intriguingly, survival analysis for biochemical recurrence-free status (BCR) indicated that patients belonging to the HDD-like cluster show a better prognosis (Fig. 4d), confirming the protective chromatin compartment alterations described above. This is a remarkable result because the signature was established using biopsies taken at the time

https://doi.org/10.1038/s41467-025-64213-4

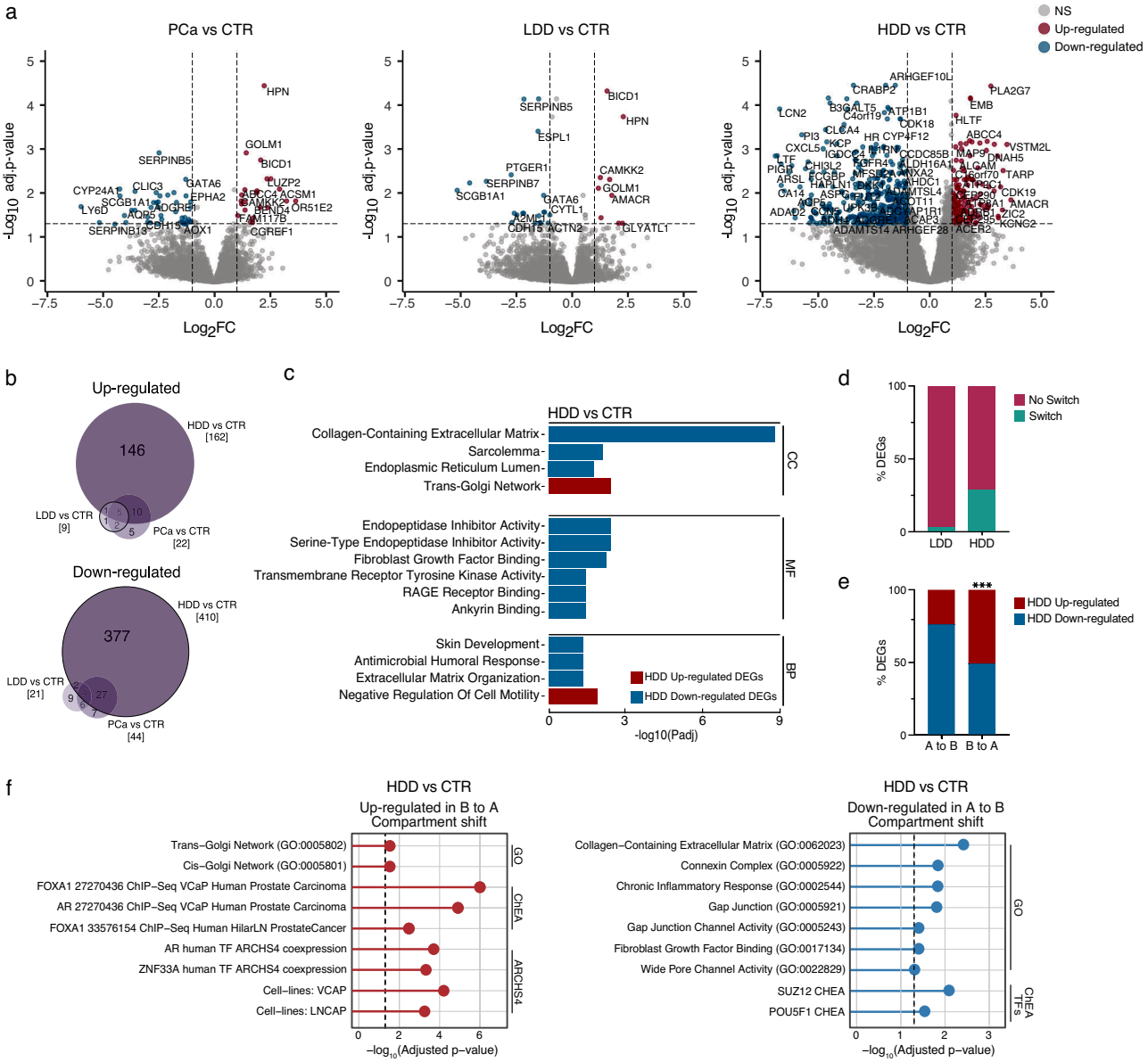


Fig. 3 | Functional features discriminating prostate cancer subtypes. a Volcano plot showing Differentially Expressed Genes (DEGs) in the comparison between PCa and CTR samples (left), LDD tumors subtype and CTR samples (center), HDD tumors subtype and CTR samples (right). Blue and red dots represent significantly down- and up-regulated genes (adjusted *p* -value < 0.05; log2FC < −1 or > 1, respectively). Statistical analysis was performed using DESeq2 with two-sided Wald tests, and *p*-values were adjusted for multiple testing using the Benjamini-Hochberg method. **b** Proportional Venn diagram of up-regulated genes (left) and down-regulated genes (right) comparing all PCa tissues vs CTR samples, LDD tumors subtype vs CTR samples and HDD tumors subtype vs CTR samples. **c** Functional classes enrichment analysis of HDD vs CTR DEGs using Cellular Component, Molecular Function and Biological Process gene ontology annotations. Bar plots representing significantly enriched terms in down-regulated genes (blue) and up-regulated (red) genes, with -$\log_{10}$ adjusted *p*-values reported (horizontal axis). The *p*-values were computed using two-sided Fisher's exact test with adjustments made by the Benjamini-Hochberg method. **d** Stacked bar plot depicting the percentage of DEGs in the comparison LDD vs CTR (left) and HDD vs CTR (right) across concordant (no switch - magenta) and discordant (switch - turquoise) compartments classification changes with respect to CTR samples. The consensus set of compartments was used for each group. **e** Stacked bar plot illustrating the relative distribution of up-regulated (red) and down-regulated (blue) DEGs (y axis), for HDD vs CTR comparison falling in the compartments switching from A in the CTR to B in HDD and vice versa. Significance enrichment detected for up-regulated DEGs in "B to A" compartment switches (one-tailed binomial test,***p-value = 1.5e-6). **f** Functional classes enrichment analysis using the intersection between significant up-regulated genes included in B to A switches (red dots) and down-regulated genes included in A to B switches (blue dots) in HDD patients with respect to CTR compartment consensus. Significant terms (adjusted *p*-value < 0.05) are shown for each database on the right, including GO (Gene Ontology), ChEA (ChIP-X Enrichment Analysis), ARCHS4 (All RNA-seq and ChIP-seq Sample and Signature Search) and ChEA - TFs (ENCODE and ChEA Consensus TFs from ChIP-X). The *p*-values for enrichments were computed with Enrichr using Fisher's exact test and Benjamini-Hochberg multiple testing correction.

of diagnosis, yet it can predict prognosis in more advanced, surgically removed tumor samples from the TCGA.

TCGA samples, extracted from prostatectomies are expected to have generally higher cellularity and tumor purity, with respect to needle biopsies. We noted a potential association between HDD-like and LDD-like groups and the TCGA sample's tumor purity scores (Fig. 4c). Nevertheless, multivariate Cox regression analysis, including the tumor purity score in the model, confirmed that the HDD-like classification is a valid independent prognostic predictor of biochemical recurrence (Fig. 4e). The

https://doi.org/10.1038/s41467-025-64213-4

HDD-like phenotype is associated with a better outcome (HR = 0.5, CI = 0.33-0.88, *p*-value = 0.01) on top of the reference clinical features such as pathological (TNM) stage, age at diagnosis and Gleason Score.

We further refined the signature to retain only the genes with expression more significantly associated with good or bad prognosis, finally resulting in the 18 genes refined signature (Supplementary Fig. S6a, S6b, see Methods). As expected, the reduced 18 genes signature improves the stratification of the TCGA cohort used to refine the signature itself (*p*-value < 0.0001; Supplementary Fig. S6c), also independently of tumor purity levels (Supplementary Fig. S6d, S6e).

To further dissect the epigenetic status of our PCa subtypes, by leveraging independent publicly available epigenomic datasets we investigated if the LDD vs HDD classification underlies a difference in

Fig. 4 | Prognostic stratification of patients driven by chromatin-derived signature. a Volcano plots showing Differentially Expressed Genes (DEGs) in the comparison between HDD and LDD subtypes. The adjusted p-value < 0.05 and $\log_2$ Fold Change value > 1 or $\log_2$ Fold Change value < −1 and are the cutoff values for significant up-regulated (red, $n = 24$) and down-regulated (blue, $n = 77$) genes, respectively. Non-significant differential expression is marked with grey color. Statistical analysis used DESeq2 with two-sided Wald tests and Benjamini-Hochberg p-value adjustment. **b** Selected Gene Set Enrichment Analysis (GSEA) profiles resulting from the analysis of HDD vs LDD pre-ranked expression scores with Hallmark gene sets from the Molecular Signatures Database (MSigDB) collection. **c** Heatmap showing RNA-seq expression values (scaled by row) across 499 prostate cancer tissues from TCGA cohort. The sample (on columns) are sorted in ascending order according to the PCI score. Rows reports the 101 DEGs from HDD vs LDD comparison and are grouped by hierarchical clustering over Euclidean distance. The top section side bar displays groups of transcriptomic profiles categorized as LDD-like (light pink, PCI score <0) and HDD-like (deep pink, PCI score > 0), along with the Gleason Score and tumor purity associated with each sample. A colored side bar on the left marks up-regulated (red) and down-regulated (blue) DEGs in HDD vs LDD comparison. **d** Kaplan-Meier plot showing Biochemical Recurrence Free (BCR) status outcome of TCGA patients stratified according to their transcriptomic similarity with HDD and LDD group. The deep pink line indicates HDD-like TCGA samples, whereas the light pink line indicates LDD-like TCGA samples. P-value was obtained from the Log rank test (p-value = 0.041). **e** Multivariate cox regression analysis using cox proportional hazards model (Hazard Ratio, HR) for Biochemical Recurrence Free (BCR). The following covariate parameters with p-value from two-sided Wald tests have been used in the model: PCI-based classification, Clinical Stage (N, T), Age, Tumor Purity, Gleason Score. HRs are plotted for each covariate, with horizontal lines representing 95% confidence interval (CIs). The vertical dashed line indicates the null value (HR = 1.0). Exact p-values are indicated on the right.

chromatin states. We used a cohort of 99 prostate cancer patients with ChIP-seq datasets for the Androgen Receptor (AR), H3K27Ac, H3K4me3, and H3K27me3 histone marks across, as well as corresponding RNA-seq data[37]. We classified the cohort patients as HDD-like or LDD-like by applying our PCI score method. When analyzing the distribution of ChIP-seq peaks concordance across patients, we observed a clear shift between HDD-like and LDD-like groups for both AR and H3K4me3 (Supplementary Fig. S6f; see Methods). Specifically, the HDD-like group exhibited stronger enrichment of the active chromatin mark H3K4me3, consistent with the higher number of conserved A compartments observed in HDD-like samples compared to LDD-like (Supplementary Fig. S4a).

