

RESEARCH

Open Access



Transcriptional analysis of metastatic hormone-naïve prostate cancer primary tumor biopsies reveals a relevant role for SOX11 in prostate cancer cell dissemination

Natalia Martin-Martin^{1,2,3†}, Saioa Garcia-Longarte^{1†}, Jon Corres-Mendizabal^{1†}, Uxue Lazcano¹, Ianire Astobiza^{1,2}, Laura Bozal-Basterra¹, Nicolas Herranz⁴, Hielke van Splunder⁵, Onintza Carlevaris¹, Mikel Pujana-Vaquerizo^{1,2}, María Teresa Blasco⁶, Ana M. Aransay^{1,7}, Antonio Rosino⁸, Julian Tudela⁹, Daniel Jimenez⁸, Alberto Martinez¹⁰, Andrei Salca⁴, Aida Santos-Martín¹¹, Sofía Rey¹¹, Aitziber Ugalde-Olano¹², David Gonzalo^{3,11}, Mariona Graupera^{2,5,13}, Roger R. Gomis^{2,6,13}, Joaquin Mateo⁴, Miguel Unda^{2,3,11}, Enrique Gonzalez-Billalabeitia^{8,14}, Ana Loizaga-Iriarte^{2,3,11}, Isabel Mendizabal^{1,3,15*} and Arkaitz Carracedo^{1,2,3,15,16*}

[†]Natalia Martin-Martin, Saioa Garcia-Longarte, Jon Corres-Mendizabal contributed equally to this work.

*Correspondence:
imendizabal@cicbiogune.es;
acarracedo@cicbiogune.es

¹ Center for Cooperative Research in Biosciences (CIC bioGUNE), Basque Research and Technology Alliance (BRTA), Bizkaia, Spain
Full list of author information is available at the end of the article

Abstract

Background: Metastatic hormone-naïve prostate cancer (mHNPC) is an infrequent form of this tumor type that is characterized by metastasis at the time of diagnosis and accounts for up to 50% of prostate cancer-related deaths. Despite the extensive characterization of localized and metastatic castration-resistant prostate cancer, the molecular characteristics of mHNPC remain largely unexplored.

Results: Here, we provide the first extensive transcriptomics characterization of primary tumor specimens from patients with mHNPC. We generate discovery and validation bulk and single-cell RNA-seq datasets and perform integrative computational analysis in combination with experimental studies. Our results provide unprecedented evidence of the distinctive transcriptional profile of mHNPC and identify stroma remodeling as a predominant feature of these tumors. Importantly, we discover a central role for the SRY-box transcription factor 11 (SOX11) in triggering a heterotypic communication that is associated with the acquisition of metastatic properties.

Conclusions: Our study will constitute an invaluable resource for a profound understanding of mHNPC that can influence patient management.

Keywords: Prostate cancer, Metastatic driver, Cancer cell reprogramming, Single-cell transcriptomics, Hormone-naïve, Tumor-stromal heterotypic communication, SOX11



© The Author(s) 2025. **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/>.

Background

Prostate cancer (PC) is the most prevalent cancer type in men, and it is predominantly diagnosed in a localized stage, when curative-intended therapies are highly efficacious. Recurrent PC is subject to different means of androgen signaling inhibition and chemotherapy. Failure of these therapeutic approaches might result in an ultimately metastatic form of the disease (metastatic castration resistant PC, or mCRPC). The molecular landscape and driver alterations of localized PC and mCRPC have been extensively characterized [1–3], contributing to build our current portrait of the disease. However, a minority of PC cases (estimated 10%) [4] are diagnosed in a disseminated stage, offering a unique biological scenario where metastatic disease emerges prior to systemic therapy (metastatic hormone-naïve PC, mHNPC).

Despite the low incidence of mHNPC, these patients account for up to 50% of mortality by PC [5]. Prostatectomy is not a standard of care procedure in mHNPC cases, which limits biological sample availability and hampers the molecular characterization of this rare form of PC. A recent comprehensive study has highlighted extensive genomic intra-tumoral heterogeneity in mHNPC primary tumors [4]. This and previous studies support that mHNPC primary tumors genetically resemble advanced mCRPC specimens rather than primary locoregional tumors [4, 6, 7] or, alternatively, present an intermediate profile [8, 9]. Alterations that are common to both mHNPC and mCRPC include inactivation of tumor suppressor genes *TP53*, *PTEN*, and *RB1*. In contrast, aberrations in androgen receptor (*AR*) gene are infrequent in mHNPC [4, 6, 7, 10] despite reduced AR activity in these tumors [11]. Altogether, genomic evidence suggests that the aggressive features of mHNPC primary tumors predate therapy exposure. Nevertheless, this rare subtype of the disease remains severely uncharacterized at the molecular level. Improved clinical management strategies in mHNPC demand the generation of molecular resources that can impact guidelines for patient stratification and treatment.

Here, we present the first comprehensive transcriptome-wide characterization of primary tumors obtained from mHNPC patients using bulk and single-cell RNA sequencing technologies. For appropriate biological comparison, we contrasted primary tumors from patients with localized disease to untreated mHNPC primary tumors using needle biopsy-derived tissue. To account for the extensive inter-patient molecular heterogeneity that characterizes PC primary lesions [1–3], we assembled and profiled two retrospective patient cohorts encompassing 110 individuals using bulk RNA-Seq. Extensive computational characterization of bulk RNA-Seq was combined with single-cell transcriptomics analysis of 15 cases, leading to the identification of *SOX11* as a potential driver of stroma remodeling and cancer cell dissemination in mHNPC.

Results

The transcriptional landscape of metastatic hormone-naïve prostate cancer primary tumors is highly divergent

To undertake a profound molecular characterization of mHNPC, we generated a series of patient cohorts with either localized disease or metastatic dissemination at diagnosis leveraging from retrospective studies in different Spanish hospitals, prospective biopsy collection, and detailed clinical annotation of publicly available resources (Fig. 1A).

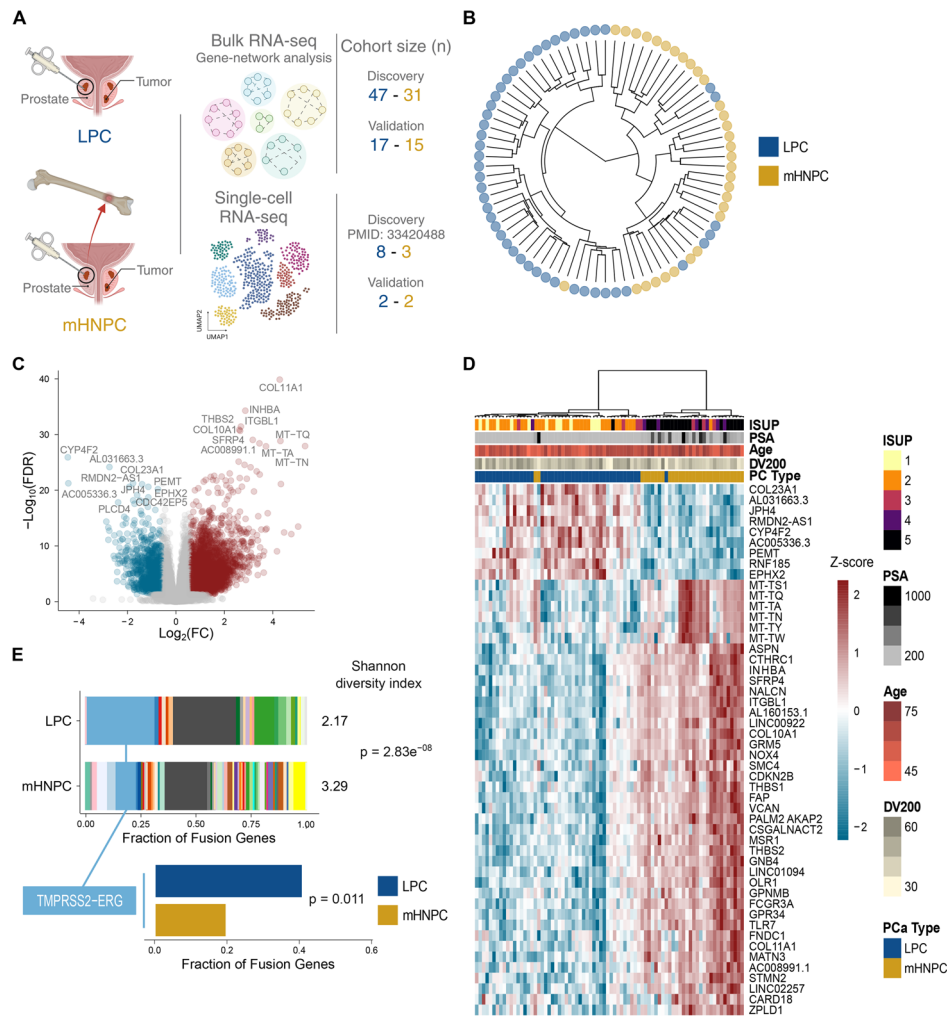


Fig. 1 Divergent transcriptional profile of primary tumor specimens from localized and metastatic hormone-naïve prostate cancer patients. **A** Overview of the cohorts of localized (LPC) and metastatic hormone-naïve (mHNPc) prostate cancer patients and transcriptomic datasets from primary tumors included in this study. **B** Unsupervised hierarchical clustering of transcriptome-wide patterns of LPC and mHNPc primary tumors (discovery cohort). **C** Volcano plot of differential gene expression analysis for the mHNPc vs. LPC comparison. **D** Heatmap and unsupervised clustering of the top 50 differentially expressed genes including technical (RNA quality score on FFPE samples, DV200) and clinical variables (age, PSA, and ISUP). **E** Gene fusion detection in LPC and mHNPc tumors from transcriptomic data. Diversity estimates according to Shannon diversity index are shown (p -value from Hutcheson's t -test). The zoom shows TMPRSS2-ERG gene fusion class with the relative fractions in each tumor type (p -values from binomial test)

Taking advantage of the availability of prostate biological material from formalin-fixed and paraffin embedded (FFPE) diagnostic needle biopsies, we first performed bulk RNA sequencing in a discovery cohort comprising 47 localized (LPC) and 31 mHNPc specimens (see workflow in Fig. 1A and Additional file 1: Table S1 for cohort characteristics). We hypothesized that comparing primary tumor biopsies from metastatic and non-metastatic disease without the influence of therapy-induced selection pressure would offer us unique insights into the molecular determinants of PC aggressiveness. Unsupervised hierarchical clustering and principal component analyses (Fig. 1B and Additional file 2: Fig. S1A) revealed highly distinct transcriptomic profiles in these two primary tumor

types, with notable interindividual variation within mHNPC tumors. We identified a large number of genes with significant transcriptional differences between disease clinical states (5147 differentially expressed genes or DEGs, $|\log_2(\text{fold change})| > 1.5$) and false discovery rate (FDR) < 0.05 , Fig. 1C and Additional file 3: Table S2) that were organized into hundreds of biological processes differentially enriched (Additional file 3: Table S3). These data unraveled a profound molecular difference between LPC and mHNPC. Histopathological differences could represent a confounding factor in this analysis. Since mHNPC exhibited higher Gleason score (or ISUP grade [12]) (Fig. 1D), we took advantage of four publicly available cohorts encompassing 780 patients with localized disease with annotated Gleason score (Additional file 3: Table S4). Top 50 DEGs failed to cluster the patients according to pathological grading, suggesting that the observed differential expression is not driven by this parameter (Additional file 2: Fig. S1B–E).

