

# Circulating tumor DNA profile recognizes transformation to castration-resistant neuroendocrine prostate cancer

Himisha Beltran,<sup>1,2</sup> Alessandro Romanel,<sup>3</sup> Vincenza Conteduca,<sup>1,4</sup> Nicola Casiraghi,<sup>3</sup> Michael Sigouros,<sup>2</sup> Gian Marco Franceschini,<sup>3</sup> Francesco Orlando,<sup>3</sup> Tarcisio Fedrizzi,<sup>3</sup> Sheng-Yu Ku,<sup>1</sup> Emma Dann,<sup>3</sup> Alicia Alonso,<sup>5</sup> Juan Miguel Mosquera,<sup>5,6</sup> Andrea Sboner,<sup>5,7</sup> Jenny Xiang,<sup>5</sup> Olivier Elemento,<sup>5,7</sup> David M. Nanus,<sup>2,5</sup> Scott T. Tagawa,<sup>2,5</sup> Matteo Benelli,<sup>3,8</sup> and Francesca Demichelis<sup>3,5,7</sup>

<sup>1</sup>Department of Medical Oncology, Dana Farber Cancer Institute, Boston, Massachusetts, USA. <sup>2</sup>Department of Medicine, Division of Medical Oncology, Weill Cornell Medicine, New York, New York, USA.

<sup>3</sup>Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Trento, Italy. <sup>4</sup>Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori (IRST) IRCCS, Meldola, Italy. <sup>5</sup>Caryl and Israel Englander Institute for Precision Medicine, New York Presbyterian Hospital, Weill Cornell Medicine, New York, New York, USA. <sup>6</sup>Department of Pathology and Laboratory Medicine, and <sup>7</sup>Institute for Computational Biomedicine, Weill Cornell Medicine, New York, New York, USA. <sup>8</sup>Bioinformatics Unit, Hospital of Prato, Prato, Italy.

**Loss of androgen receptor (AR) signaling dependence occurs in approximately 15%–20% of advanced treatment-resistant prostate cancers, and this may manifest clinically as transformation from a prostate adenocarcinoma histology to a castration-resistant neuroendocrine prostate cancer (CRPC-NE). The diagnosis of CRPC-NE currently relies on a metastatic tumor biopsy, which is invasive for patients and sometimes challenging to diagnose due to morphologic heterogeneity. By studying whole-exome sequencing and whole-genome bisulfite sequencing of cell free DNA (cfDNA) and of matched metastatic tumor biopsies from patients with metastatic prostate adenocarcinoma and CRPC-NE, we identified CRPC-NE features detectable in the circulation. Overall, there was markedly higher concordance between cfDNA and biopsy tissue genomic alterations in patients with CRPC-NE compared with castration-resistant adenocarcinoma, supporting greater intraindividual genomic consistency across metastases. Allele-specific copy number and serial sampling analyses allowed for the detection and tracking of clonal and subclonal tumor cell populations. cfDNA methylation was indicative of circulating tumor content fraction, reflective of methylation patterns observed in biopsy tissues, and was capable of detecting CRPC-NE-associated epigenetic changes (e.g., hypermethylation of *ASXL3* and *SPDEF*; hypomethylation of *INSM1* and *CDH2*). A targeted set combining genomic (*TP53*, *RBI*, *CYLD*, *AR*) and epigenomic (hypo- and hypermethylation of 20 differential sites) alterations applied to ctDNA was capable of identifying patients with CRPC-NE.**

## Introduction

Prostate cancer is a leading cause of cancer death for men worldwide (1). Despite significant advances in therapy for patients with metastatic disease, the emergence of treatment resistance remains a universal problem which ultimately contributes to the lethality of the disease. Prostate cancer is a cancer type that is driven by androgen receptor (AR) signaling, and several potent drugs targeting the AR are now commonly used to treat patients with advanced disease either in combination with gonadal androgen suppression therapy for metastatic hormone naive prostate cancer or in the castration-resistant setting. Resistance to potent AR-targeted drugs is still primarily mediated through AR signaling (2). However, an increasingly recognized proportion of

tumors develop histologic transformation to a castration-resistant neuroendocrine prostate cancer (CRPC-NE) phenotype as a mechanism of AR-independent treatment resistance (3–5). Metastatic biopsies demonstrate morphologic features of small cell carcinoma, often with low or absent expression of the AR, downregulation of downstream AR-regulated markers such as prostate-specific antigen (PSA), and expression of classical neuroendocrine markers (e.g., chromogranin, synaptophysin) (6, 7). The prognosis of CRPC-NE is poor due in part to late diagnosis and a lack of effective therapies (5, 8). Similar to other poorly differentiated neuroendocrine carcinomas (9, 10), CRPC-NE frequently harbors genomic loss of *RBI* and *TP53* (4, 6, 11, 12). However, *RBI* and *TP53* loss-of-function alterations are not specific to CRPC-NE and are also observed in a subset of castration-resistant adenocarcinomas (4). Prior clinical and preclinical studies have supported a transdifferentiation process whereby CRPC-NE evolves clonally from a luminal prostate adenocarcinoma precursor (11, 13–16). Early prostate cancer genomic alterations are retained, but other genomic and epigenetic alterations are acquired (11, 14, 17). How and when this lineage plasticity manifests in patients is not well understood and whether early detection of CRPC-NE could improve outcomes is not known.

**Authorship note:** HB and AR are co-first authors.

**Conflict of interest:** HB has received research funding from Janssen, AbbVie Stemcentrx, Astellas, Eli Lilly, and Millennium, and has served as advisor/consultant for Janssen, Astellas, Amgen, Astra Zeneca, Pfizer, and Sanofi Genzyme. DMN has served on the Data Safety and Monitoring Board for Genetech Roche.

**Copyright:** © 2020, American Society for Clinical Investigation.

**Submitted:** August 15, 2019; **Accepted:** December 18, 2019; **Published:** February 24, 2020.

**Reference information:** *J Clin Invest.* 2020;130(4):1653–1668.

<https://doi.org/10.1172/JCI131041>.

Understanding patterns of tumor evolution that occur during prostate cancer progression and treatment resistance can inform tumor biology as well as identify novel strategies for combating resistance. Serial metastatic sampling in patients with prostate cancer during progression is challenging and is also not always representative of the heterogeneity of alterations that may exist across metastases in an individual. Although rapid autopsy studies have pointed to limited intraindividual genomic (18) and epigenomic (DNA methylation) (19, 20) heterogeneity across anatomic sites of metastases at the time of lethal disease, circulating tumor DNA studies at earlier prostate cancer disease states suggests clonal dynamics that occur during tumor evolution with competing clones and subclones potentially contribute to resistance and progression (21). Most studies to date have focused on castration-resistant adenocarcinoma (CRPC-Adeno), and little is known regarding the underlying heterogeneity and tumor dynamics in the case of CRPC-NE evolution.

We hypothesized that while tumors do acquire specific CRPC-NE changes (predominantly epigenetic) during the process of lineage plasticity, early resistant or aggressive prostate cancer clones (such as those already harboring loss of *RBI* and *TP53*) may also gain clonal dominance during CRPC-NE transformation and may be early selected facilitators of transdifferentiation. Although metastatic lesions in prostate cancer may harbor evidence of subclonal heterogeneity across sites potentially due to metastasis-to-metastasis seeding as observed in prior studies (22), we predicted that tumors would become less heterogeneous as patients progress toward CRPC-NE in later stages of the disease (as seen in rapid autopsy studies, ref. 18). To test this hypothesis, we studied tumor DNA detectable in the circulation of patients with CRPC-NE and CRPC-Adeno and compared their genomic and epigenomic profiles with patient-matched tumor biopsies.

In addition to using circulating tumor DNA (ctDNA) as a tool to study heterogeneity, we explored the potential utility of this approach as a noninvasive method for the detection of CRPC-NE by querying specific genomic and epigenomic changes seen in CRPC-NE tumors (4, 5, 11) in the circulation of biopsy-confirmed patients. A circulating biomarker for CRPC-NE has potential clinical implications that could lead to future avoidance of invasive biopsy for patients (the current standard for CRPC-NE diagnosis). Furthermore, since there is a spectrum of histologies even within CRPC-NE and as patients progress from adenocarcinoma to CRPC-NE, with mixed adenocarcinoma-neuroendocrine morphologies and hybrid tumors with overlapping features sometimes observed (3, 6), identifying molecular features could provide additional information beyond or complementary to a tumor biopsy. Noninvasive detection of molecular features of CRPC-NE could therefore lead to a more refined subclassification, and possibly detect features in patients at high risk for transformation even without CRPC-NE histology, paving the way for early intervention treatment strategies.

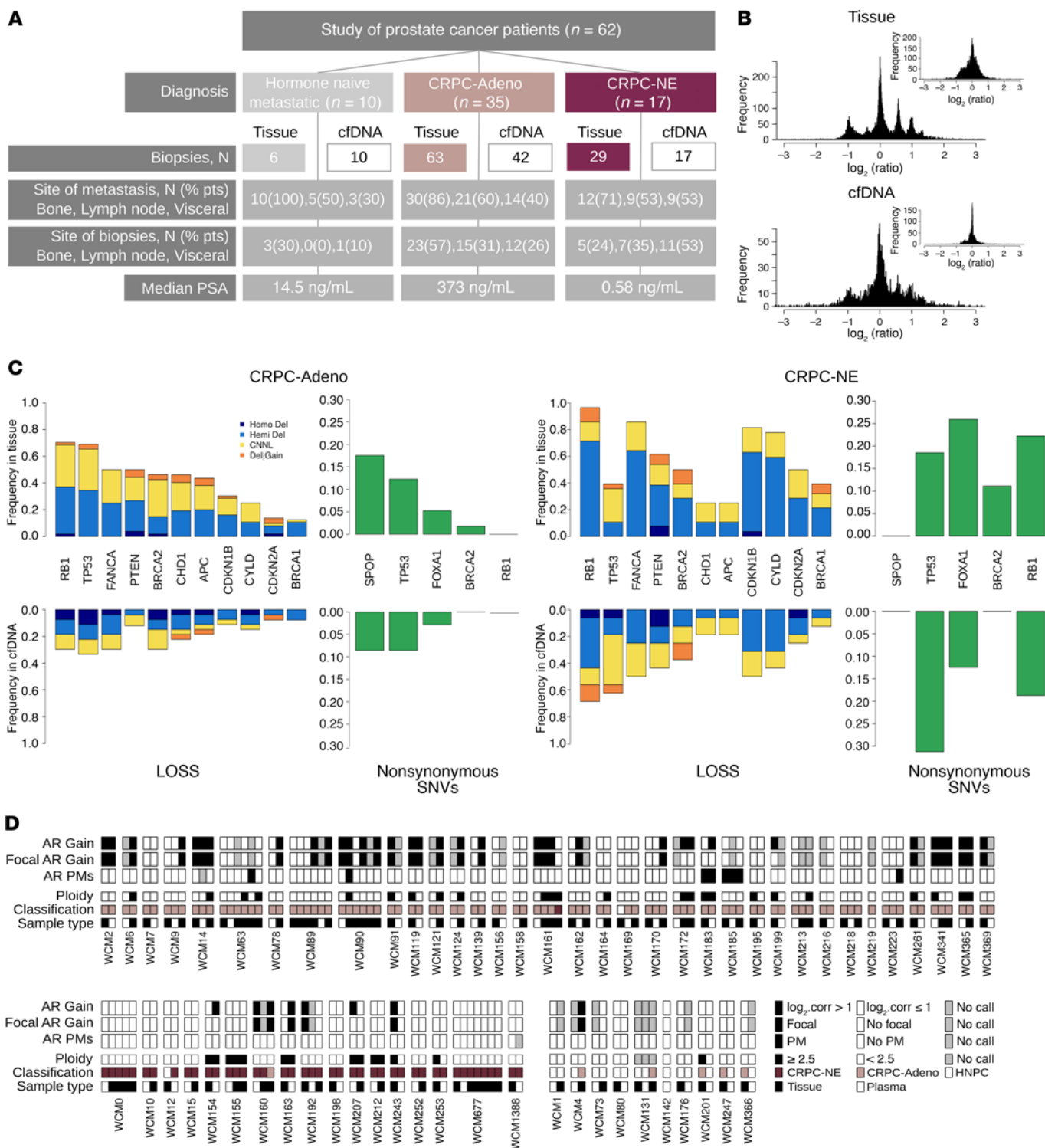
## Results

We obtained plasma blood samples and matched metastatic tumor biopsies from 62 men with metastatic prostate cancer (10 hormone-naïve metastatic prostate adenocarcinoma [mPCA], 35 CRPC-Adeno, and 17 CRPC-NE) classified histologically using