Then, we selected ChIP-seq peaks for H3K4me3, H3K27ac, or H3K27me3 with differential frequency in HDD-like vs LDD-like patients. Functional classification analysis consistently identified epithelial-to-mesenchymal transition (EMT) pathways among genes enriched for active chromatin marks (H3K27ac, H3K4me3) in LDD-like samples, and conversely enriched for the repressive mark H3K27me3 in HDD-like samples (Fig. 5a). Additionally, we found increased H3K27ac enrichment in HDD-like samples at androgen-responsive genomic regions, further supporting our findings.

We also analyzed available bulk ATAC-seq data on a limited set of TCGA patients[38]: 7 HDD-like and 18 LDD-like patients. Interestingly, although the unbalanced and limited number of samples resulting in a weak statistical power, we found differential ATAC-seq peaks detectable in HDD-like patients on 2 out of the 4 up-regulated genes of our 18 genes signature (*SC5D* and *IQGAP2*) (Supplementary Fig. S6g).

Validation of HDD signature

We then validated the clinical relevance of our refined 18 gene signature across multiple independent cohorts of prostate cancer patients. We obtained a significant prognostic stratification trend, where HDD-like patients show a better prognosis (Fig. 5b). These independent cohorts include transcriptomic profiles obtained with RNA-seq and microarrays, thus attesting the applicability of our signature to heterogenous datasets. As additional validation, we recruited another cohort of PCa patients, processing a validation cohort of 8 biopsies profiled by 4f-SAMMY-seq and RNA-seq (Supplementary Fig. S7a, S7b). We were able to classify them as 5 HDD and 3 LDD samples, based on the chromatin compartments analysis (Supplementary Fig. S7c, S7d). We also confirmed the HDD vs LDD transcriptional signature difference across the validation cohort (Fig. 5c).

Overall, our results confirm the validity of chromatin-informed patient stratification and the derived gene expression signature to discriminate patients for prognosis.

HDD signature captures epithelium and stroma reprogramming

We asked how distinct cell types in the tissue contribute to the HDD signature. Thus, we used a single-cell transcriptomic dataset of primary prostate cancer biopsies[39], i.e. an experimental setting comparable to ours, to confirm that the HDD signature comprises cell type–specific expression of epithelial and stromal genes (Fig. 5d). Specifically, three HDD upregulated genes (*SC5D*, *IQGAP2* and *ELOVL7*) showed predominant expression in benign epithelial and tumor cells, while five HDD-downregulated genes (*COL5A1*, *COL5A3*, *COL16A1*, *COL18A1*, *MXRA8*) were enriched in the stroma, reflecting compartment-specific regulation. From the same dataset we then derived transcriptomic deconvolution signatures applied using CIBERSORTx (see Methods) to estimate the relative contributions of benign and malignant epithelial cells to each of our patients' bulk transcriptomic profile. The inferred proportion of malignant epithelial cells was comparable between the HDD and LDD groups, indicating that this factor does not drive the classification (Supplementary Fig. S7e, S7f).

We next performed immunohistochemistry (IHC) on formalin-fixed, paraffin-embedded (FFPE) archived samples from the original PCa patient cohort to assess cell type–specific gene expression. In agreement with the single-cell RNA-seq data (Fig. 5d), SC5D and IQGAP2 exhibited predominant expression in epithelial cells (Supplementary Fig. S8a). We detected a trend of higher IHC scores in HDD samples compared to LDD, consistent with transcriptomic data, although these differences were not statistically significant (Supplementary Fig. S8b). Among the stroma markers, COL16A1 showed apical staining within glandular lumens and a weaker, diffuse stromal signal; COL18A1 was primarily localized to endothelial cells within the stroma; and COL5A1 was broadly detected throughout the stromal compartment (Supplementary Fig. S8a). Despite the specificity of staining, the diffuse nature of collagen signals limited robust quantitative assessment from the IHC data. Overall, the protein expression analysis confirms that our signature includes markers specific to both stromal and epithelial cells.

We next examined the specific impact of stromal-associated genes downregulation within the HDD signature on epithelial cell migration. We established an in vitro co-culture system using tdTomato-labeled WPMY-1 prostate stromal cells and VCaP primary prostate cancer epithelial cells, which exhibit an HDD-like transcriptional profile (Supplementary Fig. S9a, S9b). To mimic the HDD associated stromal signature, we simultaneously silenced four stromal-associated genes (*COL5A1*, *COL5A3*, *COL18A1*, and *MXRA8*) in WPMY-cells using RNA interference (Supplementary Fig. S9c). Control (scrambled siRNA) and gene-silenced stromal cells were then co-cultured with VCaP epithelial cells, and migration was assessed via a wound healing assay over a 45-hour period using time-lapse microscopy (Supplementary Fig. S9d). VCaP migration occurred only in the presence of stromal cells, underscoring the importance of stromal-epithelial interactions (Supplementary Fig. S9e). Notably, we observed significantly decreased migration after four-genes silencing in comparison to scrambled siRNA suggesting that the HDD stromal signature restricts the motility potential of prostate cancer cells (Fig. 5e, f and Supplementary Fig. S9f–i).

https://doi.org/10.1038/s41467-025-64213-4

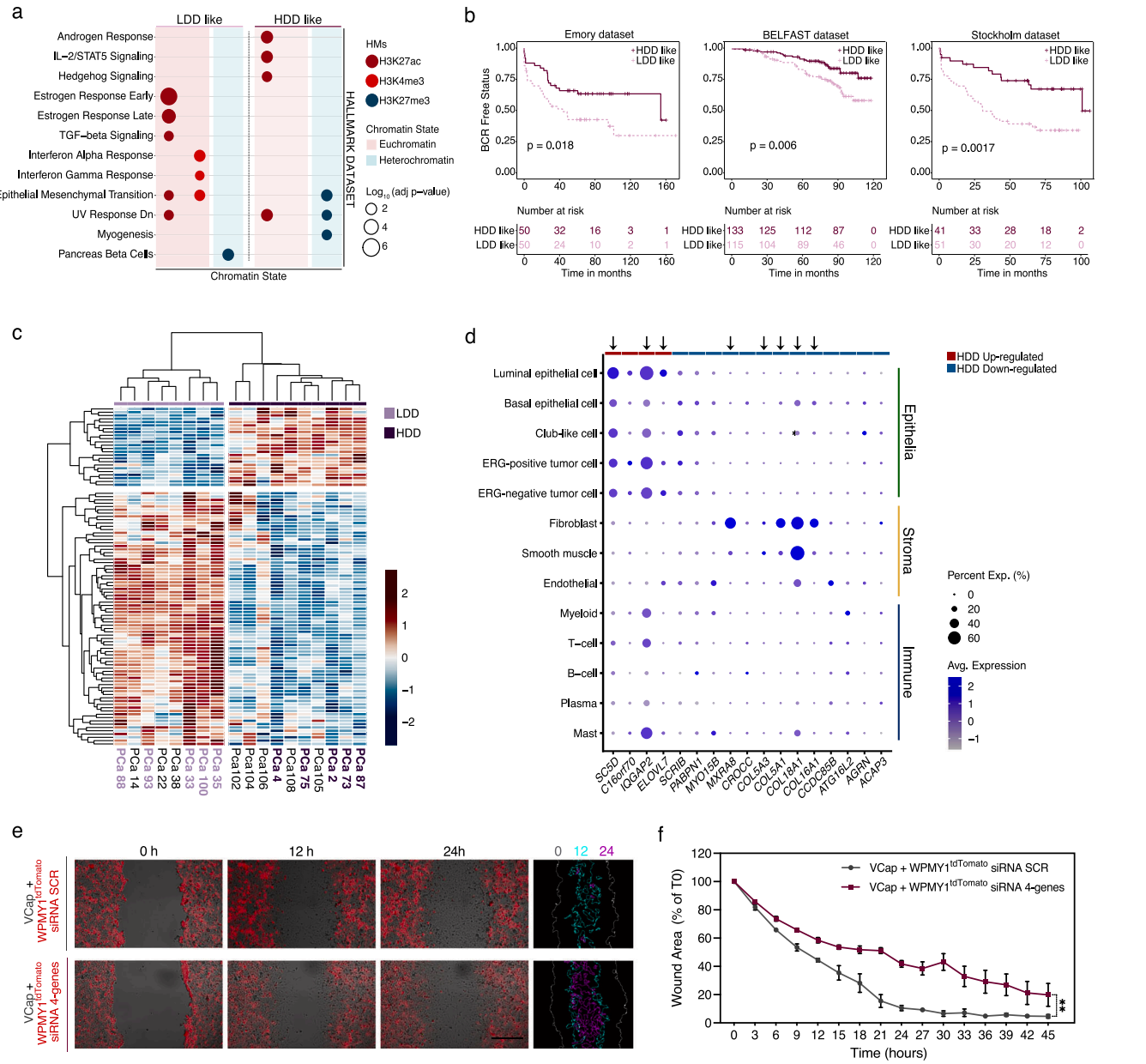


Fig. 5 | HDD transcriptional signature is associated with favorable prognosis. **a** Gene Ontology (GO) terms associated with histone mark enrichment in LDD- and HDD-like prostate cancer samples from the GSE120738 and GSE120741. Circle size indicates log-adjusted *p*-value. Color represents the corresponding histone modification: H3K27ac and H3K4me3 (active marks) in red, and H3K27me3 (repressive mark) in blue. The *p*-values for enrichments were computed with Enrichr using Fisher's exact test and Benjamini-Hochberg multiple testing correction. **b** Kaplan-Meier plots based on the 18 genes refined on 100 samples from Emory cohort (GSE54460) (left, *p*-value = 0.018), on 248 sample from Belfast cohort (GSE116918) (middle, *p*-value = 0.006), and on 92 samples from Stockholm cohort (GSE70769) (right, *p*-value = 0.0017). Deep pink lines indicate HDD-like samples, while light pink lines indicate LDD-like samples. *P*-values were computed with the Log-rank test. **c** Unsupervised clustering showing the transcriptomic expression values of the 101 DEGs from the HDD vs LDD comparison (rows) across the PCa cohort (columns), with the validation cohort marked in black. The top sidebar differentiates the transcriptomic clusters into HDD (dark-purple) and LDD (light-purple). **d** Dot plot displaying the expression of the 18-gene signature across cell clusters from scRNA-seq data of primary prostate cancer[39]. Dot size indicates the percentage of cells expressing each gene, while color intensity represents the average expression level. Arrows highlight epithelial (*SCSD*, *IQGAP2*, *ELOVL7*) and stromal (*COL5A1*, *COL5A3*, *COL16A1*, *COL18A1*, *MIXRA8*) genes, indicating cell-type–specific regulation. **e** Representative time-lapse images of wound healing assays showing VCaP prostate cancer cells (bright-field) co-cultured with WPMY-1 prostate stromal cells expressing Tomato fluorescent protein (WPMY-1^{tdTomato}, red). WPMY-1^{tdTomato} cells were treated with scrambled siRNA (SCR; top) or 4-gene-targeting siRNA pool (siRNA 4-genes; bottom). Lines indicate wound closure at 0, 12, and 24 h. Scale bar, 500 μm. **f** Percentage of wound area compared to the original size during 45 h in co-cultures of VCaP prostate cancer cells and WPMY-1^{tdTomato} fibroblasts treated with scrambled siRNA (SCR; grey line) or 4-gene-targeting siRNA cocktail (siRNA 4-genes; magenta line). Data are presented as mean ± Standard Error of the Mean (SEM) ($n = 4$ biological replicates per condition). Statistical analysis was performed using a two-way ANOVA (**p-value = 0.0021).

