nature communications

Article

https://doi.org/10.1038/s41467-025-67856-5

Prostate cancer cells converge to an inflammatory-like state upon metastatic dissemination

Received: 25 April 2025

Accepted: 10 December 2025

Published online: 23 December 2025

Check for updates

Tina Keshavarzian[1,2,10], Kira Furlano[1,3,4,10], Giacomo Grillo[1], Lisanne Mout[1], Christopher Arlidge[1], Faizan Hasan[1,2], Ankita Nand[1], Migle Mikutenaite[5], Evdoxia Karadoulama[5], Ashish Goyal[3,6], Elisabeth L. Pezzuto[3], Jessica Heilmann[3], Jakub Sykora[3], Sarah Minner[7], Thorsten Schlomm[4], Guido Sauter[7], Ronald Simon[7], Housheng Hansen He[1,2], Joachim Weischenfeldt[4,6,8], Christoph Plass[3], Clarissa Gerhäuser[3] ✉ & Mathieu Lupien[1,2,9] ✉

Identifying drivers of cancer progression to guide treatment selection is hindered by our limited understanding of tumor heterogeneity and its impact on tumor evolution. Here, we delineate the phenotypic variability across ~300,000 cells collected from multiple tumor loci in primary prostate and matched locoregional metastases using single-cell chromatin accessibility and gene expression sequencing. We find inter-patient heterogeneity to be confined to malignant populations. Within individual tumor loci, we see phenotypic heterogeneity among malignant cell populations despite a shared clonal genotypic architecture. We also observe that malignant cell populations disseminating to locoregional lymph nodes mirror the clonal architecture and phenotypic heterogeneity across primary tumor loci, while shifting from canonical prostate-cancer states to non-canonical inflammatory-like states. Our findings suggest a bottleneck imposed during the dissemination process, funneling prostate cancer cells toward an inflammatory-like cell state. These insights into the interplay between phenotypic identity and clonal architecture refine our understanding of prostate cancer progression and suggest that convergence of cancer cells towards an inflammatory-like state underlies dissemination to lymph nodes, offering a critical framework for future studies into prostate cancer metastatic potential.

Understanding how tumor heterogeneity impacts cancer evolution has been a significant obstacle in our efforts to identify the drivers of cancer development and progression. This is particularly true for high-risk prostate cancer, with a 48–80% likelihood of progression within 5 years compared to overall survival rates of 90–99% in lower-risk groups[1,2]. This disparity underscores a need to uncover the dimensions of tumor heterogeneity that fuel tumor progression, inclusive of inter-patient, intra-tumor, intra-tissue and inter-tissue.

Most studies to date have relied on investigating single primary or metastatic biopsies using bulk sequencing, guiding our understanding on the existence of intra- and inter-tissue heterogeneity. For example, bulk genetic sequencing has showcased distinct clonal populations residing within different regions of the same tissue to harbor unique

A full list of affiliations appears at the end of the paper. ✉e-mail: c.gerhauser@dkfz-heidelberg.de; mathieu.lupien@uhn.ca

genetic aberrations, which could be indicative of outcome and potential treatment sensitivities[3–9]. Additionally, nongenetic sources of variability, such as epigenetic reprogramming and transcriptional variability, play critical roles in reconfiguring cellular states to adapt to adverse conditions of metastasis and treatment resistance[10–17]. Phenotypic changes are observed throughout the onset and progression of prostate cancer to mCRPC, where nongenetic variability is showcased with prostate-specific transcription factors (TFs) recruitment to chromatin variants to drive de-differentiation and promote progression[9,18–26]. Collectively, these findings underscore the necessity to characterize phenotypic diversity within and across tumors to fully understand the functional impact of heterogeneity towards cancer progression and treatment response.

In this work, we tackle the complexity of the multiple dimensions of tumor heterogeneity in prostate cancer progression by integrating multi-regional sampling of primary and metastatic sites with high-resolution single-nucleus chromatin accessibility and transcriptomic profiling to characterize the multi-faceted nature of heterogeneity and its role in progression. We find that inter-patient variability is restricted to malignant populations, that malignant cells within individual loci exhibit phenotypic divergence despite shared clonal origins, and that metastatic dissemination involves an inflammatory-like malignant state, revealing a bottleneck that funnels prostate cancer cells toward this noncanonical phenotype.

Results

Inter-patient heterogeneity is restricted to malignant epithelial as opposed to normal tumor microenvironment populations

We characterized 38 prostate cancer specimens derived from five high-risk human patients with ISUP ≥ 3 pathology who underwent radical prostatectomy and pelvic lymph node dissection at surgery as part of the International Cancer Genome Consortium (ICGC) late-onset cohort[27] (Supplementary Dataset 1 and Fig. 1a). All patients were treatment-naïve before prostatectomy with pelvic lymphadenectomy, except one (ICGC_PCAL25), who received hormonal therapy 6 weeks before surgery. Using single-nuclei multiome sequencing (scMultiome- gene expression and chromatin accessibility from RNA- and ATAC-seq, respectively), we characterized 282,956 nuclei, with a median of 6 primary tumor loci from the prostate gland and 3 locoregional lymph node metastases across all patients. Here, we use the term "loci" to refer to tumors sampled from spatially distinct sites of the prostate or across lymph nodes.

Leveraging both RNA and ATAC-seq data from all samples with a weighted nearest neighbor model[28], 41, populations were identified based on the phenotypic characteristics across all cells (Fig. 1b and Supplementary Fig. 1a). Cell type annotations identified 30 malignant epithelial populations that accounted for 86.4% of the cells. The remaining populations included major tumor microenvironment (TME) cell types, such as fibroblasts, macrophages, B and T cells (Fig. 1c). The cell types we identified displayed expected gene expression markers, such as *AR, KRT18, and KRT8* in malignant epithelial cells and *PTPRC* in lymphocytes (Supplementary Fig. 1b). Additionally, we computed transcription factor DNA motif enrichments using ChromVar, which is the deviation score relative to an average chromatin accessibility signal computed across DNA motifs[29]. We observed significant TME-specific transcription factor DNA recognition motifs exemplified by the enrichment of the ETS1 motif in lymphocytes, the CEBPB motif in macrophages and the depletion of these motifs from accessible chromatin regions in malignant epithelial cells (Fig. 1d). Similarly, DNA recognition motifs of prostate-specific TFs, namely FOXA1 and HOXB13, were depleted in TME populations (Fig. 1d).

To quantify "inter-mixing" of populations, we utilized the Shannon entropy metric (H). We refer to population inter-mixing where populations include cells that come from multiple patients (Hp) or tissue site (Ht). In all cases, an entropy score of 0 indicates no inter-mixing with respect to the variable of interest. Conversely, maximum Hp or Ht reflects a cell population where cells from all patients or tissue sites contribute to the population's composition with equal probability. We found that individual malignant epithelial populations were mainly composed of cells originating from a single patient, with a median Hp = 0 across malignant epithelial populations, indicating no patient inter-mixing (Fig. 1e). Fibroblasts and macrophages were the most inter-mixed TME populations (Hp = 2.15 and 1.5), with fibroblasts being closest to the maximum entropy value possible (Hp max = 2.32). Similarly, we identified patient inter-mixing in the basal epithelial population (Hp = 1.2) with no inter-mixing according to tissue site (primary tumor vs lymph node; Ht = 0) (Fig. 1e and Supplementary Fig. 1b–d). This agrees with prostate basal cells being a unique cell type to the prostate gland, and showcases the validity of biological signals revealed by single-cell multiome. Lastly, B cell populations were the only normal cell type showing limited patient inter-mixing, with median Hp = 0.24 across B cell populations (Supplementary Fig. 1d), representative of patient-level phenotypic diversity previously reported in B cells[30], while having patient-specific driver alterations (Supplementary Fig. 1e). Overall, these observations suggest biological constraints on most TME cell populations towards a shared identity across different patients. In contrast, malignant populations are patient-specific and suggest patient-dependent tumor evolutionary paths, warranting analyses on a patient-by-patient basis.

Malignant cell populations exhibit intra-locus heterogeneity, yet are phenotypically similar within a prostate gland

Single primary tumor biopsies do not capture the diverse histology and genetics reported within the prostate gland and are unlikely to reflect the phenotypic diversity of prostate cancer cells in a patient[31]. Thus, to better understand phenotypic heterogeneity within a prostate but across tumor loci from spatially distinct locations, we analyzed the prostate gland on a patient-level basis (3–8 tumor sections per patient) (Fig. 2a and Supplementary Fig. 2). A median of 88% of cells were assigned to malignant epithelial populations across patients (Fig. 2a). Malignant epithelial populations exhibited almost no inter-mixing of cells from different loci according to sample entropy measurements in contrast to TMEs (median normalized Hl = 0.06 versus Hl = 0.8 across patients, Mann–Whitney U test $P = 3.05 \times 10^{-6}$) (Fig. 2b). A median of 3 malignant epithelial populations contained more than 90% of cells from the same tumor locus (Supplementary Fig. 3). Inter-mixing across loci in malignant epithelial populations was skewed to those engaged in cell division (Fig. 2c) (median normalized Hl = 0.80) suggesting that cell cycle signal can in part override locus-specificity of most malignant epithelial populations and drive inter-mixing of cells from different loci. Overall, the low degree of inter-mixing across loci replicated in each patient suggests a detectable degree of phenotypic heterogeneity in cancer cells across loci within the prostate gland.

To further assess the heterogeneity across malignant epithelial populations, we looked at correlations across populations based on their ATAC or RNA read-outs. Pan-cancer analysis of ATAC- and RNA-seq data reported chromatin accessibility at non-promoter elements to reflect cancer types with increased specificity compared to accessibility at gene promoters or gene expression levels[32]. Hence, we generated a patient-specific repertoire of chromatin-accessible DNA elements based on the normalized ATAC signal at promoters and non-promoter elements in each population. Pearson correlations had a wider range of values across non-promoter elements (median: range = 0.96, *N* elements = 120,942) compared to promoter elements (median: range = 0.44, *N* elements = 38,545) and transcripts (median range = 0.13; *N* transcripts = 36,601) (Supplementary Fig. 3b–d). These results highlight the specificity of chromatin accessibility at non-promoter elements as a key source of variation across populations, accounting for intratumoral heterogeneity.