Transcriptional data can provide valuable information about genomic reorganization in these aggressive tumors [13]. In contrast to other cancer types, gene fusions represent the most common genetic alteration in PC, being present in over 70% of the cases [14, 15]. We inferred gene fusions from the RNA-Seq data following the Trinity Cancer Transcriptome Analysis Toolkit (CTAT) [13]. In line with previous studies, $> 80\%$ of patients in our cohort presented gene fusions (Additional file 3: Table S5). Despite the similar number of individuals with gene fusions and the similar number of fusions per individual between the two groups (Additional file 2: Fig. S1F, G), the composition of fusion classes was markedly different in mHNPC, which exhibited remarkable diversity of fusion transcripts (Shannon Diversity Index, $p\text{-value} = 2.83e^{-08}$, Fig. 1E). The increased diversity in genomic structural variation is in concordance with the larger inter-patient variability observed in the mHNPC transcriptome. Notably, while LPC showed a high frequency of *TMPRSS2-ERG* gene fusions, this event was underrepresented in mHNPC patients (19% in mHNPC vs 40% in LPC, binomial test, $p\text{-value} = 0.011$, Fig. 1E). This translocation juxtaposes the androgen-regulated promoter of *TMPRSS2* with the coding region of the *ERG* oncogene. The existence of this gene fusion is associated with androgen-dependent tumors [16], and the reduction in this event in mHNPC is consistent with decreased AR dependence and earlier development of resistance to androgen deprivation therapy [17].

Gene network analysis reveals increased stromal remodeling in mHNPC

Our transcriptomics results reveal profound gene expression differences between LPC and mHNPC, which represents an important challenge for the identification of driver candidate processes. Gene network analyses constitute powerful methods to identify coordinated activities of gene sets. To prioritize among the extensive list of potential biological processes underlying the metastatic capacity of mHNPC primary tumors, we applied weighted gene correlation network analyses (WGCNA) [18] (Fig. 2A, Additional file 2: Fig. S2A–C), where highly interconnected genes are clustered in modules. WGCNA identified 49 color-coded modules, and 7 were strongly positively associated with the mHNPC phenotype (Pearson's $r > 0.6$ and FDR < 0.05 , Fig. 2B). The “purple” module exhibited the highest coordinated overexpression in mHNPC (Fig. 2B, Additional file 2: Fig. S2D and Additional file 3: Table S6). This module (including 758 genes) incorporated the largest proportion of DEGs (77.7%), Fig. 2B) highlighting the

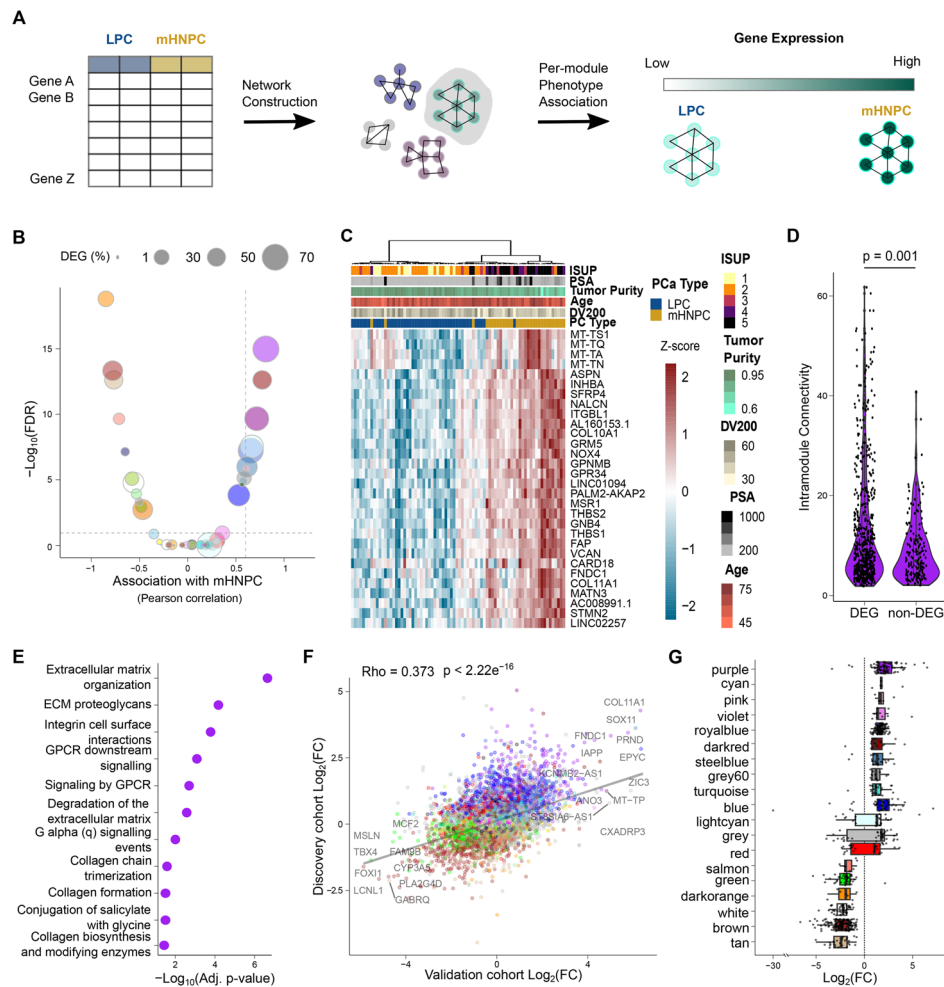


Fig. 2 Gene-network analyses point at upregulation of stroma remodeling in mHNPC. **A** Schematic representation of the gene module construction using weighted-gene network correlation analyses (WGCNA) and association with mHNPC and LPC phenotypes. **B** Volcano plot showing the results of the association of each module with the mHNPC phenotype and the percentage of differentially expressed genes (DEGs) per module (dot size). Lines represent significance thresholds at Pearson's $r > 0.6$ and $FDR < 0.05$. **C** Heatmap and unsupervised clustering of the top 30 DEGs in the purple module. **D** Intramodular connectivity between the genes in the purple module (p -value from Wilcoxon test) according to differential expression (DEG status). **E** REACTOME pathway enrichment results for the DEGs in the purple module. **F** Correlation between the $\log_2$ fold changes of the differential expression analyses in the discovery (y-axis) and validation (x-axis) cohorts for all genes (p -values from Spearman's correlation). Each gene is colored according to module membership in the discovery WGCNA analysis. **G** Boxplot of the $\log_2$ fold changes in the validation dataset for DEGs grouped by the modules identified by WGCNA in the discovery dataset. Only modules with at least 5 genes are shown

convergence of gene-by-gene and network analyses on this gene set. Unsupervised hierarchical clustering revealed that the top 30 genes of the purple module robustly discriminated mHNPC from LPC (Fig. 2C) and this list was enriched in DEGs (Fig. 1D). The DEGs showed larger intramodular connectivity than non-DEGs (Wilcoxon test, p -value = 0.0013, Fig. 2D, Additional file 3: Table S2), implying that DEGs tend to be hub genes within the network. Interestingly, functional enrichment analysis identified processes related to stroma remodeling within the purple module in mHNPC (Fig. 2E).

To validate our observations, we generated a validation patient cohort comprising primary tumor specimens from 17 LPC and 15 mHNPC patients from an independent hospital (Fig. 1A and Additional file 1: Table S1). In line with the results from the discovery cohort, transcriptome-wide variation robustly clustered the two disease types (Additional file 2: Fig. S2E, F). We identified 3180 DEGs in mHNPC tumors in the validation cohort ($|\log_2(\text{fold change})| > 1.5$) and $\text{FDR} < 0.05$, Additional file 2: Fig. S2G, Additional file 3: Table S7). Notably, there was a significant overlap between the DEGs of both datasets (Additional file 2: Fig. S2H) and the fold changes obtained in both cohorts were significantly correlated (Fig. 2F, Spearman's $\rho = 0.373$, $p\text{-value} < 2.22 \times 10^{-16}$ for all genes and Additional file 2: Fig. S2I, Spearman's $\rho = 0.833$, $p\text{-value} < 2.22 \times 10^{-16}$ for DEGs in both datasets). When projecting WGCNA module ascription into the validation dataset, we corroborated that genes in the purple module were the most robustly upregulated in mHNPC (Fig. 2G). Our computational strategy allowed us to reduce the transcriptional changes observed in mHNPC into a subset of coordinated genes associated with stroma remodeling.

Transcriptomics at single-cell resolution identifies mHNPC-specific heterotypic interaction programs coordinated by *SOX11*

Bulk transcriptomic analysis provides information of mRNA abundance with disregard to the cellular source, and obscure quantitative and functional cell type differences. To explore the compositional differences of the tumor biopsies, we employed in silico deconvolution methods that infer cell-type proportions from bulk transcriptomic data [19]. In the discovery cohort, the tumor purity estimates in mHNPC (namely, the relative abundance of the epithelial compartment) were significantly reduced, consistent with an increase in immune and stromal infiltration scores (Additional file 2: Fig. S3A and Additional file 4: Table S8). Analyses based on variance partitioning confirmed that the fraction of transcriptional variation explained by tumor purity was substantial and exhibited a larger contribution than the tumor type (mHNPC vs. LPC, Additional file 2: Fig. S3B and Additional file 4: Table S8). The validation cohort exhibited a similar trend towards a reduction in estimated tumor purity and increased non-immune stromal contribution, while the reduced statistical significance could be due to the smaller cohort size or to the differences in sample collection (RNA from the validation cohort was obtained from frozen tissue that was collected from needle biopsies for mHNPC and prostatectomies for LPC [20]) (Additional file 2: Fig. S3C, D). Changes in stroma remodeling could be ascribed to qualitative (function) or quantitative (abundance) alterations, and the inference of qualitative stromal changes from bulk RNA-Seq [21]. To study functional tumor and stromal cell alterations in mHNPC, we analyzed transcriptomic changes at single-cell resolution. We first leveraged a publicly available single-cell RNA-Seq [22] (Fig. 1A and Additional file 1: Table S1), and did not find evidence of differential composition in any of the identified 21 clusters (Fig. 3A, Additional file 2: Fig. S4A–C and Additional file 5: Table S9). To explore functional differences, we identified DEGs between mHNPC and LPC (Fig. 3B). The epithelial compartment, and specifically luminal cluster 6, exhibited the greatest number of DEGs ($n = 14$, $|\log_2(\text{fold change})| > 1.5$) and $\text{FDR} < 0.1$, Fig. 3B, Additional file 6: Table S10). To confirm that cluster 6 contained tumor cells, we performed copyKAT [23, 24] (Additional file 2: Fig. S4D). Consistent

with bulk transcriptomics analysis, stroma remodeling functions were overrepresented among the DEGs in cluster 6 (Fig. 3C). To test the coherence between single-cell and bulk differential expression analyses, we correlated the fold changes obtained for each strategy (minimum $|\log_2(\text{fold change})| > 1.5$). Notably, genes in the purple module showed the strongest correlation with cluster 6 in single-cell data and, conversely, genes within the purple module were best represented in cluster 6 (Fig. 3D, E and Additional file 6: Table S11). Taken together, the integration of bulk and single-cell analyses identifies genuine molecular alterations in an epithelial tumor cell subset that is strongly associated with mHNPC.