Discussion

Prostate cancer diagnosis, which relies on the histological examination of multiple biopsies, has revealed extensive heterogeneity and multifocality of PCa[2]. This procedure cannot predict the 20% of prostate cancer cases that progress to metastases, resulting in a high mortality rate. Importantly, most transcriptome-based analyses, which might capture the phenotypic plasticity, are unable to provide early insights into clinical progression. Thus, there is an urgent need

for alternative biomarkers to refine the stratification of prostate cancer patients.

The 3D genome architecture in the cell nucleus is crucial for the epigenetic regulation of transcription, with its reorganization frequently linked to neoplastic conditions[12,40]. In this work, using 4f-SAMMY-seq and transcriptome analysis on a limited number of cells (10,000–50,000 cells), we analyzed the epigenome structure and function of PCa patients-derived biopsies to infer their plasticity (Fig. 1).

We identified two epigenetic subtypes of PCa patients presenting different degrees of chromatin compartments and transcriptional alterations: the HDD and LDD subtypes (Fig. 2). Samples with HDD features display a marked chromatin remodeling and transcriptional repression of cancer progression pathways, suggesting a possible association between chromatin changes and inhibition of the tumor malignancy (Fig. 3).

Comparing HDD and LDD subtypes, we derived a specific expression signature that characterizes the HDD phenotypic state (Fig. 4). We confirmed that the HDD signature is significantly associated with a more favorable prognosis (Figs. 4, 5).

To develop a molecular signature exploitable in the clinic, we selected genes with prognostic value, ultimately restricting our chromatin-based RNA signature to 18 genes validated across multiple independent cohorts covering more than 900 patients with follow-up clinical annotations (Fig. 5b). We also confirmed both the chromatin compartments and transcriptomic signatures applicability to patients by analyzing a second validation cohort of prostate cancer patient's biopsies (Fig. 5c). Notably, when comparing our HDD signature with the 29 previously identified prostate cancer (PCa) signatures listed in the PCa Database (PCaDB; http://bioinfo.jialab-ucr.org/PCaDB/), only 3 out of the 18 genes were shared, namely *ACAP3* e *ATG16L2*[41] and *SCRIB*[42]. This highlights the uniqueness and potential specificity of our gene signature for early stratification of prostate cancer patients.

Compared to state-of-the-art studies on chromatin organization in prostate cancer, and in particular the most recent publication[12], our study includes several innovations. First of all, our findings can impact prognosis prediction immediately at the time of initial diagnosis as our study used diagnostic biopsies as opposed to metastasis biopsies. From a biological perspective, we discovered a tumor suppressive role of heterochromatin compartments remodeling associated with better prognosis in prostate cancer, a surprising feature yet paralleled by similar observation in earlier works on colorectal cancer[40]. Finally, our work provides key contributions in terms of clinical translatability as we validate the signature across multiple datasets, and different modalities of measuring epigenomic and transcriptomic profiles (Fig. 5).

The epigenetic remodeling of epithelial tumor cells is coupled with the loss of cell identity homeostasis through multiple steps, driven by molecular changes that depend on intrinsic and extrinsic signals[43]. The tumor microenvironment plays a central role in this process, constantly adapting its stroma cell composition and transcriptional programs to favor or counteract tumor progression[44,45]. This is particularly evident for indolent cancers such as the PCa, whose slow progression favors the crosstalk between cancer and neighboring cells. Prostate stromal tissue, constituted by fibroblasts, endothelial cells, smooth muscle cells (SMCs), and immune cells, plays a crucial role in prostate tumorigenesis[46–48]. Intriguingly, transcriptional programs can have divergent trajectories in epithelial or stromal cells. This is the case of the androgen receptor (AR) pathway, whose overexpression in epithelial cells is mostly associated with bad prognosis[49] whereas in the stroma it has a protective role by constraining cancer growth[47]. Moreover, epigenetic activation of canonical androgen response elements (AREs) has recently been linked to tumor-suppressive transcriptional programs in prostate epithelial cells, highlighting the role of the epigenetic landscape in modulating AR signaling's dual capacity to promote either tumor-suppressive or oncogenic outcomes in prostate cancer[50]. One of the significant pathways upregulated in HDD samples is the androgen receptor and, working on bulk prostatic tissue comprising 70–80% of stromal cells, we speculate that this detected AR up-regulation may be ascribed to the stroma. Further corroborating this hypothesis, the LDD PCa group correlates with a worse prognosis and its signature includes overexpression of TGFbeta that can induce a resistance to androgen deprivation therapy when expressed in the stroma[48].

We confirmed that the transcriptional signature discriminating HDD and LDD patients represents a composite response involving both stromal and epithelial cells, reflecting the complexity of the tumor microenvironment (Fig. 5d and Supplementary Fig. S8). We finally proved in cell co-cultured models that downregulation of a subset of signature genes in stroma cells can restrict the migratory capacity of primary epithelial prostate cancer cells (Fig. 5e, f).

In summary, this study we introduced an experimental approach allowing us to investigate the epigenomic landscape and chromatin compartmentalization in prostate biopsies. We identified two patient epigenetic subgroups based on their epigenome profiles. The PCa subtype associated with a favorable prognosis exhibits heterochromatin reorganization that represses tumorigenic pathways. We found that the transcriptional signature derived by this chromatin-informed patient stratification arises from a reprogramming of both epithelial and stromal cells and constitutes a independent prognostic classifier. We validated the signature across multiple independent cohorts, thus confirming its potential for translatability to assess prognosis at the time of diagnosis.

Methods

Prostate tissues cohort, ethics approval and consent to participate

This study was approved by the institutional ethics committee board of IRCCS Foundation Ca' Granda Ospedale Maggiore Policlinico (authorization n 1063). All specimens were obtained after the patients had provided written informed consent following the ethical principles of biomedical research on biospecimens. All the clinical data were provided by the Urology Division of Fondazione IRCCS Ca' Granda - Ospedale Maggiore Policlinico (Milan), preserving the confidentiality of patient personal data. Consent to publish anonymized clinical data was also obtained. Our cohort includes chemo-naïve patients followed in the Urology Division who underwent the transrectal ultrasound-guided systematic sampling of prostate tissue (TRUS biopsy). This diagnostic procedure was performed as part of routine clinical management following the detection of abnormal digital rectal examination or an elevated PSA blood level. According to the current standard for the detection of PCa, 14 ultrasound-guided biopsy cores (diagnostic biopsies), seven from each side of the prostate gland, were collected from each patient for the clinical diagnosis. During the same procedure, one additional biopsy core to be used for our research project (research biopsy), was taken from a site directly adjacent to one of the diagnostic biopsies. To ensure optimal recovery of living cells, fresh surgical prostate biopsy specimens were placed in ice-cold saline buffer and directly transported to the laboratory within 1 hour from sample collection. Tissue samples were then immediately processed to preserve the intranuclear genomic and protein architectures of cell nuclei. This setup implies a timeline in which we process research biopsies samples to isolate 4f-SAMMY chromatin fractions before receiving the results of the pathologist examination of the diagnostic biopsies. As a consequence, only a subset of samples associated to a clear PCa diagnosis by the pathologist were moved forward for the actual sequencing of 4f-SAMMY-seq fractions. Samples selected for epigenome studies consist of 25 fresh biopsies divided on the basis of the histology and the spatial distribution of the positive cores in two different groups: 18 PCa biopsies from patients with

histologically confirmed prostate cancer and 7 from patients who had no cancer in any biopsy core.

Tissues processing

The biopsy specimen was stored at 4–8 °C for maximum 3 h before dissociation, avoiding its freezing. Given the little size of needle biopsy cores (typically 20–30 mg) we empirically determined the digestion condition to achieve the optimal recovery of living cells. Briefly, the biopsy tissue was transferred in a 2 ml microcentrifuge tube and rinse twice with 1 ml of ice-cold sterile PBS 1X. Then, the tissue was cut into small pieces (~1 mm) with autoclaved surgical scissors directly in the microcentrifuge tube. The resulting minced tissue was enzymatically digested by adding 1 ml of prewarm HBSS (Gibco, 14025) containing 200 units of collagenase type I (Life Technologies, 17018-029) per ~10 mg of tissue plus 67 μg DNase I (Sigma-Aldrich, 10104159001). Tissue dissociation was carried out in a water bath at 37 °C shaking vigorously for 10 s every 5 min. To prevent tissue over-digestion, the best incubation time for each experiment was established by monitoring cell viability, cell debris and aggregates every 10 min (usually ranges from 1 to 1.5 h). After completed digestion cells were washed once by topping up to 2 ml with RPMI + 10% FBS and spun down at 300 g. The cell pellet was resuspended in RPMI + 10% FBS and dispersed by passing through a 75 μm cells strainer, followed by an additional wash of the filter with RPMI + 10% FBS. Finally, the cells were spun down at 300 g and resuspended in 1 ml of ice-cold PBS for counting with a hemocytometer.

Histological and immunofluorescence evaluation

One third portion of each research biopsy tissue specimen was embedded in Killik (Bio-Optica, 05-9801), immediately frozen in precooled isopentane (MilliporeSigma, 277258) and stored at −80°. OCT-embedded biopsy cores (one per patient) were serially sectioned with 10 μm thickness in a cryostat at −20 °C. Ten slides per patient, containing multiple sections representing distinct regions of the same tissue were prepared in parallel and stored at −80 °C. Hematoxylin and Eosin staining (H&E) was performed using H&E Staining Kit (Abcam, ab245880). The H&E-stained slides were reviewed by an expert genitourinary pathologist to assign the Gleason Score according to the International Society of Urinary Pathology grading system. The same pathologist evaluated all specimens (diagnostic and research biopsies) presented in this work. For immunofluorescence, frozen tissue sections were briefly thawed at room temperature (RT), placed in precooled acetone for 20 min at −20 °C and washed 3 times in PBS 1X for 2 min at RT. To permeabilize tissues the sections were incubated with 0.5% Triton X-100 in PBS 1X for 10 min at RT with mild agitation. After three washes in PBS 1X for 2 min at RT, coverslips were blocked by incubating with 5% Donkey serum, 3% BSA in PBS 1X for 1 h at RT. After three washes in PBS 1X for 2 min at RT, coverslips were incubated with 1:500 Lamin A/C primary antibody (Santa Cruz sc-6215) 1:500 in blocking solution overnight at 4 °C in a humidified chamber, isolating the tissue section by a hydrophobic pen. The following day, sections were washed 3 times in PBS 1X for 2 min at RT and incubated with fluorescence-conjugated secondary antibody (Alexa FluorTM 488-conjugated Donkey Anti-goat IgG Invitrogen, A-31571), 1:500 3% BSA in PBS 1X for 1 hour at RT in a dark humidified chamber. After three washes in PBS 1X for 2 min at RT, the coverslips were incubated with Hoechst solution (Thermo Fisher Scientific H3570, 1:500) for 10 min at RT in the dark, rinsed in PBS 1X and mounted on microscope slides with ProLong antifade mounting media (Thermo Fisher Scientific, P36930).