https://doi.org/10.1038/s41467-025-67856-5

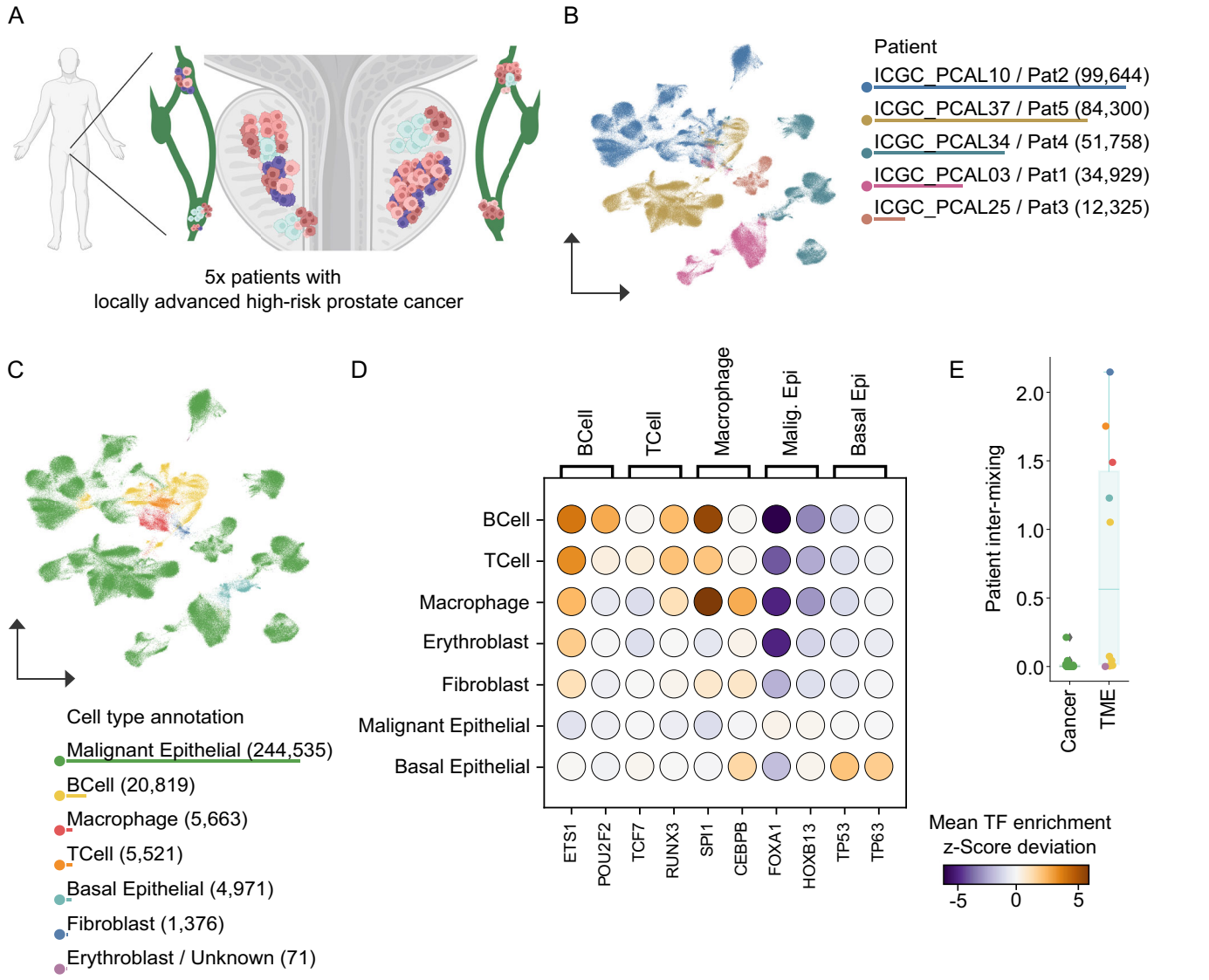
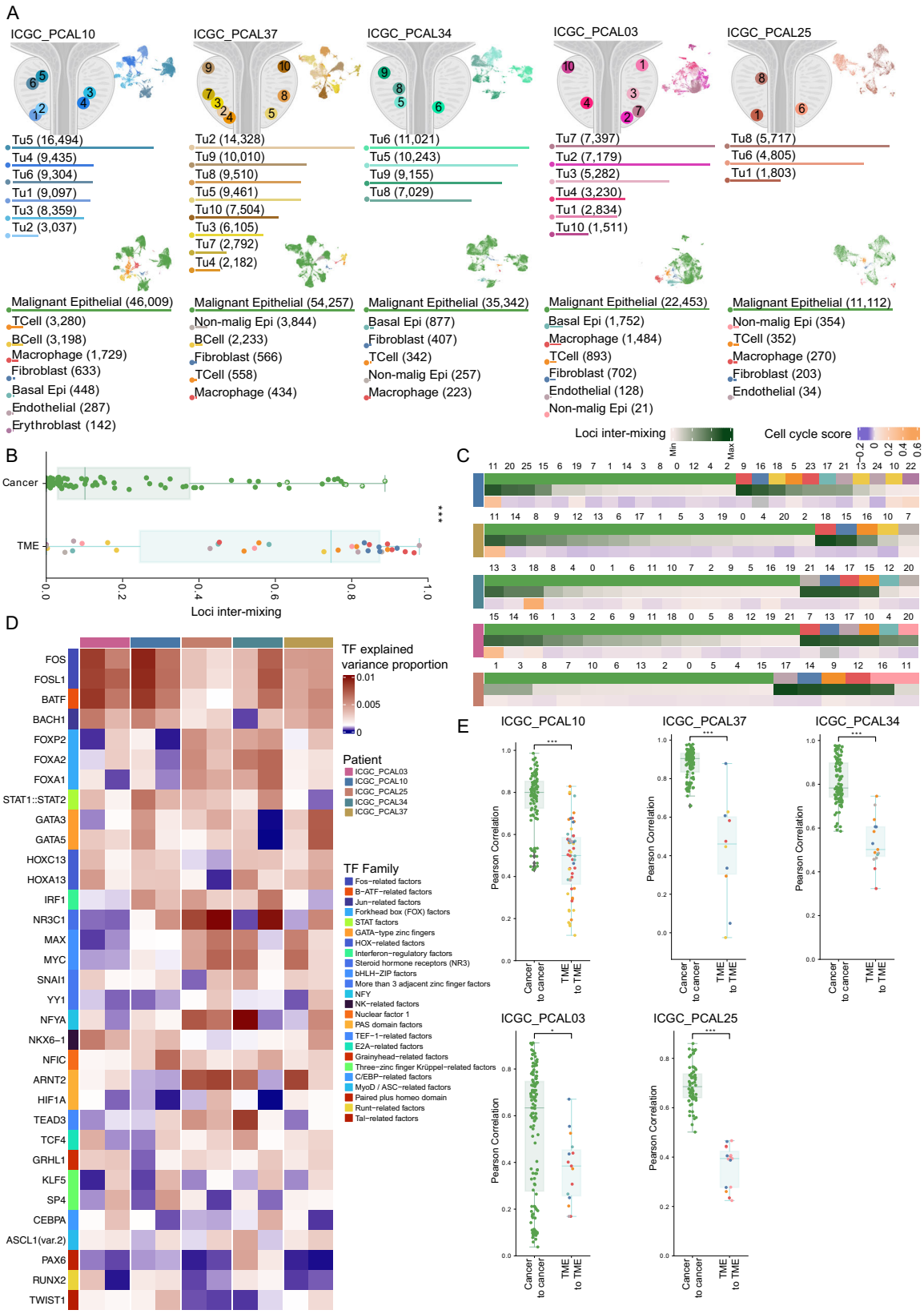


Fig. 1 | The molecular taxonomy of primary prostate cancer. A Schematic of cohort setup created in BioRender. Keshavarzian, T. (2025) https://BioRender.com/dqdlkpm. **B** Weighted-nearest neighbor (WNN) 2D UMAP representations colored by patient, **C** cell type annotation. The number of cells is shown in the figure panels. **D** Selected transcription factors' motif deviations according to Chromvar. Values represent the average z-score of motif accessibility deviations across all cells of a given cell type. Higher z-scores indicate greater enrichment of TF motifs in accessible chromatin, compared to background accessibility. **E** Boxplots of patient inter-mixing in malignant epithelial cell populations (cancer) or normal cell populations found in the tumor microenvironment (TME) according to Shannon entropy. The boxplot depicts the median (central line), interquartile range (box boundaries), and 1.5*IQR whisker ranges, with outliers shown as individual points. $N = 31$ and $N = 10$ for cancer and TME populations, respectively.

Accessible non-promoter elements can serve to regulate gene expression through the recruitment of TFs recognizing the DNA recognition motif they harbor. We reasoned that the variance in TF-DNA recognition motif enrichments across the patient-specific repertoire of accessible DNA elements can reveal sources of heterogeneity in malignant populations. Principal component analysis in each patient revealed motifs for the FOS and JUN-related TFs, recently linked to stem cell-like properties in castration resistant prostate cancer[26], consistently emerging as top contributors to the major axis of variation across malignant epithelial populations in all patients (Fig. 2d and Supplementary Fig. 4). Nuclear receptor, Forkhead, and GATA transcription factor families were also found as strong sources of variation, in line with their role as drivers of prostate cancer progression[33]. Variance across malignant epithelial populations was associated with neuroendocrine-like TFs, such as ASCL1, SNAI1, and TCF4, supporting a neuroendocrine-based source of heterogeneity within one patient (ICGC_PCAL34)[26,34,35]. Additionally, we observed PAX-family TF motifs, such as PAX6, which is implied in neuro-ectoderm development but not well characterized in prostate cancer[36]. To further reproduce and contextualize this, we linked TFs, accessibility, and gene expression using single-cell enhancer-gene regulatory networks analysis[37] on tumors from patients ICGC_PCAL34 and ICGC_PCAL03. The highly specific TFs to individual malignant epithelial clusters included FOX, HOX, FOS/JUN, GRHL, and NR-based TFs, aligned with our orthogonal analysis (Fig. 2d and Supplementary Dataset 2). Performing pathway enrichment analysis on the lists of genes specific to a given cluster, regulated by the reported TFs, revealed distinct and overlapping pathways (FDR < 0.05). The commonly enriched pathways included androgen and estrogen response, whereas more specific pathways involved EMT, hypoxia, and TGF-β signaling in ICGC_PCAL34, along with p53, TGF-β, PI3K/AKT/mTOR Signaling, mTORC1 and DNA repair in ICGC_PCAL03 (Supplementary Fig. 5 and Supplementary Dataset 2: Hallmark Pathways). Overall, these results suggest that prostate lineage-specific TFs, along with JUN/FOS, differentially engage with

https://doi.org/10.1038/s41467-025-67856-5



accessible chromatin, driving heterogeneity across malignant epithelial populations.

To contextualize malignant epithelial population heterogeneity, we computed Pearson correlations across malignant epithelial and TME populations. TME-to-TME comparisons offer a reference point for expected variability between well-defined cell types. Across patients, malignant epithelial populations were more correlated with each other than TME populations were among themselves across patients (Fig. 2e) (Mann–Whitney U test *P:* ICGC_PCAL10 = 8.77×10^{-17}, ICGC_PCAL03 = 0.02, ICGC_PCAL25 = 1.83×10^{-09}, ICGC_PCAL34 = 3.76×10^{-09}, ICGC_PCAL37 = 1.21×10^{-06}). Thus, while malignant populations exhibit distinct regulatory programs driving their heterogeneity between and

Fig. 2 | Patient-specific intra-prostate gland tumor heterogeneity analyses reveals loci-biased malignant epithelial population. A Schematic representation of the inferred spatial distribution of biopsied tumor loci within the prostate gland of each patient (left), WNN UMAP representation and total number of nuclei remaining after QC according to tumor loci (upper right) and cell type annotations (lower right). The number of cells is shown in the figure panels. The prostate diagrams were created in BioRender. Keshavarzian, T. (2025) https://BioRender.com/g6d21np. **B** Boxplots of normalized loci- inter-mixing scores in malignant epithelial cell populations (cancer) or normal cell populations found in the tumor microenvironment (TME) combined from intra-patient analyses across all patients. Starred (white star) in malignant epithelial populations represent proliferating populations. Mann–Whitney U test two-sided $P = 3.05 \times 10^{-6}$. Total number of populations from intra-patient analyses of all patients is 109. Total number of cancer populations = 75 and TME populations = 34. The total number of clusters per patient is represented in (**C**). **C** Populations called from tumor loci in the prostate gland, grouped according to cancer or TME cell types. **D** Principal component analysis results on select transcription factor motif enrichments showcasing the top two axes of variation per patient ($N = 5$ patients). Color gradients indicate variance explained per transcription factor in the principal component. **E** Pearson correlation of all populations across malignant epithelial cells (cancer to cancer), or normal cell populations (TME to TME). Populations are according to (**C**). Mann–Whitney U test two-sided *P:* ICGC_PCAL10 = 8.77×10^{-17}, ICGC_PCAL03 = 0.02, ICGC_PCAL25 = 1.83×10^{-09}, ICGC_PCAL34 = 3.76×10^{-09}, ICGC_PCAL37 = 1.21×10^{-06}. The boxplots depict the median (central line), interquartile range (box boundaries), and 1.5*IQR whisker ranges, with outliers shown as individual points. *P value < 0.05, **P value < 0.01,***P value < 0.001. Total number of correlation values per population grouping (excluding self-correlation) in malignant epithelial and TME populations, respectively: ICGC_PCAL10 = 105, 55; ICGC_PCAL37 = 120, 10; ICGC_PCAL34 = 120, 15; ICGC_PCAL03 = 120, 15; ICGC_PCAL03 = 66, 15.

within tumor loci, their core molecular identity is more preserved than what is observed across populations from the TME.