The presence of scDEGs in stromal clusters and the enrichment of mHNPC in stroma-regulatory processes suggested that molecular alterations in these tumor cells could govern stromal remodeling. We modeled intercellular communication between cluster 6 (as sender cells expressing ligands) and stromal cells (fibroblast and endothelial clusters as receiver cells expressing receptors), using NicheNet [25]. Among all processes showing differential connectivity between the two pathological conditions (LPC and mHNPC), we found an overrepresentation of ligands assigned to the purple module (91/824, 11% of interactions in LPC, and 93/356, 26.12% in mHNPC, $p\text{-value} = 1.004e^{-10}$, chi-squared test, Fig. 3F). Interestingly, some of the most differentially expressed genes replicated in bulk and single-cell cohorts were among the top ligands involved in the heterotypic communication process (Fig. 3G and Additional file 7: Table S12), including *SFRP2*, *INHBA*, *COMP*, and *IL11*.

The transcriptional regulation of this ligand set belonging to purple module prompted us to search for transcriptional regulators that were perturbed in mHNPC and

(See figure on next page.)

Fig. 3 Single cell molecular deconstruction of mHNPC. **A** Visualization of single-cell discovery dataset (36,424 cells) using Uniform Manifold Approximation and Projection (UMAP). Colors code for the assignment of each cell to different clusters by graph-based clustering (Leiden algorithm). **B** Number of differentially expressed genes (DEG) per single-cell cluster comparing mHNPC and LPC. **C** Reactome pathway enrichment analysis for DEGs in cluster 6. **D** Correlation analyses between the fold changes of differential gene expression (mHNPC vs LPC, minimum $|\log_2(\text{fold change})| > 0.6$) in bulk and single-cell for mHNPC-associated gene modules. Dark magenta did not show any genes above the fold change criteria and is not represented in the figure. Dot colors show Spearman's correlation values and dot size represents $p\text{-value}$. **E** Correlation analysis of fold changes between bulk and single-cell data for genes in the purple module in cluster 6. **F** Bar plot showing the module classification of the ligands identified in cell–cell communication between cluster 6 (sender) and non-immune stromal cells (receiver). **G** Chord diagram of mHNPC-enriched ligands in the communication between cluster 6 and stromal cells that belong to the purple module. Different stromal clusters were merged into categories for visualization. **H** Spearman correlation of the purple ligands with transcription factors in bulk transcriptome data from mHNPC patients (discovery bulk cohort). **I** Spearman correlation of the purple ligands with transcription factors in bulk transcriptome data from mHNPC patients (validation bulk cohort). **J** Projection of 28 genes associated with cluster 6 in the discovery dataset (i.e. markers that distinguish cluster 6 from other epithelial cells) in the validation cohort, exhibiting highest similarity with cluster 12 within the epithelial compartment. **K** Correlation analysis of differential expression (fold change of mHNPC vs. LPC) between bulk and single-cell RNA-Seq in cluster 12 for purple module genes. **L** Bar plot showing the results of cell–cell communication between cluster 12 (sender) and non-immune stromal cells (receiver) in the single-cell validation cohort for ligands with purple module membership. **M** Chord diagram of mHNPC-enriched ligands in the communication between cluster 12 and stroma that belong to the purple module. Different stromal clusters were merged into categories for visualization. **N** Spearman correlation of the purple ligands identified in the single-cell validation cohort with transcription factors in bulk transcriptome data from mHNPC patients (discovery bulk dataset). **O** Spearman correlation of the purple ligands identified in the single-cell validation cohort with transcription factors in bulk transcriptome data from mHNPC patients (validation bulk dataset)

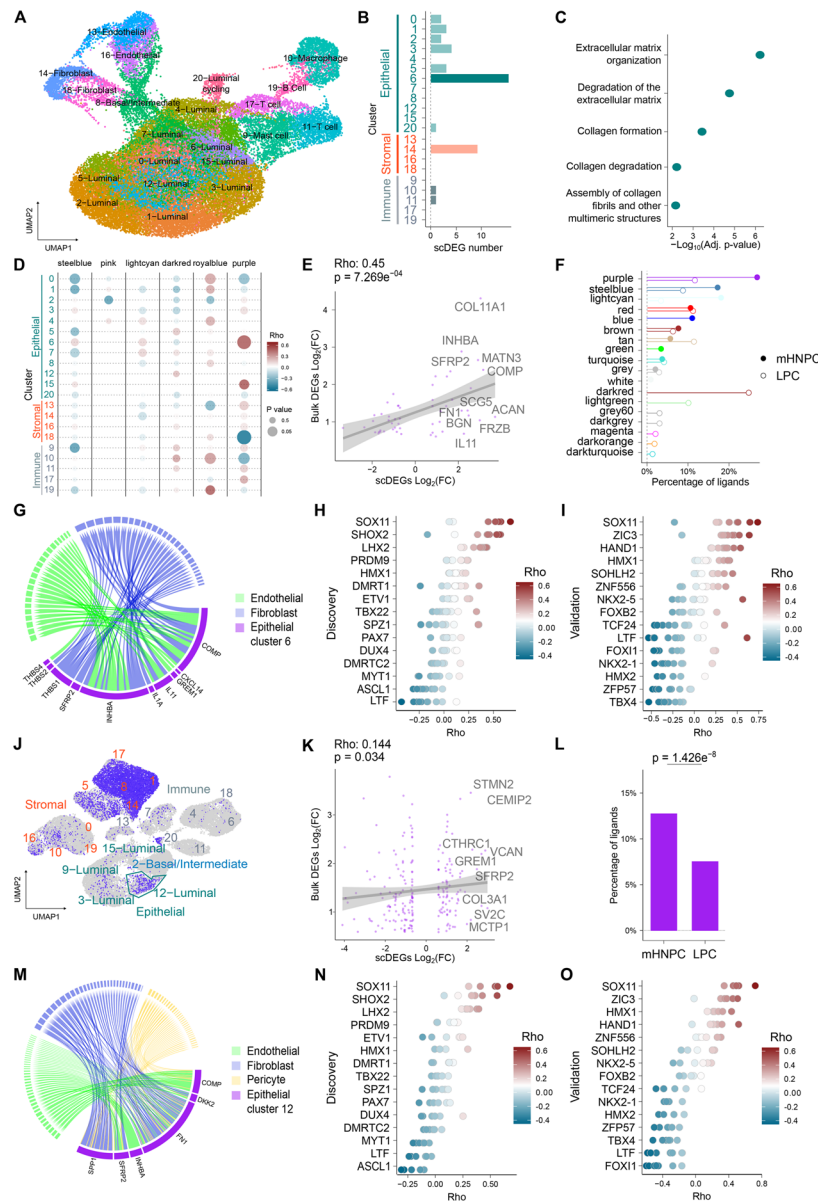


Fig. 3 (See legend on previous page.)

coordinatedly altered the expression of these genes. To this end, we designed a strategy based on gene-to-gene correlation analysis, establishing that ligand expression would correlate with the abundance of upstream transcription factors. To avoid the sparsity of single-cell data at single-gene level, we used the discovery bulk transcriptomic dataset and explored the correlation between all annotated transcription factors [26] and the shortlisted purple module ligands. Interestingly, the SRY-Box Transcription Factor 11, *SOX11*, exhibited the strongest correlation with the ligand set (Fig. 3H, Additional file 8: Table S13). Of note, the corresponding cellular communication analyses between cluster 6 and the whole stroma (including immune clusters), pointed to a similar set of ligands (10/11 shared) and *SOX11* (Additional file 2: Fig. S4E, F). Importantly, this strategy was replicated using the bulk validation cohort (Fig. 3I and Additional file 2: Fig. S4G).

We generated a validation single-cell RNA-Seq dataset from fresh needle biopsies from 2 LPC and 2 mHNPC patients (Fig. 1A and Additional file 1: Table S1). In total, we profiled 36,424 cells and identified 21 clusters corresponding to epithelial, stromal, and immune compartments (Additional file 2: Fig. S5A–C). The compositional analyses of this dataset revealed no differences in cluster abundance, consistent with findings from the discovery cohort (Additional file 5: Table S9 and Additional file 2: Fig. S5D). In this new dataset, cluster 12 exhibited the greatest similarity to the cluster labeled as number 6 in the discovery single-cell dataset: (i) luminal nature and shared cluster markers enriched in stroma remodeling, (ii) large-scale chromosomal copy number alterations, and (iii) significant association with purple WGCNA module regarding differential mHNPC expression (Fig. 3J, K, Additional file 2: Fig. S5E–G). To further characterize these clusters, we identified biological processes specifically enriched in mHNPC within cluster 6 in the discovery dataset but absent in other luminal clusters. Functional gene set enrichment analysis (GSEA) highlighted 7 such unique processes (Additional file 2: Fig. S5H). When applying GSEA to the validation dataset, cluster 12 showed the strongest enrichment of these pathways in mHNPC, with 3 out of 7 reaching significance (compared to 0–1 significant pathway in other luminal clusters) (Additional file 2: Fig. S5H). Notably, validated functional categories related to extracellular matrix organization (Extracellular Structure Organization, External Encapsulating Structure Organization and Skeletal System Development) which is a hallmark of tumor-stroma crosstalk. This functional overlap further supports the equivalence between cluster 12 and cluster 6 and reinforces their potential role in mHNPC-associated processes.