Images were acquired using a Leica TCS SP5 confocal microscope with a HCX Plan Apo ×63/1.40 objective, with a 5-zooming factor. The NIS-Elements v.5.30 software (Nikon-Lim) was employed to analyze nuclei physical parameters including area and circularity. Hoechst staining facilitated nucleus identification and delineation of the region of interest. Over 5 field of views (FOVs) were acquired and analyzed per sample, totaling the examination of more than 70 nuclei within each sample of the cohort (seven non-neoplastic controls, CTR and ten primary prostate cancer patients, PCa). The mean nuclear area or circularity was calculated for each sample and plotted alongside other samples within the same group on the charts. The analysis of Hoechst and Lamin A/C distribution across nuclei was conducted using Fiji software. All analyses were conducted in a blinded manner using a GA2 analysis module developed for semi-automatic nuclei identification. This method starts with a bright spot detection algorithm to locate each nucleus, followed by a region-growing algorithm based on Hoechst intensity and nuclear area to delineate the full extent of each nucleus. Leveraging the plot profile plugin, a 10 μm fixed line was implemented to trace and capture the fluorescence intensity distribution across each nucleus. To standardize comparisons between diverse samples, a normalization step was performed. This involved calibrating the fluorescence intensity of individual pixels along the line to the mean intensity derived from the entire set of pixel lines. For each sample the mean distribution values for both Hoechst and Lamin A/C were calculated and graphically represented alongside other samples within the same group.

Flow cytometry analysis

To quantify the relative amounts of cell populations in each biopsy, 10,000 cells from the digestion step were stained, acquired on a BD FACSCantoII Flow Cytometer and analyzed with FlowJo software in the INGM FACS facility. To avoid unspecific binding, antibodies were incubated with PBS-BSA 1% for 30 min at 4 °C. TO-PRO®−3 stain (Thermofisher, T3605) was used to assess cell viability. Tissue resident leukocytes were identified as CD326-/CD45+ (EpCAM-CD326 FITC, B347197; CD45 Pacific Blu Biolegend, 982306); epithelia cells as CD326 + /CD45-, and stromal cells as CD326-/CD45-.

4f-SAMMY-seq isolation of chromatin fractions

Chromatin fractionation on tissue-derived single cell suspension was performed with minor adaptations to the protocol described in ref. 16. Cells were counted, washed in cold PBS and resuspended in 600ul cold cytoskeleton triton buffer, CSK-Tr: 10 mM PIPES pH 6,8; 100 mM NaCl; 1 mM EGTA; 300 mM Sucrose; 3 mM MgCl2; 1X Protease Inhibitor Cocktail (Roche, 04693116001); 1 mM PMSF (Sigma-Aldrich, 93482); 1 mM DTT; 0,5% Triton X-100. After 10 min on a rotator at 4 °C, samples were centrifugated for 3 min at 900 *g* at 4 °C and supernatant was collected as S1 fraction. Pellets were washed for 10 min on the rotator at 4 °C with an additional volume of the same buffer followed by centrifugation for 3 min at 900 g at 4 °C. Chromatin was then digested by using 25 units of DNase I (Invitrogen, AM2222) in 100 ul of cytoskeleton CSK buffer: 10 mM PIPES pH 6.8; 100 mM NaCl; 1 mM EGTA; 300 mM Sucrose; 3 mM MgCl2; 1 mM PMSF; 1X Protease Inhibitor Cocktail for 60 min at 37 °C. To stop digestion, ammonium sulphate was added to samples to a final concentration of 250 mM and, after 5 min on ice, samples were pelleted at 900 *g* for 3 min at 4 °C and the supernatant was collected as S2 fraction. After a wash of 10 min on a rotator at 4 °C with 200ul of CSK buffer followed by centrifugation for 3 min at 3000 *g* at 4 °C, the pellet was further extracted 10 min on a rotator at 4 °C in 100 ul of CSK-NaCl buffer: 10 mM PIPES pH 6.8; 2 M NaCl; 1 mM EGTA; 300 mM Sucrose; 3 mM MgCl2; 1 mM PMSF; 1X Protease Inhibitor Cocktail, centrifuged at 2300 *g* 3 min at 4 °C and the supernatant was collected as S3 fraction. Pellets were washed twice for 10 min on the rotator at 4 °C followed by centrifugation at 3000 *g* 3 min at 4 °C with 200ul of CSK-NaCl buffer. Finally, pellets were solubilized in 100ul of 8 M urea for 10 min at room temperature and labelled as S4 fraction. Fractions were stored at −80 °C until DNA extraction.

4f-SAMMY-seq DNA extraction, library preparation and sequencing

Chromatin fractions (S2, S3 and S4) were diluted 1:2 in 1X TE buffer (10 mM TrisHCl pH 8.0, 1 mM EDTA) and incubated with 3 U of RNAse

cocktail (Ambion, AM2286) at 37° for 90 min, followed by 40 µg of Proteinase K (Invitrogen, AM2548) at 55° for 150 min. Genomic DNA was then isolated using phenol/chloroform (Sigma-Aldrich, 77617) extraction, followed by a back extraction of phenol/chloroform with additional volume of TE1X. DNA was precipitated in 2 volumes of cold ethanol, 0.3 M sodium acetate and 20 ug glycogen (Ambion AM9510) for 1 h on dry ice or overnight at −20 °C. Dry pellets were resuspended in 50 ul (S2) or 15 ul (S3 and S4) of nuclease-free water and incubated at 4 °C overnight. On the next day, S2 was further purified using PCR DNA Purification Kit (Qiagen, 28106) and separated using AMPure XP paramagnetic beads (Beckman Coulter, A63880) with the ratio 0.90/0.95 to obtain smaller fragments conserved as S2S (< 300 bp) and larger fragments labelled as S2L (> 300 bp) fractions. Both were resuspended in 20 ul of nuclease-free water and then reduced to 15 ul using a centrifugal vacuum concentrator. S2L, S3 and S4 fractions were sonicated in a Covaris M220 focused-ultrasonicator using screw cap microTUBEs (Covaris, 004078) to obtain a smear of DNA fragments peaking at 150–200 bp (water bath 20 °C, peak power 30.0, duty factor 20.0, cycles/burst 50, 150 s for S2L and 175 s for S3 and S4). Fractions were quantified using Qubit 4 fluorometer with Qubit dsDNA HS Assay Kit (Invitrogen, Q32854). Libraries were generated from each sample using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645L) and Unique Dual Index NEBNextMultiplex Oligos for Illumina (NEB, E6440S). Libraries were then qualitatively and quantitatively checked on and run on an Agilent 2100 Bioanalyzer using High Sensitivity DNA Kit (Agilent, 5067-4626). Libraries with distinct adapter indexes were then multiplexed and, after cluster generation on FlowCell, sequenced for 50 bases in single or paired-end reads mode on an IlluminaNovaSeq 6000 instrument at the IEO Genomic Unit in Milan.

RNA extraction, library preparation and sequencing

Ten thousand cells retrieved from the digestion step were stabilized in 200 µl of 1Thioglycerol/Homogenization Solution of the Maxwell® RSC miRNA Tissue Kit (Promega, AS1460) and stored frozen at −80 °C for total RNA automated purification using Maxwell® RSC 48 Instrument (Promega, AS8500). Total RNA was quantified by Qubit 4 fluorometer with Qubit RNA HS Assay Kit (Invitrogen, Q32852) and assessed by Agilent 2100 Bioanalyzer using Agilent RNA 6000 Pico Kit (Agilent, 5067-1513) to inspect RNA integrity. For each sample, 1 ng of total RNA was used to construct strand specific RNA-seq libraries with SMARTer Stranded Total RNA-Seq Kit - Pico Input (Takara, 634487). The yield and quality of the libraries were evaluated on Agilent 2100 Bioanalyzer using High Sensitivity DNA Kit (Agilent, 5067-4626). RNA-seq libraries were sequenced on the Illumina NextSeq™ 550 system at the sequencing facilities of Humanitas or Fondazione IRCCS Ca' Granda-Ospedale Maggiore Policlinico in Milan in paired-ends mode.

ChIP-seq and RNA-seq public datasets

The ChIP-seq data of histone post translational modifications (histone marks) for non-neoplastic prostate gland tissue were downloaded from the following datasets available on ENCODE data portal[20]: ENCSR763IDK (H3K27ac), ENCSR768PFZ (H3K36me3) and ENCSR133QBG (H3K9me3). These data were downloaded as raw high-throughput sequencing reads (FASTQ) and analyzed as described in the next sections. ChIP-seq data for H3K27ac, H3K4me3, H3K27me3 and Androgen Receptor (AR) in primary prostate cancers from ref. 37 were downloaded from Gene Expression Omnibus (GEO) database under the accession number GSE120738. We used the ChiP-seq peak calls as reported by the original authors. In addition to epigenetic data of prostate cancer we also downloaded the normalized RNA-seq count table generated from the same sample cohort using the accession number GSE120741.

TPM normalized counts of prostate cancer cell lines used for Supplementary Fig. S9a were downloaded from Human Protein Atlas (https://v23.proteinatlas.org/about/download)[51–53]. In particular we used the file rna_celline.tsv for the expression data and rna_celline_description.tsv for the metadata. We assigned androgen dependency to each cell line according to the androgen pathway activity score inferred by human protein atlas. If the score was significant, we labeled as androgen dependent, if not androgen independent. We applied our PCI score (described below) to classify them in HDD-like and LDD-like.

ChIP-seq and 4f-SAMMY-seq high-throughput sequencing reads data analyses

High-throughput sequencing reads were trimmed using Trimmomatic (v0.39)[54] with the following parameters for 4f-SAMMY-seq and ChIP-seq: 2 for seed_mismatch, 30 for palindrome_threshold, 10 for simple_threshold, 3 for leading, 3 for trailing and 4:15 for sliding window. The sequence minimum length threshold of 35 was applied to all datasets. We used the Trimmomatic "TruSeq3-SE.fa" (for single-end reads) and "TruSeq3-PE-2.fa" (for paired-end reads) as clip files. After trimming, reads were aligned using BWA (v0.7.17-r1188)[55] setting −k parameter as 2 and using as reference genome the UCSC hg38 genome (only canonical chromosomes were used) and the output saved in BAM file format. We uniformly used and aligned only a single read per DNA fragment for both single-end and paired-end sequencing samples. The PCR duplicates were marked with Picard (v2.22; https://github.com/broadinstitute/picard) MarkDuplicates option, then filtered using Samtools (v1.9)[56]. In addition, we filtered all the reads with mapping quality lower than 1. Each sequencing lane was analyzed separately and then merged at the end of the process.