Locus-specific malignant epithelial populations are composed of similar clonal architectures

High-risk prostate cancer is known for its genotypic diversity with a high frequency of structural variants and copy number variations (CNVs)[38]. For this reason, we inferred CNV profiles from the scMultiome data using NUMBAT[39] to link malignant epithelial populations to genetic clones within a patient. We detected typical prostate cancer CNV events such as chromosome 6 and 8p loss, and 8q gains (Supplementary Dataset 3), with notable variation across patients in line with prior reports[40,41]. The genetic architecture of each patient included a median of 6 clones per patient (Fig. 3a) with varying degrees of similarity between malignant clones. We measured similarity between clones using the Hamming metric, where 0 represents the same CNV label across all genomic segments between groups, and 1 indicates no shared CNV states over segments called per patient.

Patients ICGC_PCAL03 and ICGC_PCAL37 had the smallest difference between clones (mean distance between clones: ICGC_PCAL03 = 0.13, ICGC_PCAL37 = 0.06), whereas ICGC_PCAL10, ICGC_PCAL34, and ICGC_PCAL25 were more diverse (mean distance between clones: ICGC_PCAL10 = 0.39, ICGC_PCAL34 = 0.22, ICGC_PCAL25 = 0.3) (Supplementary Fig. 6a). In the latter patients, the most distinct clones contained cells from a single tumor locus, suggesting genotypic diversity to be locus- restricted. For instance, ICGC_PCAL10's clone 3 was the most distinct, with a mean distance of 0.9 from other clones, where 98.3% of clone 3 cells were from the Tu5 locus (Fig. 3b, c). This clone was the exception in ICGC_PCAL10, as the rest of the clones in ICGC_PCAL10 were similar to each other (mean distance = 0.08). Similarly, in ICGC_PCAL34, clones 4 and 6 accounted for 90% of all Tu6 locus malignant epithelial cells used in CNV inference (Fig. 3b, c). In ICGC_PCAL25, clone 4 and its subclones (5–8) were largely composed of Tu6 locus malignant epithelial cells, averaging 87% (Fig. 3b, c). Thus, the most distinct clones relative to the most common patient CNVs were restricted to a single sampled tumor section and not found in multiple loci.

Next, we wanted to understand the mapping between phenotype-based populations and genotype-based clones. We used the Jensen–Shannon divergence score to compare how malignant populations relate to each other with respect to the overlap in the probability distribution of their genotype-based clonal membership. Specifically, a score of 1 indicates that two malignant populations are composed of the same underlying sets of clones with identical proportions. Conversely, a JSD score of 0 represents no overlap in the probability distributions of clones between a pair of malignant populations. We found that despite the intra-tissue and intra-locus specificity of malignant epithelial populations based on chromatin and expression data, these populations had similar clonal architectures (median clonal distribution distance between populations across patients = 0.07) (Fig. 3 and Supplementary Fig. 6b). The similar clonal composition of phenotypically heterogeneous malignant epithelial populations from multiple tumor loci further supported a population diversity relying on nongenetic variation embedded with chromatin accessibility and transcriptional programs.

Malignant cells disseminating to locoregional lymph nodes undergo an inflammatory-like cell state transition despite genotypic similarity to primary cancer

Locoregional lymph nodes are the earliest and most common site of malignant epithelial cell disseminations in prostate cancer and are linked to worse prostate cancer outcomes[2,42]. Therefore, their involvement offers crucial insights into tumor heterogeneity and its role in the evolution towards a metastatic state. For this reason, we utilized multiple locoregional lymph nodes collected in four of our prostate cancer patients (median 3 metastatic lymph nodes per patient) to characterize their cell state composition from scMultiome assays (Fig. 4a and Supplementary Fig. 7). Except for patient ICGC_PCAL34, lymph node samples had a similar clonal architecture to each other and to their matched primary tumor loci (Fig. 4b and Supplementary Fig. 7). Additionally, locus-restricted intraglandular clones, such as clone 3 in ICGC_PCAL10, were not detected in multiple matched lymph nodes (Fig. 4b and Supplementary Fig. 7). From a phenotypic angle, malignant epithelial populations continued to show low inter-mixing of tissue sites and samples (Fig. 4c), with median Ht = 0 across all patients' malignant epithelial populations. ICGC_PCAL37, which had the lowest Gleason score (7) was an exception, with malignant populations being most inter-mixed across tissues in our cohort (median Ht = 0.4). To quantify the degree of phenotypic differences, we compared Pearson correlations of non-promoter DNA accessibility profiles between malignant epithelial populations containing >90% of cells from a single locus in either the lymph node or the prostate. In most patients, correlation ranges for inter-tissue comparisons (lymph node-specific vs prostate-specific) were not significantly different from intra-tissue comparisons (prostate-only), indicating that dissemination did not introduce a greater range of variability between populations than already present within the primary tumors (Mann–Whitney U test, P: ICGC_PCAL34 = 0.47, ICGC_PCAL10 = 0.28, ICGC_PCAL03 = 0.06, ICGC_PCAL37 = 6.78×10^{-13}) (Fig. 4d). Notable exceptions included ICGC_PCAL34, where lymph node-specific populations were more similar to each other than to prostate-specific populations ($P = 0.046$), and ICGC_PCAL37, which had only a single lymph node-specific population, precluding intra-tissue correlation estimates. Clinically, ICGC_PCAL37 had no evidence of bone metastases and remained stable for nearly 5 years before initiating systemic treatment (Supplementary Dataset 1), consistent with reduced molecular heterogeneity in lower Gleason score disease. Together, these data indicate that the

https://doi.org/10.1038/s41467-025-67856-5

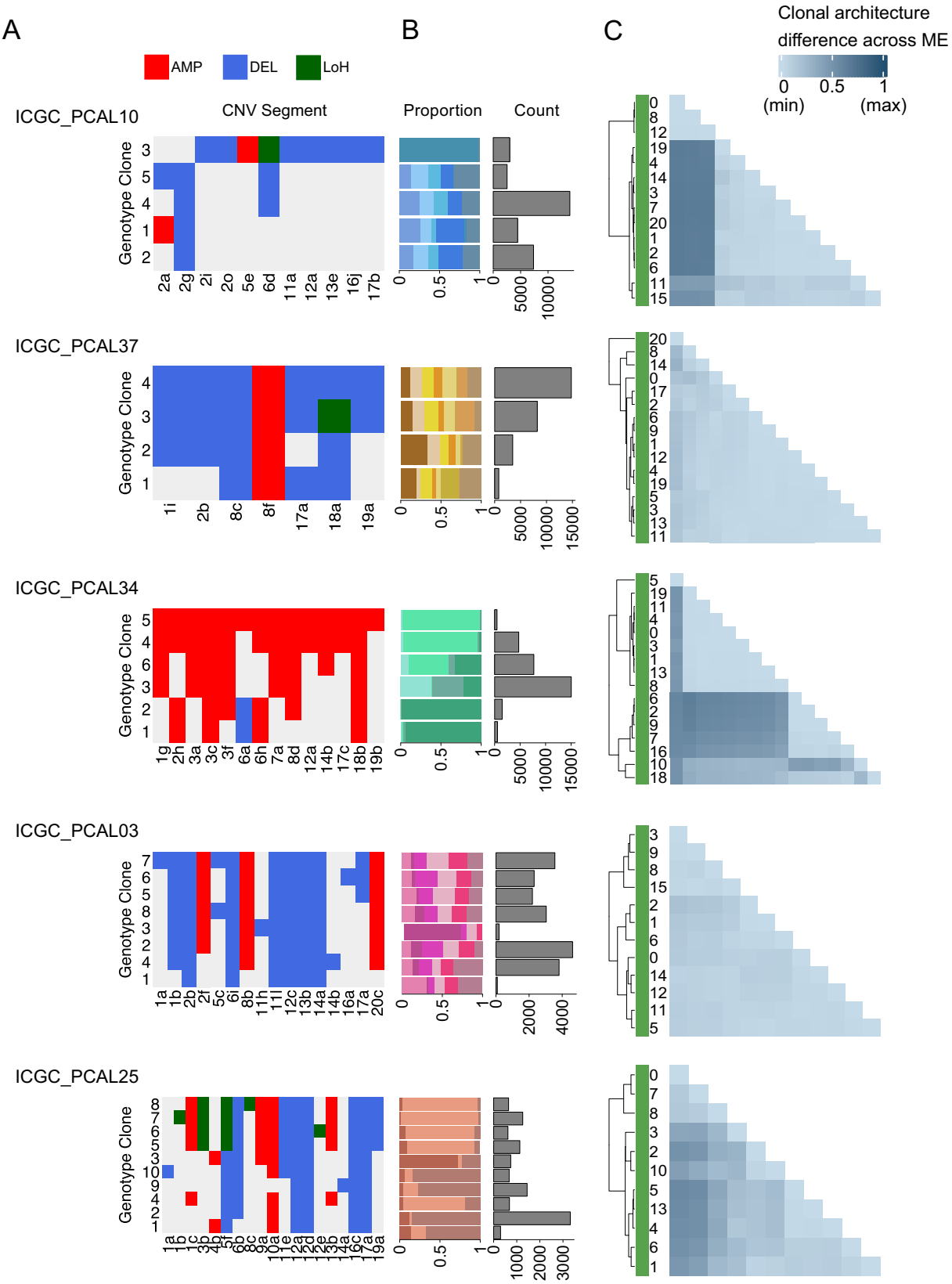
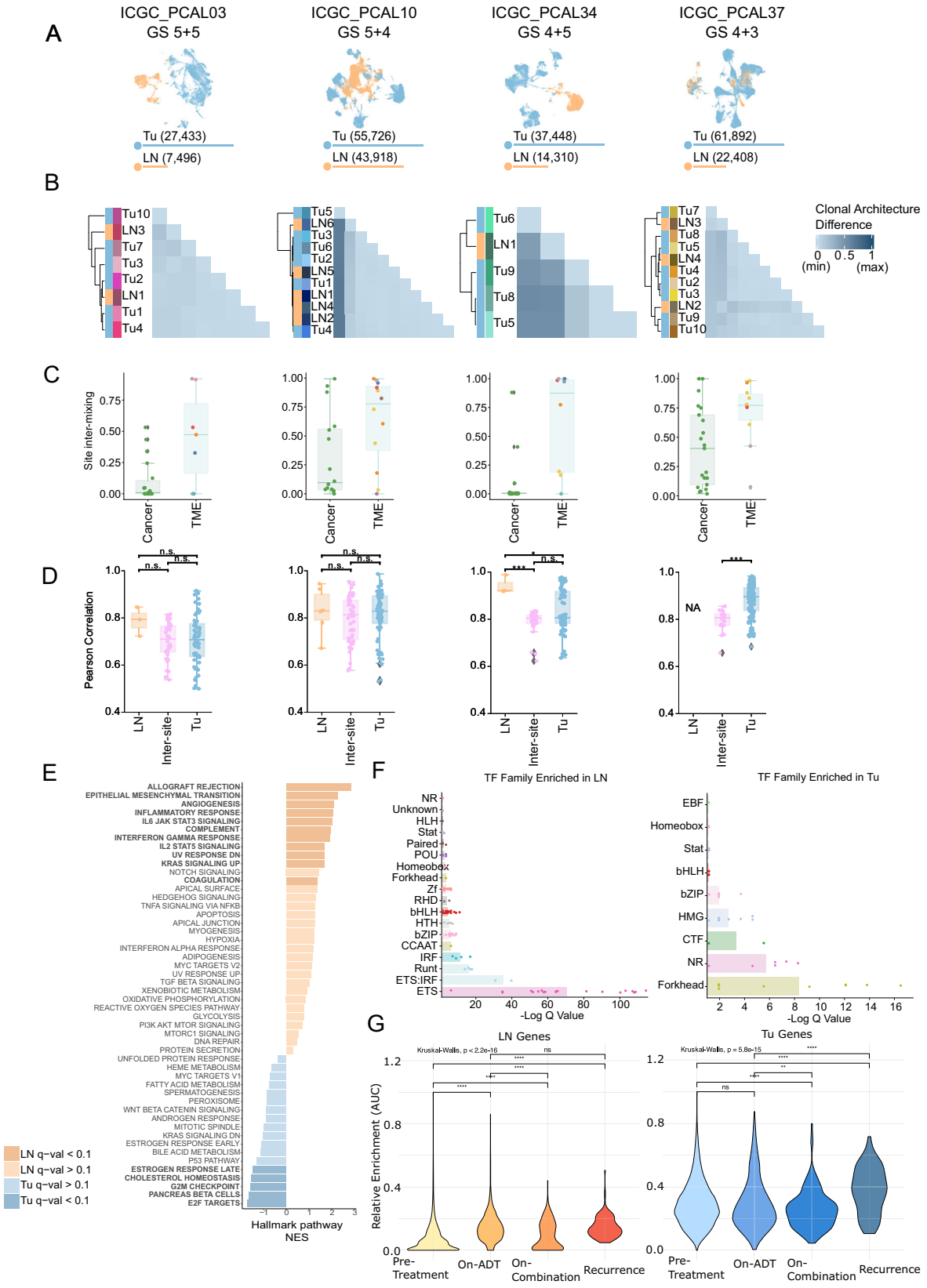