NicheNet-based cell–cell communication analysis in the validation single-cell dataset between cluster 12 and the stromal compartment corroborated the over-representation of ligands belonging to the purple module among the mHNPC-enriched molecular communication processes (178/2349, 7.5% in LPC and 256/2001, 12.8% in mHNPC, $p\text{-value} = 1.426e^{-08}$, chi-squared test, Fig. 3L). Some of the top ligands involved in this communication were concordant with the single-cell discovery dataset (*COMP*, *INHBA*, and *SFRP2*), and new ones were also detected such as *FNI* and *SPPI* (Fig. 3M and Additional file 7: Table S12). The ligands were broadly expressed across stromal and immune compartments (Additional file 2: Fig. S6A, B), similar to the ligands in the discovery dataset (Additional file 2: Fig. S6C), and target gene activation in the receiver cells did not point at a specific stromal cell type being predominantly influenced in both cohorts (Additional file 2: Fig. S6D, E). These data suggest that epithelial tumor cells in mHNPC orchestrate a coordinated stromal reprogramming rather than a process driven by a single cell type. Alternatively, it may reflect the resolution limits of single-cell data and the inherent constraints of ligand–receptor–target network structures.

Despite the expected variability in gene detection in highly sparse single-cell datasets between the two single-cell cohorts, *SOX11* was corroborated as the transcription factor with the strongest correlation with this second set of ligands in mHNPC transcriptomes (Fig. 3N, O, discovery and validation bulk datasets, respectively, and Additional file 8: Table S13). In line with the discovery single-cell dataset, the ligand analysis including the immune stroma also pointed at *SOX11* (Additional file 2: Fig. S6F, G). Collectively, by integrating bulk and single-cell datasets from four independent cohorts, we consistently

identified a tumor cell-intrinsic heterotypic communication program that is distinctive of mHNPC and potentially regulated by *SOX11*.

***SOX11* promotes metastatic dissemination in PC**

As a proof-of-concept of the potential relevance of the mHNPC transcriptional resource, we interrogated the relevance of *SOX11* for the metastatic phenotype in PC. *SOX11* belongs to the SOX transcription factor family. It plays crucial roles in stem cell function and tissue specification during embryogenesis but is largely absent in most differentiated adult tissues [27]. *SOX11* dysregulation is associated with oncogenesis, tumor progression, metastasis, and therapy resistance [28–30]. Our integrative transcriptomic analyses unveiled that *SOX11* ranks prominently among the most overexpressed genes (Fig. 2F) and, importantly, it represents the top upregulated transcription factor (Fig. 4A). The observed elevation in *SOX11* levels and activity in mHNPC prompted us to study the contribution of this transcription factor to the metastatic phenotype. To this end, we engineered DU145 PC cells (which do not exhibit detectable levels of the transcription factor) to overexpress *SOX11* (Fig. 4B, Additional file 2: Fig. S7A). Importantly, ectopic *SOX11* expression resulted in increased mRNA and protein abundance of several purple module ligands identified in our scRNA-Seq analyses as contributors to the heterotypic communication between cancer cells and the stroma (Fig. 4C, Additional file 2: Fig. S7B and Additional file 9: Table S14). Next, we inoculated control (3HA) or *SOX11*-overexpressing DU145 cells expressing orthotopically in the ventral prostate lobe of immunocompromised nude mice. Of note, both cell lines expressed GFP and luciferase for in vivo bioluminescence monitoring. Remarkably, *SOX11*-overexpressing PC cells exhibited increased metastatic capacity, with higher colonization of lumbar lymph nodes and bones (Fig. 4D, E, Additional file 2: Fig. S7C). Of interest, the effects of *SOX11* overexpression on cell number in vitro were negligible in contrast with the effect on tumor mass in vivo (Additional file 2: Fig. S7D–F). These data are consistent with the proposed role for this transcription factor in the activation of stroma remodeling processes, which would result in the manifestation of the phenotype in the presence of stroma but not in isolation (in vitro cancer cell culture).

Taking advantage of this cellular model, we characterized transcriptomics alterations that emanate from *SOX11* overexpression by RNA-Seq. Functional enrichment analysis highlighted the regulation of stroma remodeling processes in *SOX11* overexpressing cells, consistent with our data in mHNPC (Fig. 4F, Additional file 2: Fig. S7G, Additional file 10: Table S15). Next, we generated a transcriptional signature of *SOX11* activity, composed of 55 genes positively regulated by the transcription factor ($|\log_2(\text{fold change})| > 1.5$ and $\text{FDR} < 0.05$). This gene set was strongly associated with mHNPC in both the discovery and validation patient cohorts, confirming increased *SOX11* activity in this aggressive form of PC (Fig. 4G). Our comprehensive transcriptional portrait of mHNPC reveals an unprecedented role for *SOX11* in PC dissemination and supports its contribution to the acquisition of metastatic features in this disease.

The evidence pointing at alterations and functional contribution of *SOX11* in prostate cancer is limited, and current reports present conflicting results [30, 31]. On the one hand, upregulation of this gene was associated with the acquisition of aggressive features in prostate tumors after therapy, which would be conceptually aligned with

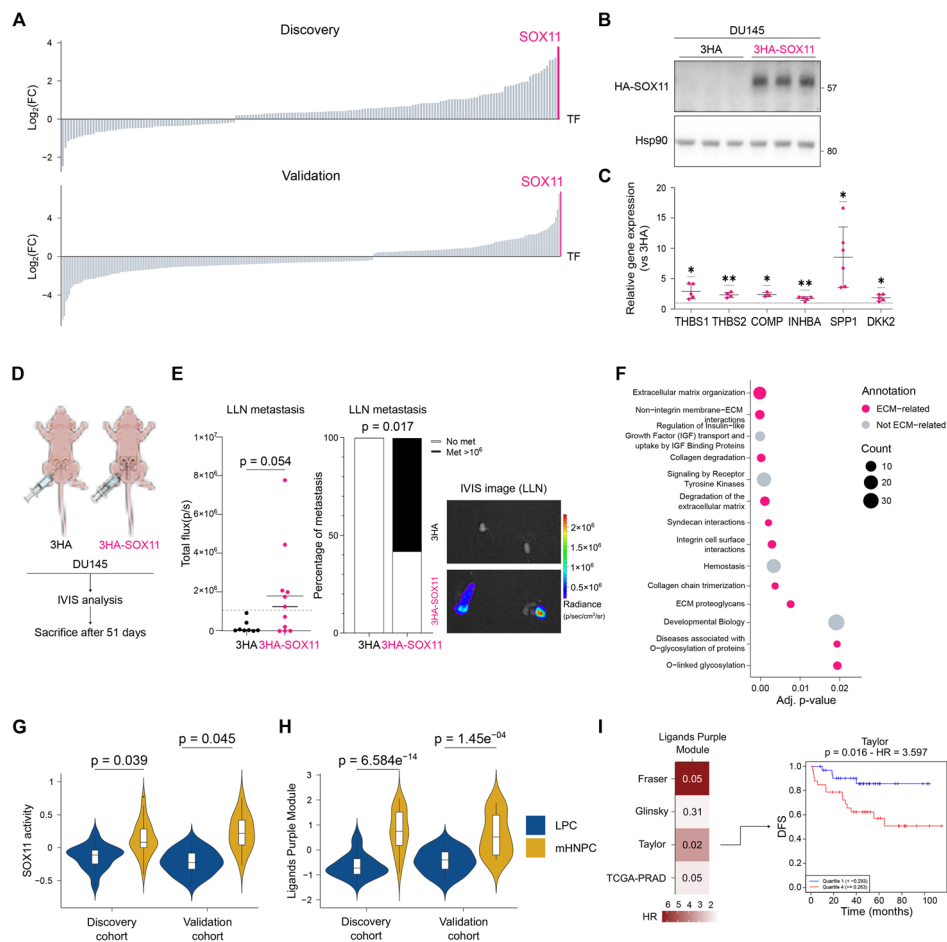


Fig. 4 Proof-of-concept study of *SOX11* as a metastatic driver in prostate cancer. **A** Waterfall plot representing the differential expression of transcription factors in mHNPc vs. LPC bulk RNA-Seq in discovery and validation cohorts. The position of *SOX11* is highlighted. **B** Representative western blot in DU145 cells showing ectopic 3HA-SOX11 expression ($n = 3$). **C** Evaluation of mRNA abundance in the indicated genes by quantitative real-time PCR upon ectopic 3HA-SOX11 expression ($n = 3-6$). One sample *t*-test. **D** Experimental design of the orthotopic injection of DU145 cells transduced with control and 3HA-SOX11-overexpressing lentiviral vectors into nude mice ($n = 5$ for 3HA, $n = 6$ for 3HA-SOX11). **E** Left, dot plot of lumbar lymph node (LLN) metastatic burden (day 51). One-tailed, Mann-Whitney test. Middle, stacked bar plot of LLN metastasis appearance. One-tailed, Fisher test. Right, representative bioluminescence images of LLN metastasis. **F** REACTOME pathway enrichment analyses of the genes upregulated upon *SOX11* overexpression. **G** Violin plots showing the levels of a *SOX11*-activity signature in the discovery and validation bulk RNA-Seq datasets. *P*-values from Wilcoxon test (two-tailed test for discovery and one-tailed for validation datasets, respectively). **H** Violin plots showing the levels of a signature composed of genes encoding for purple module ligands identified in single cell RNA-Seq and correlated with *SOX11* in the discovery and validation bulk RNA-Seq datasets. *P*-values from Wilcoxon test (two-tailed test for discovery and one-tailed for validation datasets, respectively). **I** Hazard ratio (HR) and *p*-value of a signature based on the ligands used in **H** when analyzing as endpoint biochemical recurrence in the indicated datasets and comparing patients within quartiles 1 and 4. A representative Kaplan-Meier curve is shown

our observation [30]. On the other hand, *SOX11* was postulated as a tumor suppressor based on the reduction in its expression in prostate cancer localized disease compared to benign prostate specimens [31]. Taking advantage of the availability of human prostate cancer transcriptional datasets [3, 15, 32–35], we decided to explore the expression of *SOX11* in different pathological scenarios including localized prostate cancer

and adjacent normal tissue. Analysis of four independent patient cohorts did not show a reduction in SOX11 expression in any tumor condition, while we did observe a trend towards upregulation in metastatic specimens, which would be conceptually aligned with our findings (Additional file 2: Fig. S7H).

