Divergent clonal evolution of castration resistant neuroendocrine prostate cancer

HHS Public Access

Author manuscript

Nat Med. Author manuscript; available in PMC 2016 August 08.

Published in final edited form as:

Nat Med. 2016 March ; 22(3): 298–305. doi:10.1038/nm.4045.

Divergent clonal evolution of castration resistant neuroendocrine prostate cancer

Himisha Beltran[1,2,3,#], **Davide Prandi**[4,#], **Juan Miguel Mosquera**[1,5], **Matteo Benelli**[4], **Loredana Puca**[1], **Joanna Cyrta**[1], **Clarisse Marotz**[1], **Eugenia Giannopoulou**[6], **Balabhadrapatruni V.S.K. Chakravarthi**[7], **Sooryanarayana Varambally**[7], **Scott A. Tomlins**[8], **David M. Nanus**[2,3], **Scott T. Tagawa**[2,3], **Eliezer M. Van Allen**[9,10], **Olivier Elemento**[1,6], **Andrea Sboner**[1,11], **Levi A. Garraway**[9,10,*], **Mark A. Rubin**[1,3,5,*], and **Francesca Demichelis**[1,4,11,*]

[1] Caryl and Israel Englander Institute for Precision Medicine, New York Presbyterian Hospital-Weill Cornell Medicine. New York, NY

[2] Department of Medicine, Division of Hematology and Medical Oncology, Weill Cornell Medicine. New York, NY

[3] Sandra and Edward Meyer Cancer Center at Weill Cornell Medicine. New York, NY

[4] Centre for Integrative Biology, University of Trento. Trento, Italy

[5] Department of Pathology and Laboratory Medicine. Weill Cornell Medicine. New York, NY

[6] Department of Physiology and Biophysics. Weill Cornell Medicine. New York, NY

[7] Department of Pathology, University of Alabama, Birmingham, AL

[8] Department of Pathology, University of Michigan. Ann Arbor, MI

[9] Department of Medical Oncology, Dana Farber Cancer Institute, Boston, MA

[10] The Broad Institute of MIT and Harvard, Boston, MA

[11]Institute for Computational Biomedicine, Weill Cornell Medicine. New York, NY

Abstract

An increasingly recognized resistance mechanism to androgen receptor (AR)-directed therapy in prostate cancer involves epithelial plasticity, wherein tumor cells demonstrate low to absent AR

Corresponding Authors: Himisha Beltran, MD, ; Email: hip9004@med.cornell.edu, Mark A. Rubin, MD, ; Email: rubinma@med.cornell.edu, Francesca Demichelis, PhD, ; Email: f.demichelis@unitn.it

[#]co-first authorship

[*]co-senior authorship

Author Contributions: Initiation and design of study: H.B., M.A.R., L.A.G., F.D.; Enrolled subjects and contributed samples and clinical data: H.B, S.A.T., D.M.N., S.T.T.; Performed experiments: J.M.M., L.P., J.C., C.M., B.V.S.K.C., S.V.; Statistical and bioinformatics analyses: D.P., M.B, E.G., E.M.V., O.E., A.S., F.D.; Supervision of research: H.B., M.A.R., L.A.G., F.D; Writing of the first draft of the manuscript: H.B., M.A.R., L.A.G., F.D.; All authors contributed to the writing and editing of the revised manuscript and approved the manuscript.

Accession codes: All BAM files and associated sample information are described in **Supplementary Table 11**; data are deposited in dbGap phs000909.v.p1 and accessible on the cBIO Portal for Cancer Genomics.

The authors have no relevant conflicts of interest. The authors declare no competing financial interests.

expression and often neuroendocrine features. The etiology and molecular basis for these "alternative" treatment-resistant cell states remain incompletely understood. Here, by analyzing whole exome sequencing data of metastatic biopsies from patients, we observed significant genomic overlap between castration resistant adenocarcinoma (CRPC-Adeno) and neuroendocrine histologies (CRPC-NE); analysis of serial progression samples points to a model most consistent with divergent clonal evolution. Genome-wide DNA methylation revealed marked epigenetic differences between CRPC-NE and CRPC-Adeno that also designated cases of CRPC-Adeno with clinical features of AR-independence as CRPC-NE, suggesting that epigenetic modifiers may play a role in the induction and/or maintenance of this treatment-resistant state. This study supports the emergence of an alternative, "AR-indifferent" cell state through divergent clonal evolution as a mechanism of treatment resistance in advanced prostate cancer.

Prostate cancer remains a leading cause of male cancer death worldwide[1]. The mainstay of therapy for patients with metastatic spread, including castration resistant disease, is hormonal therapy targeting the AR[2-4]. Enzalutamide and abiraterone are potent AR-targeted therapies approved for the treatment of men with CRPC[5,6]. While significantly improving survival and quality of life, most patients ultimately develop resistance to these agents[7]. Thus, predictive biomarkers to help distinguish responders from non-responders prior to starting the next line of hormonal therapy are urgently needed. We and others have observed that a subset of resistant tumors show small cell carcinoma or neuroendocrine features on metastatic biopsy (CRPC-NE)[8-10]. This phenomenon may therefore reflect an epithelial plasticity that enables tumor adaptation in response to AR-targeted therapies[11-14]. Prognosis of CRPC-NE is poor due to late recognition, heterogeneous clinical features, and lack of effective systemic therapies[15,16].

One major hurdle in the diagnosis and treatment of androgen-independent prostate cancer including CRPC-NE is our lack of understanding of the genetic and epigenetic underpinnings of this aggressive subset. To address this, we interrogated 114 metastatic tumor specimens from 81 individuals including 51 with clinical and histologic features of castration resistant adenocarcinoma (CRPC-Adeno), and 30 with CRPC-NE as confirmed by pathologic consensus criteria[8]; we studied matched normal cells from all patients, multiple tumor biopsies from 17 patients and a single tumor biopsy from 64 patients. We hypothesized that CRPC-NE could be distinguished from CRPC-Adeno based on distinct molecular alterations and that this information could improve upon and supplement the current often challenging diagnostic features reliant on morphology[17]. We also hypothesized that CRPC-NE that develops after therapy arises clonally from a CRPC-Adeno precursor, rather than from selection of pre-existing neuroendocrine clones. Lastly, we hypothesized that AR-independent prostate adenocarcinomas that share CRPC-NE-specific molecular alterations may represent tumors at high risk for progression or in transition towards CRPC-NE.

Results

We evaluated biopsies from a wide range of metastatic sites, with a predominance of bone biopsies in CRPC-Adeno compared to CRPC-NE (31% CRPC-Adeno vs. 2% CRPC-NE, *P*

< 0.05, binomial test) **(Fig. 1a)**. Clinical and pathologic features are summarized in **Supplementary Table 1** and **Supplementary Fig.1**. As expected, CRPC-NE demonstrated on average lower protein expression of the AR by immunohistochemistry **(Fig. 1b)**. We also quantified AR signaling status by measuring expression of mRNAs included in a previously defined AR signature[18] **(Supplementary Table 2)** and observed overall lower abundance in CRPC-NE compared to CRPC-Adeno **(Fig. 1b)**; however, there was significant overlap with a wide range of values observed within each subtype, suggesting that there is a spectrum of AR signaling in advanced prostate cancer that spans pathologic subtypes.

To deepen our understanding of AR independence in general —and of the CRPC-NE phenotype in particular— we first performed whole exome sequencing (WES) of 114 metastatic tumor-normal pairs. The mutational landscape of CRPC-NE was similar to CRPC-Adeno, but also consistent with published studies of CRPC-NE **(Fig. 1c)** including enrichment of *RB1* loss *(*deleted in 70% of CRPC-NE and 32% of CRPC-Adeno, $P = 0.003$, proportion test) and mutation or deletion of *TP53* (66.7% CRPC-NE versus 31.4% CRPC-Adeno, $P = 0.0043$, proportion test). Loss of *RB1* is common in primary small cell prostate and lung carcinomas, and promotes small cell carcinoma pathogenesis when concurrent with *TP53* mutation[19,20]; in our series, concurrent *RB1* and *TP53* loss was present in 53.3% of CRPC-NE vs.13.7% of CRPC-Adeno ($P < 0.0004$, proportion test).