Fig. 3 | Malignant epithelial populations show similar copy number-based genotypic clonal architectures to each other. A CNV segments, corresponding to DNA sequences of variable length across chromosomes; amplified (AMP), deleted (DEL) or associated with loss of heterozygosity (LoH), defining distinct clones. **B** Relative proportion of malignant epithelial cells assigned to each genotype clone based on their individual biopsied tumor loci origin (left panel) and total cell number (right panel). **C** Clonal distribution difference according to Jensen–Shannon divergence between malignant epithelial (ME) populations with >200 cells reported in Fig. 2. Minimum (0) indicates identical genotypic clonal architecture across cells within a malignant epithelial population and between populations. All cell counts are reported in figure panels.

https://doi.org/10.1038/s41467-025-67856-5



stability of patient-level clonal architecture, paired with comparable intra- and inter-tissue phenotypic correlations, points to a role for phenotypic variability that may account for the tissue-specificity of populations, changes in cell state not explained by genetics alone and are likely induced by the microenvironment, as malignant prostate cancer cells adapt to colonize lymph nodes.

Given the genotypic similarity but phenotypic-specificity of lymph node malignant epithelial populations, we wanted to understand if malignant cells in the lymph nodes are undergoing a coordinated phenotypic shift across patients. Thus, we extended the comparison of malignant epithelial cells at the tissue level (lymph node versus primary tumor) across all patients to determine the biological basis of

Fig. 4 | Inter-tissue heterogeneity assessment identifies phenotypic variability despite a similar genotypic clonal architecture across malignant epithelial populations. A WNN UMAP of all cells collected in the primary tumor (Tu: blue) and locoregional lymph nodes (LN: orange) of each patient (cell numbers across tissues are specified under the UMAP). The number of cells are reported in the figure panels. **B** Genotype clonal distribution difference according to Jensen–Shannon divergence between malignant epithelial (ME) cells in every sample. Minimum (0) indicates identical genotypic clonal architecture across malignant epithelial cells from each sample. **C** Inter-mixing according to tissue of origin (site) between malignant epithelial populations (cancer) or normal cell populations (TME) using Shannon entropy, boxes represent the median (central line), interquartile range (box boundaries), and 1.5*IQR whisker ranges, with outliers shown as individual points. Number of populations per patient are as follows: ICGC_PCAL03: cancer = 17, TME = 8; ICGC_PCAL10: cancer = 16, TME = 12; ICGC_PCAL34: cancer = 18, TME = 8; ICGC_PCAL37: cancer = 21, TME = 10. **D** Pearson correlation on normalized pseudo-bulked non-promoter chromatin accessibility signal across malignant epithelial populations found preferentially in Tu or LN (>90% of cells from specific tissue, >200 total *N* cells) compared within tissue or between tissue sites (inter-site). Mann–Whitney U test two-sided, *P*: ICGC_PCAL34 = 0.47, ICGC_PCAL10 = 0.28, ICGC_PCAL03 = 0.06, ICGC_PCAL37 = 6.78×10^{-13}. Boxplots showcase the median (central line), interquartile range (box boundaries), and 1.5*IQR whisker ranges, with outliers shown as individual points. n.s. not significant, $P > 0.05$. Number of population correlations according to LN, inter-site or Tu: ICGC_PCAL03 = 3, 36, 6; ICGC_PCAL10 = 6, 48, 55; ICGC_PCAL34 = 3, 39, 78; ICGC_PCAL37 = 1, 18, 153. **E** Hallmark pathway enrichment analysis based on gene rankings in Tu (blue) versus LN (orange) malignant epithelial cells. Total number of samples to obtain differential genes for enrichment analysis are $N = 27$ primary tumor samples and $N = 11$ across LN samples. **F** DNA recognition motif enrichment analysis on DNA sequences differentially accessible (absolute log fold change > 1 and FDR < 0.1) in Tu ($N = 185$ regions) versus LN malignant ($N = 779$ regions) epithelial cells. **G** Relative AUCell enrichment scores of our gene signatures across treatment stages in external single-cell RNA-seq data (3976 cells total: pre-treatment 2327; on-ADT 1072; on-combination 444; recurrence 133). Overall differences were assessed using Kruskal–Wallis tests (LN: $P < 2.2 \times 10^{-16}$; Tu: $P = 5.781 \times 10^{-15}$). Pairwise comparisons (Bonferroni-corrected Wilcoxon rank-sum test) are as follows: LN (normalized): pre-treatment vs on-ADT < 2×10^{-16}; pre-treatment vs on-combination 1.0×10^{-9}; pre-treatment vs recurrence < 2×10^{-16}; on-ADT vs on-combination < 2×10^{-16}; on-ADT vs recurrence 1; on-combination vs recurrence 2.9×10^{-10}. Tu (normalized): pre-treatment vs on-ADT 0.656; pre-treatment vs on-combination 4.5×10^{-5}; pre-treatment vs recurrence 1.2×10^{-10}; on-ADT vs on-combination 0.046; on-ADT vs recurrence 6.8×10^{-10}; on-combination vs recurrence 2.8×10^{-16}. N.S. not significant; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

disease progression. Differential gene expression analysis on malignant cells identified statistically significant upregulated pathways in lymph node tumors to be associated with the immune system, such as allograft rejection (normalized enrichment score (NES) = 2.86), inflammatory response (NES = 2.05), IL6 JAK-STAT3 signaling (NES = 2.02), as well as the epithelial-mesenchymal transition (NES = 2.27) (Supplementary Dataset 4). In contrast, pathways upregulated in primary tumor malignant epithelial cells associated with the cell cycle (E2F targets (NES = −1.71), G2M checkpoints (NES = −1.56)) and hormone-driven cancers (cholesterol synthesis (NES = −1.54) and estrogen response (NES = −1.46)) (Fig. 4e). To ensure the differential signal observed across tissues is unlikely to be due to ambient RNA, we applied ambient RNA correction on our samples[43]. Ambient RNA correction was marginally helpful in increasing TME-related signal strength, which belonged to cells that were excluded in the differential expression analyses (Supplementary Fig. 8a, b). Repeating differential gene-expression analysis on all patients using ambient RNA corrected results revealed the same trends, and the majority (8/11) of the GO terms were in agreement (Supplementary Fig. 8c).

Analysis of DNA recognition motif enrichment on differentially accessible chromatin regions detected the motifs for the ETS, IRF and RUNX TFs as the top statistically significant motifs in lymph node malignant epithelial cells (Supplementary Dataset 4). These immune-related TFs[44,45] were implicated in immune evasion of melanoma cancer cells in the lymph nodes[46]. In contrast, nuclear receptors and Forkhead DNA recognition motifs, such as those recognized by the androgen receptor (AR) and FOXA1, respectively, were enriched in differentially accessible regions of primary tumor-specific malignant epithelial cells (Supplementary Dataset 4 and Fig. 4f). We further validated the tissue-specificity of phenotypic changes in malignant epithelial populations on an intra-patient level. We computed the expression signature of the statistically significant genes per tissue at the single cell level and ranked all populations based on their median enrichment[37]. When using lymph node-upregulated genes, lymph node-specific malignant epithelial populations ranked higher than primary tumor-specific malignant epithelial populations across all patients, except for our lowest Gleason Score patient (ICGC_PCAL37) (Supplementary Fig. 9a). We observed the opposite trend for differentially upregulated genes at the primary tumor site (Supplementary Fig. 9b). In addition, fibroblasts, then immune populations ranked the highest in lymph node differential gene enrichment and surpassed all malignant epithelial populations (Supplementary Fig. 9a). These findings demonstrated inter-patient commonalities in metastatic versus primary malignant epithelial cell states, despite significant inter-patient genotypic variabilities. Taken together, malignant epithelial cells found within locoregional lymph nodes have a similar clonal architecture to the primary gland, suggesting limited constraint in the dissemination of clones. Furthermore, the malignant epithelial cells from all patients undergo shared phenotypic change toward an immune/inflammatory-like state across tissues.

Our own cohort precluded contextualizing these gene sets with respect to drug treatment, as the only androgen deprivation therapy (ADT)-treated patient (ICGC_PCAL25), the LN samples had insufficient quality for downstream single-cell profiling. Thus, we turned to external single-cell RNA sequencing data from metastatic prostate cancer[47]. This dataset covers major treatment stages of pre-treatment, on-ADT and/or anti-PD-1 combination therapy and recurrence in metastatic settings. Focusing on the relative enrichment across treatment stages of our signatures above, showcased LN-up genes to increase over treatment relative to pre-treatment (Fig. 4g). In our model, this suggests that transcriptional programs associated with lymph node metastases can emerge as the disease progresses, suggesting that an immune/inflammatory-like state may confer advantages beyond the lymph node microenvironment/metastatic niche. Concomitantly, primary prostate-upregulated genes, representing canonical prostate pathways, were also relatively enriched in post-treatment recurrence, which aligns with a renewed reliance on oncogenic AR signaling during disease relapse (Fig. 4g). Together, these results support a model of parallel pathway exploration within the primary site along the axis of progression in prostate cancer, reflecting coexisting adaptive strategies.

