

Noninvasive Detection of Neuroendocrine Prostate Cancer through Targeted Cell-free DNA Methylation

Gian Marco Franceschini¹, Orsetta Quaini¹, Kei Mizuno², Francesco Orlando¹, Yari Ciani¹, Sheng-Yu Ku², Michael Sigouros³, Emily Rothmann², Alicia Alonso³, Matteo Benelli⁴, Caterina Nardella¹, Joonghoon Auh³, Dory Freeman², Brian Hanratty⁵, Mohamed Adil⁵, Olivier Elemento³, Scott T. Tagawa⁶, Felix Y. Feng⁷, Orazio Caffo⁸, Consuelo Buttigliero⁹, Umberto Basso¹⁰, Peter S. Nelson⁵, Eva Corey¹¹, Michael C. Haffner^{5,11}, Gerhardt Attard¹², Ana Aparicio¹³, Francesca Demichelis¹, and Himisha Beltran²

ABSTRACT

Castration-resistant prostate cancer (CRPC) is a heterogeneous disease associated with phenotypic subtypes that drive therapy response and outcome differences. Histologic transformation to castration-resistant neuroendocrine prostate cancer (CRPC-NE) is associated with distinct epigenetic alterations, including changes in DNA methylation. The current diagnosis of CRPC-NE is challenging and relies on metastatic biopsy. We developed a targeted DNA methylation assay to detect CRPC-NE using plasma cell-free DNA (cfDNA). The assay quantifies tumor content and provides a phenotype evidence score that captures diverse CRPC phenotypes, leveraging regions to inform transcriptional state. We tested the design in independent clinical cohorts ($n = 222$ plasma samples) and qualified it achieving an AUC > 0.93 for detecting pathology-confirmed CRPC-NE ($n = 136$). Methylation-defined cfDNA tumor content was associated with clinical outcomes in two prospective phase II clinical trials geared towards aggressive variant CRPC and CRPC-NE. These data support the application of targeted DNA methylation for CRPC-NE detection and patient stratification.

SIGNIFICANCE: Neuroendocrine prostate cancer is an aggressive subtype of treatment-resistant prostate cancer. Early detection is important, but the diagnosis currently relies on metastatic biopsy. We describe the development and validation of a plasma cell-free DNA targeted methylation panel that can quantify tumor fraction and identify patients with neuroendocrine prostate cancer noninvasively.

INTRODUCTION

Castration-resistant prostate cancer (CRPC) is a heterogeneous disease in which treatment resistance can arise through multiple mechanisms (1–3). Although the majority of CRPC is driven by androgen receptor (AR) signaling, up to 15% to 20% of patients develop AR independence (4). AR independence has been associated with aggressive clinical features and changes in tumor phenotype, including histologic transformation from a castration-resistant adenocarcinoma (CRPC-Adeno) to neuroendocrine prostate cancer (CRPC-NE) arising

through lineage plasticity and divergent clonal evolution (3, 5–7). Patients with CRPC-NE are often managed aggressively with chemotherapy regimens similar to small-cell lung cancer (SCLC), and several CRPC-NE-directed clinical trials are also ongoing. The current diagnosis of CRPC-NE remains challenging because of the need for metastatic biopsy as well as inpatient tumor heterogeneity.

DNA sequencing of plasma cell-free DNA (cfDNA) is a noninvasive tool to detect somatic alterations in the cancer (8). However, cancer-specific mutations or copy-number changes are only modestly enriched in CRPC-NE compared with CRPC-Adeno (3, 9). Conversely, we and others have observed extensive DNA methylation changes associated with CRPC-NE (3, 10), and such changes can be detected in cfDNA (11, 12). DNA methylation in humans is mostly present at CpG dinucleotides and is associated with a wide range of biological processes, including regulation of gene expression, cell fate, and genomic stability (13). Furthermore, DNA methylation is highly tissue-specific and provides a robust signal to deconvolve the tissue of origin (14, 15), thereby allowing the enhanced detection of low cancer fractions in the circulation (16, 17), and has been successfully applied for early detection and monitoring (18, 19). As previously reported, DNA methylation at base resolution can be measured with bisulfite sequencing, which provides a fraction of methylated cytosines for each covered CpG in the form of a β value ranging from 0 (no methylation) to 1 (complete methylation). In principle, approaches such as whole-genome bisulfite cfDNA sequencing (WGBS) may offer a comprehensive picture of the patient's disease status with optimal information on methylation content. In practice, only low-pass variants of WGBS are suitable for large-scale clinical studies, given the cost of high-depth whole-genome sequencing. Low-pass sequencing suffers from low granularity and captures all regions at coarse resolution. Given that most CpG sites in this context are likely noninformative or highly redundant, we aimed to reduce the sequencing space to a minimal set

¹Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Trento, Italy. ²Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts. ³Institute for Computational Biomedicine and Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, New York. ⁴Bioinformatics Unit, Hospital of Prato, Prato, Italy. ⁵Fred Hutchinson Cancer Research Center, Seattle, Washington. ⁶Department of Medicine, Division of Medical Oncology, Weill Cornell Medicine, New York, New York. ⁷Department of Radiation Oncology, University of California San Francisco, San Francisco, California. ⁸Department of Medical Oncology, Santa Chiara Hospital, Trento, Italy. ⁹Department of Oncology, University of Turin, San Luigi Gonzaga Hospital, Orbassano, Turin, Italy. ¹⁰Department of Oncology, Istituto Oncologico Veneto IOV - IRCCS, Padua, Italy. ¹¹University of Washington, Seattle, Washington. ¹²Cancer Institute and University College London Hospitals, University College London, London, United Kingdom. ¹³Department of GU Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas.

F. Demichelis and H. Beltran contributed equally to this article.

Corresponding Authors: Francesca Demichelis, Department of Cellular, Computational and Integrative Biology, University of Trento, Via Sommarive 9, Trento 38123, Italy. E-mail: f.demichelis@unitn.it; and Himisha Beltran, Dana Farber Cancer Institute, 450 Brookline Avenue, DA1420C, Boston, MA 02215. E-mail: Himisha_Beltran@dfci.harvard.edu

Cancer Discov 2024;14:1–22

doi: 10.1158/2159-8290.CD-23-0754

This open access article is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

©2023 The Authors; Published by the American Association for Cancer Research

of CpGs that can accurately probe advanced prostate cancer phenotypes of interest, providing insights into the biologic disease state.

In this work, we present the NEuroendocrine detection and MOnitoring (NEMO) assay, a targeted DNA methylation sequencing panel designed for CRPC disease monitoring and neuroendocrine phenotype detection. We selected informative regions by performing an extensive analysis of published datasets from CRPC metastatic tumors (pathologically confirmed CRPC-Adeno and CRPC-NE samples), white blood cells (WBC), and plasma cfDNA from healthy individuals. First, we demonstrated that 53 CpG clusters are sufficient to accurately estimate the fraction of cancer-derived DNA in cfDNA (tumor content, TC), which we orthogonally validated through genomic-based tumor content estimation and *in vitro* dilutions. Next, we leveraged the DNA methylation of the collection of genomic regions that distinguish CRPC-NE from CRPC-Adeno while accounting for the variable tumor content in cfDNA. We showcase the capability of the NEMO design to detect CRPC-NE in independent datasets, including cell lines, organoids, patient-derived xenografts, and patient cohorts. The clinical utility of NEMO is demonstrated in its application to clinical cohorts from the PRIME consortium, Dana-Farber Cancer Institute (DFCI), Weill Cornell Medicine (WCM), and two completed clinical trials for aggressive variant CRPC: a phase II trial of the aurora kinase A inhibitor alisertib (20) and a phase II trial of carboplatin plus docetaxel chemotherapy (21). A binary classifier to detect pathology-confirmed CRPC-NE cases achieved an AUC of 0.93 [confidence interval (CI), 0.88–0.99] across plasma samples with any detectable tumor content and an AUC of 0.97 (CI, 0.93–1) for those with at least 50% of tumor content, with as few as 25 kbp of genomic space (~1,500 CpGs) required, highlighting the robustness of our approach.

RESULTS

Differential Methylation Landscape Along Prostate Cancer Evolution

The CRPC spectrum spans multiple phenotypes, with distinct subtypes characterized by tumor morphology and tumor expression of lineage-specific markers. As in our prior studies, we applied a morphologic definition to define CRPC-NE, which encompasses small-cell carcinoma, large-cell carcinoma, and mixed adeno-NE histologies (22). IHC staining can support the diagnosis with CRPC-NE as they typically express at least one classical NE marker (e.g., INSM1, synaptophysin) and lack expression of the AR and/or canonical AR target genes [e.g., TMPRSS2, prostate-specific antigen (PSA)]. In contrast, most prostate adenocarcinomas express AR and are negative for NE markers, although variations may be seen, including double-positive (or amphicrine) (AMPC; ARpos, NEpos) and double-negative (DNPC; ARneg, NEneg) prostate carcinoma. It is not clear whether these other subtypes are distinct entities or lie along a spectrum of lineage plasticity as intermediary states toward CRPC-NE (ref. 23; Fig. 1A), and the clinical implications of the various phenotypic subtypes are not well defined. Longitudinal studies confirming the timing and continuum of lineage progression toward CRPC-NE are needed. On the

basis of prior studies (24–26), we hypothesize that changes in DNA methylation are acquired during lineage plasticity and contribute to the phenotypic state changes that occur in advanced prostate cancer. We first analyzed our collection of tissue-based DNA methylation profiles from normal prostatic tissues (NT, $n = 7$), localized prostate cancer (PCa, $n = 7$), CRPC-Adeno ($n = 18$), and CRPC-NE ($n = 10$) obtained through enhanced Reduced Representation Bisulfite Sequencing (eRRBS; ref. 3). Of note, CRPC-NE was defined on the basis of tumor morphology using published criteria (22) and was not reliant on IHC or transcriptome profiling. By using this fairly strict definition to define CRPC-NE-associated differentially methylated sites (DMS) and differentially methylated regions (DMR; ref. 27), we could then apply this knowledge to diverse phenotypic cohorts of CRPC without *bona fide* NE morphology to understand how DNA methylation is reshaped during prostate cancer progression. First, considering DMRs, we observed that the fraction of differentially methylated genome is maximal between PCa and NT (18% of the genome; tumorigenesis), followed by CRPC-NE versus CRPC-Adeno (15.1% of the genome; phenotype transition; Fig. 1B). Conversely, a smaller fraction of the genome is differentially methylated between CRPC-Adeno and PCa (3.1% of the genome; metastasis and castration resistance). The fraction of DMSs, which demonstrate diverse density across the genome, suggests that the tumorigenesis process is associated with a higher fraction of DMSs (17.2%) compared with metastasis seeding and phenotype transition (3.1% and 7.7% of the sites, respectively; Fig. 1C). Further characterization of DMRs supported those observations (Supplementary Fig. S1A–S1C). Importantly, in all the differential methylation analyses performed, the majority of CpG sites were consistently methylated across subtypes, suggesting that only a small fraction of the sites might be informative for subtype classification.

We next focused on the differential methylation between CRPC-NE and CRPC-Adeno samples and performed transcription factor motif (TFBS) enrichment analysis around DMSs. Motifs enriched at hypermethylated CpGs in CRPC-NE included known regulators of prostate adenocarcinoma, including AR (androgen response element, ARE) and HOXB13, as well as REST, which are often downregulated in CRPC-NE (ref. 23; Fig. 1D; Supplementary Table S1). Conversely, hypomethylated CpGs were enriched for NE-associated transcription factor motifs, including ASCL1 and NEUROD1 (24, 28). We compared the observed enrichments with a parallel analysis of chromatin accessibility data (ref. 25; Assay for Transposase-Accessible Chromatin using sequencing; ATAC-seq) of preclinical models [patient-derived xenografts (PDX), organoids, cell lines] of CRPC-Adeno and CRPC-NE. There was concordance ($R = 0.57$, Pearson correlation) for the most hyper- and hypo-accessible TFBS, with similar results obtained for the CRPC-NE versus normal comparison (Supplementary Fig. S1D). Overall, this analysis suggests that the methylome provides a snapshot of the prostate cancer phenotype and related key transcriptional drivers. Only a small subset of the DMSs harbour substantial differential methylation (e.g., 7.5% and 3.1%, respectively, for hypo and hyper sites presenting an absolute difference in β greater than 0.5; Fig. 1E).

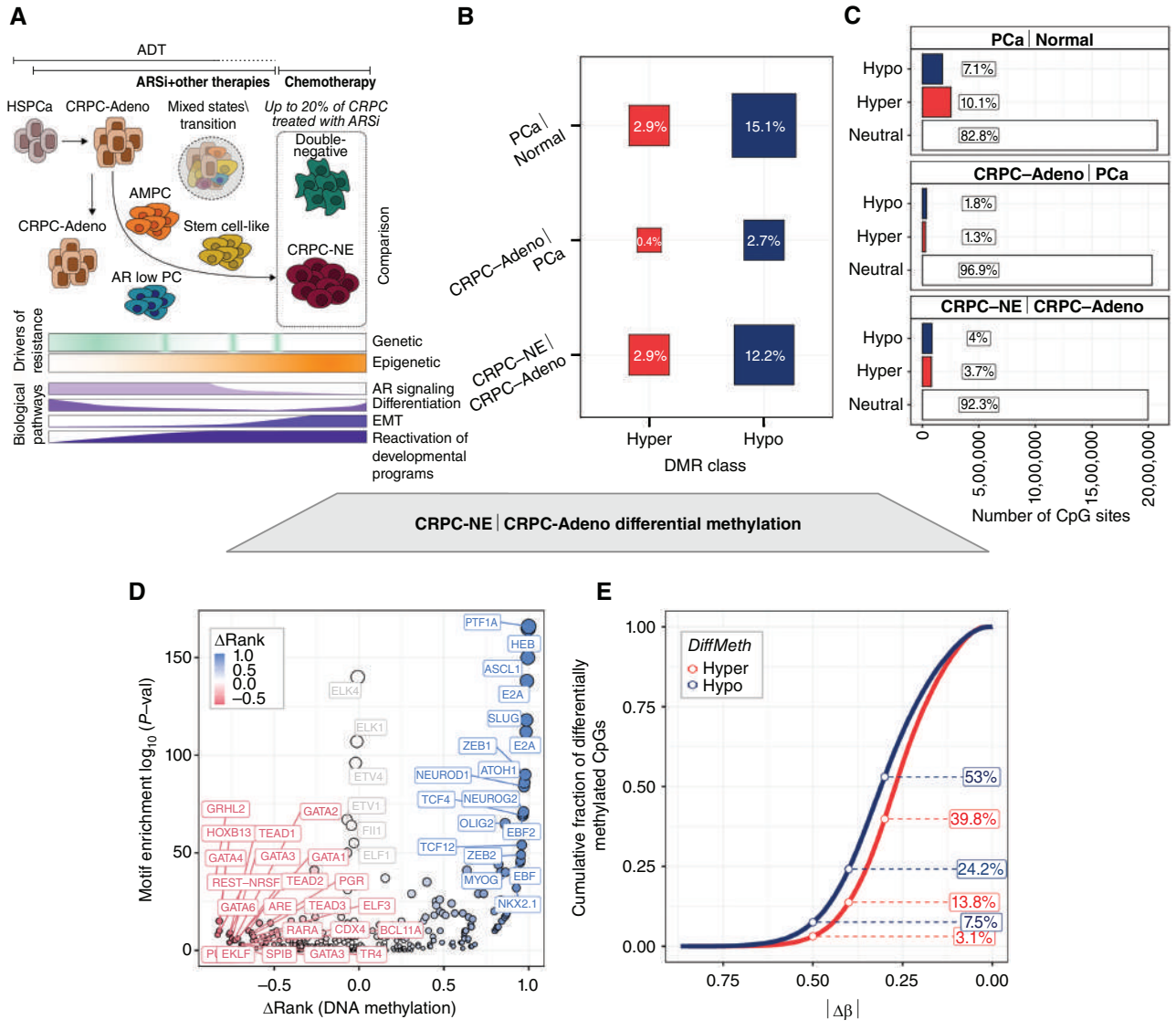


Figure 1. Genome-wide DNA methylation reflects the transition from prostate adenocarcinoma to neuroendocrine prostate cancer. **A**, Potential model of castration-resistant prostate cancer (CRPC) disease progression with emphasis on the transition from an AR-positive CRPC-Adeno toward AR-negative phenotypes. The top lines indicate a noncomprehensive set of systemic therapies. The bottom bars indicate an overview of the relative contribution of selected biological pathways to the corresponding CRPC subtype based on the current literature. The schematic includes a series of proposed CRPC subtypes, highlighting two lineage plasticity endpoints: neuroendocrine prostate cancer (CRPC-NE) and double-negative prostate cancer (DNPC). Various morphologic or transcriptomic subsets have been proposed as potential intermediary states. ADT, androgen deprivation therapy; ARSi, AR signaling inhibitors; HSPCa, hormone-sensitive prostate cancer; AMPC, ampicrine prostate cancer; EMT, epithelial-to-mesenchymal transition. **B**, Plot of the genomic burden of DMR obtained with RockerMeth differential methylation analysis for comparisons across progressive prostate cancer disease states. The reported fractions are relative to the length of the haploid genome. Normal, benign prostatic tissue. **C**, Barplot representing the number of differentially methylated CpG sites (DMS) detected across progressive prostate cancer disease states as reported in **B**. The criteria for defining differential CpG methylation is based on the AUC obtained using the single CpG site to segregate the two groups (see Materials and Methods). **D**, Dot plot of motif enrichment around DMSs between CRPC-NE and CRPC-Adeno solid tissue biopsy samples. For each motif, the difference in motif rank between the set of hypermethylated DMSs and hypomethylated DMSs is computed using the P value as the ranking variable. The y-axis reports the most significant P value obtained between the two sets of regions. Blue indicates preferentially hypomethylated motifs (likely activated); red indicates preferentially hypermethylated (likely suppressed); white indicates motifs enriched in DMSs but with no preferential directionality. **E**, Cumulative density plot of differential methylation signal in Hyper and Hypo DMSs reported in **C**. The labels indicate the fraction of differentially methylated CpG sites with an absolute difference in β ($\Delta\beta$) greater than 0.5, 0.4, and 0.3 between CRPC-NE and CRPC-Adeno samples.