Another feature distinguishing CRPC-NE from CRPC-Adeno was a paucity of somatic alterations involving the AR gene in the former ($P < 0.0001$, Wilcoxon test, **Fig. 1d-e**). Genomic amplification, activating point mutations, and splice variants involving the AR are commonly observed in CRPC-Adeno and associated with treatment resistance to AR-directed therapies[21-23]. This observation was confirmed in our cohort; 29 cases showed *AR* focal amplification or point mutation and 21 cases had alterations in known AR co-activators (*FOXA1, NCOR1/2, ZBTB16*). In contrast, AR point mutations were notably absent in CRPC-NE and gains when present were of low level and explained by tumor polyploidy. Although potentially affected by differences in prior therapies, we speculate that the absence of AR genomic alterations in CRPC-NE may be due to clonal selection of non-amplified CRPC-Adeno tumor subpopulations through selective pressure (in the context of AR-directed therapies). The ARv7 splice variant was observed in both CRPC-Adeno and CRPC-NE tumors, and although overall AR expression was lower in CRPC-NE, the ARv7 to wild type AR ratio was significantly decreased in CRPC-NE compared to CRPC-Adeno ($P = 0.0025$, Wilcoxon test) **(Supplementary Fig. 2)**. These data suggested that AR signaling is attenuated in CRPC-NE. Together with the frequent loss of *RB1* and *TP53* in this subtype, these genetic findings imply distinct biological properties associated with CRPC-NE. The overall spectrum of genomic alterations, however, was similar between the two castration resistant subtypes **(Supplementary Fig. 2)** and maintained after controlling for site of metastasis and prior chemotherapy **(Supplementary Fig. 3)**. There were no significant differences in rate of non-silent point mutations, polyploidy or copy number genomic burden between CRPC-Adeno and CRPC-NE (on average more than 30% of the genome was aberrant), and both subtypes showed a significantly higher number of clonal point mutations compared with clinically localized prostate cancer[24,25]. The median number of non-silent SNVs in metastatic samples was 41 (range: 2-729) **(Supplementary Tables 3-4)**. Five of the

six samples with the highest number of SNVs (115-663) showed genetic and/or protein expression alterations involving DNA mismatch repair genes **(Supplementary Fig. 4)** consistent with prior studies[26]. Recurrent alterations are shown in **Fig. 1c** and putative gene fusions (determined by RNASeq) in **Supplementary Table 5**.

The significant overlap between CRPC-Adeno and CRPC-NE in terms of the overall somatic copy number landscape was noteworthy in light of the marked genomic differences between adenocarcinomas and small cell carcinomas observed in other tissue subtypes (e.g., lung, gastrointestinal tract)[27-29] **(Fig. 1f)**. After correcting for admixture of non-tumor cells and for ploidy[30], we sought to identify regions of the genome differentially altered and noted deletions that are enriched in CRPC-NE and conversely preferential regions of gain in CRPC-Adeno **(Supplementary Tables 6-7)**. Putative cancer genes[31] within differentially deleted regions and concordant down regulation at the mRNA level (FDR<10% both for DNA and mRNA) are highlighted in **Fig. 1f** with the most significant gene being *CYLD* (verified by fluorescence in situ hybridization, **Supplementary Fig. 5**). The *CYLD* gene, deleted in 51% of CRPC-NE samples, encodes cylindromatosis, a deubiquinating enzyme reported as a tumor suppressor involved in negative regulation of multiple signaling pathways including nuclear factor kB (NF-kB)[32], TGF-β[33] and Notch[32]. We found that genomic loss of *CYLD* is associated with decreased mRNA expression and a modest decrease in expression of AR signaling genes in this study as well as the CRPC SU2C/PCF cohort and in cell lines **(Supplementary Fig. 5)**, suggesting that *CYLD* loss alone may be insufficient but might cooperate with other alterations to promote AR- indifference. Extending the computational framework of CLONET[24,30] to assess allele specific copy number clonality (**Supplementary Notes**), we found both focal and broad copy-neutral or copy-aberrant loss of heterozygosity across our cohort **(Supplementary Figure 5)** including focal allelic imbalance of the *DEK* gene in CRPC-NE compared to CRPC-Adeno ($P = 0.04$, binomial test). *DEK* belongs to a class of DNA topology modulators; we and others recently described an oncogenic role in prostate cancer, including in CRPC-NE[34,35].

In principle, several possible models could explain tumor evolution from a prostate adenocarcinoma to CRPC-NE **(Fig. 2a)**. This process may rely on linear expansion of subclonal tumor cell populations originating from the primary tumor with sequential acquisition of genomic alterations (linear)[36]. Alternatively, independent subclones within the primary or metastasis could give rise to parallel and distinct resistant tumor populations (independent)[37]. Finally, there may be genetic diversification in the primary or in the metastatic lesion as a mechanism of adaptation leading to selective pressure and divergent clonal evolution (divergent). While prior studies had demonstrated concordance of TMPRSS2-ERG fusion[9] and other single gene alterations[17,38] between adenocarcinoma and neuroendocrine foci in mixed tumors suggesting a common cell of origin, these complex patterns of genetic evolution have not been more rigorously evaluated. To address this and infer clonal expansion dynamics, we studied serial tumor samples from individual subjects during the course of their disease. Patient WCMC7520 underwent prostatectomy for clinically localized Gleason 9 prostate adenocarcinoma with local lymph node involvement treated initially with adjuvant androgen deprivation therapy (ADT) followed by chemotherapy at the time of metastatic disease and castration resistance **(Fig. 2b)**. 29

months after starting ADT, he developed CRPC-NE diagnosed by pelvic soft tissue biopsy. Primary, lymph node, and CRPC-NE metastases from two time-points were evaluated. Homozygous deletion of *BRCA2* and mutation of *TP53* were present in all sites suggesting a common ancestor, while DNA allele specific analysis highlighted diverse genomic states of other key genes such as *MYCN* **(Fig. 2b)**. *MYCN*, which encodes the N-myc oncogene, has been previously described as oncogenic in CRPC-NE[11]. When sites were compared, the patient's primary prostate harbored lesions that suggested divergent but clonal paths towards lymph node and both distant metastases. However, from these data we cannot definitively rule out metastasis–to–metastasis seeding as recently proposed as a mechanism of tumor progression[39]. Patient WCMC161 progressed after multiple lines of therapy for CRPC-Adeno including the development of new visceral metastases while on abiraterone and liver biopsy showing small cell carcinoma **(Fig. 2c, Supplementary Fig. 6)**. Comparison of metastases from three different time points: CRPC-Adeno (adenocarcinoma, lymph node metastasis), CRPC-Adeno (adenocarcinoma, bone metastasis), and CRPC-NE (small cell carcinoma, liver metastasis at progression on abiraterone therapy) suggested divergent clonal evolution as exemplified by the allelic state of the aberrant gene *SBDS (***Fig. 2c**) supporting that the bone and liver metastases did not arise from the earlier lymph node. Phylogenetic trees from additional multi-tumor cases outlining disease evolution from adenocarcinoma to a neuroendocrine phenotype in individual patients are shown in **Supplementary Fig. 6**, which collectively rule against a parallel evolutionary model. Based on these temporal case series, the high degree of clonality in CRPC, and the overall similarity between CRPC-Adeno and CRPC-NE genomic profiles**,** the most parsimonious model that explains the data is divergent clonal evolution of metastatic CRPC towards either an AR-driven or AR-indifferent state **(Figure 2a**, model V**)**. In other words, CRPC-NE tumors appear clonal in origin with a clonal ancestry traceable back to a CRPC-Adeno precursor.

While informative, the observed DNA changes did not appear to fully explain the clinical aggressiveness of CRPC-NE. We therefore posited that this phenotype may also be mediated by epigenetic changes. Towards this end, we generated data to evaluate CpG-rich methylation genome wide by single cytosine resolution DNA methylation (eRRBS). In contrast to the largely similar genomic data, the CRPC-NE and CRPC-Adeno subtypes showed strong epigenetic segregation by unsupervised analysis using unselected methylation sites (Fig. 3a, Supplementary Table 8, Supplementary Fig. 7). In addition, methylation of first exons and gene promoters was associated with significant changes in gene expression (**Supplementary Fig. 8**). Overall, a greater fraction of concordant epigenetic and transcriptomic events were observed with increasing significance by differential expression analysis (**Supplementary Fig. 8**) and 22% of the top dysregulated transcripts in CRPC-NE demonstrated concordant changes in DNA methylation ($P < 0.0002$). This raised the possibility that the transition to, or advent of, the CRPC-NE subtype is associated with epigenetic dysregulation. In fact, the epigenetic signal identified three cases with clinical features of AR-independence that were binned as adenocarcinoma based on standard pathology but segregated with CRPC-NE on unsupervised analysis **(Fig. 3a, Supplementary Fig. 8)**. All three of these patients demonstrated radiographic progression in the setting of a stable or low serum level of the androgen-regulated protein prostate specific antigen (PSA). These data suggest that clustering prediction based on DNA methylation may

provide additional information associated with AR independence and CRPC-NE that potentially improves upon tumor morphology.