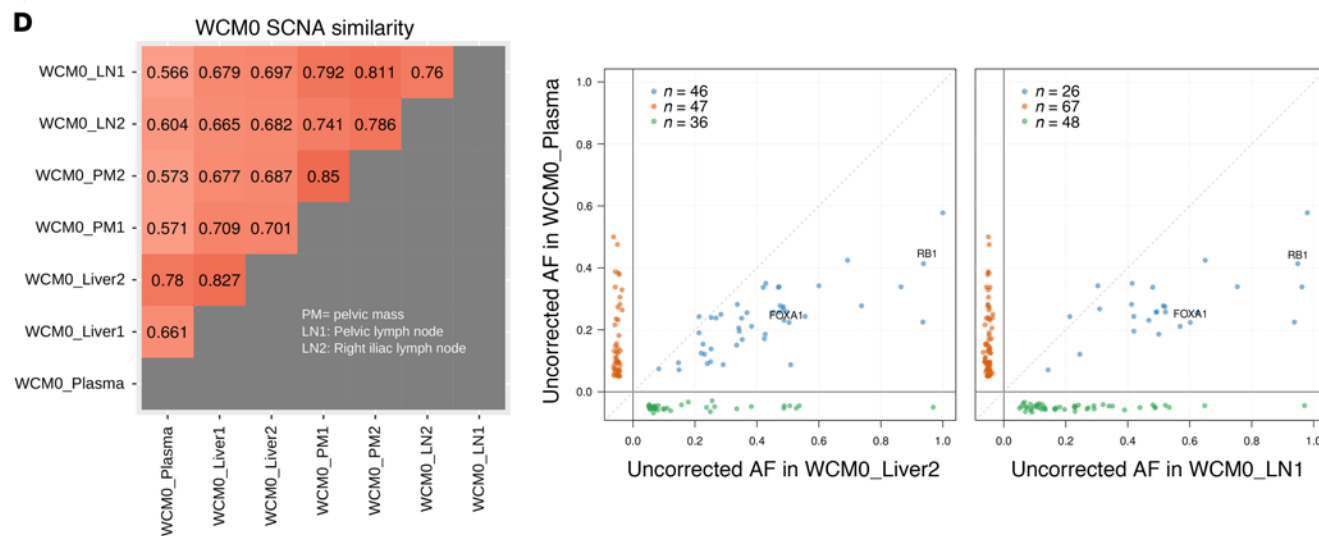
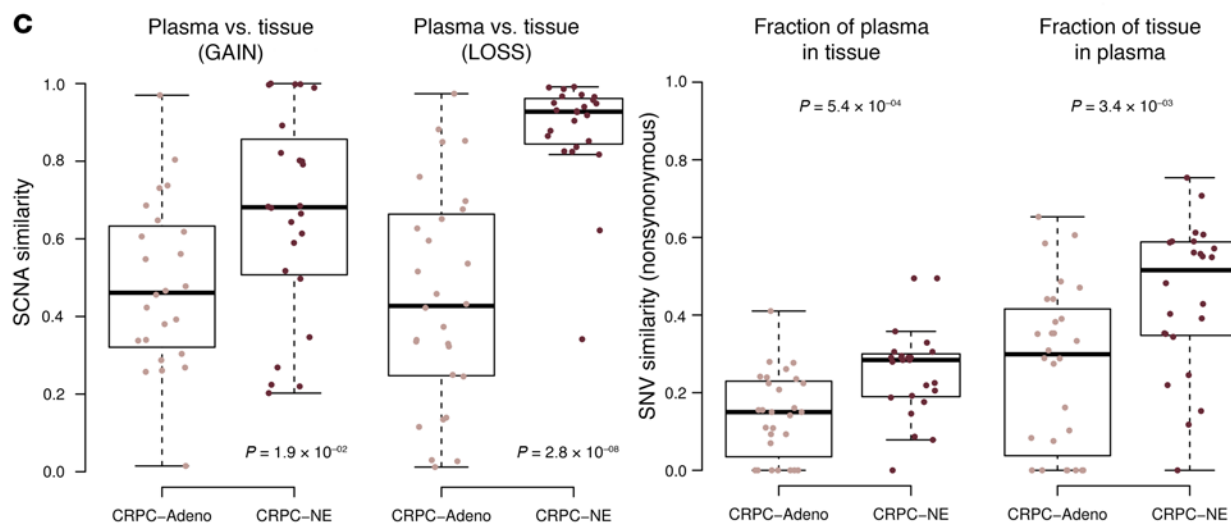
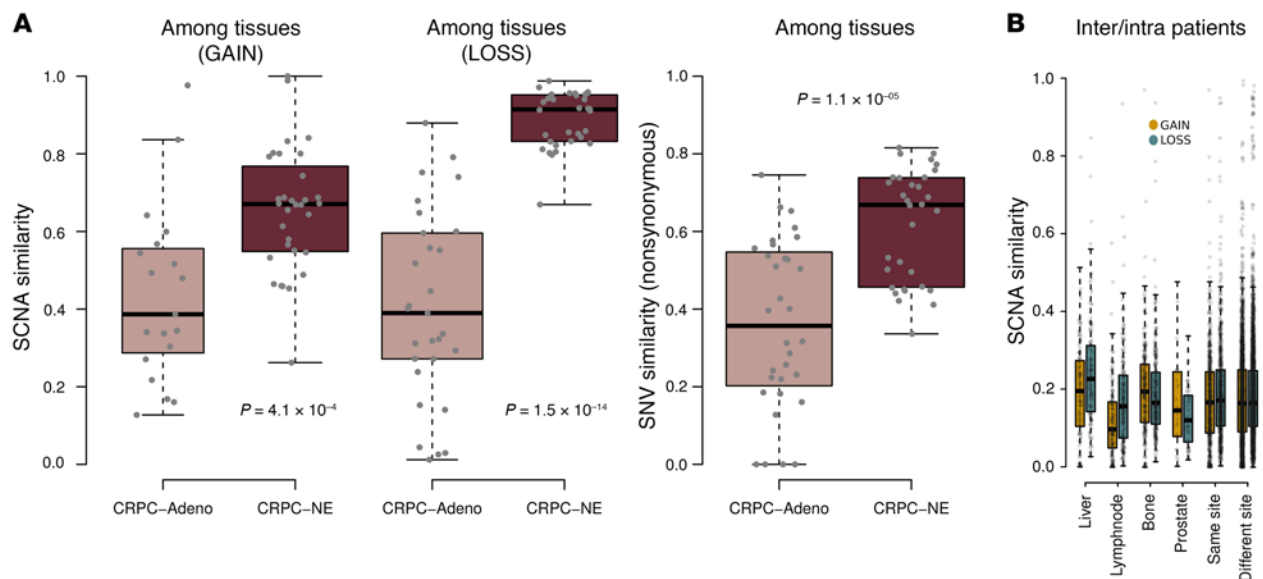
published criteria (6). Clinical features are summarized in Figure 1A and Supplemental Table 1; supplemental material available online with this article; <https://doi.org/10.1172/JCI131041DS1>. Twenty-four of these individuals had multiple tissue or plasma time points, for a total of 69 plasma samples and 98 metastatic tissues (Supplemental Tables 1–4). Metastatic sites of patients at the time of specimen collection included bone (84%), lymph node (52%), liver (32%), lung (16%), brain (6%). Overall, patients in this study had high metastatic burden, including 50% with visceral metastases and greater than 45% of patients with more than 6 bone metastases. The median number of prior therapies for CRPC was 2 (range 0–9); 42.8% patients received prior abiraterone or enzalutamide, and 65.7% received prior cytotoxic chemotherapy. The median time between prior therapy progression and blood draw was 3.2 months (range 1.1–9.1). Median progression-free survival on next line of systemic therapy (after blood/tissue collection) for CRPC patients was 4.1 months (4.3 months CRPC-Adeno, 4.0 months CRPC-NE), and median overall survival was 11.0 months (12.0 months CRPC-Adeno, 6.0 months CRPC-NE).

Tumor/normal whole-exome sequencing (WES) was performed on cell free DNA (cfDNA) extracted from plasma and germline DNA from peripheral blood mononuclear cells (PBMCs) with a median depth of coverage of  $\times 357$  for plasma samples and  $\times 105$  for germline samples. Median cfDNA tumor content (TC) was 22% (3%–94%), and TC did not associate with histology subtype (i.e., adenocarcinoma or CRPC-NE) (Supplemental Figure 1), number or type of metastatic sites, prostate-specific antigen (PSA) levels, serum neuroendocrine markers, or number of prior systemic therapies (Supplemental Figure 2, A–C). Prior exposure to cytotoxic chemotherapy was associated with higher cfDNA TC (Supplemental Figure 2D), and no statistical difference in TC was observed with respect to presence or absence of sites of metastasis (bone, lymph node or visceral) in chemotherapy-treated patients. In patients with CRPC-Adeno, a cfDNA TC of at least 50% was associated with inferior prognosis ( $P = 0.01$ , Supplemental Figure 3A), but this was not observed in the CRPC-NE subclass. Tumor mutation load and copy number burden in ctDNA did not differ significantly among tumor histologies, but both were higher in patients who received prior chemotherapy compared with those who did not (Supplemental Figure 2D). Recurrent genomic alterations observed in plasma and in tissue samples of study cohort samples detected by WES are shown in Figure 1, B and C and Supplemental Figure 4.

Overall, copy number alterations and point mutations involving commonly altered prostate cancer-associated genes (Supplemental Table 5) detected in metastatic biopsy tissue samples were evident in patient-matched ctDNA (median concordance 71%, range 0.29–1.0). This degree of concordance was similar to other targeted DNA and WES prostate cancer ctDNA studies (23, 24). As expected based on prior studies (4, 11), alterations involving *RBI*, *TP53*, and *CYLD* were more common in the ctDNA and biopsy tissues of CRPC-NE patients, whereas *AR* alterations more common in CRPC-Adeno patients. There were no significant differences in the frequency of DNA repair gene aberrations involving *BRCA1*, *BRCA2*, and *ATM* among the subtypes (Figure 1B, Supplemental Figure 4, and Supplemental Table 6). *SPOP* mutations were notably absent in CRPC-NE. The most common genomic alterations



**Figure 1. Frequencies of somatic aberrations in advanced prostate cancer driver genes. (A)** Schematic of study cohort. **(B)** WES segmented data for study cohort. WES segmented data are shown raw (inset) and ploidy- and TC-adjusted. **(C)** Distribution of somatic copy number loss and SNVs in CRPC-Adeno and CRPC-NE ctDNA and tumor tissue samples. Loss events include homozygous deletions (HomDel), heterozygous deletions (HetDel), copy number neutral losses (CNNL), and events defined by loss of one allele and gain of the other allele (Del|Gain). **(D)** AR somatic aberration status in CRPC-Adeno, CRPC-NE, and HNPC plasma and tumor tissue samples, ordered based on serial dates of collection. AR gain, focal gain, and SNV (L702H and T878A positional pileup calls) are shown together with sample ploidy and tumor class. Statistics are reported in Supplemental Table 6.