Design of a Targeted DNA Methylation Panel for Tumor Burden and CRPC-NE Detection from cfDNA

On the basis of these observations and the limitations of using WGBS of cfDNA as a scalable approach, we developed a custom-targeted panel for neuroendocrine detection and

tumor burden monitoring by probing the most informative regions selectively; we named this panel NEMO. The panel is built as a collection of region modules with distinct rationale in the context of CRPC for detecting phenotypic subtypes (Fig. 2A; Supplementary Fig. S2A and S2B). The first module

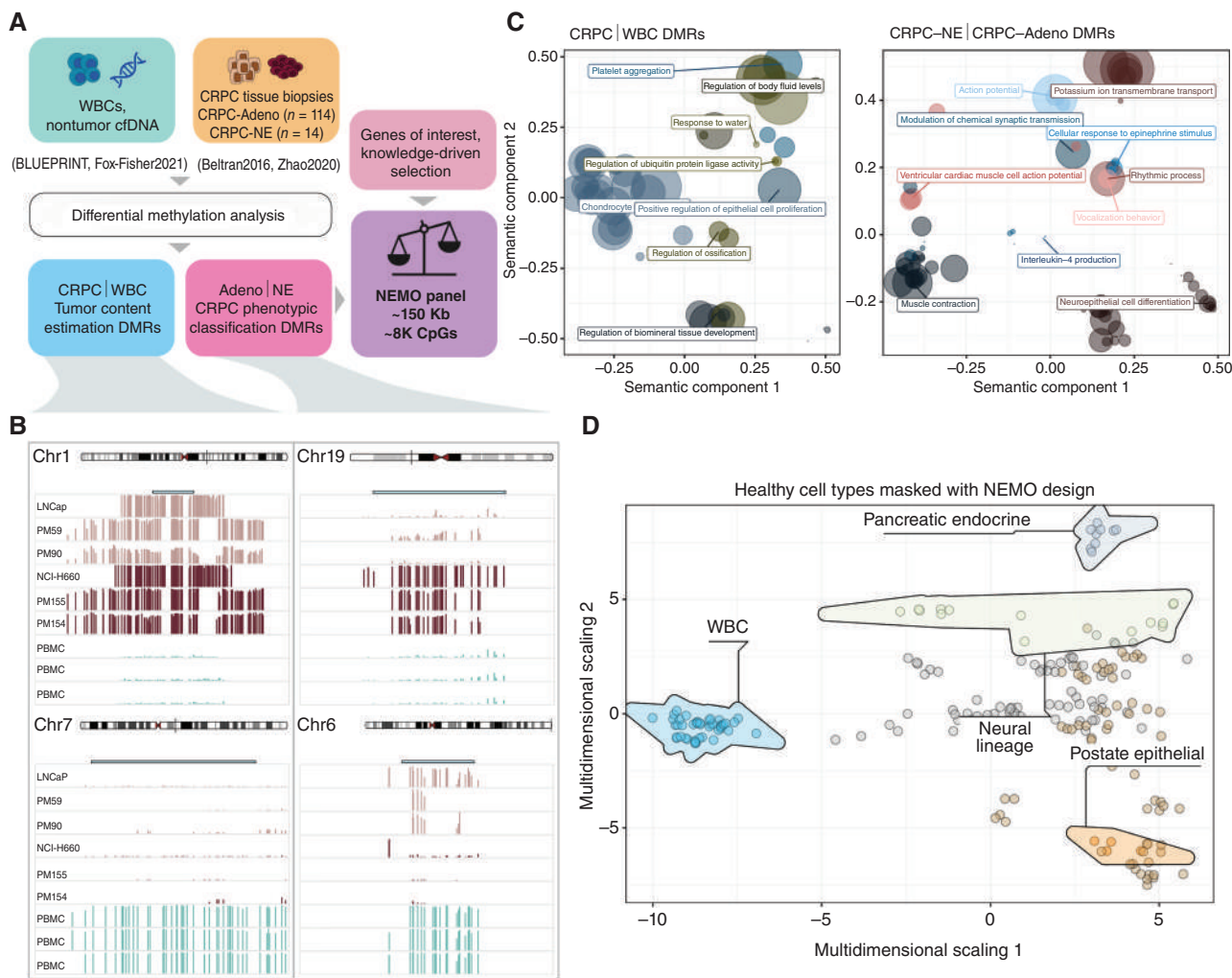


Figure 2. Design of an efficient custom sequencing panel to monitor CRPC tumor burden and detect the emergence of CRPC-NE. **A**, Schematic of NEMO panel design. DNA methylation profiles of solid tissue biopsies from patients with CRPC were collected from two independent studies. Tumor biopsies were classified as CRPC-Adeno and CRPC-NE based on tumor morphology. White blood cells (WBC) and healthy cfDNA profiles were collected from two additional studies and are considered the expected nontumor contribution in cfDNA. A series of differential methylation analyses produced informative DMRs that were prioritized following specific criteria (see Methods section). In addition, a series of knowledge-informed regions were included. The final NEMO sequencing panel design spans ~150 Kb and covers roughly 8,000 CpG sites. **B**, Example of informative DMRs. For each sample category, a subset of three representative solid tissue biopsy samples, *in vitro* models, or primary cell lines (white blood cells) are shown (cyan: PBMC, sepia: CRPC-Adeno, mauve: CRPC-NE). Blue top tracks indicate the captured regions. Left column, two examples of DMRs used for tumor content estimation, exhibiting opposite DNA methylation values in mCRPC (independent of morphology) and WBC samples. Right column, two examples of DMRs used for CRPC-NE detection, exhibiting opposite DNA methylation in CRPC-Adeno and CRPC-NE samples. **C**, Revigo semantic representation of gene ontology from genes associated with NEMO informative DMRs. The dot size is proportional to the significance of the collapsed terms. The left GO analysis was obtained from the tumor content estimation DMRs. The right GO analysis refers to the modules containing DMRs between CRPC-NE and CRPC-Adeno samples. Knowledge-driven regions have been excluded to avoid enrichment artifacts. **D**, Multidimensional scaling plot based on a WGBS atlas of normal cell types masked to retain only the regions included in the NEMO panel. Representative cell types of interest are highlighted, prioritizing the most similar healthy counterpart to the phenotypes of interest: WBCs (expected background, cfDNA), prostate epithelial (CRPC-Adeno), and pancreatic endocrine and neural lineage cells (CRPC-NE).

is dedicated to tumor content estimation from cfDNA, providing an estimation of the disease burden and an essential covariate for CRPC-NE detection. As DNA methylation is highly tissue- and disease-specific, we sought to prioritize a small set of DMRs that can distinguish between prostate cancer cells, regardless of their phenotype, and WBCs. Multiple studies established that the major contribution to the healthy cfDNA pool could be traced to WBC (14, 15). Specifically, a small set of predefined CpG clusters was tested for differential methylation in tissue biopsies from high tumor

content CRPC samples (3, 10) versus isolated WBC. After this procedure, 53 DMRs were selected for tumor content estimation. A similar approach was applied to the NE detection module for the selection of DMRs that segregate CRPC-NE and CRPC-Adeno samples, utilizing a set of CpG clusters ($n = 80$) and a refined selection of previously described DMRs and DMSs (ref. 12; $n = 919$). Examples of informative regions are reported for the tumor content estimation module and NE detection module (Fig. 2B). Furthermore, a collection of knowledge-informed regions, including DMSs near known

NE drivers and genes regulated by DNA methylation, were included ($n = 842$, see Methods section). The genes associated with the data-driven modules included in the NEMO design were enriched for terms related to the development of the epithelium (Fig. 2C, left) and neural-related terms (Fig. 2C, right). Unsupervised analysis of a collection of purified benign cells (15) highlighted the major sources of variability captured by the selected CpG sites, showing a clear separation between hematopoietic, epithelial, and neural/endocrine cell types (Fig. 2D). The total size of the knowledge-based and data-driven sequencing panel covered approximately 150 Kbp, including approximately 8,000 CpG sites within the target regions. The small size of the panel allows for high throughput and excellent scalability, even at high sequencing depths. Throughout the following sections, we extensively tested the NEMO design on newly generated data and published whole-genome sequencing data using a masking procedure that retains only the CpGs in our custom design. The masked signal obtained from whole-genome techniques is highly concordant with the one obtained with NEMO (Supplementary Fig. S2C).

DNA Methylation-based Inference of CRPC Tumor Fraction in cfDNA

Highly specific DNA methylation marks allow for precise estimation of tissue of origin using cfDNA methylation (14, 15). In the context of cancer, the contribution of prostate epithelial tissue signal to the cfDNA pool can be interpreted as a proxy of the disease burden (29), assuming that there is no other source of extensive tissue damage. Both the qualitative and quantitative assessment of tumor-derived cfDNA (also known as circulating tumor DNA, ctDNA) in circulation is important for disease monitoring. To this end, we utilized the β value of 53 DMRs as a proxy of the ctDNA fraction in circulation (i.e., tumor content). In brief, those 53 regions present opposite DNA methylation levels in isolated WBC compared with CRPC samples, resulting in a linear correlation between the β value of each region and the tumor content (Fig. 3A; Methods). A bootstrap approach using half of the regions is used to estimate a confidence interval that informs on the stability of the tumor content estimation. Testing this approach on masked data from CRPC metastatic tumor tissue biopsies and cfDNA profiles from patients with CRPC revealed a strong correlation with orthogonal tumor estimation ($R = 0.92$ and $R = 0.79$, Pearson correlation; Supplementary Fig. S3A and S3B). We applied the tumor content estimation procedure to a collection of preclinical models ($n = 24$), three pools of peripheral blood mononuclear cells (PBMC), and three healthy donor cfDNA plasma samples (HD) and obtained the expected results in most cases (Fig. 3B) with mild discrepancies likely driven by a divergence between the *in vitro* model and the DNA methylation of the original tissue biopsy. We next sought to test the accuracy of tumor content estimation through *in vitro* dilutions. Genomic DNA from the prostate cancer cell line LNCaP and a pool of peripheral blood mononuclear cells (PBMC) were mixed with a decreasing contribution of cancer DNA. The estimated tumor content was highly consistent with the ground truth ($R = 1$, $P < 2e-16$, Pearson correlation), distinguishing cancer-derived DNA from PBMCs down to $\sim 2\%$ of expected cancer DNA fraction (Fig. 3C; Supplementary

Fig. S3C). Similar concordance was observed for dilutions of a single cfDNA sample with known tumor content with cfDNA from healthy donors. Consistently, mixtures of LNCaP cells and the PM155 patient-derived organoid model resulted in a tumor content of 100%. A set of 20 cfDNA samples from patients with CRPC enrolled in the PRIME observational trial cohort was subjected to sequencing with both NEMO and the PCF-SELECT targeted genomic assay (30), a sequencing panel optimized for tumor content estimation and detection of allele-specific genomic events. The tumor content inferred by NEMO was consistent with genomic assessment (Fig. 3D, $R = 0.99$ Pearson correlation) in the range of tumor content above 15%, in which the genomic-based methodologies return reliable estimations. We further explored samples with methylation-based estimates below 15% by adapting a read-based tumor content strategy by Li and colleagues (ref. 16; see Methods). We observed total coherence between NEMO-based tumor detection and the read-based support of tumor signal, highlighting the *bona fide* detection sensitivity of tumor down to 3% (Fig. 3E; Supplementary Fig. S3D). These results and further evaluation of healthy cell types (Supplementary Fig. S3E) support the robustness of tumor content estimation with a minimal set of regions and suggest that even in a parsimonious regimen, the obtained sensitivity is comparable or higher than that from tailored genomic-based techniques.

Scoring of CRPC-NE Phenotype from cfDNA Methylation

The central scope of the NEMO panel is for detecting CRPC-NE using plasma cfDNA. We leveraged patient solid tissue biopsies as a reference, selecting the samples with the highest tumor content (TC > 80%; pathologically defined CRPC-NE = 10, CRPC-Adeno = 29) to build representative DNA methylation profiles for the two CRPC phenotypes of interest. Two additional references of background signal were obtained using isolated WBC, selecting the most significant contributors to the healthy cfDNA pool (monocytes, neutrophils, and megakaryocytes), and a collection of healthy cfDNA samples (31). To obtain the phenotype evidence (PE) score (Fig. 4A), we first estimated the contribution of CRPC-NE and CRPC-Adeno signal as the relative fraction of tumor content that maximizes the likelihood of the observed signal while assuming that the remaining non-tumoral contribution can be modeled as healthy cfDNA. In this step, we enforced consistency with the previously estimated tumor content, requiring that the sum of the inferred CRPC-Adeno and CRPC-NE components equal the value of tumor content. Once the relative contribution of CRPC-NE is estimated, the value is normalized by the total tumor content, making it independent from the tumor content and within the [0–1] range, where zero means total absence of evidence of CRPC-NE, and one means maximal evidence. As expected, the intrinsic uncertainty of tumor content estimation makes the PE score unstable in samples with less than 3% tumor content; thus, we do not provide an estimation in such cases. The PE score measured in tissue samples presented clear segregation, with modest heterogeneity in CRPC-Adeno cases (Supplementary Fig. S4A and S4B). The PE score of primary sorted healthy cells of a variety of tissues of origin demonstrate values far from the extreme representative of Adeno and NE (Supplementary Fig. S4C). Of note, although designed for prostate cancer, the