By functional enrichment analysis of differentially methylated genes, we identified epigenetically dysregulated pathways including neuronal, cell-cell adhesion, developmental, epithelial-mesenchymal transition (EMT), and stem cell programs **(Fig. 3b)**, pathways thought to be relevant for CRPC-NE progression[40]. Amongst significant findings (**Supplementary Fig. 7)**, we observed hypermethylation and under-expression of the tumor suppressor gene *SPDEF* in CRPC-NE ($P < 10^{-9}$, Wilcoxon test) **(Fig. 3c)**. *SPDEF* (prostate-derived Ets factor) is a transcriptional activator and regulator of cellular differentiation involved in suppression of tumor metastasis through inhibition of EMT in prostate cancer. We confirmed promoter methylation of the *SPDEF* gene and downregulation of mRNA expression in the neuroendocrine prostate cancer cell line NCI-H660 compared to prostate adenocarcinoma (LNCaP) (**Supplementary Fig. 9)**.

Expression of mRNA encoding the histone methyltransferase EZH2, previously associated with aggressive disease in prostate cancer[41–44], was 2 fold higher in CRPC-NE compared to CRPC-Adeno ($P < 10^{-6}$, Wilcoxon test); EZH2 protein was also more abundant in CRPC-NE (**Supplementary Fig. 9)**. These findings add to its differential status previously reported comparing CRPC-NE to localized prostate cancer[9,45]. Furthermore, EZH2-repressed target genes[46,47] were significantly down-regulated in CRPC-NE ($P < 10^{-7}$, Wilcoxon test) **(Fig. 3d)** including WNT signaling genes such as *DKK1* ($P = 0.0002$, Wilcoxon test), *NKD1* ($P = 0.0046$, Wilcoxon test), and HOX genes ($P = 0.001$, Wilcoxon test). Treatment with the EZH2 inhibitor GSK343 resulted in a preferential decrease in cellular viability of NCI-H660 compared to other non-NE prostate cancer cell lines **(Fig. 3e)** with significant down-regulation of several CRPC-NE associated genes after treatment including *NCAM* (CD56), *MYCN*[9], and *PEG10*[48] **(Supplementary Fig. 9)**. Overall these data support a key role of the epigenome in the emergence and/or maintenance of CRPC-NE.

Based on the current gaps in the clinical and molecular assessment of CRPC-NE, we used these data to develop a molecular classifier to potentially improve upon the often challenging diagnosis of CRPC-NE[8,17]. This integrated 70 gene neuroendocrine prostate cancer (NEPC) classifier was developed by exploiting expression data of genes prioritized by genomic, transcriptomic or epigenomic status **(Fig. 4a,b, Supplementary Fig. 10, Online Methods)** and demonstrated both a precision and recall of >0.99 in identifying CRPC-NE in our discovery cohort **(Fig. 4c,d, Supplementary Table 9)**. Included within the classifier were also genes we had previously described as overexpressed in CRPC-NE[9], *AURKA* ($P < 10^{-5}$, Wilcoxon test) and *MYCN* ($P < 10^{-4}$, Wilcoxon test) **(Supplementary Fig. 10)**. Interrogation of transcriptome data of 683 prostate samples **(Supplementary Table 10)** using datasets from The Cancer Genome Atlas (TCGA)[49], Grasso et al[21] (Michigan 2012), Robinson et al[23] (SU2C/PCF 2015), and internal published data[9,50] (WCMC 2011/14) revealed an elevated NEPC classifier score in up to 8% of metastatic tumors (n=191) and none of treatment naïve prostate adenocarcinoma (n=460) or benign prostate (n=32) **(Fig. 4d and Supplementary Fig. 10 and 11)**. Of those with an elevated classifier score, we reviewed the pathology and found over 80% had pathologic features of CRPC-NE (the remaining 20% were adenocarcinoma) **(Fig. 4d, Supplementary Fig. 12)**. Although

there were relatively small numbers of CRPC-NE cases in the validation cohorts, the integrated NEPC classifier demonstrated superior precision and/or recall across datasets compared to other variables such as conventional neuroendocrine markers (ie., *CGHA, SYP, NSE, CD56 transcript)* plus PSA **(Supplementary Fig. 12)**, AR mRNA expression, AR signaling status, and highly ranked differentially expressed genes (such as SPDEF) **(Fig. 4c)**. Though recognizing the influence of other factors, including differences in prior therapies, on gene expression changes we posit that castration resistant tumors with moderate or rising NEPC score may represent tumors with AR-independent features either in transition or at high risk for CRPC-NE progression under treatment pressure with AR therapies. In fact, a subpopulation of prostate adenocarcinoma cells (LNCaP) treated long term with enzalutamide acquired molecular features of CRPC-NE (ie., methylation of *SPDEF*) **(Supplementary Fig. 8 and 13)**.

Discussion

Though there were also cases identified with low AR-signaling and low NEPC classifier score and we cannot exclude a less common alternatively distinct subset, our data favors a continuum of progression from an AR-driven state towards AR-independence associated with neuroendocrine molecular features. These findings warrant larger clinical evaluation to further investigate the potential superiority and applicability of the classifier over conventional diagnostic criteria and verify whether this could be useful as a prognostic or predictive biomarker (associated with lack of response to AR therapies). Notably, this approach is amenable to metastatic prostate cancer biopsies in which tissue availability is limited and multiple immunohistochemical assays for current diagnostic methods often impractical. Incorporation of different layers helps apply the classifier to different datasets when only parts are available (DNA, RNA, or methylation) (**Supplementary Table 9**) and paves the way for future studies that might apply the classifier to types of samples (such as circulating tumor DNA[51]). If CRPC-NE alterations could be detected earlier during CRPC-Adeno disease progression, for instance, such individuals could potentially be selected for CRPC-NE-directed (such as platinum chemotherapy) rather than AR-targeted systemic therapies, or potentially co-targeting therapeutic approaches. Further, these data set the stage for dynamic testing of the reversibility of the CRPC-NE state with early intervention or epigenetic modifiers possibly including EZH2 inhibitors.

In summary, our data supports divergent evolution of CRPC-NE from one or more CRPC-Adeno cells (adaptation) rather than linear or independent clonal evolution, with selective pressure of AR-wild-type subclonal populations and acquisition of new genomic and epigenomic drivers associated with decreased AR signaling and epithelial plasticity. However, there are also other possibilities that cannot be fully excluded, such as de-differentiation of adenocarcinoma to a more progenitor-like cell state (with some cells subsequently adopting neuroendocrine features due to local effects).

ONLINE METHODS

Code availability

The source code for the study clonality analysis is accessible from the Bitbucket version controlled repository (https://bitbucket.org/; CLONET).

Cohort description and pathology classification

Tumor specimens were obtained prospectively through clinical protocols approved by the Weill Cornell Medicine (WCM) Institutional Review Board (IRB) with informed consent (IRB #1305013903, #1210013164) or retrospectively (IRB #0905010441) and germline DNA obtained from either peripheral blood mononuclear cells (PBMCs) or benign tissue. The total number of subjects enrolled in this study was 81, all of male gender; no blinding, randomization, or exclusion criteria were used. All fresh/frozen tissues were processed as previously described[26,52]. All hematoxylin and eosin stained slides were reviewed by board-certified pathologists (J.M.M., M.A.R.). Tumors were classified based on histomorphology as adenocarcinoma (A) or CRPC-NE (B-E) based on a published pathologic classification system[8] (**Supplementary Fig. 1**). Category A represents usual prostate adenocarcinoma *without* neuroendocrine differentiation, Category B represents usual prostate adenocarcinoma *with* neuroendocrine differentiation > 20%, Category C represents small cell carcinoma, Category D represents large cell neuroendocrine carcinoma, and Category E represents mixed small cell carcinoma – adenocarcinoma. Clinical and pathologic features of the cohort are summarized in **Supplementary Table 1 and Supplementary Fig. 1**.