Reads distribution profiles analysis

To compute reads distribution profiles (genomic tracks) for individual fractions of 4f-SAMMY-seq, we used Deeptools (v3.4.3)[57] bamCoverage function. For these analyses the genome was binned at 50 bp, the reads extended up to 250 bp and Deeptools RPKM normalization method was used. We considered a genome size of 2,701,495,761 bp (value suggested in the Deeptools manual https://deeptools.readthedocs.io/en/latest/content/feature/effectiveGenomeSize.html) and we excluded regions known to be problematic in terms of sequencing reads coverage using the blacklist from the ENCODE portal (https://www.encodeproject.org/files/ENCFF356LFX/).

To compute the genomic tracks for ChIP-seq IP over INPUT enrichment profiles (log_2 normalized ratio) we used the SPP R package (v1.16.0)[58] and R statistical environment (v3.5.2). The reads were imported from the BAM files using the "read.bam.tags" function, then filtered using "remove.local.tag.anomalies" and finally the comparisons were performed using the function "get.smoothed.enrichment.mle" setting "tag.shift = 0" and "background.density.scaling = TRUE". The resulting enrichment signal corresponds to a log2 normalized ratio between the pair of sequencing samples.

For computing the relative comparisons between two 4f-SAMMY-seq fractions (relative enrichment, *i.e.* log_2 normalized ratio) we used the same procedure described above for ChIP-seq IP over INPUT enrichment profiles. We defined the solubility profile as the relative enrichment ratio of 4f-SAMMY-seq sequencing reads distribution along the genome for S2L vs S3 fractions. We used S3 as baseline in this ratio as it is the second most consistent fraction among controls (Supplementary Fig. S3d).

To compute correlations between genomic tracks, we used R (v3.5.2) base function "cor" with "method = Spearman". The genomic tracks were imported in R using the rtracklayer (v1.42.2)[59] library. Then the original genomic track files with 50 bp resolution were re-binned by averaging data at 150 kb resolution using the function "tileGenome" and the correlation was computed per chromosome. The correlation values obtained for each chromosome were then summarized in one value describing the genome-wide sample correlations through a

weighted mean, where the weight of each chromosome corresponds to its length.

Open histone marks (H3K4me1, H3K36me3) peaks were called with MACS (v2.2.9.1) using a broad-cutoff of 0.1[60].

H3K9me3 peaks were defined using the EDD (v1.1.19) software[61] with parameters (binsize = 150 Kb and gap penalty = 25) processing the filtered bam files obtained as described above. The "required_fraction_of_informative_bins" parameter was set to 0.98. The blacklisted regions were defined as for the reads distribution profile analyses described above.

The enrichment boxplots shown in Fig. 2e was produced retrieving for each histone mark peak the genomic bins that fall in that genomic region. For each group of patients (CTR, LDD and HDD groups) and for each genomic bin we plot the mean 4f-SAMMY-seq solubility profile across the group. The values used were the quantile normalized tracks (across all samples) S2LvsS3 produced using SPP (see "High-throughput sequencing reads analyses"). Genomic tracks obtained by SPP or Deeptools were saved in bigWig file format.

To integrate epigenomic data from public dataset of prostate cancer we used ChIP-seq peaks BED files H3K27ac, H3K4me3, H3K27me3, and Androgen Receptor (AR) with the peak calls generated by the original authors (Supplementary file GSE120738.tar, see above). The coordinates were converted from the Genome build hg19 to hg38 using CrossMap (V0.6.6) tool[62] and UCSC chain file ftp://hgdownload.soe.ucsc.edu/goldenPath/hg19/liftOver/hg19ToHg38.over.chain.gz. By using gene expression data from the same cohort of prostate cancer samples we applied our PCI score (as described below) to classify samples into HDD-like and LDD-like groups. We then generated for each ChIP-seq target a peak concordance score summarizing the frequency (range 0–1) of each peak within each group (HDD-like and LDD-like samples). More specifically, we generated a union set of all peaks called in the original dataset for that specific ChIP target. For each sample then we intersected the peak coordinate regions with the union set reporting the coordinate if at least 1 bp was intersected with the union peaks. We finally calculated for each peak the frequency (proportion) of samples within each group (HDD-like or LDD-like) in which the peak was present. This resulted in two group-specific concordance score tracks where the signal at each region reflects the fraction of samples in that group exhibiting a ChIP-seq peak (range of values 0–1). A signal value of 1 indicates complete conservation of the peak within the group, while a value of 0 indicates complete absence. To functionally characterize the genomic regions displaying histone marks differences between LDD and HDD, we applied a low-stringency filtering approach to identify broad trends. We selected active marks (H3K4me3, H3K27ac) peaks showing at least a 20% difference in the number of samples with a peak detected and a FoldChange > 1.25 in conservation score.

Differential peak coordinates were intersected with Refseq annotated Protein Coding genes coordinates (see below the GTF file reference). Genes with intersection over the gene body or transcription start site (TSS) were included. For closed chromatin marks (H3K27me3) peaks we decreased the threshold to 10% as these peaks are generally more difficult to detect and we are interested in the general trend, rather than individual peaks significance. This relaxed threshold was used to achieve a number of genes allowing to perform functional enrichment on the general trend of changes. Functional gene enrichment was performed using the R (V4.2) package enrichR (V3.2) querying our genes over the MSigDB Hallmark_2020, filtering the resulting terms for adjusted p-value < 0.05. It must be noted that in the Stelloo et al. dataset used[37], not all the samples have all epigenomic and transcriptomic profiles. This resulted in the following number of samples with ChIP-seq concordance score tracks for HDD and LDD like groups: AR: 42 HDD, 40 LDD; H3K27ac: 44 HDD,42 LDD; H3K4me3: 29 HDD, 21 LDD; H3K27me3: 36 HDD, 34 LDD.

Genomic tracks visualization

The visualization of genomic tracks was performed with Gviz R library (v1.26.5)[63]. The track profile was calculated using the function "DataTrack" (the input bigWig file was imported using the function "import" of the rtracklayer library) and plot using the function "plotTracks". Supplementary Fig. S3e has been created plotting each track as type "polygon" and all the tracks have been set using a window of 500 except for ChIP-seq of H3K27ac where the window has been set to 5000. It is worth remarking that the "window" parameter in the "DataTrack" function is referring to the number of intervals in which the displayed region is divided to plot the profile. As such, a larger number indicate a finer grain definition of the profile. As H3K27ac is known to have sharp peaks, we aimed to plot a finer grain profile. For Fig. 1c the window was set to 1000 and for all the tracks except for ChIP-seq of H3K27ac where we used 5000, *i.e.* the same value used for the Supplementary Fig. S3e. CTR and PCa samples track overlayed with confidence interval were drawn using at the same the time the types "a" and "confint". The same settings of histone mark tracks of Fig. 1c were used for Fig. 2b.

Chromatin compartment analysis

Chromatin compartments were calculated using a revised version of the CALDER algorithm (version 1.0[64]), as implemented in the original 4f-SAMMY-seq pipeline for the sub-compartment identification[16]. Namely, for each chromosome: i) we calculated the four 4f-SAMMY-seq fractions (S2S, S2L, S3, S4) reads distribution profiles, binned at 150 kb and normalized with RPKM (see "Reads distribution profiles analysis" section); ii) for each genomic bin, defined as a vector containing the four RPKM values (one for each fraction), we calculated the Euclidean distance (dist, R stats package, method = "euclidean") with all the other bins of the same chromosome. These steps produced an NxN matrix (hereinafter referred to as biochemical similarity matrix), where N is the number of bins for the considered chromosome. Starting from the biochemical similarity matrix, we performed the CALDER procedure as described in ref. 16, to derive the eigenvector and to reconstruct chromatin compartmentalization. Specifically, we called sub-compartments but limiting their segmentation to the highest level, thus obtaining only 2 compartments corresponding to "A" and "B" compartments (Fig. 2a, b). As in the original publication on 4f-SAMMY-seq[16], we used gene density to adjust the sign of eigenvectors: the higher gene-density domains will be assigned to positive eigenvalues (A compartments), as per the convention adopted in the Hi-C data analysis field.

The consensus of compartment calls across healthy controls has been produced labeling each genomic bin (150 kb) according to its more frequent compartment across samples: e.g. if a bin is labeled as A in at least 4 out of 7 CTR samples (i.e. consensus compartment if the genomic bin is present in more than 50% of CTR), that bin will be defined as A in the consensus of controls, and vice versa for bins labelled as B in at least 4 controls (Fig. 2c). The definition of compartment shifts for each single sample was based on the concordant or discordant compartment classification for each 150 kb genomic bin (Supplementary Fig. S4a). The compartment domains shared across patients as shown in Supplementary Fig. S4a were reported using the library upsetR.

For the integration of DEGs and compartment changes, we generated A consensus and B consensus compartment separately for HDD (at least 3 out of 5) and LDD (at least 3 out of 5) samples. The definition of HDD or LDD compartment switch with respect to the CTR was based on the discordant consensus classification of A and B consensus between the CTR and HDD (or LDD).

The clustering of patients in Fig. 2d were based on Jaccard Indexes (JI) of compartments concordance for each pair of patients. The resulting matrix of pairwise JI was grouped by hierarchical clustering based on the Euclidean distance (among rows or columns) and complete linkage using the pheatmap function in the pheatmap R library.

Cell type deconvolution

For cell type deconvolution analysis, we fist normalized for each sample raw count to TPM (Transcript per Million) by using R 3.6.1 environment and applying the following steps. First, read counts for each gene were normalized (divided) by the length of the gene in kilobases (RPK, Reads per kilobase). Then RPK values in a sample are summed and divided by 1.000.000 to obtain the scaling factor. Finally, RPK values for each gene are divided by the scaling factor.

For the deconvolution of the 3 macro populations (immune, epithelia, stroma cells) we defined a custom signature matrix based on single cell expression data of prostate tissue from the Human Protein Atlas (https://www.proteinatlas.org/; V. 21.1)[51–53]. Following the cluster classification of Human Protein Atlas we considered as epithelia the subpopulations annotated as prostatic glandular cells, basal prostatic cells or urothelial cells. We considered as stroma the populations annotated as muscle cells, fibroblast or endothelial cells. We considered as immune cells infiltrate the populations annotated as t-cells or macrophages. Finally, we selected as representative genes for each population only those with a TPM > 10 for the specific population and a TPM < 2 in all the two remaining populations. We finally got a macropopulation signature matrix of 702 genes. We run the deconvolution analysis exploiting this signature matrix on our bulk data by using the relative mode of CIBERSORTX (V.1.0)[65] (Supplementary Fig. S4f).