Then, we sought to monitor the influence of SOX11 levels or activity in prostate cancer biochemical recurrence after first line therapy. We analyzed the aforementioned public transcriptomics datasets [3, 15, 32–35]. We did not observe a consistent association of *SOX11* expression (*SOX11* mRNA abundance) or SOX11 inferred activity (the aforementioned-gene signature) with biochemical recurrence-free survival in patients diagnosed with localized PC (Additional file 2: Fig. S7I, J). Our data suggest that *SOX11* presents a genuine role ascribed to the pathogenesis of mHNPC. We hypothesized that despite the differential contribution of *SOX11* to LPC and mHNPC, the ligands identified in mHNPC could be contributing to the acquisition of aggressiveness features in localized disease governed by other transcriptional processes. To test this notion, we monitored the expression of the ligands reported to contribute to the heterotypic cell–cell communication in mHNPC under the control of SOX11. A gene expression signature based on the 13 ligands identified robustly discriminated mHNPC from LPC (Fig. 4H). Strikingly, this signature exhibited robust and consistent prognostic potential in patients with localized prostate cancer. This predictive capacity for biochemical recurrence after surgery is indicative of the presence of early disseminated tumor cells with metastatic potential (Fig. 4I, Additional file 2: Fig. S7K).

Discussion

Prostate cancer (PC) research has been dominated by the analysis of the most prevalent form of the disease, which refers to localized, prostate-confined tumors (LPC). The scientific community has generated extensive clinical, biological, and molecular information about PC based on the study of LPC [1–3]. Similarly, the alterations and therapeutic vulnerabilities of metastatic PC have been established upon the study of heavily treated PC patients that ultimately develop metastasis. Whereas this strategy has been tremendously valuable in the establishment of clinical guidelines in this tumor type, the translation of this evidence to mHNPC is uncertain. Considering that mHNPC cases, albeit infrequent, represent a large fraction of the mortality by PC [4, 5], it is of the utmost importance to characterize the molecular features of this form of prostate tumor to establish therapeutic strategies tailored to this aggressive disease.

Bulk transcriptomics offers valuable insight about predominant differences among biological conditions. However, complex biological specimens can pose a challenge due to the influence of mRNAs coming from different cell types [19]. In addition, when comparing profoundly different tumors, differential expression can prove insufficient to identify relevant molecular processes. Our implementation of WGCA [18] together with the generation and integrated use of bulk and single-cell transcriptomics has enabled us to identify consistent stroma remodeling processes that can be illustrated by a set of SOX11-regulated ligands. On the one hand, the capacity of SOX11 to promote a metastatic phenotype suggests that the upregulation of this transcription factor contributes to the aggressiveness of mHNPC. On the other hand, the capacity of SOX11-regulated

secreted ligands identified in mHNPC to inform about the risk of biochemical recurrence (in the absence of overt SOX11 upregulation) suggests that the role of SOX11 in mHNPC might be exerted by other transcription factors in aggressive LPC.

Conclusions

This study provides the most comprehensive transcriptional portrait of mHNPC to date. The data generated and the various cohorts available will represent an invaluable resource to boost the molecular and biological deconstruction of mHNPC. We highlight the biological and molecular uniqueness of this aggressive form of PC and present proof-of-concept evidence of the value of this resource through the study of SOX11 as a transcription factor selectively activated in mHNPC that supports the acquisition of metastatic features.

Methods

Sample and patient disposition

Detailed information about the clinical and pathological characteristics of the cohorts included in this study are shown in Additional file 1: Table S1. Samples were labeled as LPC and mHNPC groups according to the patient metastatic status at the time of first diagnosis (M0, no distant metastasis present; M1, distant metastasis present, according to AJCC TNM staging system), upon review of electronic patient records, and based on standard of care imaging and clinical evaluations. The samples included in the discovery cohort for bulk RNA-Seq profiling (formalin fixed paraffin embedded specimens from primary tumor biopsies) were collected from patients at Basurto University Hospital in Bilbao (Spain). Sample collection was coordinated by the Basque Biobank. Samples with at least 5 (out of 12) positive needle biopsies were selected to ensure sufficient tumoral material for analysis. Tumor-rich regions were selected for RNA extraction by the pathologist (A.U-O). The ethics approvals for this project are CEIC-E 14-14 and 19-20. The patient cohort for bulk RNA-Seq for validation purposes was generated at the Morales Meseguer University Hospital in Murcia (Spain). Fresh frozen OCT-embedded tissue specimens were obtained from prostatectomy for LPC patients and from needle biopsies for mHNPC patients. The ethics approval for this project was CEIC-HMM 1/18. The single-cell validation cohort was generated from patients recruited in Basurto University Hospital in Bilbao (LPC patients) and Vall d'Hebron University Hospital in Barcelona (mHNPC patients), under IRB-approved protocols CEIC-E 14-14 and 19-20, and PRAG5248 respectively. The work with human samples was performed with informed consent and complies with the Helsinki Declaration (World Medical Association, Declaration of Helsinki: Recommendations Guiding Physicians in Biomedical Research Involving Human Subjects. Adopted by the 18th World Medical Assembly, Helsinki, Finland, June 1964).

Bulk RNA-Seq analyses on discovery cohort

We performed RNA-Seq (paired-end 150 bp reads, 86 M reads per sample in average) from formalin-fixed paraffin-embedded (FFPE) specimens derived from primary tumors in patients presenting localized and metastatic prostate cancer at the time of diagnosis

(recruited at the Basurto University Hospital, Spain). Sequencing libraries were prepared using SMARTer Stranded Total RNA-seq Kit v2 – Pico Input Mammalian kit (Takara Bio USA, Cat.# 634,411) and following the user manual (Rev. 063017). Based on the library specifications and read length of 150 bp, reads were specifically trimmed for 3 specific nucleotides using cutadapt v3.5 software as follows. First, we removed the first 3 nucleotides of R2 (cutadapt -U 3 -pair-filter=any -minimum-length=30), then removed adapters from both reads using TruSeq adapters (cutadapt -b file:cutadapt_TrueSeq_CD_R1.fa -B file:cutadapt_TrueSeq_CD_R2.fa -minimum-length=30 -pair-filter=any) and finally removed the last 3 nucleotides in R1 (cutadapt -u -3 -q 10 -pair-filter=any -minimum-length=30). After quality control using fastQC, we mapped the reads to the human genome (Ensembl GRCh38.v94) using STAR [36] version=2.7.0e with parameters: – outFilterMultimapNmax 1 – twopassMode Basic.

We identified differentially expressed genes (DEGs) between the two tumor types using DESeq2 (V1.34.0) [37] and applying variance stabilizing transformation (VST). We pre-filtered genes with less than 10 counts per smallest group size. Given the lower RNA quality in FFPE samples we calculated the recommended DV200 parameter that quantifies the percentage of fragments over 200 nucleotides. 74 out of 76 samples in our discovery cohort passed the recommended threshold of DV200 > 30–50% (Additional file 1: Table S1). We used DV200 and age as covariates in the analyses. Cell composition analyses were performed using ESTIMATE [19] R package (v1.0.13). Variance partitioning was performed using variancePartition [38] R package (v1.24.1). Enrichment analyses were performed using gprofiler2 (v0.2.1) [39]. Gene fusions were detected with the Trinity CTAT Fusion workflow [13] using docker trinityctat/starfusion (CTAT library version Apr062020). First, we ran STAR-Fusion to identify candidate fusion transcripts based on discordant read alignments. The predicted gene fusions were 'in silico validated' using FusionInspector. WGCNA [18] was run using R package "wgcna" (v1.64.1 in docker mochar/wgcna). We used VST counts as input and automatic, one-step method with the following parameters: corType="bicor", deepSplit=2, networkType="signed".

Bulk RNA-Seq data analyses on validation cohort

We performed RNA-Seq (paired-end 150 bp reads, 46 M reads per sample in average) from snap frozen OCT primary tumors collected from patients at the Morales Meseguer University Hospital. Sequencing libraries were prepared using the TruSeq Stranded mRNA LibraryPrep kit (Illumina Inc. Cat. # 20,020,594) and TruSeq RNA CD Index Plate (Illumina Inc. Cat. # 20,019,792). Reads were trimmed for Illumina universal adapter using cutadapt: cutadapt -a "AGATCGGAAGAGCACACGTCTGAACTCCAGTCA" -A "AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT" -j 0 -q 10 -m 30). Mapping and differential expression analyses were performed as described above for the discovery cohort, using RIN values in DESeq2 (instead of DV200) as covariates and applying no filtering for minimum reads in this validation cohort.

Analyses on single-cell discovery cohort

We accessed single-cell RNA-Seq data (10X Genomics) from GEO with accession ID GSE141445. The data consisted of a processed count matrix on 13 patients with filtered mitochondrial genes. We used DoubletFinder (version 2.0.3) [40] to identify doublets.

The samples were SCTransformed by v2 regularization by Seurat [41]. To integrate the samples, we used 3000 anchor genes using the Canonical Correlation Analysis reduction method. We performed the linear dimensional reduction and the Uniform Manifold Approximation and Projection (UMAP) analysis with the top 13 PCs. The clustering was performed using a resolution of 0.8. Cluster markers were identified using the FindAllMarkers function, and cell types were annotated based on well-established markers: T cells (*CD3D*, *CD3E*, *CD3G*, and *PTPRC*), B cells (*CD79A*, *CD79B*, *IGKC*, and *MS4A1*), macrophages (*CD14*, *CD68*, *CSF1R*, *FCGR3A*, and *LYZ*), mast cells (*KIT*, *MS4A2*, *TPSAB1*, and *TPSB2*), fibroblasts (*COL1A1*, *COL1A2*, *COL3A1*, *DCN*, *RGS5*, and *ACTA2*), endothelial cells (*CDH5*, *ENG*, *PECAM1*, and *VWF*), epithelial cells (*AR*, *EPCAM*, *KRT5*, *KRT8*, *KRT14*) and cell cycling marker genes (*BIRC5* and *CENPF*). For downstream computational analyses, we excluded cases that were not representing prostate biopsies or that were obtained from patients after therapy exposure. These criteria led to the selection of 3 mHNPc and 8 LPC (we excluded the treated individual SC173 and lymph node SC172). The IDs of the included patients are provided in Additional file 2: Fig. S4C.

Differential expression analysis per cluster was performed using a pseudo-bulk approach with DESeq2 (version 1.36.0) using only the singlets (Benjamini–Hochberg FDR < 0.1). UCell (version 2.0.1) [42] was used for evaluating gene signature scores, based on the Mann–Whitney U statistic. CopyKAT [23] R package (version 1.1.0) was used to identify genome-wide aneuploidy at 5 MB resolution in single cells, non-epithelial clusters 9, 10, 11, 13, 14, 16, 17, 18, and 19 were used as reference. Cell–cell communication analyses were performed using NicheNet [25] R package (v1.1.1). For prioritization based on differential expression, we used Wilcoxon test on SCT counts (log fold change of 0.25 and expression pct of 0.1). The transcription factor list was obtained from <https://humantfs.ccb.utoronto.ca/download.php>. For NicheNet target analysis, we utilized the `construct_ligand_target_matrix` function to infer potential target genes of ligands. For each compartment, we selected the ligands that were actively communicating with the compartment of interest. This approach allowed us to construct a compartment-specific ligand-target matrix, facilitating the identification of key downstream targets influenced by ligand-receptor interactions within each compartment of the tumor microenvironment.