Immunohistochemistry

Immunohistochemistry (IHC) was performed on a Bond III automated immunostainer (Leica Microsystems, IL, USA) with the following antibodies and dilutions: anti-AR (clone MU256-UC, BioGenex, CA, USA; dilution 1:800 with casein); anti-synaptophysin (clone RM-9111-S, Thermo Scientific; dilution 1:100), anti-chromogranin A (clone MU126-UC, BioGenex, CA, USA; dilution 1:400); anti-CD56 (clone NCL-SD56-504, Leica Biosystems, IL, USA); anti-PSA (clone MU014-UC, BioGenex, CA, USA; dilution 1:400); anti-PSAP (clone MU013-UC, BioGenex, CA, USA; dilution 1:250), anti-PSMA (clone M3620, Dako, CA, USA; dilution 1:100); anti-Ki67 (clone M7240, Dako, CA, USA; dilution 1:50); anti-ERG (clone ab92513, Abcam, MA, USA; dilution 1:100); anti-EZH2 (clone 612667, BD Biosciences, CA, USA; dilution 1:20); anti-MLH1 (clone 554073, BD Biosciences, CA, USA; dilution 1:400), anti-PMS2 (clone 556415; BD Biosciences; dilution 1:100); anti-MSH2 (clone NA27, Calbiochem, CA, USA; dilution 1:100) and anti-MSH6 (clone 610919, BD Biosciences, CA, USA; dilution 1:800), using the Bond Polymer Refine detection kit (Leica Microsystems, IL, USA). Antigen retrieval was performed using heat-mediated pH6 retrieval for anti-ERG, anti-PSA and anti-PSMA; pressure-cooker pH6 retrieval for anti-EZH2, no retrieval for anti-chromogranin A and anti-PSAP, and heat-mediated pH9 retrieval for all the other antibodies. Study pathologists performed semi-quantitative evaluation for protein expression in nuclear (AR, ERG, MLH1, MSH2, MSH6, PMS2), cytoplasmic (synaptophysin, chromogranin A), or both (EZH2) compartments, using a four-tier grading system: negative (0), weakly positive (1+), moderately positive (2+), and strongly positive (3+). The extent of positivity (percentage) was recorded. For evaluation of ERG and MMR

protein expression, IHC was defined as either positive or negative. For the other antibodies, the following cut-offs were considered to determine a positive expression: >20% of cells for synaptophysin and chromogranin A; >10% of nuclei for AR; and >10% of cells for EZH2.

DNA extraction, tumor purity, and exome sequencing

Slides were cut from frozen or FFPE tissue blocks and examined by the study pathologists to select high-density cancer foci and ensure high purity of cancer DNA. We employed previously developed protocols[53] successfully used in our earlier genomic studies[23,26,52]. Following this protocol, each case was quantified for tumor purity, and annotated for discrete areas of macrodissection avoiding regions of necrosis or high stromal content. All cases were also quantified for tumor purity using a recently developed algorithm CLONET[24,30]. CLONET was developed and validated to specifically deal with heterogeneous tumor samples, enabling optimal objective tumor purity and ploidy estimates by taking the germline heterozygous SNP genotype data from whole exome sequence coverage to quantify the percentage of reads supporting the considered aberration[30]. The resultant tumor purity values are used to adjust the genomic data for downstream processing and analysis. Extraction and sequencing were performed as previously described[54]. Briefly, DNA was extracted using Promega Maxwell 16 MDx. DNA was stored at −20 degrees Celsius. Whole exome capture libraries were constructed from tumor and normal tissue after sample shearing, end repair, and phosphorylation and ligation to barcoded sequencing adaptors. Ligated DNA was size selected for lengths between 200-350 bp and subjected to either exonic hybrid capture using SureSelect v2/v4 Exome bait (Agilent) or HaloPlex Exome (Agilent) **(Supplementary Table 11-12)**. The samples (70 CRPC-Adeno samples from 51 individuals and 44 CRPC-NE from 30 individuals) were multiplexed and sequenced using Illumina HiSeq for an intended mean target exome coverage of 100X for the tumor and germline samples**.** All BAM files generated for this study are available at dbGap phs000909.v.p1.

Sequence data processing and quality control

A fraction of study samples (n = 78) were pre-processed with the analytical pipeline of the Englander Institute for Precision Medicine at Weill Cornell/New York Presbyterian Hospital (IPM-Exome-pipeline v0.9)[52]. FastQC was run on the raw reads to assess the quality of the raw reads as previously described[52]. The remaining 36 samples were processed using the Broad Institute Firehose infrastructure (http://www.broadinstitute.org/cancer/cga/Firehose)[23,26]. Cross-contamination between samples from other individual subjects sequenced in the same flow cell was monitored with the ContEst algorithm[55]. Normal/tumor pairs were checked for consistency using SPIA[56]. Processing pipelines returned segmented files for somatic copy-number aberrations (SCNAs) (**Supplementary Table 6)**. No differences were observed in the results of the two pipelines ($P = 0.75$, Kolmogorov-Smirnov test).

Mutation calling and identification of significantly mutated genes

To identify and characterize somatic single-nucleotide variants (SNVs) in targeted exons, we first applied MuTect from the Broad Institute Genome Analysis Toolkit[57], based on

Bayesian statistical analysis to nominate putative SNVs upon coverage, allelic fraction, and base qualities extraction. Next, we utilized Oncotator[58] to annotate point mutations with variant- and gene-centric information relevant to cancer and last MutSig[59] to identify genes harboring a greater number of mutations than expected by chance, given the background mutation rate, the sequence context and the genomic territory. For each gene identified (**Supplementary Table 3**), it returns the p-value adjusted for multiple hypotheses testing with Benjamini-Hochberg procedure (q-value). Genes with q-value below 0.1 are considered significant. Finally, we searched for mutations enriched in one pathology class by means of binomial test adjusted with Benjamini-Hochberg procedure (q-value) (**Supplementary Table 4)**.

Tumor ploidy, purity and copy number estimates

Segmented data was used by CLONET to estimate ploidy and purity for each tumor sample as previously described[30]. Each segment is represented by the log2 of the ratio between values proportional to the tumor and normal local coverage within the genomic segment. Briefly, the ploidy of an individual tumor, defined as the mean number of sets of chromosomes of a cell, is assessed using the mean coverage observed in an individual as a proxy of the number of observed alleles; CLONET recognizes shifts in the SCNAs log2 ratio distribution reflective of an aneuploidy genome and used to then estimate tumor ploidy. Next, a local optimization approach based on putative clonal mono-allelic deletions and germline heterozygous SNP loci (called *informative SNPs* and identified by means of ASEQ[60]) is applied to assess the purity (1-admixture) of each sample; the difference between observed and expected allelic fraction (AF) at *informative SNPs* (the latter being either 0 or 1) is proportional to tumor purity. Finally, CLONET computes a purified copy number profile adjusting each segment to account for both aneuploidy and tumor purity. Purified segments with mean log2 ratio below −0.4 or greater than 0.4 were categorized as copy number loss or gain, respectively.

Allele-specific copy number analysis by CLONET

To determine the copy number landscape of a tumor sample, allele-specific copy number analysis was applied within the CLONET framework. In a 100% pure tumor sample, the empirical distribution of the AFs within a genomic segment reflects the aggregated signal from multiple cell populations. If a segment *S* is copy number neutral and both alleles are represented, sequenced DNA fragments equally sample the reference and the alternative bases, i.e., the AF around 0.5. If a segment *S* represents a 100% clonal mono-allelic deletion, the AF is either 0 or 1 depending on which allele is lost. Combinations of diverse representations of the two alleles of a diploid genome and subclonality lead to AFs in between. Low DNA purity further dilutes the signal from the expected values. CLONET uses a variable *beta* that represents the disproportion in the AF of informative SNPs within a segment *S; beta* values can be computed from the empirical distribution of the AFs. Importantly, the beta value of a genomic segment is independent from its log2 ratio and the two measures can be used to infer the allele specific copy number and the clonality state of each segment of a tumor genome. CLONET alos provides a space transformation from the beta vs log2 ratio to the cnB vs cnA (**Supplementary Notes, Supplementary Fig. 14**), where each segment is visualized at coordinates representing the number of copies of allele

A and allele B. In the absence of parental allelic information, we consider cnA ≥ cnB. Non integer values of cnA and/or cnB indicate that the copy number signal results from the convolution of the copy number states of multiple subclonal tumor cell populations. Finally, the cnB vs cnA space allows for direct comparison of allele-specific copy number profiles of different tumor samples by mean of the Euclidean distance, irrespective of tumor ploidy and tumor purity.

RNA extraction, sequencing and processing

RNA was extracted from frozen material for RNA-sequencing (RNA-seq) using Promega Maxwell 16 MDx instrument, (Maxwell 16 LEV simplyRNA Tissue Kit (Cat. # AS1280)). Specimens (34 CRPC-Adeno samples from 33 individuals and 15 CRPC-NE samples from 10 individuals) were prepared for RNA sequencing using TruSeq RNA Library Preparation Kit v2 as previously described[9,50]. RNA integrity was verified using the Agilent Bioanalyzer 2100 (Agilent Technologies). cDNA was synthesized from total RNA using Superscript III (Invitrogen). Each sample was then sequenced with the HiSeq 2500 to generate 2x75bp paired-end reads. Details of the sequencing results are reported in **Supplementary Table 13**. Reads (FASTQ files) were mapped to the human genome reference sequence (hg19/GRC37) using STAR v2.3.0e[61], and the resulting BAM files were subsequently converted into Mapped Read Format (MRF) using RSEQtools, a suite of tools for RNA-seq data processing and analysis[62]. MRF files include only the primary alignments as determined by STAR and do not include reads mapped to the mitochondrial chromosome. Quantification of gene expression was performed via RSEQtools using GENCODE v19 (http://www.gencodegenes.org/releases/19.html) as reference gene annotation set. A composite model of genes based on the union of all exonic regions from all transcripts of a gene was used resulting in a set of 20,345 protein-coding genes. Expression levels (FPKM) are estimated by counting all nucleotides mapped to each gene and normalized by the total number of mapped nucleotides (per million) and the gene length (per kilobase). Differential expression analysis was performed using the Mann-Whitney Wilcoxon test after transforming the FPKMs via $\log_2$ (FPKM+1). Multiple hypothesis testing was considered by using Benjamini-Hochberg (BH, FDR) correction. For downstream analyses (differential copy number and Integrated NEPC Score), genes with low expression in both CRPC-Adeno and CRPC-NE samples (mean + standard deviation less than 1 FPKM) were discarded.