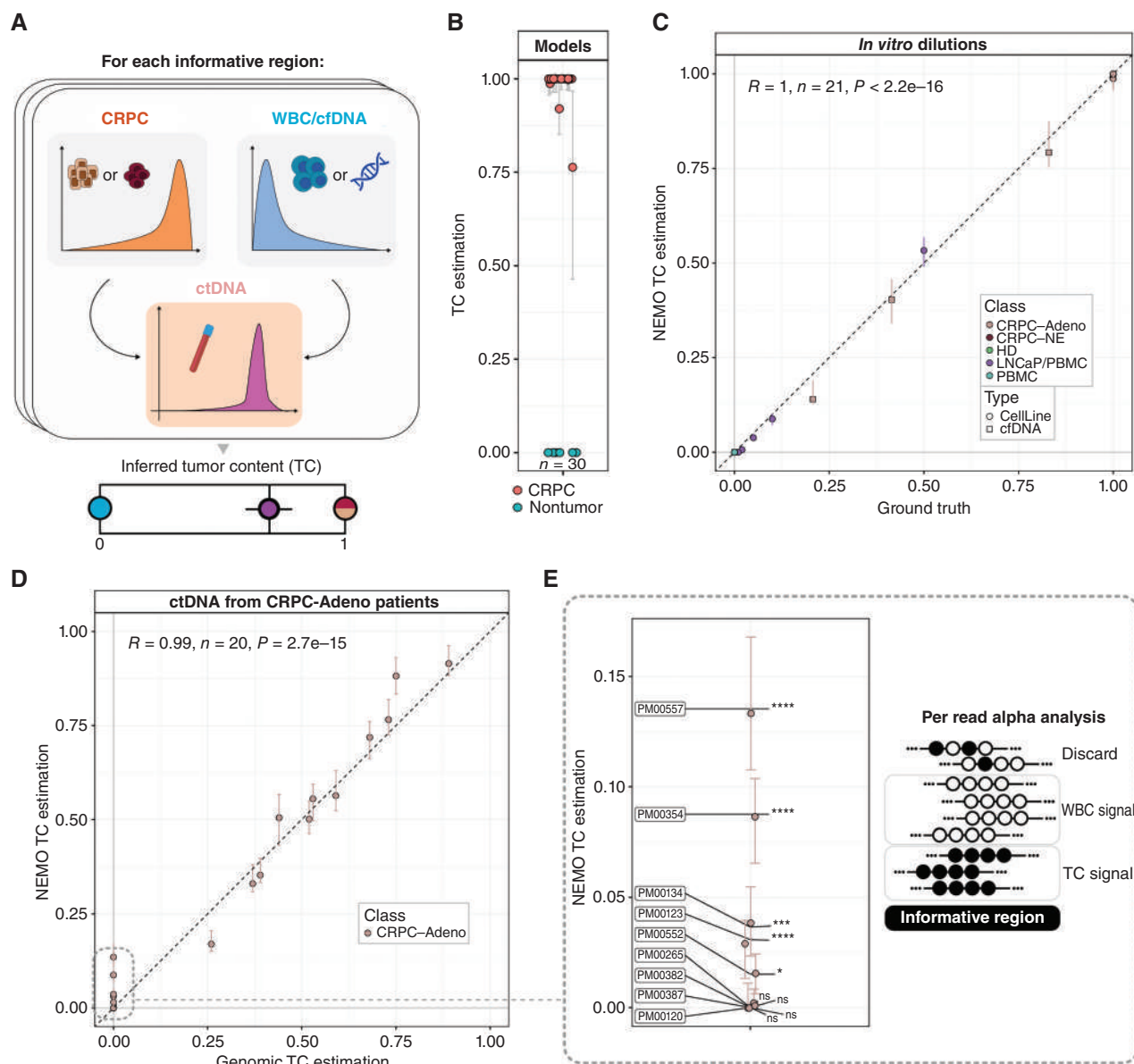


Figure 3. Inference of CRPC tumor content in circulation with a minimal set of informative regions. **A**, Schematic of the tumor content inference strategy. A set of informative regions with opposite and extreme DNA methylation values in WBCs and CRPC was used to estimate the tumor content in cfDNA samples, assuming that the nontumoral fraction of cfDNA is similar to the cfDNA methylation observed in healthy individuals. An iterative subsampling with only half of the informative regions produces a stability interval for the estimation. **B**, Tumor content estimation on a set of preclinical models, including cell lines, organoids, PDXs, and samples representative of the healthy cfDNA background (PBMC, healthy donors). **C**, Tumor content estimation on a set of serial dilutions based on preclinical models, pure cell lines, and cfDNA. PBMCs and plasma cell-free DNA from healthy donors (HD) are expected to be negative for tumor content. **D**, Tumor content estimation of CRPC ctDNA samples from a set of patients from the PRIME cohort. Genomic-based tumor content estimation was obtained by applying the PCF-SELECT panel to matched ctDNA samples collected at the same time point. **E**, Zoomed-in view of the low tumor content samples with discordant tumor content estimation between NEMO (based on DNA methylation) and PCF-SELECT (based on copy-number alterations and SNP allelic fraction). The statistical significance of an orthogonal per-read analysis based on alpha values of informative regions (see Methods) is reported near the sample.

NEMO approach might also identify other small-cell carcinomas such as SCLC (Supplementary Fig. S4D and S4E).

We applied our inference framework to a set of prostate cancer cell lines, patient-derived organoids, and PDX models, leveraging preclinical models profiled both with NEMO and masked genome-wide data. As expected, we observed that CRPC-NE organoids had high phenotype scores (Fig. 4B), with scores of adenocarcinoma models, on average, higher than metastatic

tumors from patient biopsies. This is in line with a recent study reporting increased plasticity in *in vitro* models compared to biopsies from the same subtype (32). Unexpectedly, the CRPC-NE organoid line PM155 had a relatively low PE score (~0.55). However, masked methylation data for the original patient biopsy of PM155 obtained a higher score, suggesting a possible partial divergence of that model with respect to the original sample. This observation was confirmed by integrating four

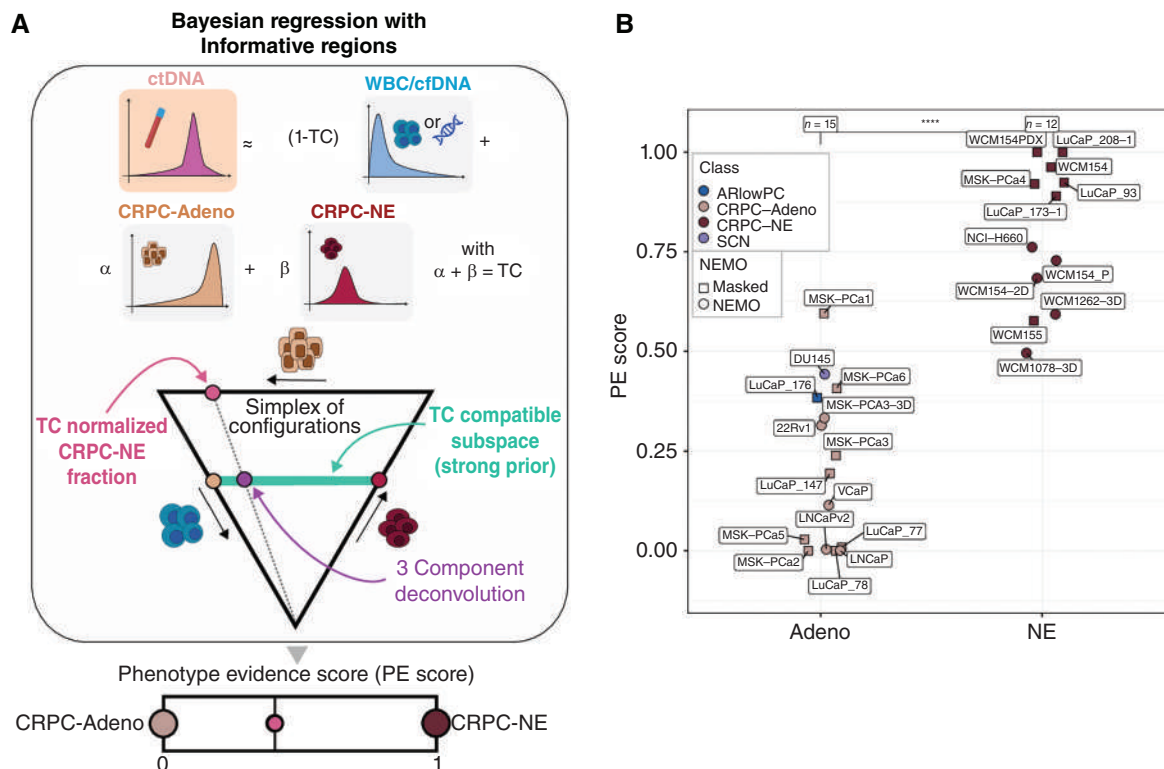


Figure 4. A tumor content-aware phenotype evidence score detects CRPC-NE in circulation. **A**, Schematic of phenotype evidence (PE) score estimation in ctDNA samples. Three reference distributions were created using a panel of high tumor-content solid tissue biopsies and PBMC/cfDNA from healthy donors. As the tumor content is known from the previous step, the expected contribution of CRPC-NE and CRPC-Adeno to the observed DNA methylation can be estimated using a Bayesian regression with a strong prior on the non-tumoral component (see Methods). This procedure is equivalent to estimating the position of each sample in a subspace on a two-dimensional simplex, representing a three-component deconvolution bound to a known tumor content. The estimation can be normalized by factoring the tumor content to obtain the relative contribution of the CRPC-NE signal over the total tumor signal, which we refer to as the Phenotype Evidence score (PE score). **B**, The PE score estimation in a collection of preclinical models of CRPC. The shape of each dot indicates whether the data has been generated with NEMO (circle) or by masking whole genome data (triangle). Statistical significance is assessed with Wilcoxon test. (continued on next page)

proposed CRPC groups defined by gene expression and chromatin accessibility (25). Intriguingly, the PE score meaningfully aligned to the spectrum of AR independence, capturing potential intermediate phenotypes such as the recently described stem-cell-like (SCL) and WNT-driven subtypes (25) (Fig. 4C), which were not included in the training data. Serial dilution of CRPC-NE patient-derived organoids (PM154, PM155) with the adenocarcinoma cell line LNCaP presented a linear decrease in PE score, as expected (Fig. 4D). Importantly, the PE score had the desirable properties of being negative in cell-free DNA from healthy individuals and stable across a wide range of tumor contents (Supplementary Fig. S4F).

We applied the PE score to a collection of cfDNA data, including unpublished WGBS cfDNA data from patients with pathologically confirmed CRPC-Adeno and CRPC-NE (masked) and cfDNA from two independent CRPC cohorts without biopsies (Wu and colleagues (19) and the PRIME cohort Supplementary Table S2A and S2B). The PE score successfully segregated pathologically-confirmed CRPC-Adeno and CRPC-NE samples, with few exceptions (79/82 and 17/20 correctly classified samples for CRPC-Adeno and CRPC-NE, respectively, using the 0.43 cutoff calculated on tissue biopsies; Fig. 4E). All patients in the unselected CRPC cohorts with no clinical evidence of NE differentiation had low PE scores despite a range of estimated

tumor contents similar to the CRPC-NE. After inspection of the corresponding patient features, we noticed that one patient with CRPC-Adeno who had a high PE score (PM341) had markedly elevated serum levels of the neuroendocrine marker chromogranin A of 1,221 ng/mL (upper limit of normal 140 ng/mL), suggesting a possible clinically undetected CRPC-NE component or transition state. Conversely, one patient with CRPC-NE with a low PE score below 10% was confirmed to have mixed histology (PM243), with metastatic tumor harboring both adenocarcinoma and neuroendocrine features. The uneven contribution of the two phenotypes to the cfDNA pool may account for the observed low score. The few remaining misclassified samples were characterized by low tumor content, possibly hampering the accuracy of our classification. Overall, our extensive analyses of multiple novel and published samples support the ability of NEMO to detect the CRPC-NE phenotype through targeted cfDNA methylation (Supplementary Tables S3 and S4).

Probing the CRPC Transcriptional Spectrum with NEMO

To further test the ability of NEMO to detect the spectrum of CRPC phenotypic subtypes, we evaluated genomic DNA methylation from a collection of 11 PDX models representing

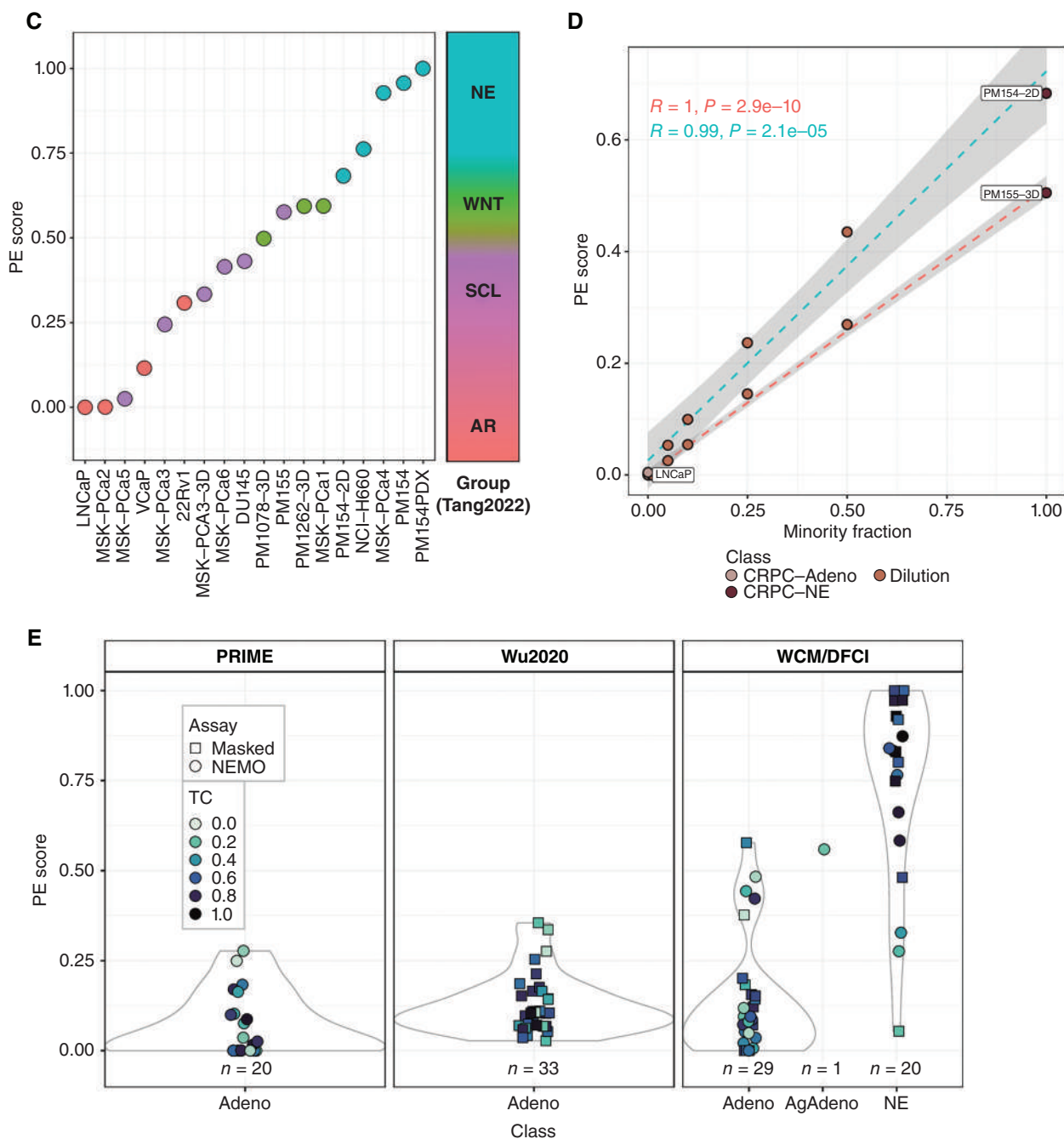


Figure 4. (Continued) **C**, Ranked dot-plot of a subset of preclinical CRPC models annotated by gene expression group based on a previous study by Tang et al. (25). SCL, stem cell-like; WNT, WNT-driven; AR, CRPC-Adeno; NE, CRPC-NE. The right bar represents the gradient of CRPC progression captured by the PE score. **D**, Linearity test of PE score estimation using *in vitro* dilutions between PM154/PM155 cell lines (minority fraction) and LNCaP. The score presents a clear linear trend with respect to the real minor component. The shaded region represents the 95% CI for a fitted linear model. **E**, Results of PE score estimation in clinical cohorts. Samples with tumor content below 3% have been excluded because of high PE score uncertainty and possible ambiguity in tumor content detection. The WCM/DFCI cohorts comprise a subset of samples previously profiled with WGBS and masked (circles), retaining only regions captured by the NEMO design. The Wu2020 cohort (19) comprises only masked samples, profiled by bisulfite genome-wide targeted NGS. The PRIME cohort comprises only samples profiled with the NEMO assay. Different shades of blue are used to represent the tumor content of each sample.

various CRPC subtypes (23), including not only CRPC-NE and CRPC-Adeno, but also double-negative prostate cancer (DNPC²) lacking both AR and NE marker expression, amplicrine prostate cancer (AMPC) expressing both AR and NE markers, and intermediate phenotypes (ARlow, NElow). These

subtypes were previously defined on the basis of transcriptomic analysis (ref. 23; Fig. 5A). The highest PE score was observed in CRPC-NE and DNPC, with ARlow and CRPC-Adeno PDXs harboring intermediate scores (Fig. 5B). Intriguingly, the AMPC PDX (ARpos/NEpos) obtained a score similar