To assess compositional differences in cell populations across conditions, we employed scCODA (v0.1.9.) (single-cell Compositional Data Analysis) [43]. For the discovery dataset, cluster 20 was automatically chosen by the tool as the reference, while for the validation dataset, cluster 15 was selected. Model convergence was assessed using trace plots, and posterior inclusion probabilities ensured result robustness. The analyses were conducted using Python 3.9.

Generation of the validation single-cell RNA-Seq dataset

The tumor was washed with cold phosphate-buffered saline (PBS; GIBCO) and minced into fragments under 0.4 mm. The tissue was chemically dissociated in a filter-sterilized with Liberase TH solution (2.5 mg/ml; REF QZy-5401135001, Thermo Scientific) in complete media composed of Advanced DMEM/F-12 (REF 12–634–010, Fisher Scientific), Pen/strep 1% (REF 15–140,122, GIBCO), MgCl₂ (5 mM, REF M4880,

Sigma-Aldrich), DNase I (200 µg/ml, REF 11284932001, Sigma-Aldrich), and Y-27632 dihydrochloride (10 µM; AbMole) at 37 °C shaking at 800 r.p.m. for 40 min. The tissue solution was manually disaggregated with a pipette every 10 min during the incubation followed by centrifugation at $2400 \times g$ for 5 min. Next, a second chemical dissociation was performed with TrypLE (REF 12605010, Thermo Scientific) supplemented with Y-27632 dihydrochloride under constant pipetting for 2 min. The inactivation of TrypLE was done by adding fetal bovine serum (FBS, GIBCO) and centrifuged at $2400 \times g$ for 5 min. The cell pellet was resuspended in PBS, filtered through 40 µm filter, and washed with PBS. The suspension was centrifuged at $2400 \times g$ for 5 min and resuspended in 200 µl of PBS. Cells were counted on a Neubauer chamber and 20,000 cells were then centrifuged at $2400 \times g$ for 5 min and resuspended in PBS for scRNA-Seq analysis according to manufacturers' indications (targeted number of cells: 10,000).

We use the 10X Chromium Controller to encapsulate and barcode single cells (Reagent Kits v3.1). The sequenced data were mapped to human reference genome GRCh38 with Cell Ranger (version 7.0.1). Seurat package (version 4.3.0) in R (version 4.2.1) was used for downstream analyses. Low quality cells expressing less than 200 unique genes, showing novelty score < 80% (the ratio of number of genes over number of UMIs) or > 20% of mitochondrial genes were excluded. Only genes expressed in more than 10 cells were shortlisted for subsequent analyses. We integrated and normalized data as explained above. For this dataset, during the normalization step we regressed out the mitochondrial percentage and we used a resolution of 0.6. Cell-type annotation was based on the following markers: T cells (*CD3D*, *CD3E*, *CD3G*, and *PTPRC*), B cells (*CD79A*, *CD79B*, *IGKC*, and *MS4A1*), macrophages (*CD14*, *CD68*, *CSF1R*, *FCGR3A*, and *LYZ*), mast cells (*KIT*, *MS4A2*, *TPSAB1*, and *TPSB2*), fibroblasts (*COL1A1*, *COL1A2*, *COL3A1*, *DCN*, *RGS5*, and *ACTA2*), endothelial cells (*CDH5*, *ENG*, *PECAM1*, and *VWF*), epithelial cells (*AR*, *EPCAM*, *KRT5*, *KRT8*, and *KRT14*), and Schwann cells (*S100B*, *NRXN1*, and *SOX10*). CopyKAT R package (version 1.1.0) was used to identify genome-wide aneuploidy at 5 MB resolution in single cells, non-epithelial clusters 0, 1, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 16, 17, 18, 19, and 20 were used as reference.

Enrichment analyses

We performed functional enrichment analyses using g:Profiler (<https://biit.cs.ut.ee/gprofiler/gost>) to identify overrepresented biological processes, molecular functions, and pathways. For each analysis, we applied multiple testing correction using the g:SCS algorithm. For background correction, we carefully selected the gene sets to ensure appropriate comparisons. For bulk RNA-Seq analysis, the background gene set included all expressed genes with a valid gene symbol. For module-based enrichments, we used all genes assigned to co-expression modules as background. Finally, for single-cell RNA-Seq analysis, the background gene set comprised all genes expressed in the specific cluster or compartment being analyzed. Gene Set Enrichment Analysis (GSEA) was performed using the GSEAPy (v1.0.4) library on single-cell RNA-Sequencing data. Genes with more than 15 counts were included in the analysis. Enrichment was assessed using 1000 permutations with `permutation_type = "gene_set"`. Pathways with at least 3 genes were retained and considered significant if they had a false discovery rate (FDR) < 0.25.

Cell culture

The human prostate carcinoma cell lines used were purchased from Leibniz-Institut DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH), which provided an authentication certificate: DU145 (ACC261). Human embryonic kidney 293FT cells were generously provided by the laboratory of Dr. Rosa Barrio. Cell lines were subjected to microsatellite-based identity validation. Cell lines were tested for mycoplasma contamination routinely using MycoAlert detection Kit (Lonza; LT07-318). DU145 and HEK 293FT were cultured in Dulbecco's modified Eagle's medium (DMEM, Gibco) supplemented with 10% (v/v) FBS (Gibco) and 1% (v/v) penicillin–streptomycin (Gibco). Cells were maintained at 37 °C and 5% CO₂ in a humidified atmosphere. Possible mycoplasma contamination was routinely monitored across all cell lines using the MycoAlert detection Kit (Lonza; LT07-318).

Generation of stable cell lines

For constitutive *SOX11* overexpression (3HA-SOX11) previously generated DU145 GFP-Luc cells were infected with a modified Lenti-EFS/P2A-blast plasmid donated by Dr. James D. Sutherland (Addgene: #208,041) in which the TURBO-ID region was substituted by a 3HA-tag and the murine form of *Sox11* (obtained from Addgene: #120,387) was cloned after the tag (between restriction sites for EcoRI and BamHI). The same plasmid without the *SOX11* insert was used as a control (3HA). The infection was performed using standard procedures: 293FT cells were transfected with the appropriate lentiviral vectors using the calcium phosphate method and the viral supernatant plus protamine sulfate (8 µg/ml) were used to infect the DU145-GFP-luc cells. Cells were selected with blasticidin (10 µg/ml) for 5 days (blasticidin was renewed after the first 3 days) [44].

Cellular and molecular assays

Proliferation assays were performed by plating 5000 DU145 cells in triplicate in 12-well dishes and fixing in formalin at the indicated time points. Cells were stained with crystal violet as previously described [45] and quantified after the resuspension of the crystals in 10% acetic acid by measuring the absorbance at 595 nm.

Protein extraction and Western blot were performed as previously described [46]. Briefly, samples were run in 4–12% gradient Nupage precast gels (Life Technologies, WG1403BX10) in MOPS buffer. Primary antibodies for HA (Cell Signalling Technologies, 3724), SOX11 (Cell Signalling Technologies, 58,207), and HSP90 (Cell Signalling Technologies, 4874S) were used at a 1:1000 dilution. Secondary anti-rabbit antibody (Jackson ImmunoResearch, 111–035-144) was used at a 1:4000 dilution.

RNA was automatically extracted using the Maxwell RSC (Promega) platform and extraction kits (Promega, AS1390), following the manufacturer's instructions. Complementary DNA (cDNA) from 1000 ng of the extracted RNA was synthesized with Maxima[™] H Minus cDNA Synthesis Master Mix (Invitrogen Ref: M1682). Synthesized cDNA was diluted 1:3 and QS5 or QS6 (Life Technologies) systems were used for performing RT-qPCR analyses. Gene expression was normalized to GAPDH expression. Applied Biosystems TaqMan probes or primers with their corresponding Universal Probe Library (Roche) probes were used: mSOX11 (Probe 11, F: ACAACGCCG

AGATCTCCAAG, R: TGAACGGGATCTTCTCGCTG), *THBS1* (Hs00962908_m1), *THBS2* (Hs01568063_m1), *COMP* (Hs00164359_m1), *INHBA* (Hs01081598_m1), *DKK2* (Hs00205294_m1), *SPP1* (Hs00959010_m1), *FN1* (Probe 25, F: GGGAGAATAAGC TGTACCATCG, R: TCCATTACCAAGACACACACT), *MEX3A* (Hs00863536_m1), *GAPDH* (Hs02758991g1).

Prostate orthotopic xenograft model of metastasis

The procedures for animal experimentation were carried out in compliance with the ethical guidelines defined by the Biosafety and Animal Welfare Committee at CIC bio-GUNE, following AAALAC recommendations. Mice were maintained in a controlled environment, with standard 12:12 light:dark cycles, 30–50% humidity, and controlled temperature at 22 ± 2 °C. Diet and water were provided ad libitum. At the experimental endpoint, mice were sacrificed by CO₂ inhalation followed by cervical dislocation.

Orthotopic models were generated by injection of either DU145 GFP-Luc 3HA-ctrl or DU145 GFP-Luc 3HA-SOX11 into the ventral lobe of the prostate of nude mice. 2×10^6 cells were injected per mice in 50 µl of PBS:Matrigel (Corning, 356,231) (70:30) mixture. Mice were imaged by IVIS right after the surgery to check that the injection was properly performed as well as to check for accidental dissemination (day 0) and were subsequently monitored by IVIS on a weekly basis. Mice were sacrificed after 51 days and prostate cancer cell dissemination to distant organs was assessed by ex vivo monitoring of selected organs by IVIS. Due to challenges in in vivo handling, including hemorrhages, the recovery of the LLNs is sometimes compromised. As a result, in this study, we were able to recover 9 LLNs from the control group and 12 from the SOX11 group. Furthermore, following an outlier analysis using Grubbs' test ($\alpha = 0.05$), we excluded one LLN from each group, resulting in a final count of 8 LLNs in the control group and 11 in the SOX11 group.

Bulk RNA-Seq data analyses on SOX11 overexpression

We performed RNA-Seq (paired-end 100 bp reads, >43 M reads per sample in average) from DU145 cells. Sequencing libraries were prepared using the TruSeq Stranded mRNA LibraryPrep kit (Illumina Inc. Cat. # 20,020,594) and TruSeq RNA CD Index Plate (Illumina Inc. Cat. # 20,019,792). Reads were trimmed for Illumina universal adapter using cutadapt: cutadapt -a "AGATCGGAAGAGCACACGTCTGAACTCCAGTCA" -A "AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT" -j 0 -q 10 -m 30). Mapping and differential expression analyses were performed as described above, applying the filtering for minimum reads but setting independent filtering parameter from DESeq2 to false. The signature for SOX11 activity (comprising 55 upregulated genes, 53 of which had gene symbol) was obtained by doing the average of the z-score values of these genes per sample.