NanoString assay—We employed a custom NanoString assay for cases without sufficient material for RNA-seq to evaluate for expression of AR signaling genes, TMPRSS2-ERG fusion transcript, and neuroendocrine associated genes. FFPE samples were cut onto 10μM thick slides, annotated by the study pathologist, and RNA was extracted using the Ambion RecoverAll™ Total Nucleic Acid Isolation Kit. RNA quality control was performed on the Agilent 2100 Bioanalyzer system by annotating total RNA concentration and percentage of RNA greater than 300 nucleotides (nt) in length. For samples with more than 50% of total RNA greater than 300 nt, 100ng input RNA was used; for samples with less than 50% of total RNA greater than 300 nt, the input RNA was proportionally increased according to the level of degradation. Samples were run on the NanoString nCounter® Analysis System according to the manufacturers directions. Briefly, total RNA was hybridized overnight at 65°, then run on the Prep Station at max sensitivity. Cartridges were then scanned on the

Digital Analyzer at 555 fields of view. Raw count data was normalized using the nSolver™ analysis software version 2.0, which normalizes samples according to positive and negative control probes and the geometric mean of the 6 housekeeping primers.

Differential copy number analysis

For each gene in a panel of more than 19,000 RefSeq genes, we computed the log2 ratio adjusted by ploidy and tumor purity and then we aggregated the mutation frequencies by specimen (**Supplementary Table 7**). Deletions and amplifications are defined using the thresholds on log2 ratio as described above. The relative frequencies of in CRPC-NE and CRPC-Adeno specimens are compared by means of binomial test adjusted for multiple hypotheses testing with Benjamini-Hochberg procedure (FDR). For each gene, **Supplementary Table 7** also reports the mean expression level in CRPC-NE and CRPC-Adeno samples together with the Mann-Whitney Wilcoxon p-value and FDR correction of the likelihood that the expression in the classes is different.

Androgen Receptor signaling

We assessed AR signaling using the expression levels of 30 genes (**Supplementary Table 2**) previously reported as defining the pathway[18]. For each specimen with expression levels available, either from RNA-seq or Nanostring, we considered its correlation to a reference sample known to have active AR signaling. This reference sample was based on LNCaP cells and was generated by taking the average values of the 30 AR-regulated genes across three replicates on the Nanostring assay. We then computed the Pearson's correlation coefficient for each specimen to this reference sample and we considered this as the “AR signaling”. To validate the approach, we tested prostate cancer cell lines with known AR activity (**Supplementary Fig. 15**).

Fusion detection and ERG rearrangement status

In order to detect ERG rearrangement status we employed several assays, including fluorescence *in situ* hybridization (FISH) break-apart assay and/or immunohistochemistry (IHC) as previously described[63]. Whole exome sequencing copy number analysis (WES) and Nanostring/RNA-sequencing (RNA-seq) were also used to assess ERG deletion and fusion transcript levels, respectively. If ERG fusion was detected at either DNA, mRNA, or protein level, the sample was considered positive. For other non-canonical and canonical fusion transcript detection, FusionSeq was employed as previously described[64].

CYLD fluorescence *in situ* hybridization (FISH)

To assess *CYLD* deletion in tissues, we developed a dual-color FISH assay consisting of a locus specific probe (RP11-327F22) plus reference probe spanning a stable region of the chromosome (RP11-488I20). All clones were tested on metaphase spreads[9,65]. *CYLD* deletion was defined by the absence at least 1 copy on average per nucleus compared to two reference signals. At least 100 nuclei were evaluated per tissue section using a fluorescence microscope (Olympus BX51; Olympus Optical, Tokyo, Japan).

Methylation profiling and data processing

Sample preparation for enhanced reduced representation bisulfite sequencing (eRRBS) was performed at the Weill Cornell Medical College Epigenomics Core Facility as previously described[66] and included 18 CRPC-Adeno samples from 10 individuals and 10 CRPC-NE from 8 individuals. In brief, the preparation steps included: 1) MspI enzyme digestion; 2) end repair of digested DNA; 3) adenylation; 4) adapter ligation, with pre-annealed 5-methylcytosine-containing Illumina adapters; 5) isolation of library fragments of 150 to 400 bp from a 1.5% agarose gel; 6) bisulfite conversion using the EZ DNA Methylation Kit (Zymo Research, Irvine, CA); 7) polymerase chain reaction (PCR) amplification; each library was prepared with FastStart High Fidelity DNA Polymerase (Roche, Indianapolis, IN) and Illumina PCR primers PE1.0 and 2.0. PCR products were isolated using Agencourt AMPure XP (Beckman Coulter, Brea, CA) beads per manufacturer's protocol (Agencourt) Amplified libraries were evaluated using a Qubit 1.0 fluorometer and Quant-iT dsDNA HS Assay Kit (Invitrogen, Grand Island, NY) for quantitation and bioanalyzer visualization (Agilent 2100 Bioanalyzer; Agilent, Santa Clara, CA). After filtering out genomic regions of somatic homozygous deletions ($\log_2$ ratio < −1.1) as inferred from the corresponding WES data (**Supplementary Table 6**), differentially methylated sites (CRPC-NE versus CRPC-Adeno) were identified by methylKit[67]. Differentially methylated genes were identified by annotating with BedTools[68] differentially methylated regions on gene promoters (defined as 2 Kbp genomic regions upstream the set of 5’ gene coordinates), first exons, gene bodies, CpG islands (UCSC table browser) and CpG shores (2 Kbp genomic regions upstream and downstream of CpG islands). GENCODE v19 was used as gene set. ToppFun[69] was used to perform functional enrichment analysis. **Supplementary Table 8** lists differentially methylated genes and includes genomic feature annotations.

Site-directed CpG methylation

Targeted CpG methylation analysis for SPDEF was performed using OneStep qMethyl PCR Kit per manufacturer's protocol (Zymo Research, Irvine, CA). 20ng of genomic DNA from cell lines was used for qMethyl PCR. Primers used for qMethyl PCR were:

Primer1 forward:CCGGTGACATCCGTGTGTTC,

Primer1 reverse:AATCGCCGGTACACTCCTTG,

Primer2 forward:GATTCTGCTCTCCCACCTCTC,

Primer2 reverse:CCAGCAGCCCTCAAAGCAAC.

Amplification parameters were: 45 cycles (Denaturation: 95oC, 30 sec, Annealing: 64oC, 30 sec, Extension: 72oC, 30 sec).

Integrated NEPC score

The Integrated Neuroendocrine Prostate Cancer (NEPC) score estimates the likelihood of a test sample to be CRPC-NE. It is calculated as the Pearson's correlation coefficient between the test vector and a reference CRPC-NE vector based on a set of 70 genes (Supplementary Table 9, Supplementary Fig. 10 and 15) using normalized FPKM values of the test sample.