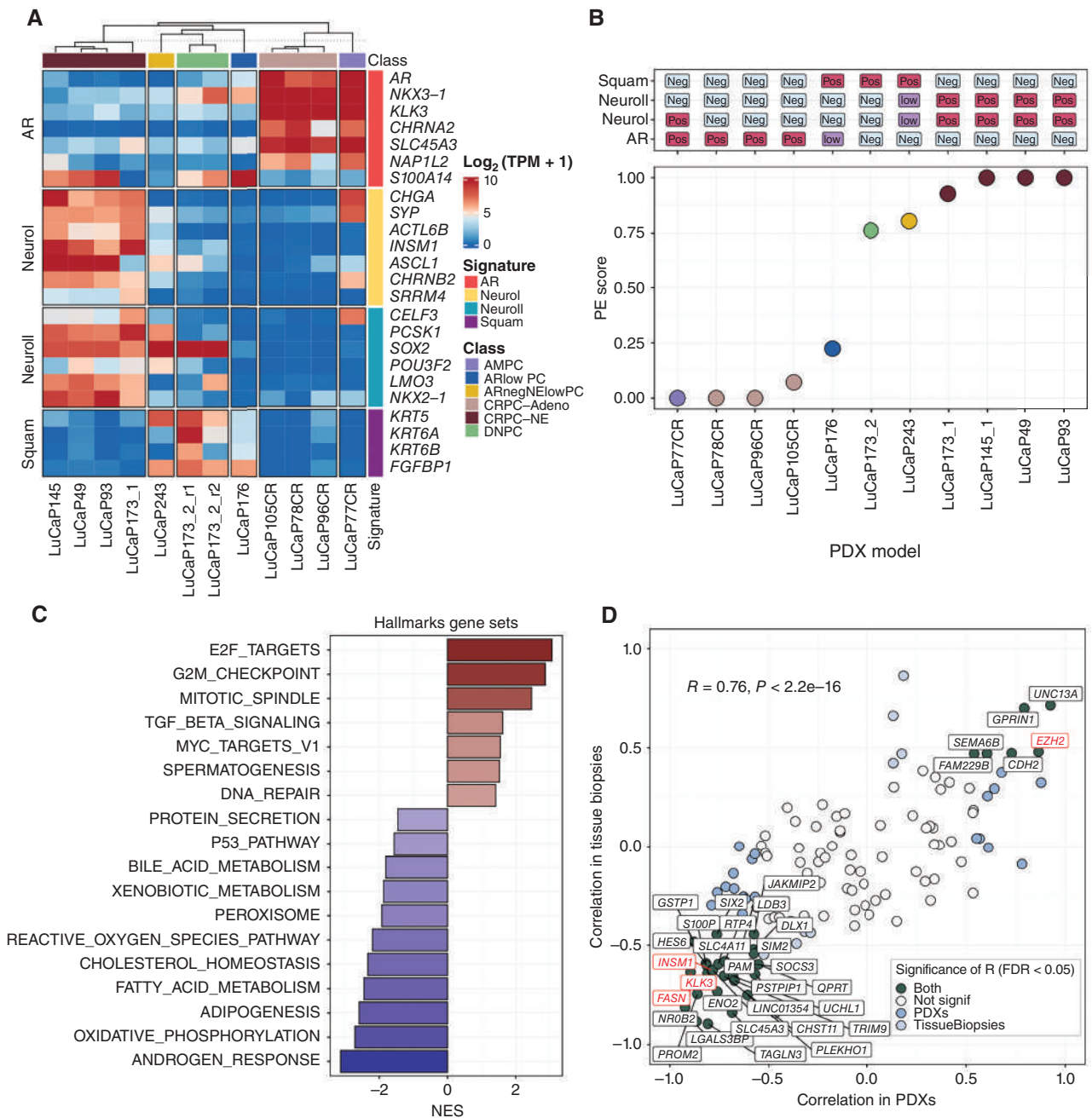
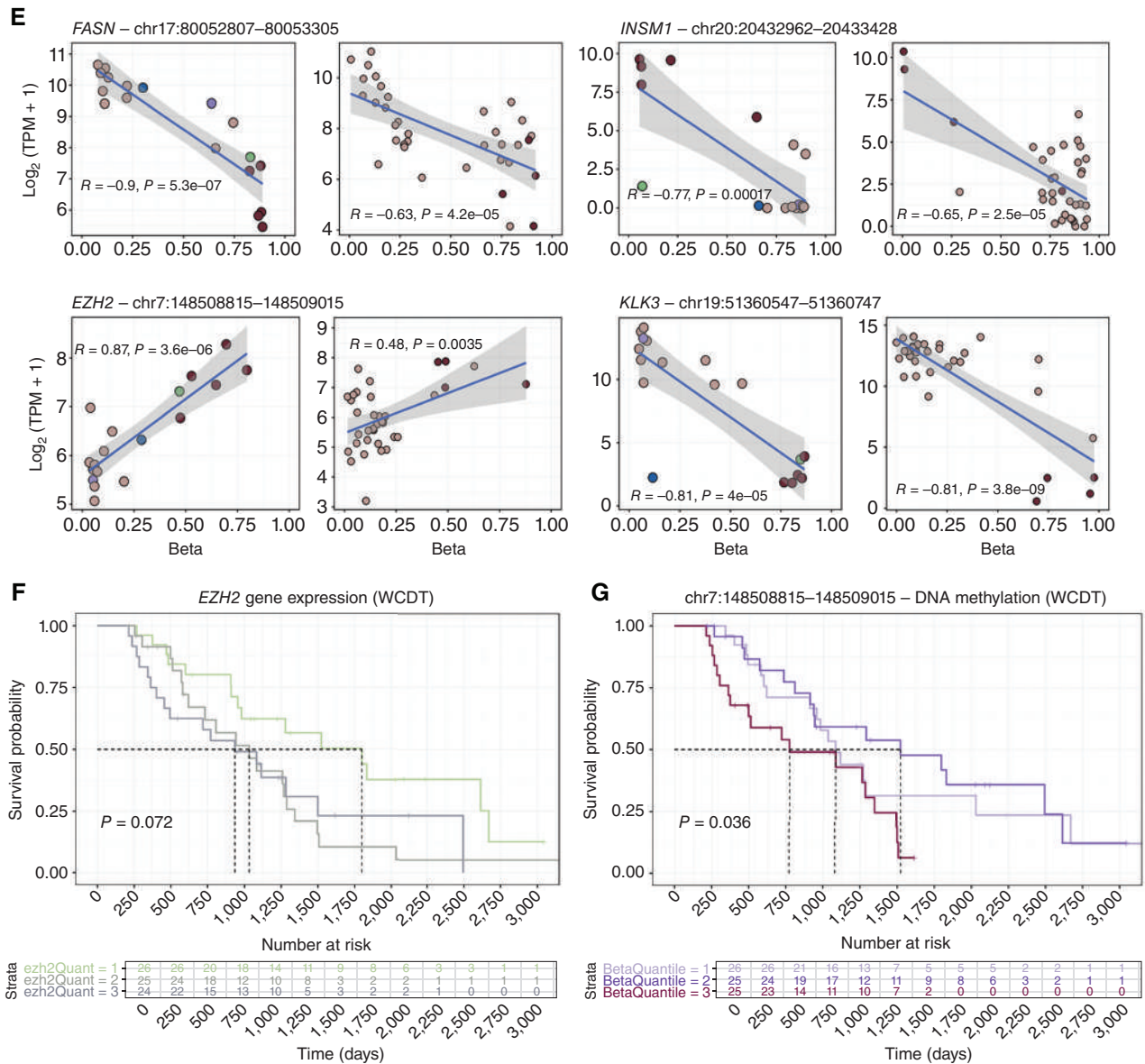


Figure 5. Probing the CRPC spectrum in PDX models reveals an association between the PE score and transcriptional activity. **A**, Heat map of gene expression for 11 LuCaP PDX models profiled with NEMO. The reported gene signatures from Labrecque et al. (23) capture the expression of genes supporting the previously described subtype classification of CRPC. **B**, PE score estimation on the 11 selected PDX models. Top annotation: discretized activity of the reference signatures reported in **A**. CRPC subtypes are colored as in **A**. **C**, GSEA analysis based on the correlation coefficient between gene expression of every single gene and PE score in PDX models. Significant results ($P_{adj} < 0.05$) for the Hallmark MSigDB collection are reported. NES: GSEA normalized enrichment score. **D**, Comparison of the Pearson correlation between gene expression and proximal DNA methylation for a selection of regions included in NEMO. The x-axis reports the correlation observed in a collection of PDX samples, whereas the y-axis reports the correlation obtained for the same region-gene pair in the WCDT cohort. Four selected genes of interest are highlighted in red. (continued on next page)

to CRPC-Adeno, suggesting that the PE score might capture features of AR independence regardless of expression of classical neuroendocrine markers (*SYP*, *CHGA*). As expected, the methylation-based PE score correlated positively with a previously described gene expression-based NEPC score (3) and negatively with AR signaling score (ref. 33; Supplementary

Fig. S5A). We further leveraged matched transcriptomic data to measure the correlation between PE score and transcriptional programs. GSEA analysis based on single gene correlation statistics showed a clear positive enrichment for terms related to proliferation, *Myc* targets, and TGFβ signaling (Fig. 5C). Conversely, a negative correlation was observed for

Downloaded from <http://aacrjournals.org/cancerdiscovery/article-pdf/doi/10.1158/2159-8290.CD-23-0754/3403116/cd-23-0754.pdf> by University of Torino user on 15 January 2024



terms related to androgen response and *TP53* pathway. Similar results were obtained with a regulon-based analysis (34), highlighting a negative correlation between the PE score and the activity of *FOXA1* and *AR* (Supplementary Fig. S5B).

As a module of the NEMO panel was designed to capture promoters and DMSs near genes of interest, we tested their ability to inform on the underlying transcriptional state. We first queried the DNA methylation profiles by EPIC array of a collection of PDX samples ($n = 18$) and their corresponding gene expression profiles to pinpoint gene regions with the most robust correlations and independently validated them in high tumor content solid tissue biopsies from the West Coast Stand Up To Cancer-Prostate Cancer Foundation Dream Team cohort

(WCDT; $n = 100$; Fig. 5D and E; Supplementary Table S5). Regions with reproducible correlation include canonical markers of CRPC-NE (*INSM1*), CRPC-Adeno (*KLK3*), and genes relevant to CRPC biology, such as *FASN* and *EZH2*. For example, when focusing on *EZH2*, a gene involved in prostate cancer plasticity and associated with aggressive disease (5, 35), there was a direct correlation between DNA methylation of the *EZH2* intronic region with increased *EZH2* transcript levels, suggesting a possible regulatory element. High *EZH2* mRNA expression or *EZH2* methylation was associated with shorter patient survival in the WCDT cohort (Fig. 5F and G; log-rank test P values of 0.072 and 0.036, respectively). A similar trend was observed across CRPC subtypes while querying for the cfDNA

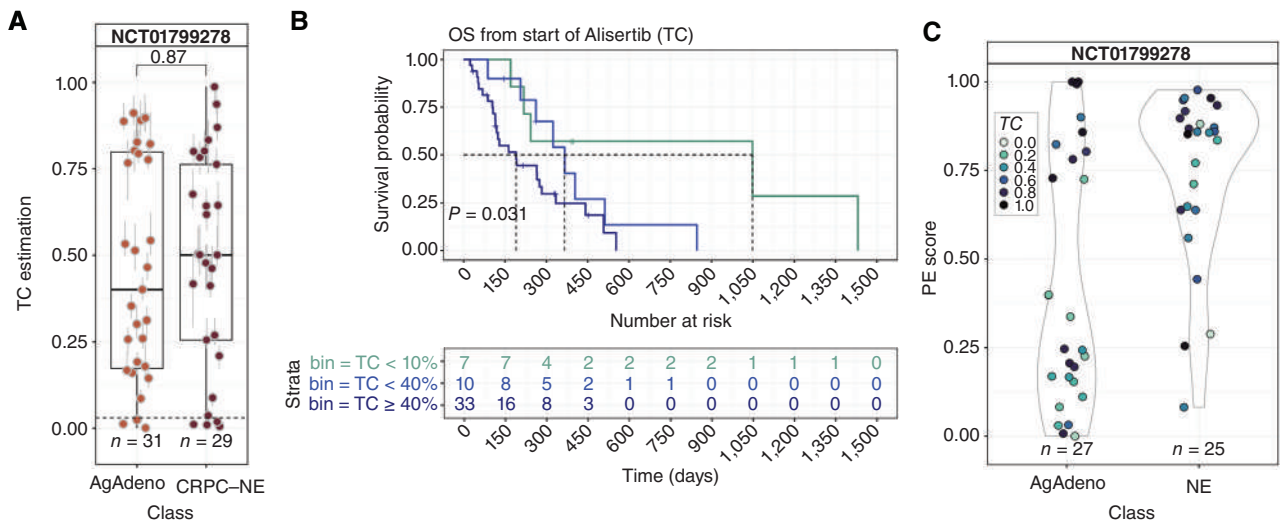


Figure 6. The application of NEMO in two phase II clinical trials reveals the prognostic value of cfDNA tumor content and detects potentially undiagnosed CRPC-NE. **A**, Boxplot of tumor content estimation from ctDNA samples in the phase II trial of the aurora kinase A inhibitor alisertib. Patients had either aggressive clinical features with adenocarcinoma histology (AgAdeno) or CRPC-NE, defined based on tumor morphology on central review of pretreatment biopsy. Eligibility criteria are listed in Supplementary Table S6. **B**, Kaplan-Meier analysis of overall survival based on estimated ctDNA tumor content in the circulation at the beginning of treatment in patients enrolled on the alisertib trial. **C**, Violin plot of PE score estimation of cfDNA from patients with AgAdeno and CRPC-NE (samples with tumor content <3% are excluded). (continued on next page)

methylation signal of the informative regions (Supplementary Fig. S5C) and correcting DNA methylation for tumor content. In summary, the PE score can capture meaningful transcriptional states and provides a model-informed analysis of specific DMRs that could be used to noninvasively obtain a relative estimation of gene expression for specific genes of interest.

Clinical Utility of Methylation cfDNA Profiling in Aggressive CRPC

To assess the clinical utility of NEMO, we profiled plasma DNA from patients with clinically defined aggressive variant prostate cancer (AgAdeno) and/or pathologically defined CRPC-NE enrolled in two prospective phase II clinical trials (Supplementary Table S6A and S6B). One trial tested the efficacy of the *AURKA* inhibitor alisertib (20), and the other tested the efficacy of carboplatin plus docetaxel chemotherapy (21). Trial eligibility criteria are listed in Supplementary Table S7. In the alisertib trial, all patients ($n = 61$) underwent baseline tumor metastatic biopsy to confirm histology as well as matched peripheral blood sampling. We profiled 60 cfDNA baseline plasma samples from the alisertib trial, of which 29 were from patients classified as CRPC-NE and 31 were AgAdeno with adenocarcinoma histology, with subtypes confirmed by central pathology review. The cfDNA tumor content estimations spanned the entire range, in line with the very aggressive features of the patients included, with no significant difference between AgAdeno and CRPC-NE cases (Fig. 6A, $P = 0.87$). We observed that the tumor content in circulation was prognostic, with patients with tumor content $\geq 10\%$ having inferior overall survival (OS) and progression-free survival (PFS) compared with those with tumor content $< 10\%$ ($P = 0.03$, Fig. 6B; Supplementary Fig. S6A). The PE scores resulted in a unimodal distribution for CRPC-NE cases, with few outliers with moderate and low scores (Fig. 6C). On the

other hand, AgAdeno samples showed a bimodal distribution, suggesting heterogeneous disease states and possibly several cases with undiagnosed CRPC-NE and/or epigenetic similarities with CRPC-NE. We tested whether those PE scores were associated with specific trial inclusion criteria within the AgAdeno group (Supplementary Fig. S6B). Interestingly, when we reviewed these cases with high PE score and adenocarcinoma histology by central review, we found that 9 of 10 had CRPC-NE reported by local pathology based on morphology ($n = 8$) or IHC staining for NE markers ($n = 1$). These data speak to differences across pathologists and point to potential value for a more objective PE score beyond pathology classification.