To compute the immunity score on our biopsies we used the software CIBERSORTX (V.1.0)[65] applying the previously computed TPM matrix of our bulk RNA (mixture file) on the LM22.txt provided in the CIBERSORTX repository. The LM22 leukocyte signature is based on a matrix of 547 genes discriminating 22 human hematopoietic cell types isolated from PBMC. We enabled batch correction by using the B-mode correction option. We next run the analysis in absolute mode by applying 500 permutations. Each sample in the mixture file gets a score that reflects the absolute proportion of each cell types in LM22 mixture file. We finally computed the absolute leukocyte score: i.e. the sum for each sample of all the 22 scores of leukocytes populations. We compared the absolute leukocyte score for each biopsy across the different groups (CTR, LDD, HDD) (Supplementary Fig. S4g).

We leveraged a single cell transcriptomic study of primary prostate cancer biopsies by ref. 39, to derive a signature discriminating between benign and tumor epithelial cells (also discriminating between ERG+ and ERG- cells). More specifically, we used the single-cell RNA-seq count matrix (dge_E.rds) available from the github repository associated to the original publication. We used their cell type annotations and a curated list of epithelial cell-type-specific marker genes. We normalized raw count data using the NormalizeData() function from the Seurat package (R.4.3.3_; Seutat.5.1) to obtain log-normalized expression values. A signature matrix was then constructed by calculating, for each marker gene, the average normalized expression across all single cells belonging to each annotated epithelial cell type. Specifically, for each gene-cell type pair, the mean expression was computed over all cells annotated as that cell type. This resulted in a matrix with genes as rows and epithelial cell types as columns (3 benign epithelial cell types LE, BE Club; 2 tumor epithelial cell types ERG + , ERG-), where each entry represents the mean normalized expression level of a given gene in a specific cell type. This signature matrix was used to perform transcriptomic deconvolution using CIBERSORTX and estimating in this way the contribution of normal epithelial and tumor cells in each patient biopsy bulk transcriptomic profile.

Transcriptomic analyses

For RNA-seq data, the overall quality of the sequenced reads was assessed using FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) (V. 0.11.8), then reads were trimmed with Trimmomatic software[54] (version 0.39) removing the adapters (ILLUMINACLIP:Picov2smart-PE.fa), primer dimers, and low quality bases at the beginning and at the end of the reads (trimmomatic PE phred33 LEADING:3 TRAILING:3 SLIDINGWINDOWD:4:15 MINLEN:36). STAR[66] (V. 2.7.0f_0328) was used to index (STAR --runMode genomeGenerate) the Human Genome (GENCODE Release 39, GRCh38 primary assembly genome[67]) and to align sequenced reads in paired-end mode (--readFiles R1.FASTQ R2.FASTQ) on the indexed reference genome. Multimapping reads and PCR duplicate were marked in the final output (--bamRemoveDuplicatesType Unique) and unaligned reads stored in a different file (--outReadsUnmapped Fastx).

The read counts on genes were calculated using as a reference a GTF file with RefSeq annotation downloaded from UCSC stored in the following directory:(http://hgdownload.soe.ucsc.edu/goldenPath/archive/hg38/ncbiRefSeq/109.20190905/hg38.109.20190905.ncbiRefSeq.gtf.gz). This file was further processed to remove non canonical and mitochondrial chromosomes, selected only curated genes (NM, NR) and finally split in protein coding (NM) and noncoding (NR) files. Reads count was performed with HTSeq-count (V. 0.13.5) on bam files (previously generated by STAR) using as a feature the union of all exons in a gene. The type of library was specified with "-s reverse" parameter. The reads that align to more than one position in the reference genome were discarded (htseq-count --non-unique none). The full matrix with raw read counts for each sample were loaded in R 3.6.1 and normalized using DESeq2[68] median of ratios. Differential expression analysis was performed with DESeq2V. 1.26[69]) using Wald test and the Benjamini and Hochberg correction for multiple tests, to compute p-values and adjusted p-values, respectively.

Gene transcript detection

Transcript fusions were detected using STAR-fusion (V.1.10.1) with default parameters. The trimmed reads were aligned in paired-end against the GRCh38 human reference genome using a precompiled Trinity Cancer Transcriptome Analysis Toolkit (CTAT) genome resource library based on GENCODE v37 annotations (https://data.broadinstitute.org/Trinity/CTAT_RESOURCE_LIB/GRCh38_gencode_v37_CTAT_lib_Mar012021.plug-n-play.tar.gz). The CTAT genome library includes STAR indices, transcript annotations, and fusion-related reference data. Only high-confidence fusion candidates passing internal STAR-Fusion filters (fusion fragment per million FFPM > 0.1) were retained for downstream analysis.

Enrichment analyses

Up-regulated genes (with adjusted p-value < 0.05 and Log_2 Fold Change > 1) and Down-regulated genes (with adjusted p-value < 0.05, and Log_2 Fold Change < −1) were used separately to query the gene-list enrichment tool Enrichr (https://maayanlab.cloud/Enrichr/)[70–72] over the main Gene Ontology classes (Biological Process - BP, Molecular Function - MF, Cellular Component - CC) updated to 2023 taking as final enriched terms only those with Benjamini-Hochberg adjusted p-value < 0.05. Genes within the recurrent compartment switch "A to B" and "B to A" were also queried over the main Gene Ontology classes taking also in this case the enriched terms with Benjamini-Hochberg adjusted p-value < 0.05. GSEA analysis[31] was performed in Preranked mode. Starting from raw p-values of all genes we generated a ranked matrix sorted by the score resulting from the following formula: $-Log_{10}(p\text{-value})$ multiplied by the sign of the Log_2 Fold Change. For the gene set references we downloaded H (Hallmark) and C5 (Ontology) GMT files from MSigDB. Finally, we performed the analysis using the parameters "Number of permutations": 1000; "Collapse": No; "Enrichment Statistic": classic; "Max size": 500; "Min size":15. Resulting geneset with a family-wise error rate (FWER) < 0.05 are finally sorted using the Normalized Enrichment Score (NES).

We then inferred pathway activity for the comparison of our tumor subtypes by using the R package decoupleR[73] (V. 2.4). We used the get_progeny (organism=human,top=100) function to get pathways

from PROGENy (V.1.20)[32] a curated collection of signaling pathways derived from perturbation experiments, where each gene has a specific weight describing its positive or negative response to a given pathway stimulation. The top 100 responsive genes in the compendium ranked by *p*-value were used for each pathway. Starting from Wald statistic values previously computed for each gene with DESeq2 (LDDvsCTR, HDDvsCTR and HDDvsLDD comparisons) we run the function decouple with parameters statistics=c("mlm","ulm","wsum") and consensus_score=TRUE, to estimate a consensus score across the top performer methods (according to benchmark done by decoupleR developers) in pathway activity prediction. Multivariate linear model (mlm), Univariate linear model (ulm), Weighted Sum (wsum) are recommended by developers of decoupleR since these methods better estimate pathway activity considering the weights associated with pathway related genes.

ssGSEA for compartment-specific gene sets

To assess the enrichment chromatin compartment derived gene sets, we performed single-sample Gene Set Enrichment Analysis (ssGSEA) using the GSVA R package (V1.46.0). TPM Gene expression data of PCa samples were loaded and log2-transormed. A conserved A and B conserved compartments were derived computing the coordinates with the same compartment type in at least 9 out of 10 PCa. Coordinates of conserved A compartment were intersected with gtf file obtaining the final gene set (gmt file format). The same was done for B compartment. The resulting A conserved gene set is composed by 7648 genes. The resulting B conserved gene set is composed by 1010 genes. The background A gene set is composed by selecting 7648 random genes from the full TPM matrix. For background of B we selected 1010 random genes. For each background we repeated the random selection 50 times. In total the final gmt file has 2 real gene sets and 100 random gene sets. Finally the ssGSEA scores were computed using gsva() function with the method set to "ssgsea" and verbose =TRUE.

DEGs and compartment switches

For the integration of DEGs and chromatin compartment switches, we applied the same threshold used for the enrichment analysis section (see above) to get significant Down-regulated and Up-regulated genes for the specific comparisons HDD vs CTR and LDD vs CTR. Additionally we annotated as NODEG the remaining list of genes with no significant changes (adjusted *p*-value > 0.05). We then annotated the genes found in the compartment switches "A to B" and "B to A" (consensus compartment switches defined as described above). Additionally, we also annotated as NO SWITCHES the genes found in the remaining compartment without switches (A CTR; A HDD; B CTR; B HDD). We finally generated a 3 by 3 contingency table counting the occurrences of intersected genes within the different categories of DEGs and Compartment switches. We performed on the resulted table a chi-squared test and using the standardized residuals to understand which combination contribute most to the association compartment switches vs expression changes. We found especially up-regulated DEGs enriched in "B to A" switches and down-regulated DEGs depleted in the same "B to A". We confirmed the significance of these enrichment, depletion with one sided binomial test. Functional gene enrichment of coherent switches-DEGs (eg. Genes into BtoA switches and also Upregulated) was performed using the R (V4.2) package enrichR (V3.2).

TCGA data

TCGA-PRAD RNA-Seq expression data were downloaded by using gdc-client tool on gdc_manifest.2022-08-08.txt metafile input downloaded from the GDC portal repository (https://portal.gdc.cancer.gov/; V.34.0; Release July, 27, 2022) by selecting from Web interface the TCGA project, prostate and RNA-seq count. The main clinical parameters (Gleason Score, T stage, N stage and age) and info about the samples (primary tumor, normal biopsy and metastatic) were selected from the same Web section of GDC portal repository by downloading clinical and sample-sheet files. ABSOLUTE tumor purity score[74] was downloaded from https://gdc.cancer.gov/about-data/publications/pancanatlas. The file is TCGA_mastercalls.abs_tables_JSedit.fixed.txt (Downloaded October, 13, 2022)[75]. Excluding primary normal biopsies, we obtained from the TCGA repository described above the 500 FPKMUpperQuartile normalized expression counts from prostate primary tumors. The Biochemical Recurrence Free (BCR) status for each patient was obtained from PCaDB (http://bioinfo.jialab-ucr.org/PCaDB/; Release May, 10 2021; eSet V.1.3), downloading the TCGA-PRAD_eSet.RDS object and applying the pData function from Biobase package. We then excluded the only sample with a Gleason Score pattern of 2 + 4 never described in literature. Despite we performed expression analyses assigning a PCI score on the entire cohort of 499 samples, the BCR data available are 466.