The gene set stems from the integration of differentially deleted/amplified and/or expressed and/or methylated genes in CRPC-NE and CRPC-Adeno. Specifically, 16 differentially deleted genes were selected among putative cancer genes[31] (see **Differential copy number analysis**). The following strategy was used to identify both differentially expressed genes that better distinguish CRPC-NE and CRPC-Adeno samples. We selected differentially expressed protein coding genes with FDR ≤ 1e-2, resulting in a total of 2425 genes, corresponding to 1301 over- and 1124 under-expressed. For each gene, we performed a Receiver Operator Curve (ROC) analysis using the normalized FPKMs as threshold parameter and calculated the Area Under the Curve (AUC). ROCs were built by considering only samples sequenced excluding two samples (7520 and 4240) that were previously published[9].leaving 34 CRPC-Adeno and 13 CRPC-NE. Only those differentially expressed genes with AUC ≥ 0.95 and with a fold-change greater than 2 or lower than 0.5 were included in the classifier, resulting in a list of 49 genes (25 over- and 24 under- expressed in CRPC-NE vs. CRPC-Adeno), 21 of which found as differentially methylated between CRPC-NE and CRPC-Adeno. Concordant information between RNA and Methylation was found for 11 genes (see **Supplementary Table 9**). In addition, we considered 2 genes (*MYCN* and *AURKA*) that we previously described as associated with CRPC-NE phenotype[9], *EZH2* (FDR = $7.9*10^{-4}$) and *DNMT1* (FDR = $6.9*10^{-5}$) for their role in controlling DNA methylation[70] and *RB1* (FDR = 0.056), reported as a key driver in the pathogenesis of CRPC-NE[9,45]. For each of the resulting 70 genes, we calculated the mean of the normalized FPKM across the 13 CRPC-NE samples with RNA-seq data and defined the resulting set of averages as reference CRPC-NE vector. The Integrated NEPC score was tested across 719 prostate samples with available transcriptome data from multiple datasets (**Supplementary Table 10)**. RNA-seq data were processed as described above. Processed SU2C-PCF[26] and Grasso et al[21] (Michigan 2012) data were downloaded from cBioPortal[71]. Since data for 4 genes (*ARHGAP8*, *BRINP1*, *C7Orf76* and *MAP10*) were not available from cBioPortal, for Michigan 2012 we used a reduced version of Integrated NEPC Score (indicated as *Integrated NEPC Score**). Samples with Integrated NEPC Score greater than or equal to 0.40 (elevated Integrated NEPC score in main text) were nominated as putative CRPC-NE (**Figure 4c, Supplementary Table 14**). In order to take into account the lower signal-to-noise ratio and the reduced version of Integrated NEPC Score in Michigan 2012 microarray data, in **Figure 4d** we consider as CRPC-NE – like those samples with Integrated NEPC Score ≥ 0.25 (significant Integrated NEPC score in **Figure 4** legend). AR signaling and Integrated NEPC Score values per sample are reported in **Supplementary Table 15**.

Cell line studies

LNCaP clone FGC and NCI-H660 cells and media were purchased from ATCC and cultured at 37°C in 5% CO2. Cell line authentication was performed (DDC Medical, Fairfield, OH) and cells were tested for mycoplasma contamination. To create a resistant cell line, the AR-dependent LNCaP line was grown in media containing 1μm enzalutamide for more than 6 months. Dose response curve was performed by plating 2.5×10^3 cells in 10μl Matrigel® (Corning®) onto an Ibidi 96 well μ-plate and covering with 40μl media containing increasing doses of enzalutamide for 10 days (media changed every 4 days). Viability was analyzed using the CellTiter-Blue assay (Promega), Western blot analysis was performed

with 50μg protein stained with 1:1000 anti-NCAM (Abcam, ab137086), 1:500 anti-SPDEF (Biorbyt orb13642), and 1:10,000 anti-GAPDH (Millipore, AB2302). Site-directed CpG methylation of SPDEF was performed as described above. Cells were blocked in 5% BSA in PBS for one hour on ice, then stained with 5μg/ml Brilliant Violet 421 anti-human CD56 (Biolegend) for one hour on ice in dark. Cells were washed twice with 8 ml PBS (8min × 500g) and resuspended in FACS sorting buffer (1x PBS, 1mM EDTA, 25 mM HEPES pH 7.0, 1% FCS (heat inactivated), 1% pluronic). 0.5 μg/ml DAPI was added immediately prior to sorting to stain dead cells. Cells were sorted on a BD FACSCalibur under the guidance of the WCMC Flow Cytometry core and analyzed with FloJo vX.0.7.

CYLD expression was silenced using ON-TARGET plus CYLD siRNA (Dharmacon L-004609-00-0005). ON-TARGET plus Non-targeting Pool (Dharmacon, D-001810-10-05) was used as control. RNAs from LNCaP cells were purified using Maxwell® 16 LEV simplyRNA Cells Kit (Promega) according to the manufacturer's instructions. cDNA was generated using the qScript™ cDNA SuperMix (Quanta BioSciences). SYBR Green-based qRT-PCR experiments were performed on a Roche LightCycler 480 II sequence detection system using Roche SYBR. The following oligonucleotides were used: Human CYLD: 5-tttgcgtgtgttgaaagtacaat -3 (forward) 5- ttcctgcgtcacactctctg-3 (reverse). Human beta-actin: 5-tccctggagaagagctacg-3 (forward) 5-gtagtttcgtggatgccaca-3 (reverse). CYLD relative expression was normalized using beta-actin.

LNCaP, NCI-H660, and DU145 cells and media were purchased from ATCC and used for GSK343 drug treatment studies using escalating doses (5, 7.5, 10uM). Cells were plated in Ibidi 96 well μ-plate embedded in 5μl Matrigel®. Matrigel droplets were then covered with 40μl media. After 48h media was replaced with fresh media containing increasing doses of GSK343 for 14 days (media changed every 4 days) in three replicates. Viability was analyzed using the CellTiter-Blue assay (Promega) according to the manufacturer's protocol. Total RNA was isolated from the cells using the RNeasy Mini Kit (Qiagen, Valencia, CA) and NanoString was performed for quantitative analysis of mRNA expression of the reported genes.

Statistical analysis

For statistical tests, we used two-sided Mann–Whitney–Wilcoxon test (referred also as Wilcoxon test in the main text) to check for significant shift between two distributions. When appropriate, we used Kolmogorov-Smirnov test to check for discrepancies in the compared distributions. The proportion test has been used to determine whether the deviations between the observed and the expected counts are significant. Finally, **Supplementary Fig. 5e** uses a t-test. The statistical test used is indicated in the respective figure legend or in the corresponding main text. All the tests are two sided. When appropriate, p-values were adjusted for multiple hypotheses testing with Benjamin-Hochberg procedure. No statistical methods were used to predetermine sample size.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to thank our patients and their families for participation in this study. We would also like to acknowledge R. Montgomery, A. Armstrong, R. Szmulewitz for contributing samples and S.S. Chae for his technical assistance. H. Beltran is the Damon Runyon-Gordon Family Clinical Investigator supported (in part) by the Damon Runyon Cancer Research Foundation CI-67-13. This work was supported by: Ann and William Bresnan Foundation (H.B., D.M.N.), Department of Defense PC121341 (H.B.), N.I.H. R01 CA116337 (H.B., F.D., M.A.R.), Starr Cancer Consortium (H.B., M.A.R.), American-Italian Cancer Foundation (L.P.), Nuovo Soldati Foundation (J. C.), N.I.H. R01CA157845 (S.V.), N.I.H. R01 CA183857 (S.A.T.), A. Alfred Taubman Medical Institute (S.A.T.), Prostate Cancer Foundation Young Investigator Award (E.M.V.), N.I.H. 1K08CA188615 (E.M.V.), #U54 HG003067 (L.A.G.), 5U01 CA111275-09 (J.M.M., M.A.R., F.D.), Associazione Italiana per la Ricerca sul Cancro AIRC, IG 13562 (F.D.), Prostate Cancer Foundation (H.B., S.A.T., M.A.R., F.D.), European Research Council CoG SPICE (F.D.). H.B., J.M.M, S.A.T., D.M.N., S.T.T., E.M.V., O.E., A.S., L.A.G., M.A.R., F.D., are supported by a Stand Up To Cancer - Prostate Cancer Foundation Prostate Dream Team Translational Cancer Research Grant. Stand Up To Cancer is a program of the Entertainment Industry Foundation administered by the American Association for Cancer Research (SU2C-AACR-DT0712).

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA: a cancer journal for clinicians. 2015; 65:5–29. doi:10.3322/caac.21254. [PubMed: 25559415]
2. Huggins C, Hodges CV. Studies on prostatic cancer: I. The effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. 1941. The Journal of urology. 2002; 168:9–12. [PubMed: 12050481]
3. Chen CD, et al. Molecular determinants of resistance to antiandrogen therapy. Nature medicine. 2004; 10:33–39. doi:10.1038/nm972.
4. Knudsen KE, Scher HI. Starving the addiction: new opportunities for durable suppression of AR signaling in prostate cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009; 15:4792–4798. doi:10.1158/1078-0432.CCR-08-2660. [PubMed: 19638458]
5. Scher HI, et al. Increased survival with enzalutamide in prostate cancer after chemotherapy. The New England journal of medicine. 2012; 367:1187–1197. doi:10.1056/NEJMoa1207506. [PubMed: 22894553]
6. de Bono JS, et al. Abiraterone and increased survival in metastatic prostate cancer. The New England journal of medicine. 2011; 364:1995–2005. doi:10.1056/NEJMoa1014618. [PubMed: 21612468]
7. Logothetis CJ, et al. Molecular classification of prostate cancer progression: foundation for marker-driven treatment of prostate cancer. Cancer discovery. 2013; 3:849–861. doi: 10.1158/2159-8290.CD-12-0460. [PubMed: 23811619]
8. Epstein JI, et al. Proposed morphologic classification of prostate cancer with neuroendocrine differentiation. The American journal of surgical pathology. 2014; 38:756–767. doi:10.1097/PAS. 0000000000000208. [PubMed: 24705311]
9. Beltran H, et al. Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. Cancer Discov. 2011; 1:487–495. doi:10.1158/2159-8290.CD-11-0130. [PubMed: 22389870]
10. Yao JL, et al. Small cell carcinoma of the prostate: an immunohistochemical study. Am J Surg Pathol. 2006; 30:705–712. doi:00000478-200606000-00005 [pii]. [PubMed: 16723847]
11. Beltran H, et al. Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. Cancer discovery. 2011; 1:487–495. doi:10.1158/2159-8290.CD-11-0130. [PubMed: 22389870]
12. Beltran H, et al. Challenges in recognizing treatment-related neuroendocrine prostate cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2012; 30:e386–389. doi:10.1200/JCO.2011.41.5166. [PubMed: 23169519]
13. Aparicio A, Tzelepi V. Neuroendocrine (Small-Cell) Carcinomas: Why They Teach Us Essential Lessons About Prostate Cancer. Oncology. 2014; 28