**Figure 2. Similarity of somatic aberration profiles across plasma and metastatic tumor tissue samples.** (A) Inpatient somatic copy number aberrations (SCNAs, Loss and Gain) and single nucleotide variants (SNVs) similarity across metastatic biopsies stratified by patient's tumor class at plasma collection (Supplemental Table 2). (B) Inter- and inpatient measures of SCNA similarity per site and across sites of metastasis. (C) Inpatient loss and gain SCNA similarities (left) and SNV (right) similarities (fraction of SNVs in plasma detected also in tissues and fraction of SNVs in tissues detected also in plasma) across tissue and plasma samples stratified by patient's tumor class at plasma collection. Only samples of patients with estimated ctDNA greater than 10% are considered. The same trends are also obtained with a more restrictive filter on TC greater than 50%. (D) SCNA similarity among plasma and tumor tissue samples of patient WCM0 (left); private and shared SNVs comparison between WCM0 plasma and selected tumor tissue samples. Reported *P* values are computed using 2-tailed Wilcoxon Mann-Whitney *U* test.

in ctDNA included *TP53* (43%), *RBI* (38%), and *FANCA* (25%) in mPCA patients; *AR* (42%) and *TP53* (37%) in CRPC-Adeno patients; and *RBI* (69%) and *TP53* (63%) in CRPC-NE patients. As observed in prior studies (25, 26), the presence of *AR* aberrations in the ctDNA of CRPC-Adeno patients was associated with inferior overall survival ( $P = 0.009$ ), which was maintained after adjustment for TC (multivariate  $P = 0.028$ , Supplemental Figure 3B). Of note, *AR* focal gains could be either persistent or dynamic when comparing serial samples in individuals with CRPC-Adeno (Figure 1D), supporting evolution of somatic *AR* aberrations. The prognostic value of other specific gene aberrations in CRPC-Adeno was consistent with a recent study by Annala et al (27). The presence of *TP53* and/or *RBI* loss-of-function alterations in the ctDNA of patients with CRPC-Adeno (but not CRPC-NE) associated with worse overall survival ( $P = 0.006$ ), including after accounting for TC (multivariate  $P = 0.006$ , Supplemental Figure 3C). Somatic alterations involving *BRCA1*, *BRCA2*, or *ATM* were also prognostic across the cohort (multivariate  $P = 0.008$ , Supplemental Figure 3D).

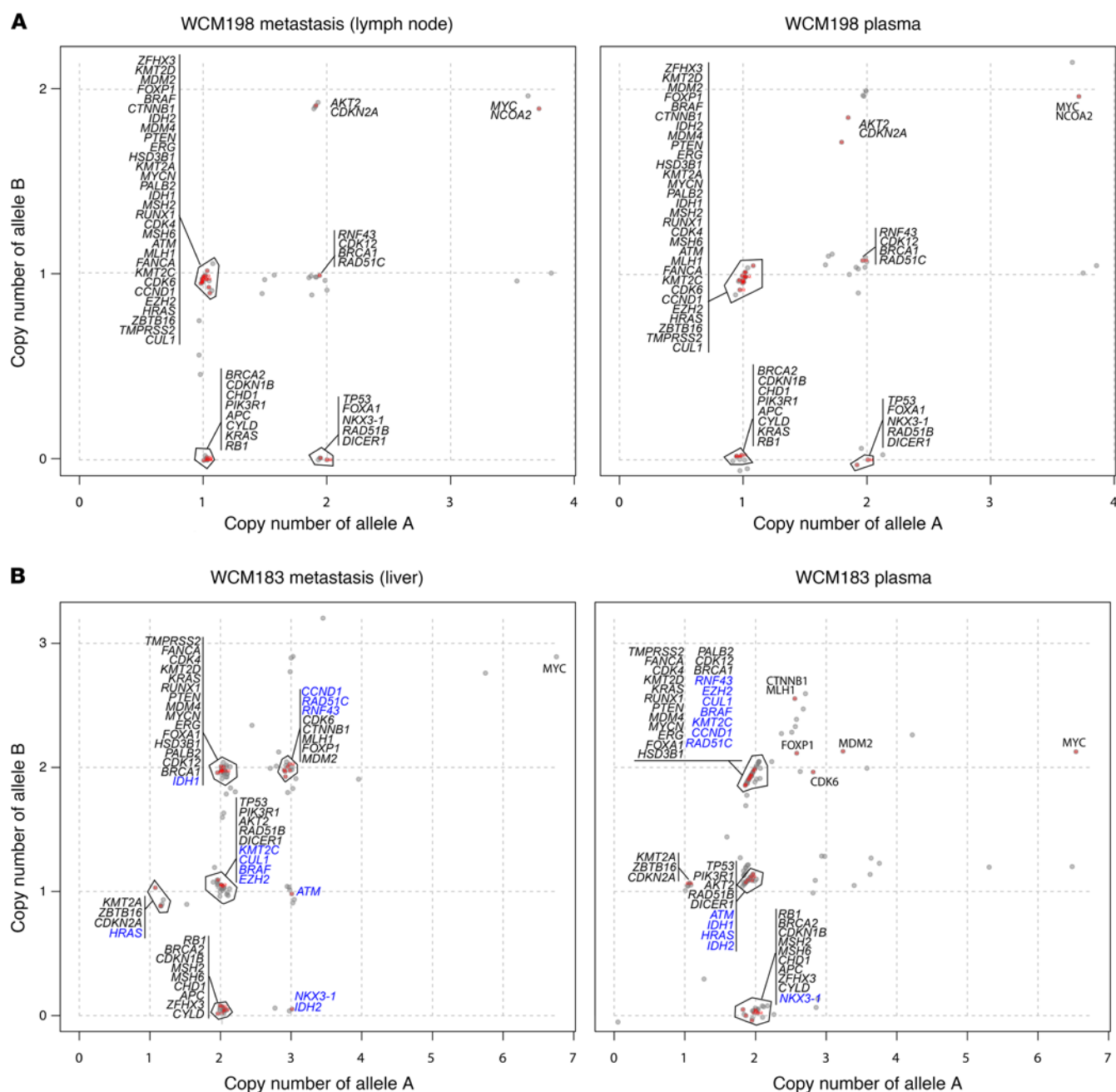
A subset of alterations detected in ctDNA was not detected in matched tissue samples (Supplemental Figure 4). Though false-negative calls cannot be excluded, it is likely that inpatient tumor heterogeneity is a prominent reason for these differences, whereby the sampled metastasis did not fully capture the genomic status of the patient's cancer burden. In order to quantify and contextualize the degree of CRPC-NE intra- and interpatient heterogeneity by comparing metastatic tissue and liquid biopsies, we first built the distributions of genome-wide heterogeneity levels among all metastatic tissue biopsies of both CRPC-Adeno and CRPC-NE using their genomic profiles (WES data). Although there was generally genomic concordance across metastatic tissues in either subtype (Figure 2A), the CRPC-NE subtype demonstrated significantly greater interindividual similarity both for somatic copy number alterations (SCNAs) and for single nucleotide variants (SNVs) (Figure 2A). Importantly, this finding was irrespective of the anatomic site of the metastatic biopsy assessed, as indicated by similarities stratified by site and across sites (Figure 2B). We also observed higher inpatient genomic similarity when comparing ctDNA and biopsy samples within individuals with CRPC-NE (Figure 2C and Supplemental Figure 5) compared with CRPC-Adeno patients. If a plasma sample represents the

genomics of all or most of an individual's metastases releasing DNA into the circulation, a higher concordance between ctDNA and single-site biopsy suggests lower inpatient tumor heterogeneity in CRPC-NE. However, differential contribution or release rates of individual metastases into the circulation cannot be fully excluded. For instance, inspection of the relative contributions into the plasma of 6 spatially distinct metastases (with 74% of average genomic similarity with one another) of one CRPC-NE patient all obtained at the same time point supported a higher contribution of liver metastasis tumor alterations in cfDNA compared with other sites (Figure 2D).

In order to further explore inpatient heterogeneity across the cohort, we leveraged allele-specific copy number analysis (28) of WES data. Although the vast majority of alterations were identified as shared between ctDNA and metastatic biopsies (Figure 3A and Supplemental Figure 6A), this higher resolution analysis allowed for the identification of distinct subclonal differences (as illustrated by case examples in Figure 3B and Supplemental Figure 6B; blue genes demonstrate different allele-specific status between tissue and plasma samples). These differences may represent private events in metastatic lesions that are therefore differentially represented in the circulation. Diversity could be restricted to a small set of lesions (Figure 3B) or to a wide range of structural differences, resulting in allele-specific diversity (Supplemental Figure 6B). Overall, these data support subclonal heterogeneity in metastatic lesions that may not be detected by single-site biopsy but could also contribute to the development of therapy resistance.

To explore how genomic alterations may change with time, we evaluated serial tumor and blood samples from individuals (Figure 4, Supplemental Figure 7, and Supplemental Table 7). In patient WCM161, for instance, we tracked metastatic biopsy time points during clinical progression from CRPC-Adeno to CRPC-NE. Interestingly, the plasma sample obtained at the time of CRPC-Adeno with lymph node and bone metastases present displayed a genomic ctDNA profile most similar (based on whole-exome-wide comparison) to the CRPC-NE liver metastasis observed on imaging and biopsied 3 months later at the time of progression on abiraterone (Figure 4A), supporting the presence of detectable resistant clones in the circulation before the development of clinical features of CRPC-NE or liver metastases.

To delve deeper into tumor dynamics, we assessed short-interval serial time points of 2 patients with metastatic CRPC progressing after multiple lines of systemic therapy (Figure 4B). Patient WCM185 is a patient with a rising PSA greater than 3000 ng/mL and bone-only metastases, a clinical picture suggestive of AR-driven progression. Patient WCM14 developed new visceral metastases despite a nonrising PSA of 20 ng/mL, clinical features commonly observed in CRPC-NE. Three weekly serial plasma time points of patient WCM185 were compared with WES data from his CRPC-Adeno tumor/plasma specimens collected over the prior 3 years. When corrected for ploidy and TC, distinct clones were identified in ctDNA, including those with and without *AR* mutations and *TP53* deletions that changed dynamically with time, suggesting multiple clones within the circulation with competing frequencies. This fast evolution in tumor clones even in the absence of therapy is consistent with clonal disequi-



**Figure 3. Allele-specific copy number quantification of matched plasma and tumor tissue samples.** (A) Example of patient demonstrating almost identical genomic status between a metastasis and a plasma sample, suggesting high homogeneity among all patient’s metastases. The lymph node metastasis of patient WCM198 and the plasma sample were collected at 4 days apart. Polygons include cancer genes with same allele-specific copy number; red dots correspond to reported gene names. (B) Example of patient with heterogeneous profiles. Patient WCM183 liver metastasis was obtained 13 days before the plasma sample. Blue indicates genes with different allele-specific copy number.

librium. On the other hand, 4 serial samples of patient WCM14, as well as his tumor tissue samples collected 5 years earlier when he had minimal metastatic disease burden, all showed consistent alterations including *MYC* gain and *RB1*, *TP53*, *PTEN*, and *BRCA2* losses across all time points, suggesting these were relatively early events. Of note, concurrent loss of *RB1* and *TP53*, frequently associated with small cell carcinomas, including CRPC-NE, was detectable before he had developed clinical features of AR independence or CRPC-NE (Figure 4B).

These data suggest that in some cases, CRPC may be dominated by a clone that arises early whereas in other cases sub-clones evolve and contribute to treatment resistance. By using ctDNA to track clones with a selective advantage with those that are less fit, we may better understand tumor dynamics during prostate cancer progression. In the setting of CRPC-NE, intrinsic drug resistance is likely a major factor that contributes to the development of clonal dominance. However, other biological factors, including a proliferative advantage,

the local microenvironment, and epigenetic alterations, also likely play a role.