TCGA-PRAD Bulk ATAC-seq data were downloaded from the Genomic Data Commons portal (https://gdc.cancer.gov/about-data/publications/ATACseq-AWG), selecting the RPKM normalized bigwig data of PRAD cohort. These ATAC-seq data were performed in replicates on 26 prostate cancer samples for a total number of 52 tracks. We also obtained the prostate cancer ATAC-seq consensus peaks always from the same database, downloading the "All cancer type-specific peak sets.zip" and selecting the PRAD_peakCalls.txt file only. Exploiting the peak call file, we computed for each sample the average ATAC-seq reads enrichment signals by using multiBigwigSummary function of deeptools. We then computed a Wilcoxon test for each of these coordinates comparing ATAC-seq enrichment between HDD-like (14 samples) vs LDD-like (38) previously classified by our PCI index. We selected as differential those regions with a *p*-value < 0.01.

PCI score (Prostate Compartmentalization Index)

Starting from our 101 DEGs of the HDD vs LDD comparison we defined two gene lists with 24 HDD Up-regulated and 77 HDD Down-regulated genes. We subset the TCGA matrix extracting only those genes matching with our full list of 101 genes. We performed a log transformation of TCGA gene expression matrix: Log_2(matrix +1). The log transformed expression of each gene across the 499 samples was standardized into a z-score. For each sample we computed 2 median values on the z-score transformed matrix: the median of the 24 HDD Up-regulated genes (HDD_UP) and the median of the 77 HDD Down-regulated (HDD_DOWN) genes. Finally, we computed the PCI score defined as the difference of median values: PCI=median(HDD_UP) - median(HDD_DOWN). The resulting PCI score with a positive value will indicate that a sample has HDD-like transcriptomic features, whereas a PCI negative value will classify a sample as LDD-like.

Survival analyses

Univariate and multivariate regression analysis were performed in R using survival and survminer packages. First, we performed a log-rank test to compare survival Kaplan-Meier curves (in terms of BCR STATUS AND BCR TIME) between HDD-like and LDD-like samples. Then we performed a multivariate Cox proportional-hazard model including the following covariates: HDD-LDD status (previously assigned according to the PCI score) of the samples, pathological stage, age at diagnosis, Gleason Score and Tumor purity score. For the multivariate Cox regression analysis, the Gleason Score was included as numeric feature by summing the primary and secondary numeric grades, as reported for TCGA samples.

Signature refinement

To refine the 101 gene signature, we applied univariate cox regression analysis on each single gene expression values across the TCGA samples, including in the model the BCR STATUS and BCR TIME. The

p-values obtained for each gene are then adjusted with Benjamini-Hochberg. Then we selected only those genes whose expression is associated with Hazar Ratio (HR) < 1 or HR > 1 with a adj.*p*-value < 0.05. We obtained 21 candidate prognostic genes. Unsupervised clustering of these genes can cluster 2 main groups of samples with HDD-like and LDD-like features with high significance in the prognostic status of the two groups (Supplementary Fig. S6a). We further refined the signature by removing 3 genes (NPIPB3, CCNE2, FMO5) that were noisier in the unsupervised clustering heatmap: these are the only genes that cluster in a discordant manner compared to their original association to HDD or LDD phenotype in our dataset (Supplementary Fig. S6a). Finally, we used the resulting refined panel of 18 genes to recompute and validate our PCI score on other independent cohorts of transcriptomic profiles of prostate cancer patients.

Validation

Through the PCaDB (http://bioinfo.jialab-ucr.org/PCaDB/ Release May, 10 2021; eSet V.1.3) we downloaded normalized expression matrices and BCR data of three independent datasets of prostate cancer: Emory (GSE54460)[76], Stockholm (GSE70769)[77], Belfast (GSE116918)[78]. For each of them we classified samples computing the PCI score based on the refined 18 genes signature.

Immunohistochemistry (IHC)

Representative 4 µm-thick sections were cut from each block of 8 patients with PCa and stained with H&E or primary antibodies for SC5D (Atlas Antibodies HPA066283), IQGAP2 (Abcam ab187153), COL5A1 (Abcam ab7046), COL16A1 (Invitrogen PA5-103382) and COL18A1 (Abcam ab275746). Specifically, positive and negative controls were included in each experiment and the staining was revealed using DAB as chromogen using a Ventana Benchmark Ultra instruments with dedicated reagents (Ventana, Roche)[79]. All slides were counterstained with hematoxylin and digitalized using an Aperio scanner at 40x magnification (Leica Microsystems). Presence of staining for all antibodies was evaluated in the epithelial (normal and tumor glands), and stromal compartments. The scoring of the staining was performed manually. Briefly, two pathologists (VV and MMa) independently analyzed the slides from all patients in a blind mode. Three regions per slide were analyzed for markers presence. For SC5D and IQGAP2 the intensity of the staining and the percentage of positive cells were used to calculate the IHC score (% of positive cells X intensity). For COL5A1, COL16A1 and COL18A1 the specific staining in different cellular compartments was evaluated.

Cell culture

Cells were obtained by ATCC and grown either in DMEM supplemented with 10% (VCaP CRL-2876) or 5% (WPMY-1 CRL-2854) of FBS at 37 °C in a humidified atmosphere of 5% (v/v) CO2. Cell lines were tested transcriptionally with following primers: *SC5D* (fw: cagtggcgtgatcttggttc; rv: atc ctg gct aac acg gtg aa); *IQGAP2* (fw: ccccgatgcaacagaagatg rv: tgaagaaccttggccactga); *COL5A1* (fw:gtggtctgaagggcacaaag; rv: tccgagtttt cccttctccc); *COL5A3* (fw: tgaatgtcgtgcagctgaac; rv:agctcctctccattggt); *COL18A1* (fw: cagaaaggagccaagggaga; rv: ccctcatgccaaatccaagg); *MXRA8* (fw: tgacttctgaccctgacacc; rv: caaatgtggggtgctgagag). WPMY-1 prostate stromal cells were transducted with the pLenti-Lifeact-tdTomato (Addgene plasmid # 64048) and labelled cells isolated by fluorescence-activated cell sorting (FACS). 45,000 tdTomato-WPMY-1 cells were seeded in each 24-well tissue culture plate the day prior transfection. For each experimental point, 24 pmoles of either negative control (SCR, Life Technologies # AM4611) or a mix of COL5A1 (Life Technologies ID: 7797, 7892); COL5A3 (Life Technologies ID: 23255, 23350); COL18A1 (Life Technologies ID: 43706, s37393); MXRA8 (Life Technologies ID: 34039, HSS147729) specific siRNAs (6 pmoles each) were transfected using the RNAiMAX Transfection Reagent (Life Technologies, #13778075). Forty-eight hours after transfection cells were analyzed by qRT-PCR to assess the silencing efficiency and co-cultured with VCaP cells.

Wound healing assay

Thirty-six hours after silencing, TdTomato-WHPY cells were trypsinized, counted and 200.000 cells/experimental point mixed with VCaP cells in a 1:1 ratio. Cells were then seeded on Matrigel-coated wells. The following day (corresponding to forty-eight hours after silencing), a scratch in the confluent monolayer at the centre of each well was made with a 200 µL pipette tip. Cells were then washed twice with sterile DPBS 1X to remove debris and displaced cells prior, then cultured in serum-free DMEM. Live imaging to study wound-healing phenomena were carried out on four independent biological experiments. An environmental controlled (OKO-cage incubator; OkoLab) Nikon widefield Ti instrument (Nikon Europe) equipped with fast spinning disk (Crest Optics) and high-sensitivity video cameras (Zyla 4.2 and iXON DU888, both from Andor- Oxford Technologies) and both 16-LED (Pe-4000; CoolLed Ltd) and 4-laser excitation (LDI, Ltd) was used for image acquisition. In particular, we used widefield modality with white led for differential interference contrast imaging (DIC) and 555 nm fluorescence led line for fluorescence contrast imaging. Images were acquired with 3 h time-frame with a total of 45 h recording time-lapse. Acquisitions were conducted over an ad-hoc designed semi-automatic pipeline within the JOBS module of NIS-Elements v.5.30 software (Nikon Europe). Analysis was conducted with ad-hoc designed pipeline for segmentation and quantification, using the GA3 module in NIS-Elements v.5.30: briefly an algorithm for automatic wound-area detection over time was employed and specific binary masks were created to segment and classify each cell object detected either as DIC cell-object or as red-fluorescent cell-object. Measurements detected from the segmentation approach were further elaborated in a blind mode, to evaluate the percentage of wound closure and the numbers of cells within wound area, in all conditions. For image and dynamics representation, flattened RGB images and movies showing merged DIC and fluorescence channels, were selected from all samples. For video representation, a 10fps frame rate was chosen.

Statistics and reproducibility

The details of statistical analyses, including sample size, statistical test, and statistical significance, are reported in each individual figure legend. Statistical analyses of imaging data were performed with Prism software (v10.4.0). Other statistical analyses were performed with R statistical environment (v3.5.2), unless other specific tools were used as indicated for individual results. No data were excluded from the analyses. The experiments were not randomized as no treatment was performed on patients samples. The investigators performing histopathology and immunohistochemistry analyses were blinded with respect to patients classification as HDD or LDD.

Reporting summary

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

Data availability

The high-throughput sequencing data generated for this study have been deposited in the database ArrayExpress with accession number E-MTAB-14226. Source data are provided with this paper.

Code availability

The bioinformatic pipeline, along with a step-by-step tutorial explaining how to process 4f-SAMMY-seq data was previously published and publicly released in ref. 16 and available at this DOI URL (https://doi.org/10.6084/m9.figshare.25437121.v1) and also in the Github repository (https://github.com/Clockris/SAMMY-method_4f).