14. Lin D, et al. High fidelity patient-derived xenografts for accelerating prostate cancer discovery and drug development. Cancer research. 2014; 74:1272–1283. doi:10.1158/0008-5472.CAN-13-2921-T. [PubMed: 24356420]
15. Wang HT, et al. Neuroendocrine Prostate Cancer (NEPC) progressing from conventional prostatic adenocarcinoma: factors associated with time to development of NEPC and survival from NEPC diagnosis-a systematic review and pooled analysis. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2014; 32:3383–3390. doi:10.1200/JCO.2013.54.3553. [PubMed: 25225419]
16. Tagawa ST. Neuroendocrine prostate cancer after hormonal therapy: knowing is half the battle. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2014; 32:3360–3364. doi:10.1200/JCO.2014.57.5100. [PubMed: 25225431]
17. Tan HL, et al. Rb loss is characteristic of prostatic small cell neuroendocrine carcinoma. Clin Cancer Res. 2014; 20:890–903. doi:1078-0432.CCR-13-1982 [pii] 10.1158/1078-0432.CCR-13-1982. [PubMed: 24323898]
18. Hieronymus H, et al. Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. Cancer cell. 2006; 10:321–330. doi:10.1016/j.ccr.2006.09.005. [PubMed: 17010675]
19. Chen H, et al. Pathogenesis of prostatic small cell carcinoma involves the inactivation of the P53 pathway. Endocrine-related cancer. 2012; 19:321–331. doi:10.1530/ERC-11-0368. [PubMed: 22389383]
20. Zhou Z, et al. Synergy of p53 and Rb deficiency in a conditional mouse model for metastatic prostate cancer. Cancer research. 2006; 66:7889–7898. doi:10.1158/0008-5472.CAN-06-0486. [PubMed: 16912162]
21. Grasso CS, et al. The mutational landscape of lethal castration-resistant prostate cancer. Nature. 2012; 487:239–243. doi:10.1038/nature11125. [PubMed: 22722839]
22. Beltran H, et al. Targeted next-generation sequencing of advanced prostate cancer identifies potential therapeutic targets and disease heterogeneity. European urology. 2013; 63:920–926. doi:10.1016/j.eururo.2012.08.053. [PubMed: 22981675]
23. Robinson D VAE, Wu Y, Schultz N, Lonigro RJ, Mosquera JM, Montgomery R, Taplin ME, Pritchard CC, Attard G, Beltran H, Abida W, Bradley RK, Vinson J, Cao X, Vats P, Kunju L, Hussain M, Feng FY, Tomlins SA, Cooney KA, Smith DC, Brennan C, Siddiqui J, Mehra R, Chen Y, Rathkopf DE, Morris MJ, Solomon SB, Durack JC, Reuter VE, Gopalan A, Gao J, Loda M, Lis RT, Bowden M, Balk SP, Gaviolaa G, Sougnez C, Gupta M, Yu EU, Mostaghel EA, Cheng HH, Mulcahy H, True LD, Plymate SR, Dvinge H, Ferraldeschi R, Flohr P, Miranda S, Zafeiriou Z, Tunari, Mateo J, Lopez-Perez R, Demichelis F, Robinson BD, Schiffman M, Nanus DM, Tagawa ST, Sigaras A, Eng KW, Elemento O, Sboner A, Heathman EI, Scher HI, Pienta KJ, Kantoff P, de Bono J, Rubin MA, Nelson PS, Garraway LA, Sawyers CL, Chinnaiyan AM. Integrative Clinical Genomics of Advanced Prostate Cancer. Cell. 2015; 161:1215–1228. [PubMed: 26000489]
24. Baca SC, et al. Punctuated evolution of prostate cancer genomes. Cell. 2013; 153:666–677. doi:10.1016/j.cell.2013.03.021. [PubMed: 23622249]
25. Barbieri CE, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nature genetics. 2012; 44:685–689. doi:10.1038/ng.2279. [PubMed: 22610119]
26. Robinson D, et al. Integrative clinical genomics of advanced prostate cancer. Cell. 2015; 161:1215–1228. doi:10.1016/j.cell.2015.05.001. [PubMed: 26000489]
27. Pietanza MC, Ladanyi M. Bringing the genomic landscape of small-cell lung cancer into focus. Nat Genet. 2012; 44:1074–1075. doi:ng.2415 [pii] 10.1038/ng.2415. [PubMed: 23011222]
28. Banck MS, et al. The genomic landscape of small intestine neuroendocrine tumors. J Clin Invest. 2013; 123:2502–2508. doi:67963 [pii] 10.1172/JCI67963. [PubMed: 23676460]
29. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012; 487:330–337. doi:nature11252 [pii] 10.1038/nature11252. [PubMed: 22810696]
30. Prandi D, et al. Unraveling the clonal hierarchy of somatic genomic aberrations. Genome Biol. 2014; 15:439. doi:s13059-014-0439-6 [pii] 10.1186/s13059-014-0439-6. [PubMed: 25160065]

31. Forbes SA, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2015; 43:D805–811. doi:gku1075 [pii] 10.1093/nar/gku1075. [PubMed: 25355519]
32. Rajan N, et al. The cylindromatosis gene product, CYLD, interacts with MIB2 to regulate notch signalling. Oncotarget. 2014; 5:12126–12140. doi:2573 [pii]. [PubMed: 25565632]
33. Lim JH, et al. CYLD negatively regulates transforming growth factor-beta-signalling via deubiquitinating Akt. Nat Commun. 2012; 3:771. doi:ncomms1776 [pii] 10.1038/ncomms1776. [PubMed: 22491319]
34. Theurillat JP, et al. Prostate cancer. Ubiquitylome analysis identifies dysregulation of effector substrates in SPOP-mutant prostate cancer. Science. 2014; 346:85–89. doi:346/6205/85 [pii] 10.1126/science.1250255. [PubMed: 25278611]
35. Lin L, et al. Mechanisms underlying cancer growth and apoptosis by DEK overexpression in colorectal cancer. PLoS One. 2014; 9:e111260. doi:10.1371/journal.pone.0111260 PONE-D-14-27910 [pii]. [PubMed: 25340858]
36. Nowell PC. The clonal evolution of tumor cell populations. Science. 1976; 194:23–28. [PubMed: 959840]
37. Maley CC, et al. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. Nat Genet. 2006; 38:468–473. doi:ng1768 [pii] 10.1038/ng1768. [PubMed: 16565718]
38. Hansel DE, et al. Shared TP53 gene mutation in morphologically and phenotypically distinct concurrent primary small cell neuroendocrine carcinoma and adenocarcinoma of the prostate. Prostate. 2009; 69:603–609. doi:10.1002/pros.20910. [PubMed: 19125417]
39. Gundem G, et al. The evolutionary history of lethal metastatic prostate cancer. Nature. 2015; 520:353–357. doi:nature14347 [pii] 10.1038/nature14347. [PubMed: 25830880]
40. Smith BA, et al. A basal stem cell signature identifies aggressive prostate cancer phenotypes. Proceedings of the National Academy of Sciences of the United States of America. 2015; 112:E6544–6552. doi:10.1073/pnas.1518007112. [PubMed: 26460041]
41. Borno ST, et al. Genome-wide DNA methylation events in TMPRSS2-ERG fusion-negative prostate cancers implicate an EZH2-dependent mechanism with miR-26a hypermethylation. Cancer discovery. 2012; 2:1024–1035. doi:10.1158/2159-8290.CD-12-0041. [PubMed: 22930729]
42. Varambally S, et al. Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. Science. 2008; 322:1695–1699. doi:10.1126/science.1165395. [PubMed: 19008416]
43. Yu J, et al. A polycomb repression signature in metastatic prostate cancer predicts cancer outcome. Cancer research. 2007; 67:10657–10663. doi:10.1158/0008-5472.CAN-07-2498. [PubMed: 18006806]
44. Karanikolas BD, Figueiredo ML, Wu L. Comprehensive evaluation of the role of EZH2 in the growth, invasion, and aggression of a panel of prostate cancer cell lines. The Prostate. 2010; 70:675–688. doi:10.1002/pros.21112. [PubMed: 20087897]
45. Clermont PL, et al. Polycomb-mediated silencing in neuroendocrine prostate cancer. Clin Epigenetics. 2015; 7:40. doi:10.1186/s13148-015-0074-4 74 [pii]. [PubMed: 25859291]
46. Cheng AS, et al. EZH2-mediated concordant repression of Wnt antagonists promotes beta-catenin-dependent hepatocarcinogenesis. Cancer Res. 2011; 71:4028–4039. doi:0008-5472.CAN-10-3342 [pii] 10.1158/0008-5472.CAN-10-3342. [PubMed: 21512140]
47. Kanduri M, et al. A key role for EZH2 in epigenetic silencing of HOX genes in mantle cell lymphoma. Epigenetics. 2013; 8:1280–1288. doi:26546 [pii] 10.4161/epi.26546. [PubMed: 24107828]
48. Akamatsu S, et al. The Placental Gene PEG10 Promotes Progression of Neuroendocrine Prostate Cancer. Cell Rep. 2015; 12:922–936. doi:S2211-1247(15)00752-4 [pii] 10.1016/j.celrep.2015.07.012. [PubMed: 26235627]
49. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015; 163:1011–1025. doi:S0092-8674(15)01339-2 [pii] 10.1016/j.cell.2015.10.025. [PubMed: 26544944]
50. Chakravarty D, et al. The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. Nat Commun. 2014; 5:5383. doi:ncomms6383 [pii] 10.1038/ncomms6383. [PubMed: 25415230]