Discussion

In this study, we leveraged scMultiome in multi-sampled primary locally advanced prostate cancer with synchronous locoregional lymph node metastases to reproducibly measure and report heterogeneity within and across patients and tissues.

Phenotypic heterogeneity in malignant cells, both between and within patients, suggests that primary tumor formation is shaped by patient-specific tumor evolutionary processes. In contrast, TME cell types, such as immune and stromal cells, are relatively conserved across patients. The similarity in TME composition aligns with prior single-cell studies for other malignancies[48–50] and supports targeting the TME as an alternative therapeutic strategy in high-risk prostate

cancer[51]. While our Shannon entropy analysis revealed relative conservation of TME composition across patients and spatial sites, it does not preclude the possibility that local, spatially restricted interactions contribute to phenotypic heterogeneity between malignant populations[52–54]. Whether TME and malignant epithelial populations with similar phenotypes congregate in distinct spatial pockets[53,54] and if local cellular interactions result in genetically distinct cells to adopt shared phenotypes remain to be explored. Addressing these will require a different setup, such as high-resolution spatial sequencing technologies, to more precisely map the interplay between the TME and malignant populations.

Despite phenotypic and genotypic inter-patient variability, we found that a common phenotypic shift occurs in malignant epithelial cells disseminating to the lymph nodes across patients. The clonal architecture in lymph node metastatic cancer cells resembled that of the malignant epithelial cells in primary tumor loci in each patient. This suggests that clones can seed and survive the dissemination process in lymph nodes without undergoing an evolutionary bottleneck based on their CNV composition, consistent with bulk sample genetic sequencing studies[4,6,55]. Additionally, barcode-based studies support a model of tumor evolution relying on phenotypic plasticity[56,57]. For instance, individual malignant prostate cancer cells, tagged with unique DNA barcodes, were shown to withstand selective pressures, such as metastatic dissemination and treatment resistance, by undergoing phenotypic transitions[58]. While paired deep whole-genome sequencing data may reveal rare mutations that could impact phenotypic transitions, our observations support a contribution for nongenetic variation, including changes in chromatin accessibility in noncoding DNA elements, in defining the phenotype of disseminating malignant epithelial cells adapted to survive in sites beyond the primary tumor.

Tissue inflammation is a driver of phenotypic plasticity associated with tumor evolution[10], with Inflammatory/invasive-like cancer cell phenotypes being linked to aggressive tumors[59–61]. Our investigation revealed phenotypic variability between primary and metastatic lymph node malignant cells marked by a transition in the enrichment of prostate-specific pathways to immune/inflammatory-like states within the cancer cells themselves, suggesting engagement of epithelial-intrinsic immune programs during dissemination. These immune/inflammatory-like pathways, including JAK-STAT signaling, have also been observed in mCRPC progression samples across diverse metastatic sites, including bone[62]. JAK-STAT signaling is also implicated as a driver of plasticity and dedifferentiation from luminal-epithelial cell states in prostate cancer models[63,64]. This suggests that activation of these pathways may represent a mechanism of progression and potentially contribute to treatment resistance, though this hypothesis warrants further validation. Additionally, we observed an increase in lymphocyte-gene expression markers, which may be showcasing phenotypic mimicry of malignant cells to TME cells (or TME-mimicry) and reliance on non-canonical prostate pathways. Mouse models of aggressive prostate and breast cancers with an intact immune system have validated malignant cells' capacity to express lymphocyte genes[65–68]. Our results further support the concept of TME-mimicry, where prostate cancer cells adopt the phenotypic characteristics of lymphocytes when disseminated to the lymph nodes. Our cohort, which is limited in size and restricted to lymph node metastases, constrains our ability to expand TME-mimicry to distal metastatic sites and to contextualize clinical outcomes. Nevertheless, with 38 samples and 282,956 single nuclei, we were able to robustly characterize malignant cell states. Because these patients were treated with curative intent, the lymph node metastases represent early locoregional dissemination, which explains the smaller cell numbers relative to primary tumors. This unique context provides a valuable window into the phenotypic variation that accompany early metastatic colonization and may have broader implications for understanding immune-evasion in prostate cancer, a disease where immunotherapy has largely been unsuccessful[69,70].

Our study highlights the need for clearer conceptual frameworks to interpret phenotypic heterogeneity in cancer, particularly in the context of malignant cell adaptation across diverse niches. In our data, malignant cells from primary tumors and matched lymph node metastases share a common CNV-based clonal architecture yet exhibit reprogramming of chromatin accessibility and transcriptional profiles. Specifically, disseminated prostate cancer cells adopt an inflammatory-like state characterized by loss of canonical prostate-lineage programs and activation of immune-associated pathways. These findings suggest that dissemination to lymph nodes imposes microenvironmental pressures that may favor a cell-state transition within the same clonal lineage towards an inflammatory-like prostate cancer cell state. We also captured a cell cycling-dependent phenotype composed of inter-locus mixed cancer cells, representing nongenetic variability independent of microenvironmental context. Together, these findings underscore the need for a framework that distinguishes between environmentally cued, adaptive transitions, and stochastic or functionally neutral fluctuations, as further discussed in ref. 71.

In conclusion, our study underscores the complexity of tumor heterogeneity and the convergence of evolutionary paths along dissemination beyond the primary site, as reflected in a cancer cell state transition shared across patients to colonize the lymph nodes.

Method

Biospecimens and quality control

Patients. Biospecimens were collected from five patients diagnosed with prostate cancer as part of the German ICGC cohort. Written informed patient consent and an ethical vote (Institutional Reviewing Board) were obtained according to the ICGC guidelines (see http://www.icgc.org). The local ethics board commission in Hamburg approved the study (PV3552 and PV4679), and the study was conducted following ethical guidelines (Declaration of Helsinki). No statistical methods were used to determine sample size since this study was exploratory.

Inclusion criteria were: primary prostate cancers, high-risk, histologically adenocarcinomas, multiple biopsies from a locally advanced specimen with a unifocal disease, and matching pelvic lymph node metastases.

Patients did not receive any neo-adjuvant radiotherapy, ADT, or chemotherapy before the prostatectomy with pelvic lymphadenectomy, except for one patient (ICGC_PCAL25) who received neoadjuvant hormone therapy with LH-RH agonist 6 weeks before surgery. One patient (ICGC_PCAL03) had a known oligometastatic disease stage at the time of operation (M1b stadium in conventional imaging with bone metastases).

All prostate specimens were analyzed according to a standard procedure, including a complete embedding of the entire prostate for histological analysis[72]. Clinical metadata were retrieved from the patient's records, and follow-up data were collected with questionnaires.

Pathology review

Specialized pathologists dissected each prostate immediately after surgery as described before[27]. In brief, dissection followed a predefined scheme to represent the position of each block relative to the entire prostate. This procedure resulted in 60–150 pieces of tissues depending on the size of the prostate. After dissection, each tissue block was placed on a separately labeled cork plate covered with a special compound for cryopreservation (OCT) before the tissue was frozen to −20 °C and subsequently stored at −80 °C. Accordingly, the dissection of the lymph node packages proceeded. For each patient, 10 distinct tumor biopsies were taken, and matching lymph node metastases were included in this study (Supplementary Dataset 1). In

total, we had a sample collection of 66 samples from 5 different high-risk locally advanced prostate cancer patients (50 primary gland biopsies and 16 matching lymph node biopsies). Gleason grading per individual biopsy was reported (Supplementary Dataset 1). Tumor cell content exceeded 80% for each sample according to pathology reviews.

Sample and library preparation for 10x single-cell multiome ATAC and gene expression

Single-cell sequencing can be prone to batch effects. Given the complexity of our cohort (multiple patients, samples, and tissues), we implemented the following strategy to ensure we are minimizing potential confounding variables of wet-lab library preparation batches and sequencing batches outlined in (Supplementary Dataset 1). Most library-preparation batches (4 samples at a time) incorporated samples from a distinct patient while including randomly chosen samples from another patient. We aimed to randomize samples derived from the same tissue sites across batches as well, such that not all metastatic samples were processed together. This approach was chosen to reduce library preparation batch effects and to better reveal biological variability from technical noise in downstream analyses.

Nuclei isolation

The nuclei were isolated following the adapted protocol for clinically relevant frozen tissues described previously[73]. We kept the samples on dry ice until use and weighed them individually before starting. The range in weight was between 7 and 24 mg. To perform single-cell multiome ATAC + gene expression analysis, we added 0.2 U/μl of Protector RNAse-Inhibitor (Sigma Aldrich PN-3335399001) to the homogenization buffer (according to the 3′GEX assay). In brief, we thawed the frozen tissue in a pre-chilled douncer with 1 ml homogenization buffer for 5 min. We then dounced with pestle A for 10 strokes on ice and continued with pestle B for another 10 strokes. We next filtered the solution through a 40 μm tip into a prechilled 1.5 ml Eppendorf tube and spun the nuclei (5 min, 500 × *g*, 4 °C) in a fixed-angle centrifuge. Then, we removed all but the pellet with 50 μl of supernatant and resuspended it in 350 μl HB to have a total volume of 400 μl. We transferred it to a 2 ml Lo-Bind Eppendorf tube and added 1 volume (400 μl) of 50% iodixanol solution to give a final concentration of 25% iodixanol and mixed by pipetting. Following, we slowly layered 600 μl of 30% iodixanol on top of 600 μl prepared 40% iodixanol and put 800 μl of 25% iodixanol/sample mixture carefully on top. After using a swinging bucket centrifuge (spinning for 20 min at 3000 × *g* with the brake off), we collected 200 μl of nuclei from the 30%–40% interface into a chilled 1.5 ml tube. We then aliquoted 32,500 nuclei (to aim eventually for 10,000 nuclei) into a 1.5 ml tube. We again spun (10 min, 600 × *g*, 4 °C) and discarded the supernatant. We next resuspended the pellet in ~6 μl diluted nuclei buffer + RNase inhibitor (see user guide, Chromium Next GEM Single Cell Multiome ATAC + gene expression from 10x).

Library preparation

Single-nucleus libraries were generated using the 10x Chromium single-cell Instrument and NextGEM Single Cell multiome ATAC + gene expression kit (10x Genomics) according to the manufacturer's protocol (Chromium Next GEM single cell multiome ATAC + gene expression Reagent Kits User Guide; CG000338 Rev E). To minimize PCR biases during ATAC library construction, we performed an extra short quantitative PCR (qPCR) to assess how many cycles are needed to amplify our libraries without reaching saturation (1:72 °C for 5 min, 2:98 °C for 30 s, 3:98 °C 10 s, 4:63 °C for 30 s, 5:72 °C for 1 min, 6: go to 3–20 times). At the end of the qPCR run, we determined the number of cycles that should be used by looking at the plot of fluorescence vs. the cycle numbers to find out what the maximum fluorescence was, and at what cycle number ¼ of that maximum fluorescence was reached and continued with the ATAC library construction by PCR with 7–9 cycles.