Proteomic analysis

Sample preparation

Protein was extracted by incubating cells in a buffer containing 7 M urea, 2 M thio-urea, and 4% CHAPS. Samples were incubated in this buffer for 30 min at RT under

agitation and digested following the FASP protocol described by Wisniewski et al. 2009 with minor modifications (PMID: 19,377,485). Trypsin was added in 50 mM ammonium bicarbonate to a trypsin:protein ratio of 1:10, and the mixture was incubated for overnight at 37 °C. Peptides were dried out in an RVC2 25 speedvac concentrator (Christ) and resuspended in 0.1% FA. Peptides were desalted and resuspended in 0.1% FA using C18 stage tips (Millipore) prior to acquisition.

Mass spectrometry analysis

The resulting peptides were loaded onto an EvoSep One (EvoSep) chromatograph coupled on-line to a TIMS ToF Pro mass spectrometer (Bruker) that uses Parallel Accumulation Serial Fragmentation (PASEF) acquisition to provide extremely high speed and sensitivity. Thirty SPD protocol (approx. 44 min. runs) was used, under default Evosep settings. Data-independent acquisition (DIA) was used for the acquisition of data.

Protein identification and quantification

The obtained data was then processed with DIA-NN [47] software for protein identification and quantification. Searches were carried out against a database consisting of human protein entries from Uniprot in library-free mode. Search parameters were 20 ppm precursor and fragment tolerance, 0.05 carbamidomethylation of cysteines as fixed modification and oxidation of methionines as variable modification.

Quantitative protein data was loaded onto Perseus software (free software from Max Plank Institute, Munich) [48]. This program was used for the differential protein abundance analyses. For this purpose, protein abundance data was $\log_2$ transformed, filtered based on reproducibility (proteins present in at least 70% of the samples of one of the groups were kept in the analysis) and imputed (missing values were substituted by abundances randomly taken from the 10% least abundant proteins in each sample). A Student's *T*-test was applied, and proteins with a $p < 0.05$ were considered as significantly differential.

Statistical analysis on in vitro and in vivo experimental data

Sample size was not predetermined using any statistical method and experiments were not randomized. Investigators were not blinded during experiments or outcome assessment. All the experiments were performed with at least three biological replicates. *N* values represent the number of independent biological experiments, the number of individual mice, or the number of patient samples.

For in vitro experiments, a one-sample *t*-test was used to compare the values normalized to the control with a hypothetical value of 1, and results are presented as mean $\pm$ standard deviation. For in vivo experiments, a one-tailed Mann–Whitney test was used, and results are presented either as the mean in a dot plot or as a stacked bar plot where a one-tailed Fisher-test analysis was applied in the contingency table. The confidence level used for all statistical analyses was 95% (p -value = 0.05). Outliers were determined by an interval spanning over the mean plus/minus two standard deviations.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-025-03623-5>.

Additional file 1: Table S1.
 Additional file 2: Figures S1–S7.
 Additional file 3: Table S2–S7.
 Additional file 4: Table S8.
 Additional file 5: Table S9.
 Additional file 6: Table S10, S11.
 Additional file 7: Table S12.
 Additional file 8: Table S13.
 Additional file 9: Table S14.
 Additional file 10: Table S15.
 Additional file 11.

Acknowledgements

We thank Monika Gonzalez, Laura Bárcena and Nuria Macias-Cámara from the Genome Platform Analyses at the CIC bioGUNE and Lidia López Jiménez from the single-cell Unit at the Josep Carreras Institute for generation of single-cell RNA-Seq libraries. We thank James D. Sutherland for the donation of the overexpression plasmid. We thank Felix Elortza and Mikel Azkargorta from CIC bioGUNE proteomics platform for proteomics sample processing and analysis. We thank Edurne Berra, Amaia Ercilla and Amaia Zabala for the critical comments on the manuscript. Illustrations were created with BioRender. We thank the Basque Biobank for Research (BIOEF) for the support with clinical sample and data management.

Authors' contributions

S.G.-L., U.L. and I.M. performed bioinformatic analyses on bulk and single-cell transcriptomic datasets and contributed to the preparation of the figures for the manuscript. I.M. supervised the work of S.G.-L. and U.L. J.C.-M. and N.M.-M. performed most in vitro and in vivo assays, performed the data analysis and contributed to the preparation of the figures of the manuscript. N.M.-M. supervised the work of J.C.-M. I.A., M.P.-V., O.C., and M.T.B. provided technical advice with the in vivo experiments. A.M.A. performed wet-lab procedures on single-cell RNA-Seq and provided technical support with bulk RNA sequencing. R.R.G. provided guidance and training on the in vivo experiments. D.G., S.R., A.S.-M., A.U.-O., M.U. and A.L.-I. generated the Basurto cohort and provided biological specimens. E.G.-B., A.R., J.T., D.J. and A.M. performed patient selection, tissue obtention and clinical annotation of the Morales Meseguer cohort. E.G.-B. supervised A.R., J.T., D.J. and A.M. N.M.-M., J.C.-M., L.B.-B., N.H., A.S. and H.V.S. performed the tissue preparation for single-cell experiments. J.M. recruited mHNPC patients in the single-cell validation cohort. M.G. supervised the single-cell preparation. A.C. conceived the study. I.M. and A.C. supervised the execution of the project and wrote the manuscript. All authors have read and approved the final version of the manuscript.

Peer review information

Andrew Cosgrove was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Funding

This study was predominantly funded by the Spanish Association Against Cancer (AECC, GCTRA18006CARR to Carracedo, Gomis, Unda, Graupera and González-Billalabeitia as PIs) and AstraZeneca Jóvenes Investigadores 2023 Award (To Carracedo, Mateo, Mendizabal and Herranz). The work of A. Carracedo is supported by the Basque Department of Industry, Tourism, and Trade (Elkartek), the BBVA foundation (Becas Leonardo), the MICINN (PID2022-141553OB-I0 (FEDER/EU), Fundación Cris Contra el Cáncer (PR_EX_2021-22), Severo Ochoa Excellence Accreditation (CEX2021-001136-S), European Training Networks Project (H2020-MSCA-ITN-2020 955534), the Fundación Jesús Serra, iDIFFER network of Excellence (RED2022-134792-T), and the European Research Council (Consolidator Grant 819242). CIBERONC was co-funded with FEDER funds and funded by ISCIII. I. Mendizabal is supported by CRIS Contra El Cancer Foundation (PR_TPD_2020-19) and a Ramón y Cajal contract (RYC2023-044682-I) funded by the MCIN. U. Lazcano is supported by the AECC Foundation (PRDVZ245829LAZC). L. Bozal was supported by the AECC Foundation (POSTD19048BOZA). H. van Splunder received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 955951. M. Graupera is supported by Worldwide Cancer Research (WCR 21-0159). R. Gomis and M.T. Blasco were supported by the BBVA Foundation, Fundación Científica AECC (PRYGN223207GOMI), and MICINN (PID2022-143093OB-I00; FEDER/EU). J. Mateo is supported by CRIS Talent Award (TALENT20-10) and a Department of Defense CDMPR Physician-Science Award (PC220307). VHIO authors would like to acknowledge the Spanish State Agency for Research (Agencia Estatal de Investigación) for the financial support as a Center of Excellence Severo Ochoa (CEX2020-001024-S/AEI/<https://doi.org/10.13039/501100011033>), the Cellex Foundation for providing research facilities and equipment, FERO Foundation, and the CERCA Programme from Generalitat de Catalunya for their support.

Data availability

All data generated in this project (raw and processed) are available in GEO and SRA including bulk discovery (GSE268308) [49], bulk validation (GSE268309) [50], single-cell RNA-Seq (GSE268307) [51], and the DU145 SOX11-overexpression bulk

dataset (GSE268408) [52]. Proteomics data is available in ProteomeXchange (PXD063665) [53]. Code for data analyses are publicly available in GitHub [54]. Scripts are also uploaded to Zenodo [55]. Creative Commons Attribution 4.0 International license is assigned to the GitHub and Zenodo scripts.

Declarations

Ethics approval and consent to participate

The ethics approvals for this project are CEIC-E 14–14 and 19–20 (Basurto University Hospital, Bilbao), CEIC-HMM 1/18 (Morales Meseguer University Hospital, Murcia) and PRAG5248 (Vall d'Hebron University Hospital, Barcelona, under IRB-approval). See Methods section for more details. Animal experiments have been approved by CIC bioGUNE Ethics committee and the local authorities with codes P-CBG-CBBA-0121 and P-CBG-CBBA-1321.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Center for Cooperative Research in Biosciences (CIC bioGUNE), Basque Research and Technology Alliance (BRTA), Bizkaia, Spain. ²Centro de Investigación Biomédica En Red de Cáncer (CIBERONC), 28029 Madrid, Spain. ³Traslational Prostate Cancer Research Lab, CIC bioGUNE-Basurto, Biobizkaia Health Research Institute Technology Park, Building 801A, 48160 Derio, Spain. ⁴Vall Hebron Institute of Oncology, Hebron University Hospital Campus, Vall d'Barcelona, Spain. ⁵Endothelial Pathobiology and Microenvironment Group, Josep Carreras Leukemia Research Institute (IJC), 08916 Badalona, Barcelona, Catalonia, Spain. ⁶Cancer Science Program, Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain. ⁷Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Instituto de Salud Carlos III, 28029 Madrid, Spain. ⁸Department of Urology, Morales Meseguer University Hospital, Murcia, Spain. ⁹Department of Pathology, Morales Meseguer University Hospital, Murcia, Spain. ¹⁰Biobanco Nodo 3, Morales Meseguer University Hospital, Murcia, Spain. ¹¹Department of Urology, Basurto University Hospital, 48013 Bilbao, Spain. ¹²Department of Pathology, Basurto University Hospital, 48013 Bilbao, Spain. ¹³Institución Catalana de Investigación y Estudios Avanzados (ICREA), Passeig de Lluís Companys 23, Barcelona, Spain. ¹⁴Medical Oncology Department, University Hospital, UCAM, 12 de Octubre, IMAS12 Madrid, Spain. ¹⁵Ikerbasque, Basque Foundation for Science, Bilbao, Spain. ¹⁶Biochemistry and Molecular Biology Department, University of the Basque Country (UPV/EHU), Bilbao, Spain.