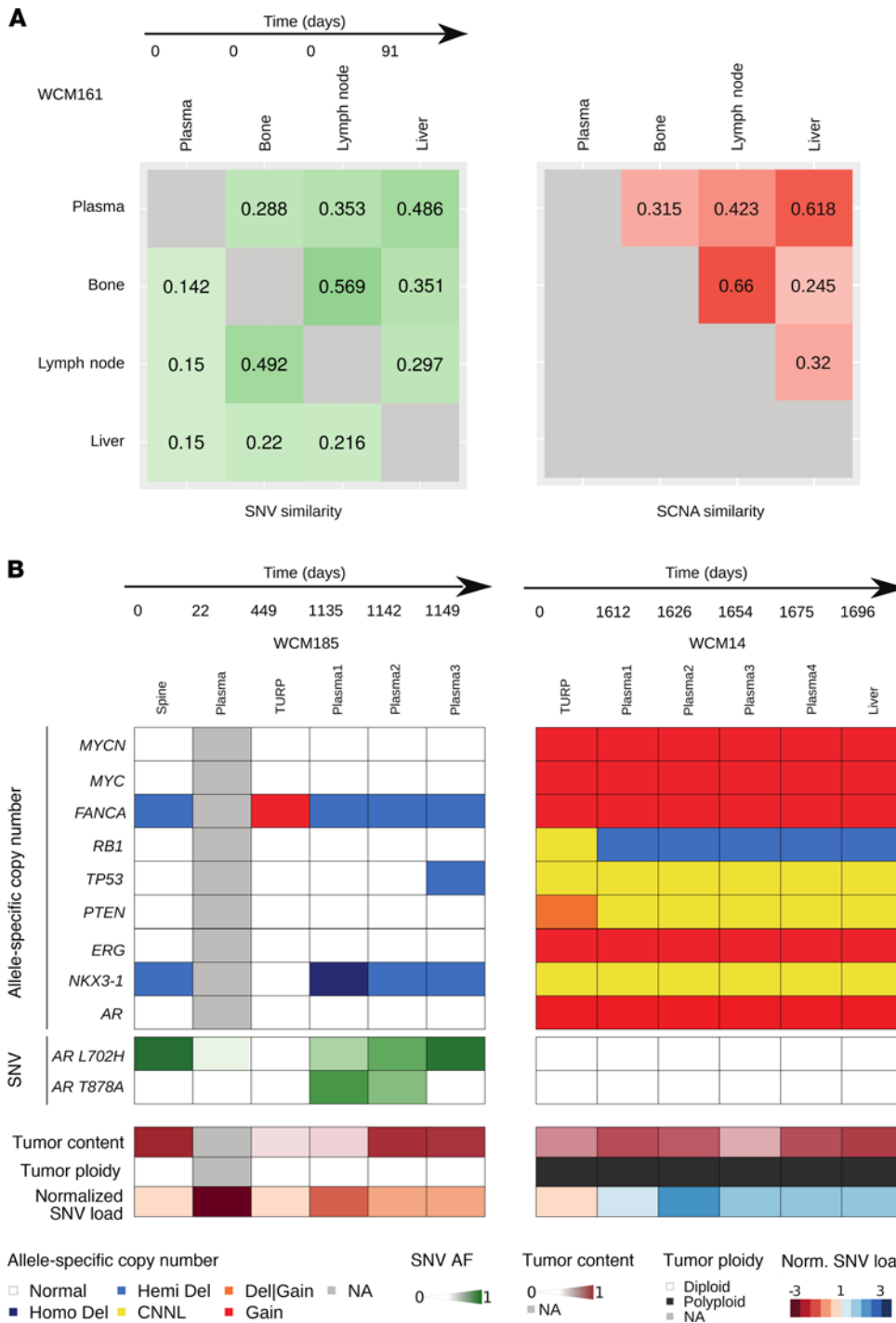
Epigenetic variability contributes to the diversity of phenotypes observed in cancer and other diseases and has been shown to play a key role in differentiation and phenotypic plasticity (29). Conserved DNA methylation patterns across metastases has also been observed in patients with lethal prostate cancer (20), supporting relative stability of DNA methylation patterns at least in late-stage disease. We previously identified significant differences between DNA methylation patterns of CRPC-NE tissues compared with CRPC-Adeno (11). In order to see if these changes are also captured by cfDNA, we performed whole-genome bisulfite sequencing of cfDNA from 5 patients with CRPC-Adeno and 6 patients with CRPC-NE and compared these methylation profiles with those of their matched tumor biopsies (Supplemental Table 8). Inference of cfDNA TC by DNA methylation using a methodology previously described (30) revealed no differences between CRPC-Adeno and CRPC-NE, and was consistent with copy number-based TC data from WES of the same samples (Figure 5A and Supplemental Figure 8B). Comparative analysis between plasma and biopsy tissue methylome data genome-wide (Figure 5C and Supplemental Figure 8A) demonstrated overall concordance of differentially methylated regions in CRPC-NE (both hypomethylated and hypermethylated sites, as shown in Supplemental Figure 8E). This included methylation of specific neuroendocrine prostate cancer classifier genes reported in Beltran et al (11) such as *ASXL3* and *SPDEF*, as well as hypomethylation of overexpressed CRPC-NE genes, including the neuroendocrine marker *INSM1* and the plasticity gene *CDH2* (Supplemental Table 12, Supplemental Figure 8C). Differentially methylated regions detectable by cfDNA were able to segregate patients with CRPC-Adeno or CRPC-NE (Figure 5B and Supplemental Figure 8D). We did not observe clustering based on the location of metastases. This potentially suggests that cfDNA methylation patterns in the circulation are distinguishable based on their resistance subclass, more so than by the anatomic sites of metastases within a given patient. To increase our study sample size, we queried DNA methylation profiles from a published independent cohort of 33 metastatic CRPC patients at baseline and after treatment with abiraterone acetate (31). Interrogation of the end-of-treatment cases within this cohort (with at least 10% TC) (30) revealed similar methylation profile distribution as our CRPC-Adeno cases treated with potent AR pathway inhibitors (Figure 5D), and the combined set demonstrated significant methylation differences between CRPC-Adeno and CRPC-NE patients ( $P = 0.0031$ ). Notably, the 2 CRPC-Adeno cases that harbored differentially methylated region (DMR) profiles compatible with CRPC-NE were from patients with visceral metastases as well as other features commonly seen in CRPC-NE (i.e., WCM119 developed radiographic progression on enzalutamide in the setting of a low serum PSA and elevated serum neuroendocrine markers; WCM14 tumor harbored concurrent *MYCN* amplification and deletion of *RBI*, which are often enriched in CRPC-NE).

Overall, we found that CRPC-NE is associated with both genomic and epigenomic features that distinguish this resistant subtype from CRPC-Adeno, and these alterations are detectable by ctDNA. Although WES and WGBS allowed for the char-

acterization of global patterns and tumor heterogeneity in a patient cohort with relatively high metastatic tumor burden, a molecular classifier for CRPC-NE would be useful if it could be applied at all stages of the disease even in situations with lower tumor burden. This would require a more sensitive targeted assay to run at deeper coverage amenable to lower TC plasma samples. We therefore tested a targeted set of genomic and epigenomic lesions that could identify CRPC-NE using cfDNA. When combining the presence or absence of CRPC-NE-associated features, including genomic deletion of mutation of *TP53*, *RBI*, and *CYLD*, lack of mutation or focal gain of *AR*, and aggregated hypo- and hypermethylation of 20 differential sites (Supplemental Table 11) (Neuroendocrine Prostate Cancer feature score, see Methods), cfDNA was robust in identifying individuals with CRPC-NE confirmed by histologic analysis of their metastatic biopsy ( $P = 0.000444$ ) (Figure 5E). Although DNA methylation in itself is a strong classifier of CRPC-NE, we reasoned that the combination with genomic alterations, likely earlier events, would enable the future identification of high-risk cases at earlier stages. Overall, these data nominate a ctDNA classifier as a potential biomarker for noninvasively detecting CRPC-NE.

## Discussion

Recent metastatic biopsy studies have uncovered the landscape of genomic and epigenomic alterations enriched in metastatic CRPC compared with primary prostate tumors (4, 32–34), yet there is still much to learn regarding the sequence of molecular events that occurs during prostate cancer progression and therapy resistance, and the degree of heterogeneity that underlies different stages of the disease. Resistance patterns have largely fallen into 2 categories: AR-driven and non-AR driven (2). The vast majority of castration-resistant tumors are AR-driven, and this may be mediated through *AR* gene amplification, mutation, splicing, other structural variants, or other means. Specific AR alterations detected by blood sampling, such as *AR* amplification in ctDNA or AR-V7 splice variant expression in circulating tumor cells, have been associated with poor response and outcomes in patients treated with the AR-targeted drugs abiraterone or enzalutamide (25, 26, 35). This study and others (25, 34, 36, 37) support continued evolution of *AR* gene alterations during disease progression. Non-AR driven disease is less common in CRPC, and the mechanisms underlying AR independence are less clear. One increasingly recognized mechanism is the development of lineage plasticity, by which the tumor cells adopt alternative lineage and oncogenic programs to grow, despite suppressed AR signaling (3). One extreme manifestation of this is the transformation from a luminal prostate adenocarcinoma to a CRPC-NE phenotype, defined by the acquisition of histologic features similar to small cell lung cancer (6) and associated with distinct molecular alterations (5, 11). It has been reported that up to 15%–20% of patients develop features of CRPC-NE in late stages of their disease (4, 5, 38). Similar to small cell lung cancer and other small cell neuroendocrine carcinomas (9), CRPC-NE is enriched with genomic loss of *TP53* and *RBI* (4, 11, 15). However, loss of these genes may also occur in nonneuroendocrine prostate cancer (33). In EGFR-mutated lung adenocarcinomas, the presence of *TP53* and *RBI* loss predisposes patients to later transfor-



**Figure 4. Heterogeneity of somatic aberration profiles in multiple plasma and tumor tissue sample series. (A)** Somatic-copy number aberration (SCNA) and single nucleotide variants (SNV) similarities among plasma and tumor tissue samples of patient WCM161. Samples are ordered by date of collection with date differences reported in days. SNV similarity is an asymmetric measure based on set inclusion and the complete matrix is shown. SCNA similarity is instead a symmetric measure based on the Jaccard coefficient and hence information is shown without redundancy. **(B)** Genomic analysis of WCM185 and WCM14 multi-sample series. The panel reports: allele-specific copy number of a selection of advanced prostate cancer driver genes (HomDel = homozygous deletion, HetDel = heterozygous deletion, CNNL = copy number neutral loss, Del|Gain = events defined by loss of one allele and gain of the other allele); AR L702H and AR T878A SNVs (determined by pileup analysis); TC estimations; ploidy state estimation; normalized SNV load, with values representing the ratio between the sample's SNV number and the sample type (plasma/tumor tissue) median SNV number. Samples are ordered by date of collection (intervals shown).

mation to small cell lung cancer after EGFR inhibitor therapy (39, 40). The timing of these events and other molecular alterations during CRPC-NE progression has not been established; our data suggests that they may also be acquired early. DNA methylation and transcriptome analyses have also pointed to specific defining features of CRPC-NE that may improve upon genomics for CRPC-NE disease detection. In the current study, we found that both genomic and epigenomic features of CRPC-NE are identifiable in patients noninvasively through plasma ctDNA analysis. The diagnosis of CRPC-NE has clinical implications as these patients may

be considered for platinum-based chemotherapy (41, 42) or other non-AR targeted approaches, rather than AR-targeted drugs.