Multiome (snATAC-seq and snRNA-seq) sequencing

Final libraries were quantified using the Qubit dsDNA HS Assay Kit (Life Technologies). The fragment size of every library was analyzed using the Bioanalyzer high-sensitivity chip. All 10x scMultiome ATAC and gene expression libraries were sequenced on the NovaSeq 6000 instruments (Illumina) according to the manufacturer's protocol (Chromium Next GEM Single Cell Multiome ATAC + gene expression reagent kits user guide; CG000338 Rev E). BCL files were converted into fastq using the command cellranger_arc mkfastq with the default parameter. The generated fastq files were processed using cellranger-arc v2.0.0 count function for read trimming, alignment, duplicate marking (ATAC), and UMI counting (GEX) (Cell Ranger ARC, 10x Genomics). The reads were aligned to the pre-built human reference genome GRCh38–v2020-A-2.0.0 (3 May 2021) provided by 10x Genomics.

Library preparation exclusion criteria and quality control

Nuclei were evaluated after staining with 4′,6-diamidino-2-phenylindole (DAPI) and visualized under a high-powered fluorescent microscope (minimum magnification: 40x). Only samples containing high-quality nuclei, defined as clump-free, debris-free, and exhibiting intact membranes, were selected for downstream experimental procedures to ensure data integrity.

After ATAC-library amplification, a quality control (QC) step was performed using qPCR to assess chromatin accessibility. Fold enrichment was calculated by comparing two accessible chromatin regions to two inaccessible chromatin regions. A minimum fold enrichment of 10-fold in accessible regions was required for sample inclusion. The following primer pairs were used:

- **Accessible regions (POS).**
  - *GAPDH:* (F) 5′-GCCAATCTCAGTCCCTTCCC-3′, (R) 5′-TAGTAGCCGGGCCCTACTTT-3′
  - *KAT6B:* (F) 5′-GAAGAGGCGGACCCAGCGGT-3′, (R) 5′-TTCCTGCCGGTCATCTCGCTT-3′

- **Inaccessible regions (NEG).**
  - *SLC22A3:* (F) 5′-GGAGAGGGTGGACAGATTGA-3′, (R) 5′-TCAGCCTTGCTGCTACAGTG-3′
  - *QML93:* (F) 5′-CACTGGTTGTCTTTGCAGGA-3′, (R) 5′-CCTGGGTCATATTGGGACAC-3′

Samples failing to meet these criteria—such as those with poor nuclei integrity or insufficient fold enrichment in accessible over inaccessible chromatin regions—were excluded from further analysis.

Therefore following samples were excluded: ICGC_PCAL03_Tu5, ICGC_PCAL03_Tu6, ICGC_PCAL03_Tu8, ICGC_PCAL03_Tu9, ICGC_PCAL10_LN3, ICGC_PCAL10_Tu8, ICGC_PCAL10_Tu9, ICGC_PCAL25_LN1, ICGC_PCAL25_LN2, ICGC_PCAL25_Tu2, ICGC_PCAL25_Tu3, ICGC_PCAL25_Tu4, ICGC_PCAL25_Tu5, ICGC_PCAL25_Tu7, ICGC_PCAL25_Tu9, ICGC_PCAL25_Tu10, ICGC_PCAL34_Tu2, ICGC_PCAL34_Tu3, ICGC_PCAL34_Tu4, ICGC_PCAL34_Tu7, ICGC_PCAL34_Tu10, ICGC_PCAL37_Tu6.

Samples with fewer than 1500 cells remaining after cellranger-arc v2.0.0 and QC (described below) were also removed: ICGC_PCAL03_LN2 (838 cells post QC), ICGC_PCAL10_Tu7 (52 cells post QC), ICGC_PCAL10_Tu10 (1339 cells post QC), ICGC_PCAL34_Tu1 (319 cells post QC), ICGC_PCAL37_LN1 (670 cells post QC), ICGC_PCAL37_Tu1 (317 cells post QC). In total, 38 samples (27 individual primary gland tumor biopsies and 11 matching lymph node biopsies) passed all QC criteria for inclusion in the analyses presented in this manuscript (Supplementary Dataset 1 batches and QC).

scMultiome data pre-processing

For each sample, output files from Cell Ranger ARC were processed with Seurat and Signac, with the following versions: $R = 4.1.0$,

Signac = 1.4.0, Seurat = 4.1.0. Cells were filtered if one of the following criteria was met: nFeature_ATAC > 500, nFeature_RNA > 500, percent.mt <20, TSS.enrichment > 2. Cells were removed if both of the following conditions were met: nFeature_RNA > 8000 and nFeature_ATAC > 80000. Notably, we computed TSS enrichment using the Signac function (`TSSEnrichment`) and found that median TSS scores reported by ArchR on the same cells are on average, 2.5× higher than those reported by Signac, showcased by recomputing TSS scores using ArchR and Signac across 11 samples from ICGC_P-CAL34 (Supplementary Dataset 1). 99% with TSS over 2 in Signac had TSS over 4 using ArchR, and the remaining 1% had a median of Signac TSS 3.6. Lastly, we did not opt for doublet correction and chose to filter on the top percentile of counts for RNA and ATAC, as doublet correction methods on RNA and ATAC individually did not produce results that agreed across samples[74,75]. Lastly, ATAC data from each sample were then reprocessed using a binned genome approach with 5 kb window sizes across hg38 using the Signac FeatureMatrix function. Downstream analyses were performed on the resulting Seurat object, containing an RNA and tiled genome matrix and were merged across patients.

Inter-patient analysis

Inter-patient data analysis was done using Python (version 3.9.13). Due to the large size of the patient-wide dataset containing all samples, we transitioned from using R to Python-based packages. Subsequent to QC steps and cell filtering described above, Seurat RNA and ATAC assays from each patient were converted to a Python-readable format using R packages MuDataSeurat and SeuratDisk. We then used the Muon package to load and merge each assay across all samples[76] and proceeded with downstream processing.

For RNA, we removed mitochondrial and ribosomal transcripts prior to cluster/population calling. We had already used mitochondrial-based metrics to remove low-quality cells. However, mitochondrial and ribosomal reads can represent technical variability unrelated to the biological variability of interest[77]. Thus, for a more conservative approach in understanding inter-patient heterogeneity, we decided against the inclusion of mitochondrial and ribosomal reads across patients to minimize potential technical variation across patients influencing population-calling (clustering) results. We used scanpy default functions to normalize and scale data (`normalize_total`), retrieve and highly variable genes (`highly_variable_genes`), perform dimensionality reduction with PCA, and build the nearest-neighbor graph with defaults.

For ATAC, we normalized with TF-IDF and retrieved the top 5000 highly variable features, performed dimensionality reduction with LSI, and built the nearest neighbor graph on the top 50 LSI dimensions, excluding dimensions with higher than 0.7 Pearson correlation coefficient with depth (nCount_ATAC). We then used the Muon implementation of the weighted-nearest neighbor approach to build an integrated graph of the RNA and ATAC data, which captures the complementary information from both gene expression and chromatin accessibility in each cell[28,76]. We used Leiden clustering resolution with 1.3 on the obtained multi-modal graph. After clustering, we called peaks for each population using the MACS3 algorithm and built an inter-patient-wide feature set as described below (pseudobulk population calling). We then repeated dimensionality reduction steps for ATAC, recomputed WNN, clustering, and population assignments. Lastly, we did not proceed with batch correction, as nonmalignant cell populations showed intermixing across patients. This is indicative of the clustering algorithms preserving genuine biological differences. Batch correction could be potentially detrimental to removing meaningful differences across malignant epithelial populations[78].

Intra-patient analyses

All intra-patient analyses were performed in *R* version = 4.1.0, Signac = 1.4.0, and Seurat = 4.1.0. For prostate-specific analyses (Fig. 2), we merged samples coming from the primary tumor sections within a prostate gland of a given patient. For intra-patient analyses across tissues (Fig. 4), we merged all samples from metastatic lymph nodes and primary prostate tumor sections. For each merged object, we processed RNA, ATAC, and the subsequent Multiome object as described above, but used the equivalent *R*-based Seurat/Signac implementations. For RNA processing, we used `SCTransform` to normalize the filtered count matrix and `PCA` to run louvain graph-based clustering on the first 50 principal components, and for ATAC, we used `RunTFIDF` to normalize the data, found highly variable regions using `FindTopFeatures`, and `RunSVD` to project ATAC data to lower dimensions. We built the nearest neighbor graph on the top 50 PCA and top 50 LSI dimensions, excluding dimensions with higher than 0.7 correlation with ATAC depth (ATAC_nCount). The weighted nearest neighbor graph was built using `FindMultiModalNeighbors`. Populations were called on this graph using `FindClusters` for each patient with multiple resolutions, we tested and annotated clusters across resolutions 0.4, 0.6, and 0.8; and found 0.4 to cover the major cell types across all patients.

Pseudobulk peak calling

We wrote a Snakemake pipeline for calling and merging peaks for any group of cells in our dataset. First, we generated a fragment TSV file for every cell in a given group (in this case, our called populations) per sample. Second, MACS3 was then used to call peaks on the fragment files with the following parameters: -f BED -q 0.01 -g 2.7e9 --keep-dup all --nomodel --nolambda --shift −75 --extsize 150 --bdg --call-summits[79] and peaks were iteratively merged[80]. We then re-computed the FeatureMatrix in Signac and either transferred to Python for inter-patient analysis or added them to the Seurat objects for intra-glandular/intra-patient analyses. Third, the new FeatureMatrices were used to obtain signal over promoter and non-promoter elements and compute ChromVar-based transcription factor enrichment scores[29]. For each peak catalog, we assigned whether a peak was a promoter element if it was within +−1 kb TSS in gencode transcript annotations, otherwise, the peak was annotated as non-promoter, which was used in subsequent Pearson correlation analyses described below. We used a merged peakset across all samples for inter-patient and patient-specific peaksets for intra-patient analyses.

Population annotations

We used a multi-faceted manual and automated approach to go about population annotations, independently utilizing RNA and ATAC data. For every population, we computed the marker genes and chromVar transcription factor motif enrichments by looking at the top upregulated genes relative to the entire sample of interest (Seurat's FindAllMarkers). For every cell in our dataset, we computed the singleR assignments using the publicly available default datasets and a prostate-specific dataset[81,82]. We also computed each cell's ModuleScore across gene expression markers reported in relevant papers[24,82–84] using Seurat's `AddModuleScore`, with additional cell-cycle markers obtained from https://github.com/hbc/tinyatlas. We then visualized all singleR's cell type assignment proportions per population and visualized ModuleScores to facilitate manual cell type assignments. For every population, we checked the alignment of the highly expressed genes with expected genes per cell type with literature search and available atlases (proteinatlas.org). We also utilized ATAC-based transcription factor-deviation scores per population, as ATAC signal tends to be less prone to biological variability such as cell cycle, and TF-motif enrichment computations rely on more global genome-wide statistics. We cross-referenced TFs based on their enrichment relative to the expected average (which would contain more malignant epithelial cell TFs) and annotated populations according to

TFs. Furthermore, more than one expert was exposed to the population markers and visualizations and independently annotated populations according to the steps above to ensure agreement. This rigorous, multi-layered approach provided confidence that our findings represent biological cell types within and across patient samples.

Assessing the degree of phenotypic heterogeneity in locus-specific populations

We assessed heterogeneity between populations in two ways. First, we reported how inter-mixed populations were based on variables of interest, such as patient, tumor locus, and tissue, using Shannon entropy. Additionally, to compare scores across patients when reporting the median across patients, we normalized the scores to be between 0 and 1 when 0 indicating no entropy and 1 indicating maximum entropy relative that can be achieved given the number of loci and cells per loci. Second, we derived Pearson correlation values on DESeq2 normalized pseudobulked signal across the three feature sets of non-promoter DNA elements, as well as promoter elements and gene expression. We obtained normalized pseudobulked values per population by summing raw counts of each population according to the three feature sets and normalizing them using DESeq2 vst[85].