In the phase II trial of carboplatin plus docetaxel, we evaluated 41 baseline plasma samples by NEMO. Tumor content was measurable in all 41 samples, with CRPC-NE (i.e., pathologically confirmed small-cell prostate cancer) and aggressive variant prostate cancer (AVPC; defined clinically, Supplementary Table S7) showing comparable values (Fig. 6D, $P = 0.12$). Of note, not all patients with AVPC had a metastatic biopsy prior to enrolling on the trial to confirm NE or adenocarcinoma histology. Consistently, cfDNA tumor content was strongly associated with overall survival from the start of chemotherapy (Fig. 6E, P -value < 0.001). The PE score successfully captured 6 of 8 patients with CRPC-NE, whereas 6 of 24 AVPC presented with high scores potentially compatible with undiagnosed CRPC-NE and/or epigenetic similarities with CRPC-NE (Fig. 6F). When looking at the clinical inclusion criteria for the study, high PE score positively associated with the presence of bulky lymphadenopathy or pelvic mass (Supplementary Fig. S6C). We next assessed the association of cfDNA tumor content, PE score, and the additional inclusion criteria used in both trials with clinical outcomes, and tumor content was independently associated with overall survival after

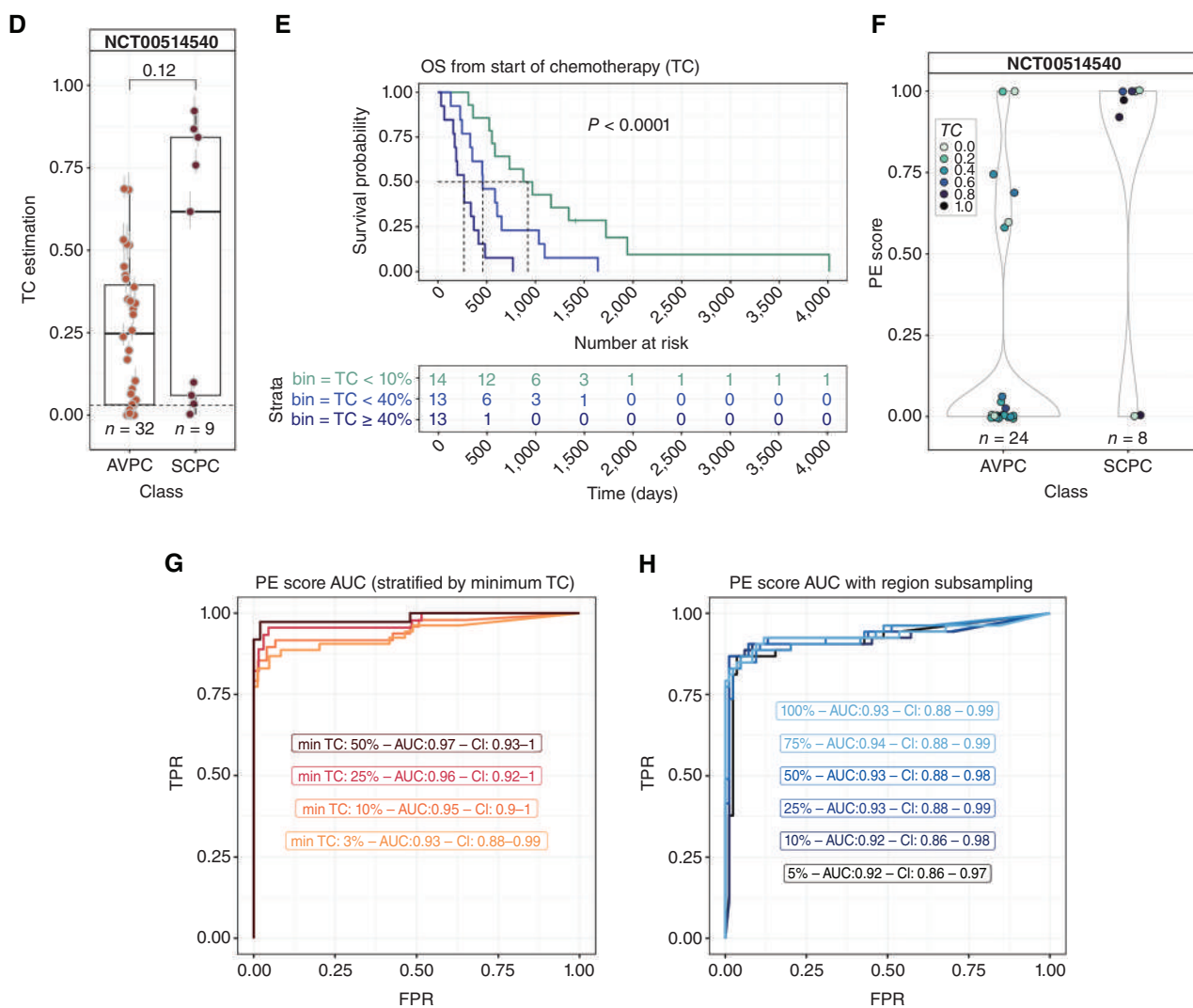


Figure 6. (Continued) **D**, Boxplot of tumor content estimation from ctDNA samples in the docetaxel plus carboplatin chemotherapy trial. Patients had clinically defined aggressive variant prostate cancer (AVPC) or small-cell prostate carcinoma (i.e., CRPC-NE). A pretreatment biopsy to confirm histology was not required. Eligibility criteria are listed in Supplementary Table S6. **E**, Kaplan-Meier analysis of overall survival based on estimated tumor content in circulation at the beginning of treatment for the chemotherapy cohort. **F**, PE score estimation of cfDNA from patients with AVPC and small-cell prostate cancer (SCPC; samples with tumor content <3% are excluded). **G**, Global AUC of binary classification (CRPC-NE, CRPC-Adeno) based on ctDNA samples with biopsy-confirmed pathology and excluding AVPC and AgAdeno samples. Different shades represent the minimum tumor content required before measuring the PE score segregation performance. **H**, Global AUC of binary classification based on ctDNA samples as in G. Different shades represent the downsampling of informative regions used for PE score calculation with respect to the total ensemble of informative regions. The more lenient threshold of 3% tumor content has been used for this analysis.

controlling for additional variables (Supplementary Fig. S6D). There was no significant association between PE score and response to treatment in either of these two phase trials, which was not unexpected as the alisertib trial was negative for its primary endpoint and the carboplatin-docetaxel trial was not specifically geared towards targeting CRPC-NE.

Finally, we measured the binary classification performance of the PE score across all available CRPC-Adeno and CRPC-NE plasma samples resulting in an AUC of 0.93 (CI: 0.88–0.99, Sensitivity: 0.85, Specificity: 0.95, nAdeno = 83, nNE = 53; samples with tumor content >3%) for histology-confirmed CRPC-NE and a Youden index optimal cutoff of 0.43, consistent with the PE scores observed in the preclinical models. The classification

performance raised to 0.97 (CI, 0.93–1) in high tumor content (>50%) plasma samples (Fig. 6G). The binary classification performance remained robust after a classification module downsampling test to as low as 5% of informative regions corresponding to about 1,500 CpGs and 25 Kbp of genomic space (Fig. 6H; Supplementary Fig. S6E), supporting a highly scalable design.

DISCUSSION

Most precision oncology efforts in CRPC have focused on genomic subtyping to select patients for PARP inhibitors, immunotherapy, or other biomarker-driven therapies (36). Beyond genomics, CRPC is also characterized by diverse

phenotypic subtypes that drive response and resistance to systemic therapies. One emerging phenotypic subtype is CRPC-NE, reported in up to 15% to 20% of patients with late-stage CRPC (2, 33). The diagnosis of CRPC-NE currently relies on metastatic tumor biopsy that is invasive and not often clear based on tumor heterogeneity and variability in histopathology interpretation. Although CRPC-NE is enriched for *RBI* and *TP53* tumor suppressor losses, these alterations are not specific to CRPC-NE and can also occur in CRPC-Adeno (3). Specific molecular biomarkers for improved identification of CRPC-NE could lead to early detection and early intervention strategies to improve patient outcomes.

In this work, we demonstrate an efficient and noninvasive method for the estimation of disease burden (tumor content) combined with the detection of CRPC-NE (PE score) enabled by targeted sequencing of cfDNA methylation. Through extensive analyses of publicly available datasets and knowledge-driven biomarker selection, we prioritized informative regions of the human genome that reflect the CRPC phenotypic state. Interrogation of patient-derived whole-genome DNA methylation profiles highlighted that only a minority of CpGs bear highly informative differential DNA methylation sites across prostate cancer disease progression. Grounding on those observations, we designed a custom panel to capture only regions of interest. The reduced size of the panel allows for higher coverage sequencing with a lower read throughput compared with whole-genome approaches, both desirable properties in the setting of cfDNA sequencing. A DNA methylation-based approach for tumor content estimation revealed consistency when compared with genomic-based tumor content inference and ground truth (*in vitro* dilutions), leveraging only a few tens of informative regions. Importantly, the precise estimation of tumor content does not rely on the presence of clonal copy-number changes or single-nucleotide variants (SNV), which are strictly required for genomic-based approaches. Although tumor content estimation can be detected by other strategies, those do not offer phenotype scoring desired in the CRPC context.

The PE score captures the putative fraction of CRPC-NE over the total tumor content detected, segregating cancer-derived plasma samples based on their known pathologic diagnosis. Although the ability to quantify any tumor-related features does depend on the presence of tumor material in the circulation, we intentionally built a PE score that, provided there are tumor molecules available (ie, lower limit of detection 3%), is analytically independent of the percentage of tumor content. The clinical context in which this assay might be most useful is in later-stage metastatic prostate cancer when CRPC-NE is most often suspected and where ctDNA fraction is typically high (>10%–15%). However, the panel sensitivity could also facilitate future investigation in earlier disease states to evaluate if CRPC-NE clones may be detected even before clinical suspicion.

Except for a few cases, the PE score provided excellent segregation between CRPC-Adeno and CRPC-NE samples. Discordance between the pathologic classification from a single biopsy and the PE score in ctDNA in some cases is expected and likely reflects the ability of liquid biopsies to detect intrapatient tumor heterogeneity and distinct metastatic sites that contribute differently to the ctDNA pool. Analysis of patient-derived organoid and xenograft models suggests that the PE score can

also capture other AR-independent CRPC phenotypes beyond CRPC-NE. Querying single informative regions near genes of interest also suggested the potential use of DNA methylation as a proxy of gene expression, as exemplified by the case of *EZH2*. Although these results warrant further exploration, possibly aided by multi-omics profiling of plasticity models and patients (37), we envision that DNA methylation of selected informative sites may provide a glimpse into the transcriptional state of clinically relevant genes.

The application of NEMO to two clinical trial cohorts revealed meaningful prognostic value of methylation-based cfDNA tumor content and identified the phenotypic spectrum within aggressive variant CRPC. There have been very few clinical trials dedicated to CRPC-NE. Although the trial eligibility was not specific for biopsy-confirmed CRPC-NE, these trials represented a unique opportunity to evaluate aggressive variant prostate cancer that shares clinical features with CRPC-NE. We showed that tumor content is independently prognostic in this aggressive disease population, as has been seen in other traditional CRPC settings (38). Although this disease is very challenging to treat and neither histology nor PE score were sufficient to predict response to alisertib or docetaxel plus carboplatin chemotherapy, the application of NEMO to these trials could inform patient selection for future trials testing CRPC-NE-directed therapies.

Other DNA methylation-based approaches for noninvasive CRPC-NE detection have recently been described. First, a method based on cell-free methylated DNA immunoprecipitation sequencing (39), an immunoprecipitation approach that measures DNA methylation, resulted in an excellent classification performance (11). However, this approach does not provide explicit tumor content estimation or precisely measure hypomethylated regions and likely requires a larger number of reads. Read-based analyses of whole-genome sequencing showed great discriminatory capacity (40) and have been successfully applied in this context even using a more affordable ultra-low-pass sequencing approach (41). Although we obtain comparable performance, there are several advantages of a targeted approach. First, we can obtain high-resolution information about the DNA methylation state and even characterize broad and focal copy number alterations (Supplementary Fig. S7A and S7B) and potentially SNVs for target regions of interest when compatible with bisulfite conversion. Our analyses also suggest that a smaller design has the potential to decrease further the number of reads required while retaining performance. As shown by our comprehensive initial data survey, most DNA methylation across the genome is uninformative, indicating that any nontargeted genome-wide approach will eventually suffer the scarcity of information obtained with a limited number of reads. On the other hand, all bisulfite-based assays, including NEMO, suffer from extensive DNA damage during conversion, leading to a partial loss of the input cfDNA. Altogether, we regard the mentioned approaches as complementary to custom-targeted sequencing, each with its strengths and weaknesses. Future work integrating DNA methylation with other emerging cfDNA epigenetic platforms such as 5-hydroxycytosine (42) and H3K27Ac (37) and/or fragmentomics (43) could provide complementary insights into the changes in functional state that associate with CRPC-NE.

In a real-world setting, there are potential scenarios in which NEMO application may be useful. CRPC-NE might be suspected after considering whether the patient is early or late-stage CRPC, types of prior therapy (e.g., prior AR pathway inhibition), and if there are aggressive clinical features (e.g., rapid progression with low or nonrising PSA, new visceral metastases). In this context, the metastatic disease burden is typically high, as reflected by tumor content estimation, and the PE score obtained by NEMO expresses the degree of AR independence from cancer cells contributing to the ctDNA pool. We suspect that most CRPC-Adeno will have low PE score, as we observed in the two CRPC cohorts. An extreme score (PE score <0.3 , PE score >0.7) suggests the dominant presence of either adenocarcinoma or NE phenotype. Conversely, an intermediate score should be handled with caution, considering the possibility that the two phenotypes are coexisting or the trans-differentiation process is ongoing.

The current diagnosis of CRPC-NE is based on tumor biopsy. There are no clear clinical guidelines on when to perform a biopsy, and not all patients with aggressive clinical features will have CRPC-NE. One potential clinical application of the NEMO platform would be for an elevated PE score detected in cfDNA to help prompt when to perform a new biopsy for pathological confirmation of CRPC-NE and subsequent treatment selection for CRPC-NE-directed therapy. We could also envision the use of cfDNA as a replacement for a biopsy in the future. As the PE score can also detect aggressive variants beyond CRPC-NE, further studies demonstrating its clinical utility for this population are warranted to validate its use as a biomarker to guide treatment selection. Serial monitoring studies are also needed to understand better how and when CRPC-NE emerges dynamically in patients developing lineage plasticity and how this might improve clinical decision-making and patient outcomes. Understanding how DNA methylation changes chart with dynamics of chromatin accessibility and histone marks during the transition from CRPC-Adeno to CRPC-NE would also provide biological insights into the potential cooperation of epigenetic events during the lineage plasticity process. Other potential applications of NEMO could be for predicting response to existing CRPC therapies (e.g., AR and PSMA-targeted therapies, chemotherapy, others), selecting patients for trials targeting CRPC-NE (e.g., DLL3-targeted therapies), or approaches geared toward targeting or modulating the lineage plasticity process (e.g., epigenetic strategies).

METHODS

Genome-wide Datasets for NEMO Design and Testing

We leveraged a collection of DNA methylation profiles from published studies to design NEMO. All files were obtained from DNA methylation counts, reporting CpG position, coverage, and the fraction of methylated reads (i.e., β). Each dataset was initially available in hg19 or lifted from hg38 to hg19 using the *rtracklayer liftover* function using a chain file from UCSC (<https://hgdownload.cse.ucsc.edu/goldenpath/hg38/liftOver/hg38ToHg19.over.chain.gz>). As part of the data harmonization process, all DNA methylation information was collapsed to the forward strand. CpG positions with coverage lower than 10 were discarded before the analysis to restrict the number of points to high-quality measurements. After the nomination of informative regions included in the NEMO panel, whole-genome DNA methylation data is restricted to the masked information within NEMO regions. The following datasets were utilized.

Tissue sample datasets:

- (i) Lin and colleagues 2013 (44). Tissue eRRBS data. Seven normal prostatic tissues and 7 localized prostate adenocarcinoma.
- (ii) Beltran and colleagues 2016 (3). Tissue eRRBS data. Pathologically confirmed 18 CRPC-Adeno and 10 CRPC-NE biopsies. CRPC-NE was defined based on tumor morphology (22) and was not reliant on IHC or transcriptome profiling. Tumor purity estimation was assessed from matched whole-exome sequencing data by CLONETv2 (45).
- (iii) Zhao and colleagues 2020 (10). Tissue WGBS data. DNA methylation coverage files were obtained from the authors. Pathologically confirmed 95 CRPC-Adeno, 5 treatment-induced small-cell neuroendocrine prostate cancer biopsies (CRPC-NE). Small-cell carcinoma was defined on the basis of tumor morphology. Tumor purity estimation was obtained from tabular data from the companion work (46), using the average applied computational methodologies (pathology, DNA, RNA).
- (iv) BLUEPRINT consortium: Methylation coverage files were downloaded from the RnBeads resource (ref. 47; <https://rnbeads.org/>). Representative WBC types most abundant in healthy cfDNA were retained, including megakaryocytes, monocytes, and granulocytes ($n = 10$).
- (v) Loyer2023 (15), WGBS from FACS sorted healthy cell types.
- (vi) Whole-genome bisulfite sequencing and EPIC array data from LuCaP PDX lines are publicly available on Gene Expression Omnibus (GEO), accession numbers GSE227086, GSE227814, and GSE227695.

cfDNA datasets:

- (i) Fox-Fisher 2021 (31), healthy cfDNA WGBS ($n = 23$). This dataset is a reference for the expected DNA methylation profile of non-cancer-derived cell-free DNA.
- (ii) Beltran2020 (12): cell-free DNA WGBS data on patient-derived plasma samples, processed as described in the original work, from patients with pathologically confirmed matched biopsies 6 CRPC-NE, 5 CRPC-Adeno (based on patient diagnosis at blood draw). CRPC-NE was defined on the basis of tumor morphology (22). Tumor purity estimation was obtained from matched genomic samples applying CLONETv2.
- (iii) ctDNA WGBS cohort (this work): processed as described in Beltran and colleagues 2020 (12), from pathologically confirmed 6 CRPC-NE, 9 CRPC-Adeno (based on patient diagnosis at blood draw). CRPC-NE was defined on the basis of tumor morphology (22). Tumor content estimation was obtained from matched genomic data applying CLONETv2 or, if the genomic data was unavailable or of insufficient quality, using PAMES (48).
- (iv) Wu2020 cohort (19): cfDNA samples from 19 patients with metastatic CRPC, including samples collected prior to starting enzalutamide or abiraterone ($n = 19$), during treatment ($n = 3$), or at progression ($n = 16$) were subjected to target enrichment NGS for 5.5 million pan-genome CpG sites. All patients had rising PSA at the time of plasma collection. We obtained DNA methylation counts resulting from alignment with bismark (49) to the hg19 genome from the corresponding authors. Only targeted sequencing samples with suitable depth were included. For this cohort, a more lenient filter on CpG coverage (≥ 5) was applied during preprocessing to mitigate the limited overlap with our design.