In this study, we also identified several prostate cancer alterations shared between adenocarcinoma and CRPC-NE, supporting a same cell of origin, though certain genomic alterations (such as *AR*, *TP53*, and *RBI*) had different prognostic value in patients based on their histologic subtype. We took a genome-wide approach in order to characterize the clonal and subclonal heterogeneity in advanced prostate cancer. Similar to prior studies using targeted DNA or WES (23, 24), there was overall high



concordance of alterations shared between cfDNA and matched tumor biopsies within individual patients. Allele-specific analysis allowed for a more refined estimation of subclonal diversity. In general, comparison across patients, metastases, and between plasma and biopsies pointed to less heterogeneity in CRPC-NE compared with CRPC-Adeno. This observation, as well as data from published rapid autopsy studies (18, 20), points to higher intraindividual homogeneity across metastases in later stages of the disease. A proposed model of prostate cancer progression toward CRPC-NE, building upon published work in the field, is depicted in Figure 6. Although prior studies have supported a monoclonal origin of metastatic prostate cancer (43, 44) with metastatic lesions traceable back to a founding clone within the primary tumor (18), tumors do subsequently acquire alterations with disease progression and treatment resistance. Polyclonal spread (22, 45) in later stages further leads to inpatient heterogeneity, particularly of subclonal alterations (as we detected here through allele-specific copy number analysis of ctDNA). Evolution of these subclones may contribute to cancer therapy resistance (46, 47). During the transition toward CRPC-NE and AR independence, our data support likely selection of a dominant tumor clone (potentially harboring combined loss of *RBI* and *TP53*) that persists and dominates due to selective pressures of therapy, resulting in more stringent bottlenecks in the subclonal makeup of the tumor. Therefore, although collective clinical and preclinical data support CRPC-NE arising through transdifferentiation from a prostate adenocarcinoma precursor, our data also suggest that there may be dominance or clonal selection of a resistant clone during this process, and that these are not 2 mutually exclusive models of clonal evolution. Important future studies should address if earlier disease states such as select hormone-naive, high-risk localized or metastatic prostate cancers harbor CRPC-NE alterations that manifest later through clonal selection.

We also found that DNA methylation profiles are distinct in CRPC-NE, which is detectable both in tissues and in cfDNA. DNA methylation patterns have also been reported to be consistent across sites of metastases in men with lethal CRPC-Adeno (19, 20), supporting the use of methylation as a relatively stable biomarker for detecting specific resistance patterns. How and when these patterns shift during the development of CRPC-NE is not known, nor is the reversibility of any of these epigenetic changes. Serial dynamic monitoring of both genomic and epigenomic CRPC-NE features is now required to better understand when CRPC-NE alterations occur and how frequently in the context of specific therapies. Methylation profiles point to differential epigenetic regulation of key genes involved in CRPC-NE and may also point to new therapeutic avenues.

A limitation of this study is the relatively small sample size due to stringent patient selection with the requirement for sufficient cfDNA (and relatively high metastatic tumor burden) to perform WES/WGBS, as well as the need for matched metastatic biopsies. A targeted assay of select genes would be required to more sensitively evaluate for CRPC-NE features at disease states with lower cfDNA TC. Due to this patient selection as well as enrichment of this study with patients with CRPC-NE, treatments were heterogeneous across the cohort and outcomes may have been influ-

enced by therapy cross-resistance or other factors. The prognostic and predictive role of the identified genes in different stages of the disease would need to be tested and confirmed in prospective cohorts. Although we focus on the CRPC-NE subtype, we recognize that other resistance subtypes also exist even within non-AR driven disease (38). Nonetheless, these new observations highlight the feasibility and promise of a combined genomic/epigenomic cfDNA approach to identify men with advanced CRPC-NE and provides insights into the degree of inpatient heterogeneity that exists at different stages of disease progression.

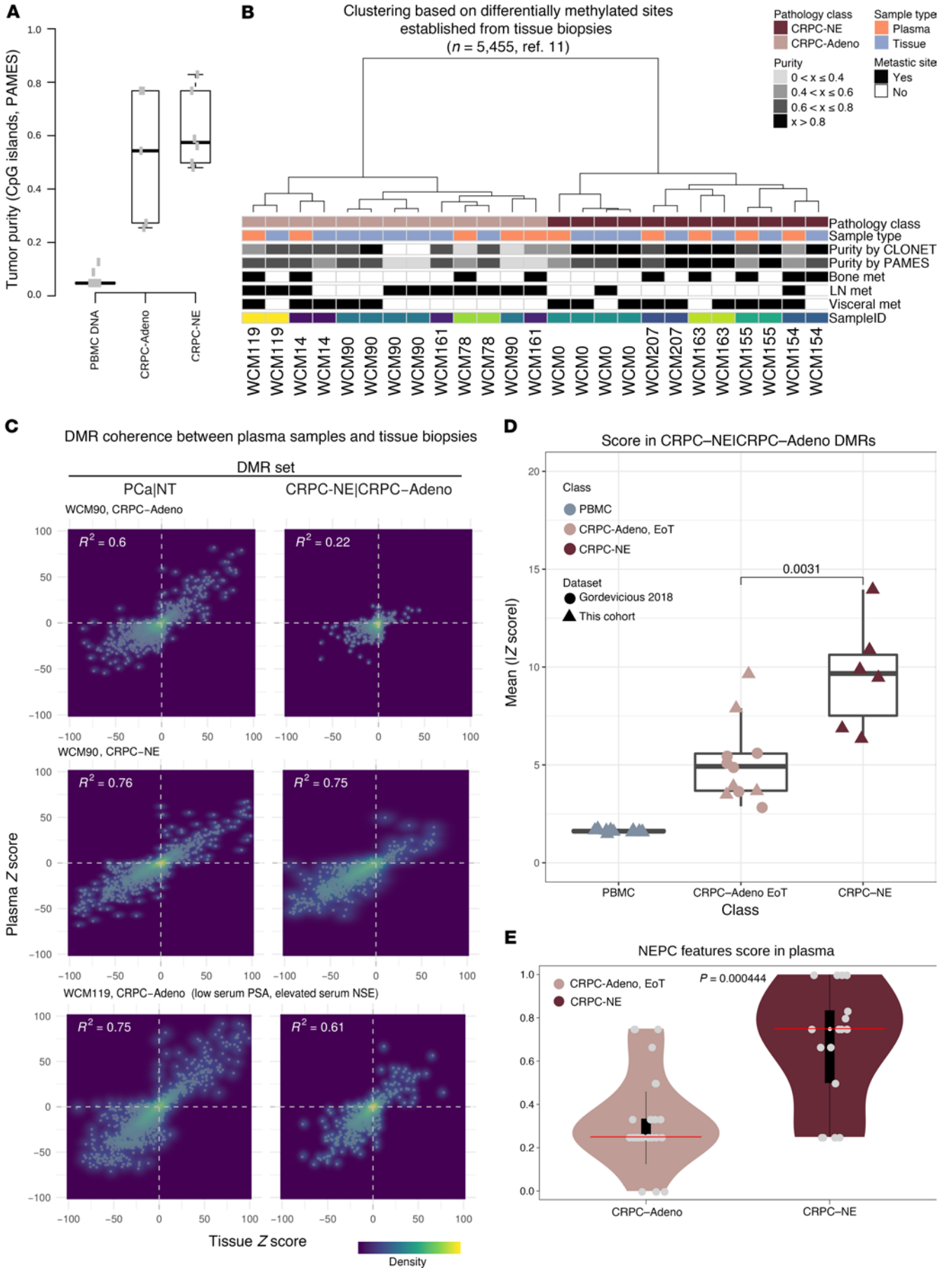
## Methods

### Clinical cohort

Patients were prospectively enrolled on an IRB-approved protocol (#1305013903) with written informed consent. Patients were eligible for this study if they had metastatic prostate cancer with at least one metastatic tumor biopsy and matched plasma sample (EDTA or Streck Cell-Free DNA BCT) and at least 2–50 ng total cfDNA present in 1 mL plasma. WES of metastatic biopsy was previously reported for 53 of 63 patients (11, 48). Peripheral blood samples were collected within 30 days of treatment initiation (in ~85% of cases). The choice of systemic therapies before plasma sample collection was at the discretion of the treating physician. Treatments were administered continuously until evidence of disease progression or unacceptable toxicity, or, in the case of chemotherapy according to regime-specific treatment protocols, for a planned number of cycles. We included patients treated for at least 1 month, defined as the duration of time from initiation to discontinuation of therapy. Clinical data were obtained from the electronic medical record. Pathology review was performed and reported as adenocarcinoma or CRPC-NE using published morphologic criteria (6).

### Plasma processing and DNA extraction and quantification

Whole blood was centrifuged at 1600g for 10 minutes at 4°C within 3 hours after blood collection. The plasma layer was transferred to 2-mL microcentrifuge tubes and centrifuged at 16,000g for 10 minutes at 4°C to ensure removal of any cellular debris. The plasma was then collected and stored at –80°C. cfDNA was extracted from plasma using the NeoGeneStar Cell-Free DNA Purification kit per manufacturer's instructions. Briefly, for 2-mL plasma samples, cfDNA was isolated via proteolytic digestion with 1 µg RNA carrier at 55°C for 30 minutes. cfDNA capture on the superparamagnetic particles was accomplished via addition of 3 volumes (6 mL) LYS buffer, 0.8 mL isopropanol, and 30 µL NGS Beads (NeoGeneStar), with a 30-minute room temperature incubation. Samples were washed twice with NeoGeneStar Wash Buffer (diluted with an equal volume of absolute ethanol) and twice with 75% ethanol, air dried, and eluted in 30 µL 10 mM Tris, 0.1 mM EDTA (pH 8.5). Germline DNA was extracted from extracellular blood components (PBMCs) using the Promega Maxwell 16 MDx kit per the manufacturer's instructions. Blood aliquots were vortexed briefly and mixed with LYS buffer and proteinase K solution, vortexed briefly again, and incubated at 56°C for 20 minutes. Lysates were then transferred to the first well of a multiwell cartridge prefilled with reagents, 65 µL elution buffer was loaded, and the automated system successfully isolated high-concentration genomic DNA. cfDNA was quantified with Qubit and quality was also assessed using Agilent's High Sensitivity DNA kit.



**Figure 5. Differential methylation signal is detected in the circulation of patients.** (A) PAMES purity estimation of plasma WGBS plasma and PBMC samples. Top 10 most informative hypermethylated CpG Island were used. (B) Ward's hierarchical clustering of 25 samples using 1-Pearson's correlation coefficient as distance measure. The annotation tracks include information on sample tumor purity and on the site of the relative sequenced tissue biopsy for all tissue samples and on the presence or absence of lymph node, bone, or visceral metastases in the corresponding patient for the plasma samples. (C) Evaluation of DMR concordance in matched plasma and tissue samples. Two distinct DMR sets were nominated applying the Rocker-meth algorithm on tissue samples, and single-sample Z scores were computed for each DMR. Comparison among values detected in ctDNA and tissue biopsy are reported for 3 representative patients: WCM90, CRPC-Adeno; WCM0, CRPC-NE; and WCM119, CRPC-Adeno with radiographic progression on enzalutamide, PSA 3.5, and elevated serum NSE. Color density is proportional to point density to the power of 1/4 to improve visualization. First, order linear regression  $R^2$  is reported. (D) Comparison of average absolute Z score based on CRPC-NE|CRPC-Adeno DMR in plasma samples from this cohort and from a set of patients treated with abiraterone acetate (reported in ref. 31). To maximize the compatibility of clinical history and disease stage, only end-of-treatment samples with an estimated TC greater than 10% were included. Significance was assessed using 2-tailed unpaired Wilcoxon Mann-Whitney  $U$  test without continuity correction. (E) NEPC feature scores are plotted as assessed in plasma data of CRPC-Adeno and CRPC-NE patients.