References

1. The Global Cancer Observatory (GCO): https://gco.iarc.fr/en.
2. Mottet, N. et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. *Eur. Urol.* **71**, 618–629 (2017).
3. Hamdy, F. C. et al. Fifteen-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer. *N. Engl. J. Med* **388**, 1547–1558 (2023).
4. Loeb, S. et al. Overdiagnosis and overtreatment of prostate cancer. *Eur. Urol.* **65**, 1046–1055 (2014).
5. Trifiletti, D. M., Sturz, V. N., Showalter, T. N. & Lobo, J. M. Towards decision-making using individualized risk estimates for personalized medicine: A systematic review of genomic classifiers of solid tumors. *PLoS One* **12**, e0176388 (2017).
6. Matulay, J. T. & Wenske, S. Genetic signatures on prostate biopsy: clinical implications. *Translational Cancer Research;7, Supplement 6 (July 30, 2018): Translational Cancer Research (Prostate Cancer: Current Understanding and Future Directions)* (2018).
7. Flavahan, W. A., Gaskell, E. Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, eaal2380 (2017).
8. Parreno, V. et al. Transient loss of Polycomb components induces an epigenetic cancer fate. *Nature* **629**, 688–696 (2024).
9. Krijger, P. H. L. & De Laat, W. Regulation of disease-associated gene expression in the 3D genome. *Nature Reviews Molecular Cell Biol.* **17,** 771–782 (2016).
10. Willemin, A., Szabó, D. & Pombo, A. Epigenetic regulatory layers in the 3D nucleus. *Molecular Cell* **84**, 415–428 (2024).
11. Fischer, A. H. et al. The cytologic criteria of malignancy. *J. Cell Biochem.* **110**, 795–811 (2010).
12. Zhao, S. G. et al. Integrated analyses highlight interactions between the three-dimensional genome and DNA, RNA and epigenomic alterations in metastatic prostate cancer. *Nat. Genet.* **56**, 1689–1700 (2024).
13. Klein, E. A. et al. A 17-gene assay to predict prostate cancer aggressiveness in the context of Gleason grade heterogeneity, tumor multifocality, and biopsy undersampling. *Eur. Urol.* **66**, 550–560 (2014).
14. 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).
15. Cuzick, J. et al. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. *Lancet Oncol.* **12**, 245–255 (2011).
16. Lucini, F. et al. Biochemical properties of chromatin domains define genome compartmentalization. *Nucleic Acids Res.* **52**, e54 (2024).
17. Dhanasekaran, S. M. et al. Delineation of prognostic biomarkers in prostate cancer. *Nature* **412**, 822–826 (2001).
18. Hessels, D. & Schalken, J. A. The use of PCA3 in the diagnosis of prostate cancer. *Nat. Rev. Urol.* **6**, 255–261 (2009).
19. Varambally, S. et al. Golgi protein GOLM1 is a tissue and urine biomarker of prostate cancer. *Neoplasia* **10**, 1285–1294 (2008).
20. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
21. Sarker, D., Reid, A. H. M., Yap, T. A. & de Bono, J. S. Targeting the PI3K/AKT pathway for the treatment of prostate cancer. *Clin. Cancer Res* **15**, 4799–4805 (2009).
22. Scaglia, N., Frontini-López, Y. R. & Zadra, G. Prostate Cancer Progression: as a Matter of Fats. *Front. Oncol.* **11**, 719865 (2021).
23. Ahmad, F., Cherukuri, M. K. & Choyke, P. L. Metabolic reprogramming in prostate cancer. *Br. J. Cancer* **125**, 1185–1196 (2021).
24. Wolfgang, C. D., Essand, M., Lee, B. & Pastan, I. T-Cell Receptor Chain Alternate Reading Frame Protein (TARP) Expression in Prostate Cancer Cells Leads to an Increased Growth Rate and Induction of Caveolins and Amphiregulin. *Cancer Res.* **61**, 8122–8126 (2001).
25. Cocchiola, R. et al. The induction of Maspin expression by a glucosamine-derivative has an antiproliferative activity in prostate cancer cell lines. *Chem. Biol. Interact.* **300**, 63–72 (2019).
26. Sun, Z. & Yan, B. Multiple roles and regulatory mechanisms of the transcription factor GATA6 in human cancers. *Clin.Genetics* **97,** 64–72 (2020).
27. Liu, S., Wang, W., Zhao, Y., Liang, K. & Huang, Y. Identification of Potential Key Genes for Pathogenesis and Prognosis in Prostate Cancer by Integrated Analysis of Gene Expression Profiles and the Cancer Genome Atlas. *Front. Oncol.* **10**, 809 (2020).
28. Kelly, K. A. et al. Detection of early prostate cancer using a hepsin-targeted imaging agent. *Cancer Res* **68**, 2286–2291 (2008).
29. Piao, S. et al. High Expression of PDLIM2 Predicts a Poor Prognosis in Prostate Cancer and Is Correlated with Epithelial-Mesenchymal Transition and Immune Cell Infiltration. *J. Immunol. Res* **2022**, 2922832 (2022).
30. Pulliam, T. L. et al. Regulation and role of CAMKK2 in prostate cancer. *Nat. Rev. Urol.* **19**, 367–380 (2022).
31. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
32. Schubert, M. et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat. Commun.* **9**, 20 (2018).
33. Ghavi-Helm, Y. et al. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet* **51**, 1272–1282 (2019).
34. Sebestyén, E. et al. SAMMY-seq reveals early alteration of heterochromatin and deregulation of bivalent genes in Hutchinson-Gilford Progeria Syndrome. *Nat. Commun.* **11**, 6274 (2020).
35. Hug, C. B., Grimaldi, A. G., Kruse, K. & Vaquerizas, J. M. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* **169**, 216–228 e19 (2017).
36. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**, 1011–1025 (2015).
37. Stelloo, S. et al. Integrative epigenetic taxonomy of primary prostate cancer. *Nat. Commun.* **9**, 4900 (2018).
38. Eksi, S. E. et al. Epigenetic loss of heterogeneity from low to high grade localized prostate tumours. *Nat. Commun.* **12**, 7292 (2021).
39. Song, H. et al. Single-cell analysis of human primary prostate cancer reveals the heterogeneity of tumor-associated epithelial cell states. *Nat. Commun.* **13**, 141 (2022).
40. Johnstone, S. E. et al. Large-Scale Topological Changes Restrain Malignant Progression in Colorectal Cancer. *Cell* **182**, 1474–1489.e23 (2020).
41. Li, R. et al. Extended application of genomic selection to screen multiomics data for prognostic signatures of prostate cancer. *Brief Bioinform* **22**, bbaa197 (2021).
42. Ramos-Montoya, A. et al. HES6 drives a critical AR transcriptional programme to induce castration-resistant prostate cancer through activation of an E2F1-mediated cell cycle network. *EMBO Mol. Med.* **6**, 651–661 (2014).
43. Bracken, C. P. & Goodall, G. J. The many regulators of epithelial-mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* **23**, 89–90 (2022).
44. de Visser, K. E. & Joyce, J. A. The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth. *Cancer Cell* **41**, 374–403 (2023).
45. Risom, T. et al. Transition to invasive breast cancer is associated with progressive changes in the structure and composition of tumor stroma. *Cell* **185**, 299–310.e18 (2022).

46. Pakula, H. et al. Distinct mesenchymal cell states mediate prostate cancer progression. *Nat. Commun.* **15**, 363 (2024).
47. Liu, Y. et al. Stromal AR inhibits prostate tumor progression by restraining secretory luminal epithelial cells. *Cell Rep.* **39**, 110848 (2022).
48. Wang, H. et al. Antiandrogen treatment induces stromal cell reprogramming to promote castration resistance in prostate cancer. *Cancer Cell* **41**, 1345–1362.e9 (2023).
49. Rodriguez-Bravo, V. et al. The role of GATA2 in lethal prostate cancer aggressiveness. *Nat. Rev. Urol.* **14**, 38–48 (2017).
50. Chen, X. et al. Canonical androgen response element motifs are tumor suppressive regulatory elements in the prostate. *Nat. Commun.* **15**, 10675 (2024).
51. Uhlén, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
52. Sjöstedt, E. et al. An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science* **367**, eaay5947 (2020).
53. Karlsson, M. et al. A single-cell type transcriptomics map of human tissues. *Sci. Adv.* **7**, eabh2169 (2021).
54. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
55. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
56. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
57. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
58. Kharchenko, P. V., Tolstorukov, M. Y. & Park, P. J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol.* **26**, 1351–1359 (2008).
59. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).
60. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
61. Lund, E., Oldenburg, A. R. & Collas, P. Enriched domain detector: a program for detection of wide genomic enrichment domains robust against local variations. *Nucleic Acids Res.* **42**, e92 (2014).
62. Zhao, H. et al. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006 (2014).
63. Hahne, F. & Ivanek, R. Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol. Biol.* **1418**, 335–351 (2016).
64. Liu, Y. et al. Systematic inference and comparison of multi-scale chromatin sub-compartments connects spatial organization to cell phenotypes. *Nat. Commun.* **12**, 2439 (2021).
65. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).
66. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
67. Schneider, V. A. et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res.* **27**, 849–864 (2017).
68. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
69. 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).
70. Xie, Z. et al. Gene Set Knowledge Discovery with Enrichr. *Curr. Protoc.* **1**, e90 (2021).
71. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinforma.* **14**, 128 (2013).
72. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
73. Badia-I-Mompel, P. et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinforma. Adv.* **2**, vbac016 (2022).
74. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
75. Taylor, A. M. et al. Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell* **33**, 676–689.e3 (2018).
76. Long, Q. et al. Global transcriptome analysis of formalin-fixed prostate cancer specimens identifies biomarkers of disease recurrence. *Cancer Res.* **74**, 3228–3237 (2014).
77. 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).
78. Jain, S. et al. Validation of a Metastatic Assay using biopsies to improve risk stratification in patients with prostate cancer treated with radical radiation therapy. *Ann. Oncol.* **29**, 215–222 (2018).
79. Righi, I. et al. Immune Checkpoints Expression in Chronic Lung Allograft Rejection. *Front. Immunol.* **12**, 714132 (2021).

Acknowledgements

We thank Beatrice Bodega, Marina Lusic, Mattia Forcato, Gioacchino Natoli, Martin Schaefer and all members of our laboratories for critical feedback on earlier versions of the manuscript. We thank Claudio Tripodo (IFOM and University of Palermo) and Giorgia Zadra (CNR) for feedback on early phases of the project. For the sequencing we acknowledge support from the Genomics Unit of the European Institute of Oncology (IEO), Department of Experimental Oncology, Humanitas Sequencing facility and Istituto Nazionale Genetica Molecolare (INGM) (Marco Ghilotti). We thank Ilaria Rancati (IFOM cell culture facility) for support with Maxwell instrument. This work was supported by Interomics Flagship Project (CNR) and MFAG (grant #18535) to CL; AIRC Start-Up (grant #16841) to FF; FRRB INTERSTRAT-CAD (grant #CP2_14/2018) to FF; AIRC post-doctoral fellowships No. 26942 to FG and No. 22351 to ES; AIRC pre-doc fellowship No. 31282-2024 to GL; PIR01_00011 "IBISCo", PON 2014-2020 to MM. Schematic representations of Fig. 1a, b and Supplementary Fig. S3a, 9d have been created using BioRender.com.

Author contributions

VR designed and performed experiments as well as data analysis. G.L. and C.P. designed and performed data analysis. F.G. and RQ performed immunofluorescence and A.F. performed the IF quantification. V.R., M.V. and C.C. performed the wound healing assay. M.M., E.S. and E.D.P.S. participated in data analysis. MC acquired and analyzed FACS data. VV supervised IHC and sequencing experiments. E.M., G.A., F.R. and E.D.L. provided patient samples and participated to the data interpretation. EC and MMa performed IHC and histopathology evaluation. V.R., G.L. and C.P. wrote the manuscript. FF and CL conceived the study and wrote the manuscript. All authors approved the submitted version.

Competing interests

A patent application is being filed for the signature presented in this manuscript by V.R., G.L., C.P., F.F. and C.L.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-64213-4.

Correspondence and requests for materials should be addressed to Valentina Rosti, Francesco Ferrari or Chiara Lanzuolo.

Peer review information *Nature Communications* thanks anonymous reviewer(s) for their contribution to the peer review of this work. [A peer review file is available].

Reprints and permissions information is available at http://www.nature.com/reprints

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025