A full list of datasets analyzed and produced in this study is reported in Supplementary Table S8.

Differential Methylation Analysis across PCa Evolution

Differential methylation analysis focused on regions (DMR) was performed using Rockermeth with default parameters, except for

na_threshold set to 0.3 to tolerate the missing values commonly observed in sequencing data. We set the positive group for each run to the most advanced disease state, following the divergent clonal evolution hypothesis. FDR of 0.05, and the presence of at least 6 detected CpG sites inside the region were applied as filters to retain significant DMRs. To compute the DMR burden across the genome, the sum of the length of all DMRs for each class was divided by the haploid genome size. For the comparison with DMSs reported in Beltran2016, the table with significant CpGs was obtained from the publication and overlapped with DMRs.

Differential Methylation-based Motif Analysis

For the motif analysis on DMSs, the following procedure was applied. First, for any given comparison and starting from the Rockermeth (27) intermediate AUC result, CpG sites with an AUC lower than 0.2 or higher than 0.8 were considered hypomethylated and hypermethylated, respectively. The coordinate of each CpG was expanded by 150bp on both sides, and overlapping genomic windows were collapsed. This refinement step aims to mimic the common size of peaks obtained in experiments like chromatin immunoprecipitation sequencing or ATAC-seq, in contrast to the broader size of Rockermeth DMRs. HOMER motif enrichment was performed on hyper and hypo region sets separately, with the following parameters:

```
findMotifsGenome.pl {} hg19 -p 2 -size 200 -len 8 -gc -mask
```

Notice that HOMER applies zero or one occurrence per sequence counting. After obtaining the motif enrichment for both sets, each motif was ranked based on the enrichment *P*-value obtained. The difference between the rank in hyper and hypo region sets was scored for each motif and normalized in the $[-1, +1]$ range. A motif enriched only in hypo or hyper regions will have an extreme value (near +1 or -1), whereas a motif presenting a similar enrichment in both sets of regions will have a difference in rank near 0. Notably, motifs can be significantly enriched in both sets.

Comparison with Chromatin Accessibility Data

Two independent experimental datasets were queried to compare the differential enrichment of motifs obtained from the previous procedure. Both data sources were generated using ATAC-seq. The comparison with the CRPC-NE versus normal prostate was carried out using supplementary data from the original work (50), which reported the motif ranking in transformed prostate cell (PARCB models) lines versus the original healthy prostatic cells. Motif ranks obtained from DNA methylation analysis were normalized to match the range to the one reported in the study. Similarly, ATAC-seq data from organoids reported in Tang2022 (25) were downloaded and reprocessed using a standardized pipeline (nf-core/atacseq, v1.2, <https://nf-co.re/>). After obtaining a consensus peak set, raw counts were processed using DESeq2 (51), and differential accessibility analysis was performed. Differentially accessible peaks ($FDR < 0.01$, $|\log_2FC| > 1$) were separately fed as input for motif enrichment analysis with HOMER. The rank difference in motif enrichment was scored as in the DNA methylation-based analysis described previously. A set of TFs of interest was obtained by annotating the most differentially enriched motifs based on literature and adding known determinants of CRPC-NE biology.

Panel Design

The NEMO panel is designed to concomitantly assess multiple tumor features in the circulation of prostate cancer patients, from the tumor fraction (i.e., tumor content estimation, TC set), to its molecular characterization, specifically concerning the adenocarcinoma versus the neuroendocrine components (NE set). A hybrid knowledge-based data-driven approach was implemented for the panel design. In total, the panel includes 1,980 informative regions across modules (Supplementary Table S9).

Tumor Content Module

The accurate assessment of the circulating tumor fraction is essential to correctly interpret the DNA methylation value at sites that guide the CRPC-NE versus CRPC-Adeno classification, as the WBC-derived signal could confound the detection of phenotype-specific DNA methylation patterns. Previous reports (14, 31, 52) highlighted how the healthy component of cfDNA is vastly dominated by hematopoietic-derived signals, including granulocytes, monocytes, lymphocytes, and other progenitors. We seek CpG clusters that are extremely different in mCRPC (with CRPC-Adeno and CRPC-NE considered as a single entity) compared with WBCs. This selection is similar to the one operated for tissue deconvolution. We account for only two primary signal sources: metastatic prostate cancer and healthy cfDNA. In this setting, an ideal CpG site could have a $\beta = 0$ in all WBC populations and a $\beta = 1$ in all mCRPC cases, assuming DNA methylation is measured in pure samples. The opposite signal would be equally informative.

Briefly, the selection of the NEMO panel TC set sites is as follows:

- (i) Start from a set of CpG clusters reported in CancerLocator (53).
- (ii) Collect correspondent DNA methylation levels from mCRPC solid tissue biopsies and WBC using the published dataset, including Beltran2016, Zhao2020, and BLUEPRINT.
- (iii) Retain cancer samples with tumor purity greater than 0.75. Retain regions with at least 3 CpGs with coverage of 10X or greater and minimal variance of β in WBC profiles [$sd(\beta) < 0.1$].
- (iv) Find the top hypermethylated and hypomethylated regions comparing mCRPC with WBC. The platform (WGBS, eRRBS) is modeled as a covariate in the differential analysis, performed using limma GLM and empirical Bayes procedure on *M*-values transformed methylation values (54).
- (v) Using an iterative procedure, exclude CpG clusters within 1 Mb from a cluster that has already been included in the set, thus ensuring the spread of the signal across independent regions.

Previous works based on tissue-based analyses of tumor versus normal cells demonstrated that even a dozen robust and independent CpGs or CpG islands (CGI) are sufficient to inform tumor content (48, 55). We opted to limit this set to tens of CpG clusters, reasoning that a small group of regions would likely recover good signals and allow for robust estimation. To check the consistency of differential methylation, two independent sets of high-purity TCGA-PRAD samples (TC > 75%) and WBC samples were used (both based on HM450 array, TCGA consortium, and Hannum dataset (56)). We use this independent test to filter out a few regions that perform poorly in primary prostate cancer and retain the rest, resulting in 26 hypermethylated and 27 hypomethylated regions with opposite methylation states in mCRPC and WBC.

Phenotype Evidence Score Module

The set of genomic regions to assess NE evidence in the circulation of mCRPC patients was defined by a data-driven approach considering (i) informative CpG clusters proximal to promoters and (ii) Intergenic genomic regions based on differentially methylated regions (DMR), which have been previously reported as widely differentially methylated, likely due to modifications in regulatory enhancers (26). Given the rarity of CRPC-NE DNA methylation profiles, we decided to use all the available solid tissue biopsies for those steps, leaving further validation to *in vitro* models, PDXs, and cfDNA samples with known disease states.

CpG Clusters

Starting from the same CpG clusters mentioned above, we sought to find regions that maximize the differential signal between a collection of CRPC-Adeno and CRPC-NE metastatic tissue biopsies.

We used a similar procedure as the one used in the tumor content set, retaining all available tissue samples from CRPC-NE and CRPC-Adeno patients. As CRPC-NE samples have significantly higher cellularity than CRPC-Adeno, we account for different tumor content using CellDMC (57). After setting stringent thresholds on P value and $\Delta\beta$, only a few tens of regions were scored as significant. This is possibly due to the limited extent of the CpG set used or the fact that CpG clusters are constructed by design around promoters associated with the CpG island and covered by the HM450 array. At the same time, the DNA methylation changes observed in the neuroendocrine phenotype seem to fall mostly in intergenic (open sea CpG) regions (our data and ref. 26). For this reason, we complement this with an additional region set described in the next paragraph.

DMRs

To maximize the search space for informative DMRs and include intergenic regions, we turned to our recent genome-wide analysis reported in Beltran and colleagues 2020 (12), where CRPC-Adeno and CRPC-NE metastatic tissue biopsies were processed using the Rockermeth (27) DMR caller, which finds genomic segments that display coherent DNA methylation changes between two groups of samples. Each segment's average DNA methylation value can be measured for every sample and later corrected for heterogeneous tumor content in a framework similar to the one employed for CpG clusters, accounting for diverse tumor content values (assessed by CLONET or by PAMES) using CellDMC. As expected, this correction excludes several DMRs that display a slight change in β ($|\Delta\beta| < 20$), a value close to the observed difference in cellularity between the two sample groups. Once this set of NE-specific DMRs is nominated, we checked their behavior on the independent cohort (10) and observed that DMRs with the strongest differential signal are largely coherent. Therefore, to retain the best signal, we included only DMRs with $|\Delta\beta| > 40$. Because this set of DMRs is large in terms of bps covered (a single DMR might span up to 1 Mb), we only include CpGs with clearly differential signals which define single DMRs, avoiding capturing portions of the regions devoid of informative CpG sites. This procedure led to 919 regions. Finally, the CpG sites used in the integrated NE-classifier described in our previous study (3) are also included in the NEMO panel ($n = 40$).

Additional Regions of Interest

Beyond the differentially methylated sites and regions used for NE classification, we added two additional sets of regions of interest. These sets encompass differentially methylated CpG sites described in Beltran2016, with additional filters for either (i) extreme segregation between groups and differential methylation or (ii) annotation with a curated list of genes involved in mCRPC biology and NE transdifferentiation. Selected genes include *AR*, *STX1*, *EZH2*, *ENO2*, *NCAM1*, *SYP*, *FOXA2*, *CHGA*, *CHGB*, *SLAH2*, *ASCL1*, *NEUROD1*, *INSM1*, *SPDEF*, *ASXL3*, *DLL3*, *TMPRSS2*, *KLK3*, *SOX2*, *CDH2*, *GSTP1*, *SEZ6*, *BAALC*, *ELAVL4*, *GATA2*, *CAND2*, *ETV5*, and *TRIM9*. The selection is based on NE-related literature, common biomarkers, and genes reported as differentially expressed and methylated in a previous study (3). Those region sets are small and less suitable for classification purposes but might inform the transcriptional state of single genes.

DNA methylation data's availability is minimal compared with gene expression; we also designated a small set of NEMO regions by leveraging published transcriptomic data. Briefly, two independent datasets with matched DNA methylation and gene expression data were used to nominate a subset of genes generally regulated by the promoter's DNA methylation, following the procedure described in EPISCOPE (58). This analysis selects genes likely regulated by DNA methylation in normal tissues and thus expected to have DNA methylation changes following sharp deregulations. Next, we query this subset of genes for either inclusion in the abovementioned set of interest or its deregulation between CRPC-NE and CRPC-Adeno. For

gene deregulation, we leveraged solid tissue samples (3) and an established isogenic model that recapitulates NE trans-differentiation (50). We require that a gene is sharply upregulated in at least one of the two comparisons and then add the differential CpGs in its promoters to this set. This strategy retrieves a series of genes that have already been included in the previous selection steps but also allows for the inclusion of additional genes for which the DNA methylation data on a limited set of samples is not conclusive.

NEMO Datasets

In Vitro Dilutions and Preclinical Models. *In vitro* dilutions were performed with genomic DNA (gDNA) or cfDNA to the desired ratio after quantification using Qubit dsDNA High Sensitivity Assay (Invitrogen). Dilutions were performed before fragmentation (for gDNA) and bisulfite conversion. gDNA from LNCaP cell line was mixed with gDNA from healthy donors PBMCs to a final amount of 500 ng as follows: 250 ng and 250 ng, 125 ng and 375 ng, 50 ng and 450 ng, 25 ng and 475 ng, 10 ng and 490 ng, 5 ng and 495 ng, and 0.5 ng and 499.5 ng. The same dilutions were performed with gDNA from a patient-derived organoid cell line (PM154) and gDNA from LNCaP cell line. cfDNA from a healthy donor (purchased by Cambridge Bioscience) was mixed to a total amount of, respectively, 25 and 20 ng with cfDNA from a prostate cancer patient with high tumor content (PCF-SELECT estimation: 0.86) as follows: 12.5 ng and 12.5 ng and 15 ng and 5 ng. Similarly, three purchased plasma samples were used to obtain healthy plasma cell-free DNA (HD samples). Cell lines were purchased from the ATCC and/or authenticated using the ATCC STR profile as reference. All cell lines used had a number of passages below 10, and *Mycoplasma* tests were conducted periodically with InvivoGen Plasmotest. Flash frozen tissues from LuCaP-PDX samples were obtained from collaborators at the University of Washington. DNA extracted from CRPC organoids were obtained from cultures at WCM and DFCI.

Sample Collection and Inclusion and Ethics

For the PRIME cohort samples, plasma samples were obtained from a cohort of consented patients with metastatic CRPC prospectively enrolled at multiple clinical centers under an Ethics protocol approved by the Trento Hospital Ethics Committee (APSS Trento Ethics Committee approval # Int. 2562), the lead clinical center, and subsequently at all satellite sites. Deidentified plasma samples from WCM and DFCI were collected after written informed consent from patients with metastatic CRPC with or without pathologically confirmed NEPC (#WCM IRB 1305013903, #DFCI IRB 19-883). Plasma samples from the Alisertib cohort were collected as part of a single-arm, multi-institutional open-label phase II trial (NCT01799278) (20). Plasma samples from the docetaxel plus carboplatin chemotherapy trial cohort were collected as part of the phase II clinical trial at MD Anderson Cancer Center (NCT00514540; ref. 21). Genomic DNA from pools of healthy donor PBMCs (males, age >40 years old with no medical history of cancer or anticancer treatments) were collected as part of the PRIME study and used as control and for serial dilutions (approved by the University of Trento Ethics Committee, UniTrento #2017-010).

Plasma Preparation and cfDNA Extraction

Plasma was separated from whole blood with a standard double spin protocol and stored at -80°C . The input plasma volume ranged between 1 mL and 2 mL depending on the cohort. For cfDNA extraction of samples processed in Trento, plasma was thawed at room temperature (RT) and immediately processed with the QIAamp Circulating Nucleic Acid Kit (QIAGEN) according to the manufacturer's protocol. The extracted cfDNA was eluted in 30 μL Tris HCl 10 mmol/L pH 8 and quantified with Qubit dsDNA High Sensitivity Assay (Invitrogen); the quality was assessed using the Bioanalyzer High Sensitivity DNA Kit (Agilent). For samples processed at

Weill Cornell Medicine (WCM, DFCI, and Alisertib trial), 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 of RNA carrier at 55°C for 30 minutes. cfDNA capture on the superparamagnetic particles was accomplished via the addition of three volumes (6 mL) of LYS buffer, 0.8 mL isopropanol, and 30 µL NGSTM Beads. The capture of cfDNA was carried out via 30-minute room temperature incubation, 2x wash and 2 × 80% EtOH, air drying, and elution in 30 µL 10 mmol/L Tris, 0.1 mmol/L EDTA (pH 8.5).

gDNA Extraction and Fragmentation

For PDX samples, 20–25 5-µm slides of fresh-frozen tissue were cut using the HM525 NX cryostat (Thermo Fisher Scientific) and used for subsequent gDNA extraction. gDNA from PDX was extracted using the DNeasy Blood & Tissue Kit (QIAGEN), and gDNA from cell lines was extracted using the NucleoSpin Tissue Kit (Macherey-Nagel) according to the manufacturer's instructions and eluted in 100 µL Tris HCl 10 mmol/L pH 8. The gDNA was quantified using the Qubit dsDNA High Sensitivity Assay (Invitrogen). gDNA for libraries preparation was fragmented to a target size of 350 bp using the Covaris focused ultrasonicator M220.

Bisulfite Conversion, Library Preparation, and Capture Sequencing

On the basis of the manufacturer's instructions, bisulfite conversion was performed on 5–25 ng of cfDNA and 200 ng of fragmented gDNA using the EZ DNA Methylation-Lightning Kit (Zymo). gDNA was requantified after bisulfite conversion using Qubit ssDNA High Sensitivity Assay (Invitrogen) and 100 ng of converted gDNA was used as input, whereas the whole converted cfDNA was used for libraries preparation for target sequencing with xGen Methyl-Seq DNA Library Prep Kit (IDT), following the kit's protocol with few modifications. All indexing PCR reagents were substituted with the HiFi HotStart Uracil+ ReadyMix (KAPA), and 11–12 PCR cycles were performed to achieve library yields higher than 500 ng for downstream experiments. For probes hybridization, 500 ng of each library of up to 12 samples were pooled together and captured using the xGen Hybridization and Wash Kit (IDT) following the xGen hybridization capture of DNA libraries protocol (version 4). The multiplexed libraries were dried together with the custom xGen Blocking Oligos and Human Cot DNA. Probes hybridization occurred during overnight incubation with custom xGen Lockdown Probes at 65°C. The captured DNA was amplified for 11 PCR cycles according to the panel size. Pre- and post-capture libraries were quantified using Qubit dsDNA High Sensitivity Assay (Invitrogen), and the quality was assessed with the Bioanalyzer High Sensitivity DNA Kit (Agilent). Libraries were sequenced on Illumina NovaSeq 6000 or MiSeq sequencing machines, with variable nominal depth based on the run and sample type (~100X for genomic DNA, ~1000X for cfDNA samples).