#### Whole-exome sequencing of ctDNA and matched PBMC DNA

Based on the expected circulating DNA fragment size distribution and the range of input DNA identified from each plasma sample, we opted for the Roche SeqCap EZ Library SR for library preparation. Germline genomic DNA was sheared using the Covaris E220 Evolution instrument. For cfDNA, input for library prep ranged from 2.8–50 ng and for germline DNA, input ranged from 50–100 ng. The libraries of both germline DNA and cfDNA were prepared using the KAPA HTP Library Preparation Protocol (Kapa Biosystems) containing end repair, A-base addition, and ligation of sequence adaptors. The sample libraries were normalized and pooled following PCR amplification. The pooled libraries were then hybridized with whole-exome SeqCap EZ probe pool. Finally, the pooled, indexed, and amplified capture sample libraries were sequenced using the Illumina HiSeq4000 sequencer at 100 cycles. All plasma and matched normal samples were run with single-end protocol unless differently specified. Illumina bcl2fastq2 Conversion Software was used to demultiplex samples into individual samples and convert per-cycle binary base call (BCL) files into FASTQ files for downstream data analysis. WES sample processing and the data generation procedure of tissue biopsies sequenced in this study follow the same protocol as previously reported (11); all tissues were run with paired-end protocol.

#### Titration experiment to determine optimal ctDNA input for WES

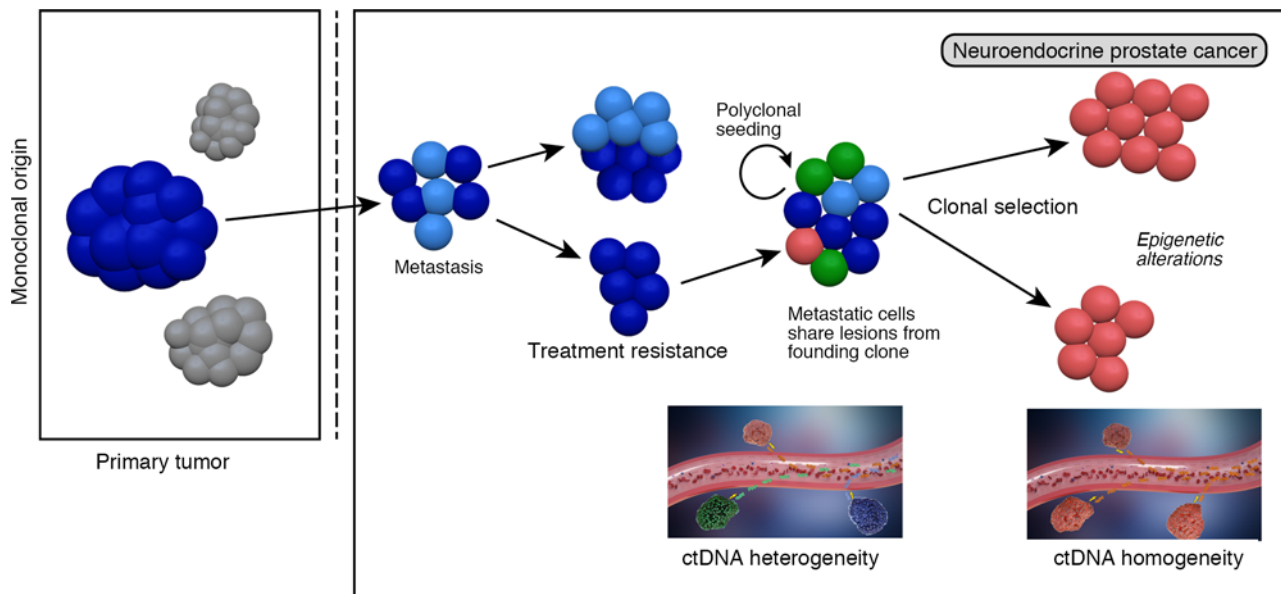
Before profiling the whole cfDNA study cohort, we studied the effect of input amount for cfDNA on the sequencing-based genomic profiles. One patient with known genotype features (WCM163) was selected for this experiment. Specifically, the cfDNA was diluted into 5 concentrations using 5 ng, 10 ng, 20 ng, 50 ng, and 100 ng. For matched germline DNA, 100 ng was used for all comparisons. All downstream library preparation and sequencing were conducted as stated above. Somatic copy number calls of approximately 1000 cancer-associated genes were compared (Supplemental Figure 9, A and B).

#### WES processing pipeline

*Data preprocessing.* Tumor tissue BAM files were generated through the WCM Englander Institute of Precision Medicine pipeline (49) and preprocessed at the University of Trento. The FastQC tool ([www.bioinformatics.babraham.ac.uk/projects/fastqc](http://www.bioinformatics.babraham.ac.uk/projects/fastqc)) was run on the raw reads of cfDNA samples and matched germline DNA samples to assess their quality. Quality metrics include average base quality, sequence duplication rate, and the k-mer enrichment along the length of the reads. These measures were used to assess whether the sequencing and the de-multiplexing of the samples were performed correctly. After initial quality control, adapter sequences were trimmed using Trimmomatic (50). Short reads were then aligned to GRC37/hg19 reference using Burrows Wheeler alignment (51). Picard (<http://broadinstitute.github.io/picard/>) and SAMtools (52) were used to generate single-sorted BAM files for each patient sample. BAM files were then realigned (to correct for possible misalignments due to indels) and recalibrated (to adjust for over- and underestimated base quality scores in the data) using GATK standard pipelines (53). Finally, SAMtools were used to adjust BAM MD tags (strings for mismatch positions). The alignment quality of the BAM files was obtained by several metrics related to the average coverage and capture rate to calculate how many aligned reads fall within a capture region of the Nimblegen SeqCap EZ Exome V3 kit. For any given sample, the capture rate was given by the percentage of mapped reads that overlap any capture region in the kit and the total number of mapped reads. Average coverage was computed from the captured regions of the Nimblegen kit. Metrics from 62 processed patients' data showed that average coverages for plasma DNA and germline DNA were  $\times 361$  and  $\times 109$ , respectively; average capture rates for plasma DNA and germline DNA were 81% and 77%, respectively (Supplemental Table 3). All subsequent steps were applied to all samples, tissues, and cfDNA, unless differently specified.

*Identity check for individual samples.* In order to verify the correct matches for individual samples, we applied a solid genotype distance-based test, SPIA (54), exploiting the related R package SPI-Assay. Genotypes of 334 selected SNPs were computed using ASEQ (55). The 334 SNPs were previously selected based on the minor allele frequency (MAF) and uniform distribution across the genome.

*Optimized de-duplication procedure for WES plasma data from single-end protocol and with deep coverage.* WES data generated from both germline and cfDNA study samples showed high fractions of read duplicates in the sample cohort (mean 59% and 79%, respectively). Anticorrelation between duplication percentage and input DNA was observed ( $-0.83$  Pearson coefficient,  $P = 9 \times 10^{-4}$ ). Duplication levels in off-target regions were comparable to the on-target region. Provided that without using specific technologies (e.g., molecular barcoding) it is not possible to measure and quantify the exact proportions of PCR and natural duplication (due to high coverage), we opted for an ad hoc computational solution that limits the impact of nonnatural read duplicates in downstream copy number analyses. First, we measured the empirical distribution of read duplicates at different coverage intervals from all samples of our cohort. Then, for each coverage interval, we collected the 50th, 75th, and 90th percentiles and used these percentiles across all coverage intervals to create 3 different duplication thresholds (Supplemental Figure 9C). Finally, we applied these thresholds when the coverage statistics at a single base resolution were computed for downstream analysis. Specifically, when coverage statistics are computed for a position  $x$  in the genome, the raw



**Figure 6. Proposed model of prostate cancer progression toward CRPC-NE.** Metastatic prostate cancer lesions harbor shared alterations that are often traceable back to a primary tumor clone, supporting a monoclonal origin of metastatic prostate cancer. Tumors acquire alterations with disease progression and treatment resistance, and these alterations may be subclonal. Polycyclonal spread in later stages further leads to intrapatient heterogeneity. During the transition toward CRPC-NE, there is likely selection of a dominant clone that persists. Intraindividual tumor heterogeneity and tumor homogeneity are captured through ctDNA analyses. DNA methylation profiles dramatically shift with CRPC-NE.

coverage  $C$  is first computed considering all reads and is then normalized considering the duplication threshold  $T$  specific for  $C$ . Coverage statistics were computed for all 3 thresholds (i.e., 50th, 75th, and 90th percentiles). Coverage distributions obtained applying the different duplication thresholds are shown in Supplemental Figure 9D.

**WES segmentation.** Segmentation of patient samples was performed using the recent tool FACETS (56), a computational method that extends the standard circular binary segmentation (CBS) algorithm to a joint segmentation that combines read counts and SNP allelic fraction data. To deal with plasma samples, the preprocessing module of FACETS was extended, including our de-duplication procedure, and segmentation was performed considering separately all 3 thresholds (50th, 75th, and 90th percentiles) and complete de-duplication. Comparison of segmentation results on a large cancer gene list ( $N = 920$ ) showed that although the segmentation signal from de-duplicated samples strongly correlates with the raw segmentation signal (where no de-duplication was applied) in all cases (Supplemental Figure 9E), the stronger the de-duplication level applied, the higher the divergence of the signal from the diagonal, indicating a loss of detection power. Based on this observation we decided to use 75th percentile threshold for cfDNA sample WES segmentation and related somatic copy number downstream analyses (Supplemental Table 9). Study cohort WES-segmented data were adjusted for tumor ploidy and TC, revealing an overall plasma signal similar to that for tissue-based data (Figure 1B), with peaks distinctive of mono- and biallelic deletions and gains of 1, 2, 3 or more gene copies.

**Tumor ploidy and tumor content assessment.** To assess tumor ploidy and TC (estimation of cfDNA tumor fraction, ctDNA) from plasma DNA samples, we used an extended version of CLONET (28), a tool we developed to deal with highly heterogeneous tissue samples. It now embeds a refined mathematical approach for local TC estimation (25) to further increase the performance in computation time and in the

ability to detect somatic aberrations in low TC samples. Specifically, a new method for the calculation of CLONET  $\beta$  values (i.e., the percentage of reads from cells harboring 2 alleles) was implemented and used in this work. For each segment  $S$ , previously identified, we computed the corresponding  $\beta$  value as follows:  $\beta = \max\{i \mid W(d > D_i)\} + (\max\{i \mid W(d < D_i)\} - \min\{i \mid W(d > D_i)\}) \times P$ , where  $P = (\text{median}(d) - \min(d) / (\max(d) - \min(d)))$  and  $W(d > D_i)$  represents the Mann-Whitney statistics (using significance cutoff at 1%) comparing the mirrored AF distribution  $d$  distribution of all informative SNPs in segment  $S$  (i.e., SNPs with heterozygous genotype in germline patient's sample) and the reference distribution  $D_i$  simulating a monoallelic deletion with  $\beta = i$  (where  $i \in [0,1]$ ) and AF noise estimated from the germline patient's samples at SNP positions. We applied the extended CLONET pipeline to estimate TC and ploidy of all cfDNA and tumor tissue samples. CLONET analysis estimated TC for 55 of 69 plasma samples (80% of the total): 30 samples with TC  $\geq 20\%$  (43% of the total), 43 samples with TC  $\geq 10\%$  (62% of the total) and 12 samples with TC  $< 10\%$ . Manual inspection of the somatic copy number profiles of the remaining 15 samples suggest low or absent tumor signal. Ploidy estimates were available for 68 samples with 12 samples having a ploidy  $\geq 2.5$ . For tumor tissue samples, CLONET analysis estimated TC for 90 of 98 samples: 85 samples with TC  $\geq 20\%$  (87% of the total), 87 samples with TC  $\geq 10\%$  and 1 sample with TC  $< 10\%$ . Ploidy estimates were available for 96 samples, with 34 samples having a ploidy  $\geq 2.5$ . CLONET allele-specific somatic copy number profiles were computed for all plasma and tissue samples at gene-level resolution, using a gene model consisting of 19,027 genes (Supplemental Table 9).