To assess how transcription factor families drive variance, we conducted PCA on median chromVar motif deviations per population. We then binned the feature loadings into 30 bins across 633 motifs, covering the top and bottom ~5th percentiles of features contributing to a given axis of variation.

To contextualize biological programmes driving intra-prostate tumor heterogeneity, we used the SCENIC+ pipeline on tumors from ICGC_PCAL34 and ICGC_PCAL03[37]. Pre-processing of ATAC data for the SCENIC+ pipeline was done using the polars_1xx branch of the pycisTopic package for scalability purposes. ScRNA preprocessing and the SCENIC+ pipeline were run according to recommendations highlighted at https://scenicplus.readthedocs.io/en/latest/. eRegulons specificity scores for each malignant epithelial cluster were calculated using *scenicplus.RSS.regulon_specificity_scores()* to select the top 10 most specific and confident eRegulons per malignant epithelial cluster. The genes of the highly specific eRegulons were used to create gene lists for pathway enrichment using the *gseapy* Python package.

CNV inference

We used the NUMBAT algorithm[39,86] to perform CNV inference. We performed NUMBAT per sample to ensure tumor vs. normal calls were properly cross-referenced with our RNA/ATAC-based cell type annotations. We then used NUMBAT on a patient-wide basis to get a sense of all CNV segments with higher accuracy. However, due to the large number of cells from each patient, we kept only the cells annotated as malignant epithelial and had to further subsample the number of cells used in the per-patient analysis as follows, for the pipeline to run. To allow computational feasibility, we kept a maximum of 3000 cells per sample in ICGC_PCAL03, ICGC_PCAL10, ICGC_PCAL37, and 6000 malignant epithelial cells per sample in ICGC_PCAL34 and ICGC_PCAL25. Lymphocyte populations from ICGC_PCAL10 and ICGC_PCAL37 were used as background, since they had the largest number of cells from the TME. For visualization purposes (Fig. 2b), we used the most common CNV call in all cells assigned to a given clone. To compute the difference between clones, we used the Hamming metric, which measures the distance between two vectors of equal length. The CNV calls per cell were then used to compute this metric between all cells and clones, to broadly assess the difference between clones in a given patient.

Assessing the degree of clonal architecture heterogeneity in populations

We assessed how similar the clonal makeup of each population was according to the Jensen–Shannon (JS) divergence metric. We used JS divergence as it provides a symmetrical and finite measurement of the difference between two probability distributions. For every malignant epithelial population with more than 200 cells, we computed the proportion of cells that were assigned to each clone, and the JS distance was calculated per population pairs.

Differentially accessible regions and transcripts identification and analyses

To identify differentially accessible peaks and transcripts, we pseudobulked raw counts from malignant cells from a given patient per sample across all features. Only cells annotated as malignant epithelial were kept. We then used DESeq2's ~patient+tissue to retrieve differential peaks and transcripts of primary prostate and metastatic lymph nodes, FDR thresholded at <0.1. We then computed transcription factor DNA motif enrichment scores using HOMER on differentially accessible peaks with absolute LFC > 1 and $q < 0.1$[87].

Ambient RNA correction

To gain confidence in our differential analyses, we repeated our pipeline but used ambient RNA corrected count matrices using CellBender[43]. We optimized parameters per sample, except for the FPR, which we had to keep consistent across all, according to the authors (FPR = 0.05). Overall, CellBender retained a median of 40% more cells than CellRanger and did not have an apparent benefit to the malignant epithelial cell expression patterns (Supplementary Fig. 8a, b). We kept only the cells that passed QC according to our scMultiome data pre-processing as outlined, and repeated our differential analysis as described above.

Validation of gene signatures in an external single-cell RNAseq dataset

To validate the AUCell enrichment scores of gene signatures, we utilized an external dataset[47]. Specifically, we downloaded the Seurat object (primecut.integrated.geneexp.seurat.rds) from the publicly available repository hosted at https://data.mendeley.com/datasets/5nnw8xrh5m/1. This dataset contains gene expression profiles of castration-sensitive prostate cancer (mCSP) cells from various metastatic sites. We subset the dataset to exclude TME cell types, isolating a total of 3976 malignant epithelial mCSPC cells. These cells were distributed across four metastatic sites: lymph nodes ($n = 1096$), bone ($n = 1263$), lung ($n = 641$), and liver ($n = 976$). We calculated AUCell enrichment scores for our lymph node up- and down-regulated genes across the treatment stages in all metastatic sites to assess the dynamics of our gene signatures in response to treatment. We calculated AUCell enrichment scores for lymph node up- and down-regulated genes across treatment stages in all metastatic sites to assess gene signature dynamics. Scores were normalized between 0 and 1, and a Kruskal–Wallis test determined whether at least one treatment group differed significantly. Pairwise comparisons were conducted using the Wilcoxon rank-sum test. This approach allowed us to investigate both site-specific and systemic changes in gene signature dynamics during mCSPC progression and treatment.

Reporting summary

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

Data availability

The processed data generated in this study have been deposited in the Gene Expression Omnibus (GEO) database under accession code GSE271675. Raw human sequencing data (FASTQ files) are deposited in the European Genome-Phenome Archive (EGA) under Study EGAS50000000524 and Dataset EGAD50000000745. Access is restricted under German and EU data protection regulations and is governed by the DKFZ Data Access Committee (DAC; EGAC00001000279) and the DKFZ Epigenomics Data Transfer

Agreement (EGAP00001000277). Qualified researchers can request access through the EGA portal. Requests are typically evaluated within 2–4 weeks. Approved access is granted for 3 years, in accordance with the Data Transfer Agreement, and may be renewed upon request. The publicly available single-cell RNA-seq dataset used for AUCell validation in this study is available at Mendeley (https://doi.org/10.17632/5nnw8xrh5m.1)[47]. SingleR assignments used the publicly available default datasets through the package[81,82], and the prostate-specific dataset is available at GEO SuperSeries GSE120716[81,82].

The remaining data are available within the article, Supplementary Information or Source data file. Source data are provided with this paper.

Code availability

Scripts for generating processed data and plotting can be found on GitHub (https://github.com/LupienLab/scMultiomePCa), and CodeOcean (doi:10.24433/CO.2339341.V1)[88]. The analytical pre-processing pipelines are also on the CoBE platform (www.pmcobe.ca).

References

1. Chang, A. J., Autio, K. A., Roach, M. 3rd & Scher, H. I. High-risk prostate cancer-classification and therapy. *Nat. Rev. Clin. Oncol.* **11**, 308–323 (2014).
2. Rebello, R. J. et al. Prostate cancer. *Nat. Rev. Dis. Primers* **7**, 9 (2021).
3. Boutros, P. C. et al. Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat. Genet.* **47**, 736–745 (2015).
4. Warner, E. W. et al. Multiregion sampling of de novo metastatic prostate cancer reveals complex polyclonality and augments clinical genotyping. *Nat. Cancer* **5**, 114–130 (2024).
5. Zhang, N. et al. High clonal diversity and spatial genetic admixture in early prostate cancer and surrounding normal tissue. *Nat. Commun.* **15**, 3475 (2024).
6. Singhal, U. et al. Integrative multi-region molecular profiling of primary prostate cancer in men with synchronous lymph node metastasis. *Nat. Commun.* **15**, 4341 (2024).
7. Espiritu, S. M. G. et al. The evolutionary landscape of localized prostate cancers drives clinical aggression. *Cell* **1**, 27 (2018).
8. Stopsack, K. H. et al. Oncogenic genomic alterations, clinical phenotypes, and outcomes in metastatic castration-sensitive prostate cancer. *Clin. Cancer Res.* **26**, 3230–3238 (2020).
9. Kron, K. J. et al. TMPRSS2–ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer. *Nat. Publ. Group* **49**, 1336–1345 (2017).
10. Alonso-Curbelo, D. et al. A gene-environment-induced epigenetic program initiates tumorigenesis. *Nature* **590**, 642–648 (2021).
11. Mack, S. C. et al. Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature* **506**, 445–450 (2014).
12. McDonald, O. G. et al. Epigenomic reprogramming during pancreatic cancer progression links anabolic glucose metabolism to distant metastasis. *Nat. Genet.* **49**, 367–376 (2017).
13. Moorman, A. et al. Progressive plasticity during colorectal cancer metastasis. *Nature* **637**, 947–954 (2025).
14. Hanahan, D. Hallmarks of cancer: new dimensions. *Cancer Discov.* **12**, 31–46 (2022).
15. Magnani, L. et al. Genome-wide reprogramming of the chromatin landscape underlies endocrine therapy resistance in breast cancer. *Proc. Natl. Acad. Sci.* **110**, E1490–E1499 (2013).
16. Grosselin, K. et al. High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat. Genet.* **51**, 1060–1066 (2019).
17. Deblois, G. et al. Epigenetic switch-induced viral mimicry evasion in chemotherapy resistant breast cancer. *Cancer Discov.* https://doi.org/10.1158/2159-8290.CD-19-1493 (2020).
18. Pomerantz, M. M. et al. Prostate cancer reactivates developmental epigenomic programs during metastatic progression. *Nat. Genet.* **52**, 790–799 (2020).
19. Grillo, G. et al. Transposable elements are co-opted as oncogenic regulatory elements by lineage-specific transcription factors in prostate cancer. *Cancer Discov.* **13**, 2470–2487 (2023).
20. Mazrooei, P. et al. Cistrome partitioning reveals convergence of somatic mutations and risk variants on master transcription regulators in primary prostate tumors. *Cancer Cell* **36**, 674–689.e6 (2019).
21. Stelloo, S. et al. Integrative epigenetic taxonomy of primary prostate cancer. *Nat. Commun.* **9**, 4900 (2018).
22. Pomerantz, M. M. et al. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat. Genet.* **47**, 1346–1351 (2015).
23. Takeda, D. Y. et al. A somatically acquired enhancer of the androgen receptor is a noncoding driver in advanced prostate cancer. *Cell* **174**, 422–432 (2018).
24. Chen, S. et al. Single-cell analysis reveals transcriptomic remodellings in distinct cell types that contribute to human prostate cancer progression. *Nat. Cell Biol.* **23**, 87–98 (2021).
25. Labrecque, M. P. et al. Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. *J. Clin. Investig.* **129**, 4492–4505 (2019).
26. Tang, F. et al. Chromatin profiles classify castration-resistant prostate cancers suggesting therapeutic targets. *Science* **376**, eabe1505 (2022).
27. Gerhauser, C. et al. Molecular evolution of early-onset prostate cancer identifies molecular risk markers and clinical trajectories. *Cancer Cell* **34**, 996–1011.e8 (2018).
28. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
29. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
30. Yang, Y. et al. Pan-cancer single-cell dissection reveals phenotypically distinct B cell subtypes. *Cell* **187**, 4790–4811.e22 (2024).
31. Haffner, M. C. et al. Genomic and phenotypic heterogeneity in prostate cancer. *Nat. Rev. Urol.* **18**, 79–92 (2021).
32. Thurman, R. E. et al. The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
33. Lu, X. et al. NKX2-1 drives neuroendocrine transdifferentiation of prostate cancer via epigenetic and 3D chromatin remodeling. *Nat. Genet.* **57**, 1966–1980 (2025).
34. Rodarte, K. E. et al. Neuroendocrine differentiation in prostate cancer requires ASCL1. *Cancer Res* **84**, 3522–3537 (2024).
35. Lee, G. T. et al. TCF4 induces enzalutamide resistance via neuroendocrine differentiation in prostate cancer. *PLoS ONE* **14**, e0213488 (2019).
36. Zhang, X. et al. Pax6 is a human neuroectoderm cell fate determinant. *Cell Stem Cell* **7**, 90–100 (2010).
37. Bravo González-Blas, C. et al. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat. Methods* **20**, 1355–1367 (2023).
38. Hieronymus, H. et al. Copy number alteration burden predicts prostate cancer relapse. *Proc. Natl. Acad. Sci. USA.* **111**, 11139–11144 (2014).
39. Gao, T. et al. Haplotype-aware analysis of somatic copy number variations from single-cell transcriptomes. *Nat. Biotechnol.* **41**, 417–426 (2022).
40. Fraser, M. et al. Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* **541**, 359–364 (2017).
41. The Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* **163**, 1011–1025 (2015).