Data Processing

FASTQ files were processed with the methyl-seq pipeline version 1.6 from the nf-core repository (59) (<https://nf-co.re/>). Using the command:

```
nextflow run nf-core/methylseq -r 1.6 -profile docker -aligner bismark -input $FASTQ_PATH -genome hg19 -accel
```

β values for each CpG are obtained from Bismark coverage files as $M/(U+M)$ where M is the count of methylated reads, and U is the count of unmethylated reads. The on-target rate was estimated with *deeptools* (60). Conversion efficiency was estimated using CHH/CHG sites, as no methylation is expected in those contexts in any tissue of our interest. Downstream analyses were performed with custom scripts with the methods detailed below. One sample from the

Alisertib cohort and 9 samples from the MD Anderson cohort were excluded from the analysis due to insufficient coverage.

Tumor Content Estimation

An optimally informative CpG cluster region would reflect the tumor content (TC) of a cfDNA sample directly: observing a β of 0.7 would imply that the TC is around 0.7. Clearly, this observation can be extended to many CpG sites and works when inverting the differential methylation sign. Averaging several CpG sites spread across the genome would lead to a stable estimation of TC, likely balancing out the possible local fluctuations dictated by copy number changes, biological variability, and coverage. To estimate the tumor content for a sample, we start with the β values measured in the dedicated regions. We here outline the procedure for hypermethylated regions in CRPC compared with WBC, and the same holds for regions with inverse profiles. This strategy is inspired by PAMES (48), with a few adaptations for the cfDNA setting.

First, we notice that the β in WBC is very close to zero, whereas the β in CRPC samples is high but not as extreme. This behavior is expected as tissue biopsies might contain immune cell infiltrates that would dilute the CRPC-specific signal. Consequently, it is reasonable to assume that the DNA methylation state of those regions is opposite in WBC and CRPC cells and that the remaining variability is either technical or dictated by cell admixture. Thus, the DNA methylation level of any of those regions offers a proxy of tumor content. A first raw estimation of the tumor content is obtained as the median of the DNA methylation level across informative regions:

$$TC_R = \text{med}(\beta_i^r - \text{med}(\beta_i^w))$$

where i is the index running through the informative regions, and TC_R is the first raw tumor content estimation. We chose the median over the mean to increase stability in the case of outliers. Typically, the two peaks characterizing the bimodal nature of DNA methylation across all CpG sites are expected at 0 and 1. However, platform-specific offsets can produce biases toward less extreme values (61). To account for such shifts, we rely on the average methylation level in WBCs: assuming concordant DNA methylation levels at selected loci, those samples provide a reliable estimation of the expected technical offsets. Following this idea, we correct the first estimation as follows:

$$TC = \frac{TC_R}{h-o}$$

where h and o are the platform offsets estimated in WBC for the fully methylated and demethylated regions, respectively. This procedure is equivalent to stretching over the available signal range to populate the range [0–1]. Regions that fall below zero or above one are set to zero and one, respectively. Finally, to check the stability of the estimation of every single sample, we iteratively used 50% of regions and performed 100 permutations using only that subsection. High variability across permutations (i.e., >25%) would suggest conflicting signals in single regions, pointing to possible cases of poor estimation or to samples with unexpected tissue of origin. Finally, we utilize a collection of healthy cfDNA profiles as the reference background for TC estimation in plasma-derived samples instead of WBC profiles. This modification has only a marginal effect on high tumor-content samples but prevents spurious tumor content estimation driven by small fractions of healthy cfDNA of nonhematopoietic origin. Notably, we observed saturation of signal in few cases, with our TC estimation reaching almost 100%. Those events are rare and present only in the most aggressive cohorts, suggesting that the high quantity of ctDNA in circulation shadows the healthy cfDNA background.

Read-based Tumor Content Detection

To validate our tumor content estimation in the lower spectrum ($TC < 15\%$), we adapted a read-based metric previously introduced in CancerDetector (16) Alpha value. Complementing the standard

β value, the Alpha value of each read overlapping an informative region is computed as the fraction of methylated CpG over the total CpG. This metric allows for a more precise distinction between different read configurations that might generate the same β values. We leverage the Alpha value of tumor content informative regions to first exclude any read that does not exhibit a totally coherent behavior, retaining reads with at least 4 CpG sites, at least 100-bp overlap with informative regions, and an Alpha value of 0 or 1. Next, based on the direction of differential methylation, we compute the fraction of reads supporting tumor-derived signal versus the total. We compare the fraction of supporting reads for all informative regions between each ctDNA sample with cfDNA from healthy donors (HD). A significantly higher fraction of tumor-supporting signal compared to HDs (Wilcoxon signed rank test) indicates the presence of a subpopulation of tumor-derived reads, which is expected to be proportional to the tumor content.

PE Score Estimation

We sought to construct a NE-evidence score to detect CRPC-NE. As tumor content is a major confounder of DNA methylation signal, we use a strategy that is aware of this covariate and leverages the previously obtained TC estimation. We first constructed a reference of highly pure CRPC-Adeno and CRPC-NE solid tumor samples (TC > 0.8, NE = 10, Adeno = 29) and collected healthy cfDNA as a reference for noncancer cfDNA. Next, we reconstruct the observed signal as a linear combination of 3 components: cfDNA (known, 1-TC), CRPC-Adeno, and CRPC-NE. Intuitively, we seek to reconstruct the observed signal as the combination of healthy and cancer-derived signals, knowing that the total contribution of the latter should be equal to the previously estimated tumor content. This result can be obtained with the following framework. First, for each region, we estimate the mean (μ) and SD (σ) of β values for each region in the three reference groups. For each informative region, the observed cfDNA profile can be modeled as a linear combination of three components, and the global estimation of the relative fractions is achieved using a Bayesian linear regression (implemented in the R package “brms,” with: “nchain = 1,” 500 burn-in steps and 2,000 sampling steps) with strongly informative priors on the cfDNA fraction and noninformative priors on the CRPC-Adeno and CRPC-NE components.

$$\begin{aligned} f_{cf} &\sim \text{Norm}(1 - TC, 0.01) \\ f_{Ad} &\sim \text{Norm}(TC/2, 0.5) \\ f_{NE} &\sim \text{Norm}(TC/2, 0.5) \\ \beta_{obs} &= f_{cf}\mu_{cf} + f_{Ad}\mu_{Ad} + f_{NE}\mu_{NE} \end{aligned}$$

Furthermore, the regression is weighted on the basis of the inverse of the SD of each region in CRPC-Adeno and CRPC-NE samples, increasing the contributions of regions that are more stable. After the regression coefficients are estimated, negative values are set to 0, and the remaining coefficients are normalized to sum to one, following widely adopted a posteriori constraints applied in cell-type deconvolution (62). Once the fraction of NE contribution is obtained, it is normalized over the total tumor content to obtain the PE score, which ranges from 0 to 1 and is independent of the tumor content. A tissue-informed binary threshold of 0.43 was obtained by computing the PE score on all the solid tissue biopsies analyzed and setting the threshold between the means of the two distributions. Notably, a single CRPC-NE outlier with a low PE score has a transcriptomic profile compatible with CRPC-Adeno, as reported in a recent study (32). We noticed that propagating the uncertainty in tumor content estimation partially hampers the reliability of the PE score below 10% and makes it uninformative below 3%. We opted not to provide any PE score assessment for cfDNA samples with a TC below 3%, although we suggest caution in interpreting the PE score between 5% to 10% of tumor content. To test the analytical stability of the PE score (Supplementary Fig. S4F), we collected a set of representative ctDNA samples at high tumor content and heterogeneous PE score from this study, and we mixed DNA methylation signals with increasing

fractions of healthy cfDNA signal, following the equation $\beta_d = \beta_{ct}(d) + \beta_{cf}(1-d)$, where d indicates decreasing dilution factors (0–1, step 0.1). Importantly, the tumor content estimation and the PE score estimation procedures are tailored to the NEMO design, providing the required performance in this specific setting. We chose to develop this tailored approach over standard deconvolutions to better leverage the stepwise estimations of the NEMO modules. However, these strategies were not developed as generally applicable tumor content estimation or deconvolution algorithms. Average coverage for NEMO regions in all the samples reanalyzed in this work are reported in Supplementary Table S10.

Lung Cancer Analysis

To evaluate the performance of the NEMO assay in a different setting, we evaluated published DNA methylation data of lung adenocarcinoma (LUAD) and small-cell lung cancer (SCLC) cell lines (<https://depmap.org>) and solid tissue biopsies (TCGA-LUAD, SCLC (63)). As all those samples were profiled using the Illumina HM450 methylation array, we limited the analysis to a subset of probes contained in the NEMO design ($n = 866$). Differential methylation analysis was performed using cell lines and PBMC samples and prioritizing differentially methylated CpGs to estimate tumor content ($n = 50$) and PE score ($n = 200$). Next, the NEMO statistical framework was applied to solid tissue biopsy data based on this adapted subset of informative regions and using the average of cancer cell and PBMC lines as the reference.

Integrative Gene Expression Analysis of PDX Samples and WCDT Cohort

Gene expression data for the PDX samples were retrieved from Recount3 (refs. 23, 64; SRP183532), except for the LuCaP 243 PDX model, for which counts have been obtained from the Nelson Lab and integrated with the other PDX models. The GSEA analysis was performed with clusterprofiler (65): each gene was correlated with the PE score and ranked by correlation coefficient. In addition to the 12 PDX models profiled with NEMO, we utilized a larger set of 18 PDX DNA methylation profiles (EPIC array, GSE227814, and GSE227695). The genome-wide array data were masked using the NEMO design, obtaining the average DNA methylation value for each informative region with at least one probe. Gene expression data for the WCDT cohort were obtained from the GDC portal (<https://portal.gdc.cancer.gov/>, WCDT-MCRPC). For the association analysis between DNA methylation and gene expression, only the common genes measured in both PDXs and WCDT cohorts were retained. First, the correlation of each informative region was tested against the gene expression of the corresponding annotated gene, following the annotation obtained with HOMER (66). For each gene, only the regions with a range of β values greater than 0.5 across PDXs were tested, and only the region with the most significant correlation was retained. The same regions were then tested for correlation in the WCDT cohort, retaining only samples with tumor content greater than 75%. In both cases, P values were corrected with the FDR procedure. Regulon activity was computed using VIPER (67) and Dorothea pan-cancer regulons (34). For the survival analyses based on *EZH2* expression and DNA methylation, only WCDT samples with tumor content greater than 50% were retained. Survival data were retrieved from https://github.com/DavidQuigley/WCDT/tree/master/clinical_metadata. In the case of DNA methylation, the β value was corrected by regressing out the nontumoral component using our healthy cfDNA reference, as infiltrating WBCs might confound the real underlying DNA methylation value. Briefly, DNA methylation values were transformed with an arcsine transformation ($f(x) = \arcsin(2x - 1)$), as described in the InfiniumPurify framework (68). Next, the corrected β was obtained as $\beta_T = \frac{\beta_{Obs} - \beta_C(1 - TC)}{TC}$ and remapped in the original space with the inverse transformation.

The same procedure was applied to cfDNA samples with TC greater than 20% for Supplementary Fig. S5C (from the PRIME, WCM/DFCI, and both phase II clinical trial cohorts).

Genomic Analysis of NEMO Samples

We focused on the PRIME cohort in which orthogonal genomic data is available to test whether regions included in NEMO and additional off-target reads could inform on the copy-number state of regions of interest, both within and outside our design. The 20 samples were profiled with PCF select (ref), a targeted panel optimized to capture genomic alterations in the CRPC settings, which serves as a gold standard in this context. Importantly, the PCF select panel contains a set of 116 genes with only partial overlap with NEMO. First, we sought to leverage off-target reads from our assay to obtain broad copy-number estimations using IchorCNA (ref. 69; Supplementary Fig. S7A). The command was run with a bin size of 500 Kb and the following parameters:

```
-ploidy "c(2,3)" -normal "c(0.5,0.6,0.7,0.8,0.9)" -maxCN 5 -include-HOMD False -chrs "c(1:22, \ "X\ ")" -chrTrain "c(1:22)" -estimateNormal True -estimatePloidy True -estimateScPrevalence True -scStates "c(1,3)" -txnE 0.9999 -txnStrength 10000
```

A panel of normal samples was created by mixing the 3 cfDNA from healthy donors (HD). As for some genes of interest which often present focal gain or losses (*AR*, *TP53*), a target region was available within the NEMO panel; we calculated regional log₂R as previously described (30). We aggregate those estimations prioritizing the log₂R of the target amplicon when available. The comparison between the log₂R is reported in Supplementary Fig. S7B, highlighting the good concordance between the two assays.

Data Analysis and Statistics

Unless otherwise stated, all pairwise tests are unpaired and two-tailed Wilcoxon rank sum tests. All reported boxplot are based on first quartile, median, and third quartile, with whiskers extending up to 1.5 IQR. Unless differently specified, all correlation coefficients are Pearson R rounded to the second digit. Survival data for the Alisertib cohort and MD Anderson cohort were received from corresponding authors of the original studies (refs. 20, 21; notice that survival data was not available for all samples). Kaplan-Meier curves and log-rank *P* values were computed and plotted using the *survival* and *survminer* R packages, and survival data are included in Supplementary Table S6A and S6B. A forward feature selection procedure for Cox proportional HR was used for the multivariate analysis reported in Supplementary Fig. S6D to minimize AIC and including TC, PE score, and all inclusion criteria in each trial as predictive variables. Reported analyses and plots were created with R 4.1.2 with tidyverse packages. Given the retrospective nature of this study, no randomization was applied. Considering the need to measure the classification performance, no sample blinding was applied. No power analysis was conducted, and size was dictated by sample availability. Replication of patient-derived samples was not conducted due to limited input material availability.

Data Availability

An R package to compute tumor content and PE score from DNA methylation counts is available at (<https://github.com/demichelislab/NEMOcode>). Raw counts and processed values are available at Zenodo (<https://doi.org/10.5281/zenodo.8431083>), including region-wise and CpG-wise DNA methylation values for samples sequenced with the NEMO assay and reprocessed masked samples, and in the GEO repository (GSE245745). Raw WGBS data have been deposited in dbGaP (WGBS: phs001752v2).

Authors' Disclosures

G. Franceschini reports a patent for the NEMO platform pending. M. Benelli reports personal fees from Novartis outside the submitted

work. B. Hanratty reports a patent for DNA Hypomethylation as a Predictive Marker in Cancer Therapy pending and a patent for Inferring Transcription Factor Activity from DNA Methylation and its Application as a Biomarker pending. M. Adil reports a patent for 18/340699 pending. O. Elemento reports personal fees and other support from Volastra Therapeutics; other support from Owkin; personal fees from Champion Oncology; and other support from Freenome outside the submitted work. S.T. Tagawa reports grants from NCI during the conduct of the study; grants and personal fees from Sanofi, Astellas, Pfizer, Merck, Gilead, POINT Biopharma, Novartis, Janssen, and Bayer; personal fees from Daiichi Sankyo and EMD Serono; grants and personal fees from Telix; and grants and personal fees from Amgen outside the submitted work. F.Y. Feng reports personal fees from Janssen, Myovant, Roivant, Bayer, Novartis, SerImmune, Artera, Bristol Meier Squibb, ClearNote Genomics, Astellas, and Novartis, and personal fees from Genentech outside the submitted work. O. Caffo reports grants from Accelerated Award 2018 from AIRC and Cancer Foundation UK during the conduct of the study; personal fees from AAA, Astellas, Bayer, Janssen, AstraZeneca, Ipsen, Novartis, MSD, and Pfizer, and personal fees from Recordati outside the submitted work. U. Basso reports grants from Associazione Italiana per Ricerca sul Cancro AIRC during the conduct of the study; personal fees from BMS, Novartis, MSD, Janssen, Merck, and Astellas; personal fees from Pfizer outside the submitted work; and grants and personal fees from Ipsen. P.S. Nelson reports grants and personal fees from Janssen; personal fees from Merck and Pfizer, and personal fees from Venable Fitzpatrick outside the submitted work. E. Corey reports grants from NIH and grants from NIH outside the submitted work; and Eva Corey was a consultant of DotQuant and received funding under Institutional SRA from AstraZeneca, Bayer Pharmaceuticals, MacroGenics, Foghorn, KronosBio, Janssen Research, and Forma Therapeutics. M.C. Haffner reports a patent for Inferring Transcription Factor Activity from DNA Methylation and its Application as a Biomarker pending. G. Attard reports grants from Cancer Research UK, Prostate Cancer Foundation, and grants from John Black Charitable Foudnation during the conduct of the study; grants, personal fees, and non-financial support from Janssen, Novartis, and Astellas; personal fees from Pfizer and AstraZeneca; grants and personal fees from Blue Earth Therapeutics and Veracyte, and grants and non-financial support from Artera outside the submitted work; in addition, G. Attard has a patent for Blood biomarkers issued and a patent for Abiraterone with royalties paid. F. Demichelis reports grants from NCI and grants from Fondazione AIRC during the conduct of the study; in addition, F. Demichelis has a patent for WO 2021/162981 A2 (priority date: 11.02.2020 US) pending and a patent for P032355GB pending. H. Beltran reports grants from NCI, grants from Prostate Cancer Foundation, and grants from Department of Defense during the conduct of the study; grants and personal fees from Daiichi Sankyo, grants from Novartis, Bristol Myers Squibb, and Circle Pharma; personal fees from Janssen, Astellas, Merck, Pfizer, Foundation Medicine, Blue Earth Diagnostics, LOXO, Bayer, Amgen, Oncorus, Curie Therapeutics, and Sanofi, and personal fees from Astra-Zeneca outside the submitted work; in addition, H. Beltran has a patent for NEMO platform as coinventor issued to WO 2021/162981 A2. No disclosures were reported by the other authors.