#### Detection of somatic single nucleotide variants

To identify and characterize somatic single-nucleotide variants (SNVs) in WES captured regions, we run MuTect (57) on BAM files where read duplicates were removed using Picard (<http://broadinstitute.github>).

io/picard). Putative somatic SNVs were nominated as genomic positions called by MuTect and having in the plasma/tumor sample a depth of coverage at least  $\times 30$ , an allelic fraction  $\geq 0.05$  and at least 2 reads supporting the alternative base, and having an allelic fraction  $< 0.01$  in the matched normal sample. Oncotator (58) was finally used to annotate retained SNVs with variant- and gene-centric information relevant to cancer. A complete annotated list of called missense SNVs is available in Supplemental Table 10.

### Similarity measure based on copy number profiles

For each patient, we compared the somatic copy number aberration (SCNA) profiles of tumor tissues and cfDNA samples (Supplemental Table 9). Comparison was performed using an ad hoc notion of similarity, which exploits TC corrected SCNA profiles (Figure 1B) and measures the percentage of concordant aberration signal. Specifically, given a sample  $P$  and a sample  $T$ , the algorithm computes the following steps: (a) SCNA profile signal of sample  $P$  is centered around the  $P$  mean signal; (b) SCNA profile signal of sample  $T$  is centered around the  $T$  mean signal; (c) SCNA signals of  $P$  and  $T$  samples are synchronized around their median difference to avoid the presence of systematic signal shifts due to technical or processing bias; specifically, SCNAs signal  $P$  is normalized as  $P = P - \text{median}(P - T)$ ; (d)  $S_{\text{loss}}$ , a measure of loss similarity, is obtained by measuring the fraction of genes having signal below the detection threshold  $-THR$  in both  $P$  and  $T$  profiles over the total number of genes having a signal below the detection threshold  $-THR$  in at least 1 of the 2 samples; (e)  $S_{\text{gain}}$ , a measure of gain similarity, is obtained by measuring the fraction of genes having signal above the detection threshold  $THR$  in both  $P$  and  $T$  profiles over the total number of genes having a signal above the detection threshold  $THR$  in at least 1 of the 2 samples; and (f) a global similarity measure is obtained as  $S = (S_{\text{loss}} + S_{\text{gain}}) / 2$ .

A model with 19,027 genes was considered along with a detection threshold  $THR$  equal to 0.3.

### Similarity measure based on SNV profiles

For each patient and each plasma and tumor tissue sample's pair private and shared somatic SNVs were then compiled. For each patient, an inclusion matrix considering all plasma-tissue and tissue-tissue pairs was computed, including nonsynonymous SNVs only. Specifically, for each pair ( $P1$ ,  $P2$ ) of samples the inclusion matrix reports both the fraction of mutations detected in  $P1$  that are also detected in  $P2$  and the fraction of mutations detected in  $P2$  that are also detected in  $P1$ .

### Estimation of clonality divergence

Nondiscretized estimations of allele-specific copy number analysis (Supplemental Table 9) were used to calculate the Clonality Divergence Index. For each sample and each genomic segment, the Euclidean difference between the estimated raw allele-specific copy number and the closest expected clonal allele-specific copy number was assessed. The average of the minimum distances is the Clonality Divergence Index. Formally, considering a sample  $P$  and segment  $S$  and its estimated allele-specific copy number coordinates ( $cnA$ ,  $cnB$ ), in a Euclidean space with  $cnA$  and  $cnB$  real numbers, we compute the value:

$$D_s = \min_{e \in C} \overline{ep}$$

(Equation 1)

Where  $\overline{ep}$  is the Euclidian distance of  $p = (cnA, cnB)$  and a point in  $C = \{(a,b) \mid a,b \in M\}$ , with  $M$  equal to the least integer greater than or equal to the maximum raw  $cnA$  or  $cnB$  estimated across all  $P$  segments. The Clonality Divergence Index of  $P$  is the mean of all  $D_s$  computed for all segments  $S$  of  $P$ . A sample with all and only segments with perfect integer ( $cnA$ ,  $cnB$ ) would have a Clonality Divergence Index equal to zero.

### Whole-genome bisulfite sequencing (WGBS) data generation and processing

cfDNA (5 ng) and germline DNA (100 ng) were sonicated using a Covaris S220 to approximately 180–220 bp (Covaris) and bisulfite converted using the EZ DNA Methylation–Gold Kit (catalog D5005, Zymo Research Corporation). The single-stranded DNA obtained was processed for library construction using the Accel-NGS@Methyl-seq DNA Library kit (catalog 36024) as per manufacturer instructions (Swift BioSciences). Briefly, truncated adapter sequences were incorporated to the single-stranded DNA in a template-independent reaction through sequential steps using the Adaptase module (Swift BioSciences). DNA was then enriched using PCR with primers compatible with Illumina sequencing, 9 cycles for cfDNA, and 6 cycles for genomic DNA. The libraries were clustered at 12 pM on a pair-end read flow cell and sequenced for 125 cycles on Illumina HiSeq 2500 or 4000.

Primary processing of sequencing images was done using Illumina's Real Time Analysis software (RTA). CASAVA 1.8.2 software was then used to demultiplex samples and generate raw reads and respective quality scores. The WGBS raw data was quality filtered and adapter trimmed using Flexbar with the following parameters: minimal overlap of adaptor and read sequence, 6; minimal read length after adaptor removal/trimming, 21; and allowed mismatches and gaps per 10bp overlap, 2. Reads were aligned to unmasked human genome build GRCh37/hg19 and methylation calls were generated with Bismark (59) as described in the data analysis section of the Weill Cornell Epigenomics Core in-house bisulfite sequencing analysis pipeline (60). The average conversion rate in WGBS samples is 99.6%, the average CpG coverage is  $\times 14.3$ , and the average mapping efficiency is 76% (Supplemental Table 8). For plasma-related analysis, only CpG sites covered by at least 10 reads and read in at least 10% of the study samples were considered for downstream analysis. For each sample, the percentage of methylation per site ( $\beta$  value) was computed. ctDNA fraction was estimated by purity assessment from clonal methylation sites (PAMES), using 10 hypermethylated prostate-specific CpG islands (30). CpG wise differential methylation analysis (CRPC-NE versus CRPC-Adeno) was performed by AUC analysis. Hyper- and hypomethylated sites were identified as those sites demonstrating  $AUC = 1$  and  $AUC = 0$ , respectively. Genomic annotation of methylation sites was performed by the tool annotatePeaks included in the HOMER package (Supplemental Table 11) (61).

### Concordance of differential methylation in matched plasma and tissue samples

DMRs were nominated using an in-house algorithm. The method is divided into 3 main modules. Given methylation values for each CpG site in a control and a test set of samples, the following procedure was applied: (a) for each detected site, the area under the receiver operator characteristics (AUROC) was computed, evaluating the segregation between groups based solely on target CpG signal; (b) sequential AUROC values were divided into segments with a common methyla-

tion status (hypermethylated, neutral, or hypomethylated) following a 3-state heterogeneous shifting level model, adapted from Magi et al. (62); (c) significance of differential methylation between control and test group was assessed for each DMR with Wilcoxon Mann-Whitney  $U$  test on average  $\beta$  values, and  $P$  values are corrected for multiple hypothesis testing with standard Benjamini-Hochberg procedure. We named this algorithm Rocker-Meth (Receiver operating characteristic curve analyzer of DNA methylation data; available at <https://github.com/cgplab/ROckerMeth>).

Starting from published enhanced reduced representation bisulfite sequencing (eRRBS) of solid tumor biopsy samples and healthy prostate tissues (11, 63), we performed 2 orthogonal differential methylation analyses using default parameter values with the exception of  $na\_threshold$  set to 0.3. First, we identified DMRs between primary prostate cancer tissues (test,  $n = 7$ ) and normal prostatic tissue (control,  $n = 7$ ) (PCa|NT). Second, we compared CRPC-NE (test,  $n = 10$ ) with CRPC-Adeno (control,  $n = 18$ ), obtaining a set of DMRs that reflects methylation changes upon transdifferentiation (CRPC-NE|CRPC-Adeno). FDR of 0.01, an absolute average  $\beta$  difference of 20 between test and control, and the presence of at least 6 detected CpG sites inside the region were applied as filters. Shared DMRs were isolated using Bedtools multiinter clustering (64) and discarded. To capture single-sample alterations of DNA methylation,  $Z$  scores for each DMR in each pair of tissue plasma-matched samples were computed using the following equation:

$$Z_{i,z} = \frac{(\bar{\beta}_i - \text{median}_j(\bar{\beta}_j))}{\text{mad}_j(\bar{\beta}_j)} \quad (\text{Equation 2})$$

Where  $i$  and  $j$  indicate the index over tumor and normal samples,  $z$  is the DMR index,  $(\bar{\beta}_i)$  and  $(\bar{\beta}_j)$  are the averaged (i.e., median) beta values of CpG sites within a DMR  $z$  for tumor and normal samples, respectively, and  $\text{median}_j$  and  $\text{mad}_j$  are the median and the maximum absolute deviation of the  $(\bar{\beta}_j)$  across normal samples  $j$ . Normal prostatic tissue was used as reference. To control for potential sources of noise due to nontumoral cfDNA, we also profiled the methylome of PBMCs of all 11 patients using WGBS sequencing. For both DMR sets we evaluated the  $Z$  scores detected across PBMC samples and excluded the DMRs that consistently deviated ( $\text{abs}(Z \text{ score}) > 5$ ) from the reference in more than 80% (at least 9 of 11) of PBMC samples. A total of 7606 and 3843 DMRs were obtained for PCa|NT and CRPC-NE|CRPC-Adeno, respectively (Supplemental Table 10). To maximize the compatibility of different assays, only CpG sites detected in at least more than half of plasma samples (ctDNA and PBMC) were retained in single-sample  $Z$  score computation, excluding sites that were detected only in eRRBS tissue samples. Furthermore, all DMRs containing less than 6 detected CpG sites in WGBS samples or with an undetectable signal in 1 or more WGBS plasma samples were excluded from pairwise tissue-plasma analysis and DMR score computation to ensure consistency between samples.

*Integration of plasma methylomes from an independent CRPC cohort.* To evaluate the robustness of our findings, we analyzed a published cohort of cfDNA methylomes of patients with CRPC treated with abiraterone and prednisone (31) generated with the HM450 array, including only end-of-treatment time point data to maximize the similarity in disease state and clinical history. Using the previ-

ously defined set of CRPC-NE|CRPC-Adeno DMRs, we computed single-sample  $Z$  scores for end-of-treatment time points (using normal tissue samples from TCGA-PRAD as reference) (65). In order to obtain a comparable set of samples, only samples with TC (30) above or equal to 10% were retained in the analysis using a threshold similar to prior WES ctDNA studies (24).

### Neuroendocrine Prostate Cancer (NEPC) score

The NEPC score includes genomic- and methylation-based features. Genomic features include *RBI* deletion or mutation, *TP53* deletion or mutation, *AR* focal gain or mutation, and *CYLD* deletion. *TP53*, *RBI*, and *CYLD* genomic-based features for the NEPC feature score are set to 1 for a specific sample if the aberration is present; *AR* genomic feature is instead set to 1 for a specific sample in case of absence of aberrations. Methylation-based signal includes 2 components based on top 20 hypo- and 20 hyper-methylated sites based on previous tissue analysis (Supplemental Table 11). Briefly, we ranked differentially methylated sites (Supplemental Table 12, ref. 11) by FDR and by  $\beta$  difference (CRPC-NE vs CRPC-Adeno) and retained only those sites found in WGBS data. To avoid redundancy, we also required that each informative CpG site was not within 10 kilobases from all other selected sites. Next, to have comparable methylation levels from tissue and plasma samples, a purity correction of  $M$  values was applied using PAMES. Then, for each top hyper- and hypomethylated site, we identified the threshold that best discriminated between CRPC-Adeno and CRPC-NE tissue samples by receiver operating characteristic (ROC) curve analysis. Last, plasma methylation data were dichotomized as follows: for each plasma sample, methylation feature score is equal to 1 if the majority of sites show methylation level above (hyper) or below (hypo) the site-specific, tissue-based threshold. The NEPC feature score was computed for a specific sample by counting the number of features set to 1 and normalizing by the number of available features for that sample. Features with no call were set to NA and were not considered in the score computation.