Received: 7 August 2024 Accepted: 16 May 2025

Published online: 03 June 2025

References

1. Cancer Genome Atlas Research N. The molecular taxonomy of primary prostate cancer. *Cell*. 2015;163:1011–1025.
2. Fraser M, Sabelnikova VY, Yamaguchi TN, Heisler LE, Livingstone J, Huang V, et al. Genomic hallmarks of localized, non-indolent prostate cancer. *Nature*. 2017;541:359–64.
3. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell*. 2010;18:11–22.
4. Warner EW, Van der Eecken K, Murtha AJ, Kwan EM, Herberts C, Sipola J, et al. Multiregion sampling of de novo metastatic prostate cancer reveals complex polyclonality and augments clinical genotyping. *Nat Cancer*. 2024;5:114–30.
5. Helgstrand JT, Roder MA, Klemann N, Toft BG, Lichtensztajn DY, Brooks JD, et al. Trends in incidence and 5-year mortality in men with newly diagnosed, metastatic prostate cancer—a population-based analysis of 2 national cohorts. *Cancer*. 2018;124:2931–8.
6. Gilson C, Ingleby F, Gilbert DC, Parry MA, Atako NB, Ali A, et al. Genomic profiles of de novo high- and low-volume metastatic prostate cancer: results from a 2-stage feasibility and prevalence study in the STAMPEDE Trial. *JCO Precis Oncol*. 2020;4:882–97.
7. Velez MG, Kosiorek HE, Egan JB, McNatty AL, Riaz IB, Hwang SR, et al. Differential impact of tumor suppressor gene (TP53, PTEN, RB1) alterations and treatment outcomes in metastatic, hormone-sensitive prostate cancer. *Prostate Cancer Prostatic Dis*. 2022;25:479–83.
8. Abida W, Armenia J, Gopalan A, Brennan R, Walsh M, Barron D, et al. Prospective genomic profiling of prostate cancer across disease states reveals germline and somatic alterations that may affect clinical decision making. *JCO Precis Oncol*. 2017;2017:PO.17.00029.
9. Hamid AA, Gray KP, Shaw G, MacConaill LE, Evan C, Bernard B, et al. Compound genomic alterations of TP53, PTEN, and RB1 tumor suppressors in localized and metastatic prostate cancer. *Eur Urol*. 2019;76:89–97.
10. Stopsack KH, Nandakumar S, Wibmer AG, Haywood S, Weg ES, Barnett ES, et al. Oncogenic genomic alterations, clinical phenotypes, and outcomes in metastatic castration-sensitive prostate cancer. *Clin Cancer Res*. 2020;26:3230–8.
11. Hamid AA, Huang HC, Wang V, Chen YH, Feng F, Den R, et al. Transcriptional profiling of primary prostate tumor in metastatic hormone-sensitive prostate cancer and association with clinical outcomes: correlative analysis of the E3805 CHAARTED trial. *Ann Oncol*. 2021;32:1157–66.
12. van Leenders G, van der Kwast TH, Grignon DJ, Evans AJ, Kristiansen G, Kweldam CF, et al: The 2019 International Society of Urological Pathology (ISUP) Consensus Conference on Grading of Prostatic Carcinoma. *Am J Surg Pathol*. 2020;44:e87–e99.

13. Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol.* 2019;20:213.
14. Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, et al. The genomic complexity of primary human prostate cancer. *Nature.* 2011;470:214–20.
15. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science.* 2005;310:644–8.
16. Hermans KG, van Marion R, van Dekken H, Jenster G, van Weerden WM, Trapman J. TMPRSS2:ERG fusion by translocation or interstitial deletion is highly relevant in androgen-dependent prostate cancer, but is bypassed in late-stage androgen receptor-negative prostate cancer. *Cancer Res.* 2006;66:10658–63.
17. Finianos A, Gupta K, Clark B, Simmens SJ, Aragon-Ching JB. Characterization of differences between prostate cancer patients presenting with de novo versus primary progressive metastatic disease. *Clin Genitourin Cancer.* 2017;15:30247–1.
18. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
19. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4:2612.
20. Ugalde-Olano A, Egia A, Fernandez-Ruiz S, Loizaga-Iriarte A, Zuniga-Garcia P, Garcia S, et al. Methodological aspects of the molecular and histological study of prostate cancer: focus on PTEN. *Methods.* 2015;77–78:25–30.
21. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun.* 2015;6:8971.
22. Chen S, Zhu G, Yang Y, Wang F, Xiao YT, Zhang N, et al. Single-cell analysis reveals transcriptomic remodellings in distinct cell types that contribute to human prostate cancer progression. *Nat Cell Biol.* 2021;23:87–98.
23. Gao R, Bai S, Henderson YC, Lin Y, Schalck A, Yan Y, et al. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat Biotechnol.* 2021;39:599–608.
24. Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, et al. Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell.* 2018;33(676–689):e673.
25. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods.* 2020;17:159–62.
26. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. *Cell.* 2018;172:650–65.
27. Sarkar A, Hochedlinger K. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell Stem Cell.* 2013;12:15–30.
28. Dy P, Penzo-Mendez A, Wang H, Pedraza CE, Macklin WB, Lefebvre V. The three SoxC proteins—Sox4, Sox11 and Sox12—exhibit overlapping expression patterns and molecular properties. *Nucleic Acids Res.* 2008;36:3101–17.
29. Tsang SM, Oliemuller E, Howard BA. Regulatory roles for SOX11 in development, stem cells and cancer. *Semin Cancer Biol.* 2020;67:3–11.
30. Zou M, Toivanen R, Mitrofanova A, Floch N, Hayati S, Sun Y, et al. Transdifferentiation as a Mechanism of Treatment Resistance in a Mouse Model of Castration-Resistant Prostate Cancer. *Cancer Discov.* 2017;7:736–49.
31. Yao Z, Sun B, Hong Q, Yan J, Mu D, Li J, et al. The role of tumor suppressor gene SOX11 in prostate cancer. *Tumour Biol.* 2015;36:6133–8.
32. Cortazar AR, Torrano V, Martin-Martin N, Caro-Maldonado A, Camacho L, Hermanova I, et al. CANCEERTOOL: A Visualization and Representation Interface to Exploit Cancer Datasets. *Cancer Res.* 2018;78:6320–8.
33. Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature.* 2012;487:239–43.
34. Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, Montgomery K, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci U S A.* 2004;101:811–6.
35. Tomlins SA, Mehra R, Rhodes DR, Cao X, Wang L, Dhanasekaran SM, et al. Integrative molecular concept modeling of prostate cancer progression. *Nat Genet.* 2007;39:41–51.
36. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29:15–21.
37. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550.
38. Hoffman GE, Schadt EE. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics.* 2016;17:483.
39. Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res.* 2020;9:ELIXIR-709.
40. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* 2019;8(329–337):e324.
41. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell.* 2021;184(3573–3587):e3529.
42. Andreatta M, Carmona SJ. UCell: Robust and scalable single-cell gene signature scoring. *Comput Struct Biotechnol J.* 2021;19:3796–8.
43. Büttner M, Ostner J, Müller CL, Theis FJ, Schubert B. scCODA is a Bayesian model for compositional single-cell data analysis. *Nat Commun.* 2021;12:6876.
44. Barroso-Gomila O, Trulsson F, Muratore V, Canosa I, Merino-Cacho L, Cortazar AR, et al. Identification of proximal SUMO-dependent interactors using SUMO-ID. *Nat Commun.* 2021;12:6671.
45. Torrano V, Valcarcel-Jimenez L, Cortazar AR, Liu X, Urošević J, Castillo-Martin M, et al. The metabolic co-regulator PGC1alpha suppresses prostate cancer metastasis. *Nat Cell Biol.* 2016;18:645–56.
46. Zabala-Letona A, Arruabarrena-Aristorena A, Martin-Martin N, Fernandez-Ruiz S, Sutherland JD, Clasquin M, et al. mTORC1-dependent AMD1 regulation sustains polyamine metabolism in prostate cancer. *Nature.* 2017;547:109–13.
47. Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods.* 2020;17:41–4.

48. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*. 2016;13:731–40.
49. Martin-Martin N, Garcia-Longarte S, Corres-Mendizabal J, Lazcano U, Astobiza I, Bozal-Basterra L, et al. Transcriptional analysis of metastatic hormone-naïve prostate cancer primary tumour biopsies reveals a relevant role for SOX11 in prostate cancer cell dissemination [Discovery dataset]. 2025. GSE268308. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE268308>.
50. Martin-Martin N, Garcia-Longarte S, Corres-Mendizabal J, Lazcano U, Astobiza I, Bozal-Basterra L, et al. Transcriptional analysis of metastatic hormone-naïve prostate cancer primary tumour biopsies reveals a relevant role for SOX11 in prostate cancer cell dissemination [Validation dataset]. 2025. GSE268309. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE268309>.
51. Martin-Martin N, Garcia-Longarte S, Corres-Mendizabal J, Lazcano U, Astobiza I, Bozal-Basterra L, et al. Transcriptional analysis of metastatic hormone-naïve prostate cancer primary tumour biopsies reveals a relevant role for SOX11 in prostate cancer cell dissemination [scRNA-Seq]. 2025. GSE268307. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE268307>.
52. Martin-Martin N, Garcia-Longarte S, Corres-Mendizabal J, Lazcano U, Astobiza I, Bozal-Basterra L, et al. Transcriptional analysis of metastatic hormone-naïve prostate cancer primary tumour biopsies reveals a relevant role for SOX11 in prostate cancer cell dissemination [DU145-SOX11 overexpression_dataset]. 2025. GSE268408. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE268408>.
53. Martin-Martin N, Garcia-Longarte S, Corres-Mendizabal J, Lazcano U, Astobiza I, Bozal-Basterra L, et al. Transcriptional analysis of metastatic hormone-naïve prostate cancer primary tumour biopsies reveals a relevant role for SOX11 in prostate cancer cell dissemination [DU145-SOX11 overexpression_Proteomics_dataset]. 2025. PXD063665. <https://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD063665>.
54. Martin-Martin N, Garcia-Longarte S, Corres-Mendizabal J, Lazcano U, Astobiza I, Bozal-Basterra L, et al. Transcriptional analysis of metastatic hormone-naïve prostate cancer primary tumour biopsies reveals a relevant role for SOX11 in prostate cancer cell dissemination [Script repository]. 2025. Github. https://github.com/imendizabalCIC/Transcriptional_Landscape_mHNPC.
55. Martin-Martin N, Garcia-Longarte S, Corres-Mendizabal J, Lazcano U, Astobiza I, Bozal-Basterra L, et al. Transcriptional analysis of metastatic hormone-naïve prostate cancer primary tumour biopsies reveals a relevant role for SOX11 in prostate cancer cell dissemination [Script repository]. 2025. Zenodo. <https://zenodo.org/records/15356690>.

Publisher's Note

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