42. Porcaro, A. B. et al. Advanced age is an independent prognostic factor of disease progression in high-risk prostate cancer: results in 180 patients treated with robot-assisted radical prostatectomy and extended pelvic lymph node dissection in a tertiary referral center. *Aging Clin. Exp. Res.* **35**, 1881–1889 (2023).
43. Fleming, S. J. et al. Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nat. Methods* **20**, 1323–1335 (2023).
44. de Bruijn, M. & Dzierzak, E. Runx transcription factors in the development and function of the definitive hematopoietic system. *Blood* **129**, 2061–2069 (2017).
45. Tamura, T., Yanai, H., Savitsky, D. & Taniguchi, T. The IRF family transcription factors in immunity and oncogenesis. *Annu. Rev. Immunol.* **26**, 535–584 (2008).
46. Reticker-Flynn, N. E. et al. Lymph node colonization induces tumor-immune tolerance to promote distant metastasis. *Cell* **185**, 1924–1942 (2022).
47. Hawley, J. E. et al. Anti-PD-1 immunotherapy with androgen deprivation therapy induces robust immune infiltration in metastatic castration-sensitive prostate cancer. *Cancer Cell* **41**, 1972–1988 (2023).
48. Liu, Z. et al. Single-cell chromatin accessibility analysis reveals the epigenetic basis and signature transcription factors for the molecular subtypes of colorectal cancers. *Cancer Discov.* **14**, 1082–1105 (2024).
49. Regner, M. J. et al. A multi-omic single-cell landscape of human gynecologic malignancies. *Mol. Cell* **81**, 4924–4941 (2021).
50. Terekhanova, N. V. et al. Epigenetic regulation during cancer transitions across 11 tumour types. *Nature* **623**, 432–441 (2023).
51. Bejarano, L., Jordão, M. J. C. & Joyce, J. A. Therapeutic targeting of the tumor microenvironment. *Cancer Discov.* **11**, 933–959 (2021).
52. Arora, R. et al. Spatial transcriptomics reveals distinct and conserved tumor core and edge architectures that predict survival and targeted therapy response. *Nat. Commun.* **14**, 5029 (2023).
53. Wang, X. et al. Spatial transcriptomics reveals substantial heterogeneity in triple-negative breast cancer with potential clinical implications. *Nat. Commun.* **15**, 10232 (2024).
54. Oliveira, M. F. et al. High-definition spatial transcriptomic profiling of immune cell populations in colorectal cancer. *Nat. Genet.* **57**, 1512–1523 (2025).
55. Reiter, J. G. et al. Lymph node metastases develop through a wider evolutionary bottleneck than distant metastases. *Nat. Genet.* **52**, 692–700 (2020).
56. Rehman, S. K. et al. Colorectal cancer cells enter a diapause-like DTP state to survive chemotherapy. *Cell* **184**, 226–242 (2021).
57. Berthelet, J. et al. The site of breast cancer metastases dictates their clonal composition and reversible transcriptomic profile. *Sci. Adv.* **7**, eabf4408 (2021).
58. Serio, R. N. et al. Clonal lineage tracing with somatic delivery of recordable barcodes reveals migration histories of metastatic prostate cancer. *Cancer Discov.* https://doi.org/10.1158/2159-8290.CD-23-1332 (2024).
59. Guilhamon, P. et al. Single-cell chromatin accessibility profiling of glioblastoma identifies an Invasive cancer stem cell population associated with lower survival. *Elife* **10**, e64090 (2021).
60. Richards, L. M. et al. Gradient of developmental and injury response transcriptional states defines functional vulnerabilities underpinning glioblastoma heterogeneity. *Nat. Cancer* **2**, 157–173 (2021).
61. Hamed, A. A. et al. Gliomagenesis mimics an injury response orchestrated by neural crest-like cells. *Nature* **638**, 499–509 (2025).
62. Westbrook, T. C. et al. Transcriptional profiling of matched patient biopsies clarifies molecular determinants of enzalutamide-induced lineage plasticity. *Nat. Commun.* **13**, 5345 (2022).
63. Chan, J. M. et al. Lineage plasticity in prostate cancer depends on JAK/STAT inflammatory signaling. *Science* **377**,1180–1191 (2022).
64. Deng, S. et al. Ectopic JAK-STAT activation enables the transition to a stem-like and multilineage state conferring AR-targeted therapy resistance. *Nat. Cancer* **3**, 1071–1087 (2022).
65. Morel, K. L. et al. Low tristetraprolin expression activates phenotypic plasticity and primes transition to lethal prostate cancer in mice. *J Clin Invest.* **135**, e175680 (2024).
66. Tshering, L. F. et al. Immune mechanisms shape the clonal landscape during early progression of prostate cancer. *Dev. Cell* **58**, 1071–1086 (2023).
67. Kurzejamska, E. et al. C/EBPβ expression is an independent predictor of overall survival in breast cancer patients by MHCII/CD4-dependent mechanism of metastasis formation. *Oncogenesis* **3**, e125 (2014).
68. Gao, R. et al. Cancer cell immune mimicry delineates onco-immunologic modulation. *iScience* **24**, 103133 (2021).
69. Boettcher, A. N. et al. Past, current, and future of immunotherapies for prostate cancer.*Front. Oncol.* **9**, 884 (2019).
70. Cha, H.-R., Lee, J. H. & Ponnazhagan, S. Revisiting immunotherapy: a focus on prostate cancer. *Cancer Res* **80**, 1615–1623 (2020).
71. Whiting, F. J. H., Househam, J., Baker, A.-M., Sottoriva, A. & Graham, T. A. Phenotypic noise and plasticity in cancer evolution. *Trends Cell. Biol.* **34**, 451–464 (2024).
72. Erbersdobler, A., Hammerer, P., Huland, H. & Henke, R. P. Numerical chromosomal aberrations in transition-zone carcinomas of the prostate. *J. Urol.* **158**, 1594–1598 (1997).
73. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
74. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e4 (2019).
75. Thibodeau, A. et al. AMULET: a novel read count-based method for effective multiplet detection from single nucleus ATAC-seq data. *Genome Biol.* **22**, 252 (2021).
76. Bredikhin, D., Kats, I. & Stegle, O. MUON: multimodal omics analysis framework. *Genome Biol.* **23**, 42 (2022).
77. Subramanian, A., Alperovich, M., Yang, Y. & Li, B. Biology-inspired data-driven quality control for scientific discovery in single-cell transcriptomics. *Genome Biol.* **23**, 267 (2022).
78. Fan, J., Slowikowski, K. & Zhang, F. Single-cell transcriptomics in cancer: computational challenges and opportunities. *Exp. Mol. Med.* **52**, 1452–1465 (2020).
79. Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
80. Corces, M. R. et al. The chromatin accessibility landscape of primary human cancers. *Science* **362**, eaav1898 (2018).
81. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
82. Henry, G. H. et al. A cellular anatomy of the normal adult human prostate and prostatic urethra. *Cell Rep.* **25**, 3530–3542 (2018).
83. 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).
84. Dong, B. et al. Single-cell analysis supports a luminal-neuroendocrine transdifferentiation in human prostate cancer. *Commun. Biol.* **3**, 778 (2020).
85. 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).
86. Li, R. et al. Numbat-multiome: inferring copy number variations by combining RNA and chromatin accessibility information from single-cell data. *Brief. Bioinform.* **26**, bbaf516 (2025).

87. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
88. Keshavarzian, T. Dynamics of tumor heterogeneity from primary to metastatic dissemination in prostate cancer. *Code Ocean* https://doi.org/10.24433/CO.2339341.V1 (2024).

Acknowledgements

The authors thank the Princess Margaret Genomics Centre and the Princess Margaret Bioinformatics group for providing support and infrastructure for the computational analysis of this work as well as high-throughput sequencing support. We thank members of the M.L. lab for their fruitful discussions and feedback. This work is supported by the Canadian Institute of Health Research (CIHR) (FRN-153234, 158225, 168933 & 191847 to M.L.), the Ontario Institute for Cancer Research (OICR) Investigator Award through funding provided by the Government of Ontario (IA-031 to M.L.), the Princess Margaret Cancer Foundation, the Cancer Digital Intelligence (CDI) and the Ontario Graduate Student awards by the Government of Ontario to T.K. M.L. holds the Joey and Toby Tanenbaum/Brazilian Ball Chair. K.F. was funded by a clinical scientist fellowship awarded by the German Cancer Research Center Clinician Scientist Program, which is supported by the Dieter Morszeck Foundation. Background graphics of prostate and surrounding lymph nodes used in Figs. 1a; 2a and Supplementary Fig. 7a were created with BioRender.com; all required publication licenses were obtained and referenced in figure legends.

Author contributions

T.K. conceptualization, analysis, supervision, validation, investigation, visualization, methodology, writing–original draft, writing–review and editing, data interpretation K.F. conceptualization, sample library preparation, supervision, investigation, data interpretation, writing–review and editing G.G. experimental support L.M. experimental support C.A. sample library preparation F.H. experimental support A.N. software resources M.M. technical support E.K. technical support A.G. experimental support E.P. experimental support J.H. technical support J.S. experimental support S.M. sample collection and pathological assessment T.S. sample collection G.S. sample collection R.S. sample collection, pathological assessment, and data curation H.H.H. conceptualization J.W. technical support C.P. conceptualization, resources, supervision C.G. sample data curation, conceptualization, project administration, supervision M.L. conceptualization, resources, supervision, funding acquisition, project administration, writing–review and editing.

Competing interests

G. Grillo reports that at the time of publication, he was an employee of T-One Therapeutics s.r.l. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-67856-5.

Correspondence and requests for materials should be addressed to Clarissa Gerhäuser or Mathieu Lupien.

Peer review information *Nature Communications* thanks Srinivasa Rao and the other 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.

[1]Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada. [2]Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada. [3]Division of Cancer Epigenomics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [4]Department of Urology, Charité–Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany. [5]Biotech Research & Innovation Centre (BRIC), The Finsen Laboratory, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark. [6]Cancer-related Advanced Research and Education, TCG CREST, Kolkata, India. [7]Department of Pathology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [8]Department of Oncology, The DCCC Brain Tumor Center, Rigshospitalet, Copenhagen, Denmark. [9]Ontario Institute for Cancer Research, Toronto, ON, Canada. [10]These authors contributed equally: Tina Keshavarzian, Kira Furlano. ✉e-mail: c.gerhauser@dkfz-heidelberg.de; mathieu.lupien@uhn.ca