### Statistics

Association of plasma genomics and binary clinical variables was performed using 2-tailed Wilcoxon Mann-Whitney  $U$  test with significance level set at 5%. Association of plasma genomics and continuous clinical variables was performed using Pearson correlation statistics with significance level set at 5%. Statistical comparison of genomic variables across a sample's classes was performed using 2-tailed Wilcoxon Mann-Whitney  $U$  test with significance level set at 5%. Univariate overall survival and progression-free survival analyses were performed using the Kaplan-Meier estimator (log-rank test). Multivariate overall and progression-free survival analyses were performed using a proportional hazard model with stepwise model selection by Akaike information criterion using forward and backward directions.

### Study approval

This study was approved by the Weill Cornell Medicine IRB (nos. 1305013903, 1210013164). Written informed consent was obtained from participants before inclusion in the study.

### Author contributions

HB, AR, FD initiated and designed the study. HB, VC, JMM, DMN, and STT enrolled patients and contributed samples and clinical

data. MS, AA, JM, and JX performed experiments. AR, NC, GMF, FO, TF, SYK, ED, AS, OE, MB, and FD performed statistical and computational analyses. HB and FD supervised the research. All authors contributed to the writing and editing of the manuscript and approved the manuscript. The order of the co-first authors was determined by their complementary contributions and the balance of institutional representation on the manuscript.

## Acknowledgments

This study was supported by grants from the Prostate Cancer Foundation (to HB, FD), National Cancer Institute (SPORE P50-CA211024 to HB and FD and 2R01CA125612-05A1 to FD), Department of Defense (PCRP W81XWH-13-1 to HB), and the Ann and William Bresnan Foundation (to HB and DMN). This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research

and innovation programme (grant agreement no. 648670). All BAM files generated for this study and associated sample information are described in Supplemental Tables 3 and 6, and are accessible through dbGaP (phs001752.v1.p1) at [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001752.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001752.v1.p1).

Address correspondence to: Himisha Beltran, Department of Medical Oncology, Dana Farber Cancer Institute, 1 Jimmy Fund Way, Smith 758, Boston, Massachusetts 02115, USA. Phone: 617.582.9421; Email: [himisha\\_beltran@dfci.harvard.edu](mailto:himisha_beltran@dfci.harvard.edu). Or to: Alessandro Romanel or Francesca Demichelis, Department of Cellular, Computational and Integrative Biology, University of Trento, Via Sommarive 9, 38123 Trento, Italy. Phone: 39.0461.285217; Email: [alessandro.romanel@unitn.it](mailto:alessandro.romanel@unitn.it) (A. Romanel). Phone: 39.0461.285305; Email: [f.demichelis@unitn.it](mailto:f.demichelis@unitn.it) (F. Demichelis).

- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69(1):7-34.
- Watson PA, Arora VK, Sawyers CL. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat Rev Cancer*. 2015;15(12):701-711.
- Beltran H, et al. The role of lineage plasticity in prostate cancer therapy resistance. *Clin Cancer Res*. 2019;25(23):6916-6924.
- Abida W, et al. Genomic correlates of clinical outcome in advanced prostate cancer. *Proc Natl Acad Sci U S A*. 2019;116(23):11428-11436.
- Aggarwal R, et al. Clinical and genomic characterization of treatment-emergent small-cell neuroendocrine prostate cancer: a multi-institutional prospective study. *J Clin Oncol*. 2018;36(24):2492-2503.
- Epstein JJ, et al. Proposed morphologic classification of prostate cancer with neuroendocrine differentiation. *Am J Surg Pathol*. 2014;38(6):756-767.
- Beltran H, et al. Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. *Cancer Discov*. 2011;1(6):487-495.
- Conteduca V, et al. Clinical features of neuroendocrine prostate cancer. *Eur J Cancer*. 2019;121:7-18.
- Rickman DS, Beltran H, Demichelis F, Rubin MA. Biology and evolution of poorly differentiated neuroendocrine tumors. *Nat Med*. 2017;23(6):1-10.
- George J, et al. Comprehensive genomic profiles of small cell lung cancer. *Nature*. 2015;524(7563):47-53.
- Beltran H, et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat Med*. 2016;22(3):298-305.
- Tan HL, et al. Rb loss is characteristic of prostatic small cell neuroendocrine carcinoma. *Clin Cancer Res*. 2014;20(4):890-903.
- Zou M, et al. Transdifferentiation as a mechanism of treatment resistance in a mouse model of castration-resistant prostate cancer. *Cancer Discov*. 2017;7(7):736-749.
- Ku SY, et al. Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. *Science*. 2017;355(6320):78-83.
- Mu P, et al. SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. *Science*. 2017;355(6320):84-88.
- Park JW, et al. Reprogramming normal human epithelial tissues to a common, lethal neuroendocrine cancer lineage. *Science*. 2018;362(6410):91-95.
- Berger A, et al. N-Myc-mediated epigenetic reprogramming drives lineage plasticity in advanced prostate cancer. *J Clin Invest*. 2019;130:3924-3940.
- Kumar A, et al. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. *Nat Med*. 2016;22(4):369-378.
- Yegnasubramanian S, et al. Hypermethylation of CpG islands in primary and metastatic human prostate cancer. *Cancer Res*. 2004;64(6):1975-1986.
- Aryee MJ, et al. DNA methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases. *Sci Transl Med*. 2013;5(169):169ra10.
- Carreira S, et al. Tumor clone dynamics in lethal prostate cancer. *Sci Transl Med*. 2014;6(254):254ra125.
- Gundem G, et al. The evolutionary history of lethal metastatic prostate cancer. *Nature*. 2015;520(7547):353-357.
- Wyatt AW, et al. Concordance of circulating tumor DNA and matched metastatic tissue biopsy in prostate cancer. *J Natl Cancer Inst*. 2017;109(12):d1x118.
- Adalsteinsson VA, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun*. 2017;8(1):1324.
- Romanel A, et al. Plasma AR and abiraterone-resistant prostate cancer. *Sci Transl Med*. 2015;7(312):312re10.
- Azad AA, et al. Androgen receptor gene aberrations in circulating cell-free DNA: biomarkers of therapeutic resistance in castration-resistant prostate cancer. *Clin Cancer Res*. 2015;21(10):2315-2324.
- Annala M, et al. Circulating tumor DNA genomics correlate with resistance to abiraterone and enzalutamide in prostate cancer. *Cancer Discov*. 2018;8(4):444-457.
- Prandi D, Demichelis F. Ploidy- and purity-adjusted allele-specific DNA analysis using CLONETv2. *Curr Protoc Bioinformatics*. 2019;67(1):e81.
- Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature*. 2007;447(7143):433-440.
- Benelli M, Romagnoli D, Demichelis F. Tumor purity quantification by clonal DNA methylation signatures. *Bioinformatics*. 2018;34(10):1642-1649.
- Gordevičius J, et al. Cell-free DNA modification dynamics in abiraterone acetate-treated prostate cancer patients. *Clin Cancer Res*. 2018;24(14):3317-3324.
- Armenia J, et al. The long tail of oncogenic drivers in prostate cancer. *Nat Genet*. 2018;50(5):645-651.
- Robinson D, et al. Integrative clinical genomics of advanced prostate cancer. *Cell*. 2015;161(5):1215-1228.
- Quigley DA, et al. Genomic hallmarks and structural variation in metastatic prostate cancer. *Cell*. 2018;175(3):889.
- Antonarakis ES, et al. AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N Engl J Med*. 2014;371(11):1028-1038.
- Viswanathan SR, et al. Structural alterations driving castration-resistant prostate cancer revealed by linked-read genome sequencing. *Cell*. 2018;174(2):433-447.e19.
- Henzler C, et al. Truncation and constitutive activation of the androgen receptor by diverse genomic rearrangements in prostate cancer. *Nat Commun*. 2016;7:13668.
- Bluemn EG, et al. Androgen receptor pathway-independent prostate cancer is sustained through FGF signaling. *Cancer Cell*. 2017;32(4):474-489.e6.
- Lee JK, et al. Clonal history and genetic predictors of transformation into small-cell carcinomas from lung adenocarcinomas. *J Clin Oncol*. 2017;35(26):3065-3074.
- Offin M, et al. Concurrent RB1 and TP53 Alterations define a subset of EGFR-mutant lung cancers at risk for histologic transformation and inferior clinical outcomes. *J Thorac Oncol*. 2019;14(10):1784-1793.

41. Aparicio AM, et al. Platinum-based chemotherapy for variant castrate-resistant prostate cancer. *Clin Cancer Res.* 2013;19(13):3621-3630.
42. Aparicio AM, et al. Combined tumor suppressor defects characterize clinically defined aggressive variant prostate cancers. *Clin Cancer Res.* 2016;22(6):1520-1530.
43. Haffner MC, et al. Tracking the clonal origin of lethal prostate cancer. *J Clin Invest.* 2013;123(11):4918-4922.
44. Liu W, et al. Copy number analysis indicates monoclonal origin of lethal metastatic prostate cancer. *Nat Med.* 2009;15(5):559-565.
45. Hong MK, et al. Tracking the origins and drivers of subclonal metastatic expansion in prostate cancer. *Nat Commun.* 2015;6:6605.
46. Burrell RA, Swanton C. Tumour heterogeneity and the evolution of polyclonal drug resistance. *Mol Oncol.* 2014;8(6):1095-1111.
47. Parikh AR, et al. Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers. *Nat Med.* 2019;25(9):1415-1421.
48. Beltran H, et al. A phase II trial of the aurora kinase a inhibitor alisertib for patients with castration-resistant and neuroendocrine prostate cancer: efficacy and biomarkers. *Clin Cancer Res.* 2019;25(1):43-51.
49. Beltran H, et al. Whole-exome sequencing of metastatic cancer and biomarkers of treatment response. *JAMA Oncol.* 2015;1(4):466-474.
50. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-2120.
51. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760.
52. Li H, et al. The sequence alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-2079.
53. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-1303.
54. Demicheli F, et al. SNP panel identification assay (SPIA): a genetic-based assay for the identification of cell lines. *Nucleic Acids Res.* 2008;36(7):2446-2456.
55. Romanel A, Lago S, Prandi D, Sboner A, Demicheli F. ASEQ: fast allele-specific studies from next-generation sequencing data. *BMC Med Genomics.* 2015;8:9.
56. Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* 2016;44(16):e131.
57. Cibulskis K, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol.* 2013;31(3):213-219.
58. Ramos AH, et al. Oncotator: cancer variant annotation tool. *Hum Mutat.* 2015;36(4):E2423-E2429.
59. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics.* 2011;27(11):1571-1572.
60. Garrett-Bakelman FE, et al. Enhanced reduced representation bisulfite sequencing for assessment of DNA methylation at base pair resolution. *J Vis Exp.* 2015;(96):e52246.
61. Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010;38(4):576-589.
62. Magi A, et al. EXCAVATOR: detecting copy number variants from whole-exome sequencing data. *Genome Biol.* 2013;14(10):R120.
63. Lin PC, et al. Epigenomic alterations in localized and advanced prostate cancer. *Neoplasia.* 2013;15(4):373-383.
64. Quinlan AR. BEDTools: The Swiss-Army tool for genome feature analysis. *Curr Protoc Bioinformatics.* 2014;47:11.12.1-11.1234.
65. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell.* 2015;163(4):1011-1025.