Authors' Contributions

G. Franceschini: Conceptualization, data curation, software, formal analysis, validation, investigation, visualization, methodology, writing—original draft, writing—review and editing. **O. Quaini:** Resources, data curation, methodology, writing—review and editing. **K. Mizuno:** Resources, writing—review and editing. **F. Orlando:** Resources, investigation, methodology, writing—review and editing. **Y. Ciani:** Investigation, writing—review and editing. **S. Ku:** Resources, writing—review and editing. **M. Sigouros:** Resources, methodology, writing—review and editing. **E. Rothmann:** Resources, project administration, writing—review and editing. **A. Alonso:** Resources,

methodology, writing–review and editing. **M. Benelli:** Methodology, writing–review and editing. **C. Nardella:** Project administration, writing–review and editing. **J. Auh:** Resources, writing–review and editing. **D. Freeman:** Resources, writing–review and editing. **B. Hanratty:** Resources, data curation, writing–review and editing. **M. Adil:** Resources, data curation, writing–review and editing. **O. Elemento:** Resources, writing–review and editing. **S.T. Tagawa:** Resources, writing–review and editing. **F.Y. Feng:** Resources, writing–review and editing. **O. Caffo:** Resources, writing–review and editing. **C. Buttiglieri:** Resources, writing–review and editing. **U. Basso:** Resources, writing–review and editing. **P.S. Nelson:** Resources, writing–review and editing. **E. Corey:** Resources, formal analysis, writing–review and editing. **M.C. Haffner:** Resources, formal analysis, writing–review and editing. **G. Attard:** Resources, formal analysis, writing–review and editing. **A. Aparicio:** Resources, formal analysis, writing–review and editing. **F. Demichelis:** Conceptualization, resources, data curation, formal analysis, supervision, funding acquisition, validation, investigation, visualization, methodology, writing–original draft, project administration, writing–review and editing. **H. Beltran:** Conceptualization, resources, data curation, formal analysis, supervision, funding acquisition, validation, investigation, methodology, writing–original draft, project administration, writing–review and editing.

Acknowledgments

We would like to thank the patients who generously donated blood and tissue that made this research possible. We also thank the CIBIO NGS facility and the Weill Cornell Epicore facility for their help with experimental setup and sequencing. This study was supported by grants from the Prostate Cancer Foundation (to H. Beltran, F. Demichelis), NCI (SPORE P50-CA211024 to H. Beltran, F. Demichelis; R37CA241486 to H. Beltran; P30CA015704, R01CA234715 and R01CA266452, to P.S. Nelson), from Fondazione AIRC per la Ricerca sul Cancro (AIRC) and Cancer Research UK (CRUK) (AIRC ID 22792 and CRUK ID A26822, Accelerator Award to F. Demichelis, O. Caffo, C. Buttiglieri, U. Basso, G. Attard), Department of Defense (W81XWH-17-1-0653 to HB), the Doris Duke Foundation, the Safeway Foundation, and the V Foundation (to M.C. Haffner). The establishment of LuCaP PDXs was supported by the Pacific Northwest Prostate Cancer SPORE (P50CA97186), the P01 NIH grant (P01CA163227), and the Institute for Prostate Cancer Research. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The publication costs of this article were defrayed in part by the payment of publication fees. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 USC section 1734.

Note

Supplementary data for this article are available at Cancer Discovery Online (<http://cancerdiscovery.aacrjournals.org/>).

Received July 11, 2023; revised October 31, 2023; accepted December 15, 2023; published first December 18, 2023.

REFERENCES

- Lorenzin F, Demichelis F. Evolution of the prostate cancer genome towards resistance. *J Transl Genet Genom* 2019;3:5.
- Bluemn EG, Coleman IM, Lucas JM, Coleman RT, Hernandez-Lopez S, Tharakan R, et al. Androgen receptor pathway-independent prostate cancer is sustained through FGF signaling. *Cancer Cell* 2017;32:474–89.
- Beltran H, Prandi D, Mosquera JM, Benelli M, Puca L, Cyrta J, et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat Med* 2016;22:298–305.
- Davies AH, Beltran H, Zoubeidi A. Cellular plasticity and the neuroendocrine phenotype in prostate cancer. *Nat Rev Urol* 2018;15:271–86.
- Ku SY, Rosario S, Wang Y, Mu P, Seshadri M, Goodrich ZW, et al. *Rb1* and *Tp53* cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. *Science* 2017;355:78–83.
- Mu P, Zhang Z, Benelli M, Karthaus WR, Hoover E, Chen C-C, et al. *SOX2* promotes lineage plasticity and antiandrogen resistance in *TP53*- and *RBI*-deficient prostate cancer. *Science* 2017;355:84–8.
- 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.
- Heitzer E, Haque IS, Roberts CES, Speicher MR. Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat Rev Genet* 2019;20:71–88.
- Aggarwal RR, Quigley DA, Huang J, Zhang L, Beer TM, Rettig MB, et al. Whole-genome and transcriptional analysis of treatment-emergent small-cell neuroendocrine prostate cancer demonstrates intraclass heterogeneity. *Mol Cancer Res* 2019;17:1235–40.
- Zhao SG, Chen WS, Li H, Foye A, Zhang M, Sjöström M, et al. The DNA methylation landscape of advanced prostate cancer. *Nat Genet* 2020;52:778–89.
- Berchuck JE, Baca SC, McClure HM, Korthauer K, Tsai HK, Nuzzo PV, et al. Detecting neuroendocrine prostate cancer through tissue-informed cell-free DNA methylation analysis. *Clin Cancer Res* 2022;28:928–38.
- Beltran H, Romanel A, Conteduca V, Casiraghi N, Sigouros M, Franceschini GM, et al. Circulating tumor DNA profile recognizes transformation to castration-resistant neuroendocrine prostate cancer. *J Clin Invest* 2020;130:1653–68.
- Greenberg MVC, Bourc’his D. The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol* 2019;20:590–607.
- Moss J, Magenheimer J, Neiman D, Zemmour H, Loyfer N, Korach A, et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat Commun* 2018;9:5068.
- Loyfer N, Magenheimer J, Perez A, Cann G, Bredno J, Klochendler A, et al. A DNA methylation atlas of normal human cell types. *Nature* 2023;613:355–64.
- Li W, Li Q, Kang S, Same M, Zhou Y, Sun C, et al. CancerDetector: ultrasensitive and non-invasive cancer detection at the resolution of individual reads using cell-free DNA methylation sequencing data. *Nucleic Acids Res* 2018;46:e89.
- Klein EA, Richards D, Cohn A, Tummala M, Lapham R, Cosgrove D, et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann Oncol* 2021;32:1167–77.
- Silva R, Moran B, Baird A-M, O’Rourke CJ, Finn SP, McDermott R, et al. Longitudinal analysis of individual cfDNA methylome patterns in metastatic prostate cancer. *Clin Epigenetics* 2021;13:168.
- Wu A, Cremaschi P, Wetterskog D, Conteduca V, Franceschini GM, Klefogiannis D, et al. Genome-wide plasma DNA methylation features of metastatic prostate cancer. *J Clin Invest* 2020;130:1991–2000.
- Beltran H, Oromendia C, Danila DC, Montgomery B, Hoimes C, Szmulewicz RZ, 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;25:43–51.
- Aparicio AM, Harzstark AL, Corn PG, Wen S, Araujo JC, Tu S-M, et al. Platinum-based chemotherapy for variant castrate-resistant prostate cancer. *Clin Cancer Res* 2013;19:3621–30.
- Epstein JI, Amin MB, Beltran H, Lotan TL, Mosquera J-M, Reuter VE, et al. Proposed morphologic classification of prostate cancer with neuroendocrine differentiation. *Am J Surg Pathol* 2014;38:756–67.
- Labrecque MP, Coleman IM, Brown LG, True LD, Kollath L, Lakely B, et al. Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. *J Clin Invest* 2019;129:4492–505.
- Baca SC, Takeda DY, Seo J-H, Hwang J, Ku SY, Arafah R, et al. Reprogramming of the FOXA1 cistrome in treatment-emergent neuroendocrine prostate cancer. *Nat Commun* 2021;12:1979.

25. Tang F, Xu D, Wang S, Wong CK, Martinez-Fundichely A, Lee CJ, et al. Chromatin profiles classify castration-resistant prostate cancers suggesting therapeutic targets. *Science* 2022;376:eabe1505.
26. Balanis NG, Sheu KM, Esendebe FN, Patel SJ, Smith BA, Park JW, et al. Pan-cancer convergence to a small-cell neuroendocrine phenotype that shares susceptibilities with hematological malignancies. *Cancer Cell* 2019;36:17–34.
27. Benelli M, Franceschini GM, Magi A, Romagnoli D, Biagioni C, Migliaccio I, et al. Charting differentially methylated regions in cancer with Rocker-meth. *Commun Biol* 2021;4:1–15.
28. Cejas P, Xie Y, Font-Tello A, Lim K, Syamala S, Qiu X, et al. Subtype heterogeneity and epigenetic convergence in neuroendocrine prostate cancer. *Nat Commun* 2021;12:5775.
29. Barefoot ME, Loyfer N, Kiliti AJ, McDeed AP, Kaplan T, Wellstein A. Detection of cell types contributing to cancer from circulating, cell-free methylated DNA. *Front Genet* 2021;12:671057.
30. Orlando F, Romanel A, Trujillo B, Sigouros M, Wetterkog D, Quaini O, et al. Allele-informed copy number evaluation of plasma DNA samples from metastatic prostate cancer patients: the PCF_SELECT consortium assay. *NAR Cancer* 2022;4:zcac016.
31. Fox-Fisher I, Piyanzin S, Ochana BL, Klochendler A, Magenheim J, Peretz A, et al. Remote immune processes revealed by immune-derived circulating cell-free DNA. *eLife* 2021;10:e70520.
32. Bolis M, Bossi D, Vallerga A, Ceserani V, Cavalli M, Impellizzeri D, et al. Dynamic prostate cancer transcriptome analysis delineates the trajectory to disease progression. *Nat Commun* 2021;12:7033.
33. Abida W, Cyrta J, Heller G, Prandi D, Armenia J, Coleman I, et al. Genomic correlates of clinical outcome in advanced prostate cancer. *Proc Natl Acad Sci U S A* 2019;116:11428–36.
34. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res* 2019;29:1363–75.
35. Berger A, Brady NJ, Bareja R, Robinson B, Conteduca V, Augello MA, et al. N-Myc-mediated epigenetic reprogramming drives lineage plasticity in advanced prostate cancer. *J Clin Invest* 2019;129:3924–40.
36. Mateo J, McKay R, Abida W, Aggarwal R, Alumkal J, Alva A, et al. Accelerating precision medicine in metastatic prostate cancer. *Nat Cancer* 2020;1:1041–53.
37. Baca SC, Seo J-H, Davidsohn MP, Fortunato B, Semaan K, Sotudian S, et al. Liquid biopsy epigenomic profiling for cancer subtyping. *Nat Med* 2023;29:2737–41.
38. Reichert ZR, Morgan TM, Li G, Castellanos E, Snow T, Dall'Olio FG, et al. Prognostic value of plasma circulating tumor DNA fraction across four common cancer types: a real-world outcomes study. *Ann Oncol* 2023;34:111–20.
39. Shen SY, Burgener JM, Bratman SV, De Carvalho DD. Preparation of cfMeDIP-seq libraries for methylome profiling of plasma cell-free DNA. *Nat Protoc* 2019;14:2749–80.
40. Ulz P, Perakis S, Zhou Q, Moser T, Belic J, Lazzari I, et al. Inference of transcription factor binding from cell-free DNA enables tumor subtype prediction and early detection. *Nat Commun* 2019;10:4666.
41. De Sarkar N, Patton RD, Doebley A-L, Hanratty B, Adil M, Kreitzman AJ, et al. Nucleosome patterns in circulating tumor DNA reveal transcriptional regulation of advanced prostate cancer phenotypes. *Cancer Discov* 2023;13:632–53.
42. Sjöström M, Zhao SG, Levy S, Zhang M, Ning Y, Shrestha R, et al. The 5-hydroxymethylcytosine landscape of prostate cancer. *Cancer Res* 2022;82:3888–902.
43. Herbets C, Annala M, Sipola J, Ng SWS, Chen XE, Nurminen A, et al. Deep whole-genome ctDNA chronology of treatment-resistant prostate cancer. *Nature* 2022;608:199–208.
44. Lin P-C, Giannopoulou EG, Park K, Mosquera JM, Sboner A, Tewari AK, et al. Epigenomic alterations in localized and advanced prostate cancer. *Neoplasia* 2013;15:373–83.
45. Prandi D, Demichelis F. Ploidy- and purity-adjusted allele-specific DNA analysis using CLONETv2. *Curr Protoc Bioinformatics* 2019;67:e81.
46. Quigley DA, Dang HX, Zhao SG, Lloyd P, Aggarwal R, Alumkal JJ, et al. Genomic hallmarks and structural variation in metastatic prostate cancer. *Cell* 2018;174:758–69.
47. Müller F, Scherer M, Assenov Y, Lutsik P, Walter J, Lengauer T, et al. RnBeads 2.0: comprehensive analysis of DNA methylation data. *Genome Biol* 2019;20:55.
48. Benelli M, Romagnoli D, Demichelis F. Tumor purity quantification by clonal DNA methylation signatures. *Bioinformatics* 2018;34:1642–9.
49. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 2011;27:1571–2.
50. Park JW, Lee JK, Sheu KM, Wang L, Balanis NG, Nguyen K, et al. Reprogramming normal human epithelial tissues to a common, lethal neuroendocrine cancer lineage. *Science* 2018;362:91–5.
51. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
52. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell* 2016;164:57–68.
53. Kang S, Li Q, Chen Q, Zhou Y, Park S Lee G, et al. CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. *Genome Biol* 2017;18:53.
54. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47.
55. Luo H, Zhao Q, Wei W, Zheng L, Yi S, Li G, et al. Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer. *Sci Transl Med* 2020;12:eaax7533.
56. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* 2013;49:359–67.
57. Zheng SC, Breeze CE, Beck S, Teschendorff AE. Identification of differentially methylated cell types in epigenome-wide association studies. *Nat Methods* 2018;15:1059–66.
58. Teschendorff AE, Zhu T, Breeze CE, Beck S. EPISCORE: cell type deconvolution of bulk tissue DNA methylomes from single-cell RNA-Seq data. *Genome Biol* 2020;21:221.
59. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* 2020;38:276–8.
60. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 2016;44:W160–5.
61. Hicks SC, Irizarry RA. methylCC: technology-independent estimation of cell type composition using differentially methylated regions. *Genome Biol* 2019;20:261.
62. Teschendorff AE, Breeze CE, Zheng SC, Beck S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in epigenome-wide association studies. *BMC Bioinformatics* 2017;18:105.
63. Mohammad HP, Smitheman KN, Kamat CD, Soong D, Federowicz KE, Van Aller GS, et al. A DNA hypomethylation signature predicts anti-tumor activity of LSD1 inhibitors in SCLC. *Cancer Cell* 2015;28:57–69.
64. Wilks C, Zheng SC, Chen FY, Charles R, Solomon B, Ling JP, et al. recount3: summaries and queries for large-scale RNA-seq expression and splicing. *Genome Biol* 2021;22:323.
65. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* 2021;2:100141.
66. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;38:576–89.
67. Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH, et al. Network-based inference of protein activity helps functionalize the genetic landscape of cancer. *Nat Genet* 2016;48:838–47.
68. Zheng X, Zhang N, Wu H-J, Wu H. Estimating and accounting for tumor purity in the analysis of DNA methylation data from cancer studies. *Genome Biol* 2017;18:17.
69. Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun* 2017;8:1324.