51. Carreira S, et al. Tumor clone dynamics in lethal prostate cancer. Sci Transl Med. 2014; 6:254ra125. doi:6/254/254ra125 [pii] 10.1126/scitranslmed.3009448.
52. Beltran H, et al. Whole-Exome Sequencing of Metastatic Cancer and Biomarkers of Treatment Response. JAMA Oncol. 2015; 1:466–474. doi:2294965 [pii] 10.1001/jamaoncol.2015.1313. [PubMed: 26181256]
53. Esgueva R, et al. Next-generation prostate cancer biobanking: toward a processing protocol amenable for the International Cancer Genome Consortium. Diagn Mol Pathol. 2012; 21:61–68. doi:10.1097/PDM.0b013e31823b6da6. [PubMed: 22555088]
54. Van Allen EM, et al. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. Nature medicine. 2014; 20:682–688. doi:10.1038/nm.3559.
55. Cibulskis K, et al. ContEst: estimating cross-contamination of human samples in next-generation sequencing data. Bioinformatics. 2011; 27:2601–2602. doi:10.1093/bioinformatics/btr446. [PubMed: 21803805]
56. Demichelis F, et al. SNP panel identification assay (SPIA): a genetic-based assay for the identification of cell lines. Nucleic Acids Res. 2008; 36:2446–2456. doi:gkn089 [pii] 10.1093/nar/gkn089. [PubMed: 18304946]
57. Cibulskis K, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013; 31:213–219. doi:nbt.2514 [pii] 10.1038/nbt.2514. [PubMed: 23396013]
58. Ramos AH, et al. Oncotator: cancer variant annotation tool. Hum Mutat. 2015; 36:E2423–2429. doi:10.1002/humu.22771. [PubMed: 25703262]
59. Lawrence MS, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013; 499:214–218. doi:nature12213 [pii] 10.1038/nature12213. [PubMed: 23770567]
60. Romanel A, Lago S, Prandi D, Sboner A, Demichelis F. ASEQ: fast allele-specific studies from next-generation sequencing data. BMC Med Genomics. 2015; 8:9. doi:10.1186/s12920-015-0084-2 10.1186/s12920-015-0084-2 [pii]. [PubMed: 25889339]
61. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29:15–21. doi:bts635 [pii] 10.1093/bioinformatics/bts635. [PubMed: 23104886]
62. Habegger L, et al. RSEQtools: a modular framework to analyze RNA-Seq data using compact, anonymized data summaries. Bioinformatics. 2011; 27:281–283. doi:btq643 [pii] 10.1093/bioinformatics/btq643. [PubMed: 21134889]
63. Park K, et al. Antibody-based detection of ERG rearrangement-positive prostate cancer. Neoplasia. 2010; 12:590–598. [PubMed: 20651988]
64. Sboner A, et al. FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. Genome Biol. 2010; 11:R104. doi:gb-2010-11-10-r104 [pii] 10.1186/gb-2010-11-10-r104. [PubMed: 20964841]
65. Mosquera JM, et al. Concurrent AURKA and MYCN gene amplifications are harbingers of lethal treatment-related neuroendocrine prostate cancer. Neoplasia. 2013; 15:1–10. [PubMed: 23358695]
66. Lin PC, et al. Epigenomic alterations in localized and advanced prostate cancer. Neoplasia. 2013; 15:373–383. [PubMed: 23555183]
67. Akalin A, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. 2012; 13:R87. doi:gb-2012-13-10-r87 [pii] 10.1186/gb-2012-13-10-r87. [PubMed: 23034086]
68. Dale RK, Pedersen BS, Quinlan AR. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. Bioinformatics. 2011; 27:3423–3424. doi:btr539 [pii] 10.1093/bioinformatics/btr539. [PubMed: 21949271]
69. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. 2009; 37:W305–311. doi:gkp427 [pii] 10.1093/nar/gkp427. [PubMed: 19465376]
70. Vire E, et al. The Polycomb group protein EZH2 directly controls DNA methylation. Nature. 2006; 439:871–874. doi:nature04431 [pii] 10.1038/nature04431. [PubMed: 16357870]

71. Cerami E, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012; 2:401–404. doi:2/5/401 [pii] 10.1158/2159-8290.CD-12-0095. [PubMed: 22588877]

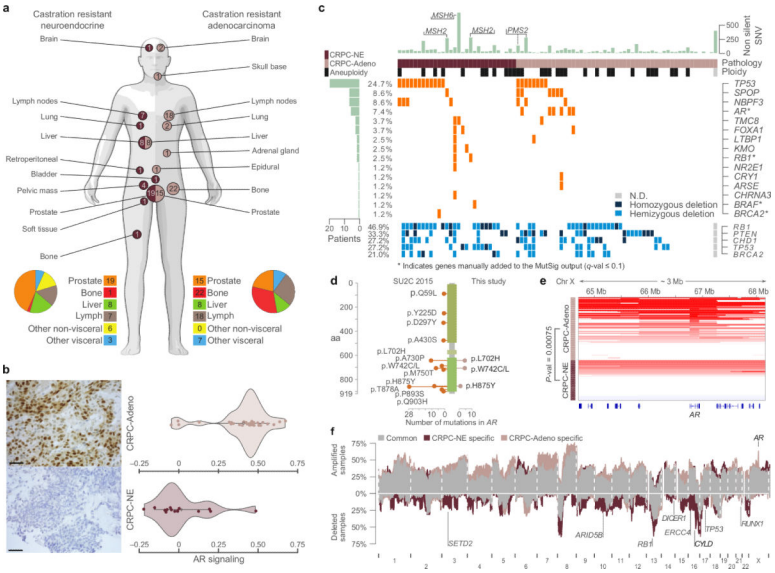

Figure 1. Clinical and mutational profile of the cohort

(**a**) Schematic illustrating sites of biopsy for CRPC-NE (dark pink) and CRPC-Adeno (light pink) subgroups. Numbers in circles indicate numerosity of samples from each site. (**b**) AR signaling (right) based on abundance of mRNA transcripts included in the AR signaling signature described in ref 19. Violin plots show the density of AR signaling. Each dot represents a sample; diamonds and solid lines represent the mean and 95% confidence interval, respectively. Representative immunohistochemistry (left) shows AR protein expression. Scale bars, 50 μm. (**c**) Significantly mutated genes. Each row represents a gene and each column an individual subject. Top light green bars correspond to the total number of non-silent SNVs in an individual. Left light green bars indicate the number of subjects harboring non silent corresponding mutations in the genes indicated on the right. Bottom panel reports the copy number status of selected genes. (**d**) Genomic location of AR mutations in samples from SU2C-2015 and this study. (**e**) Copy number status of AR locus. Color intensity and location are indicative of level and focality of amplification. (**f**) Frequency of copy number aberrations; concordant fractions (gray), CRPC-NE specific (dark pink) and CRPC-Adeno specific (light pink). Data adjusted for tumor ploidy and purity. Highlighted genes are significantly preferentially aberrant in one class and demonstrate concordant differential mRNA levels (for DNA and mRNA: FDR <= 10% for deletions and p-value <= 1% for amplifications).

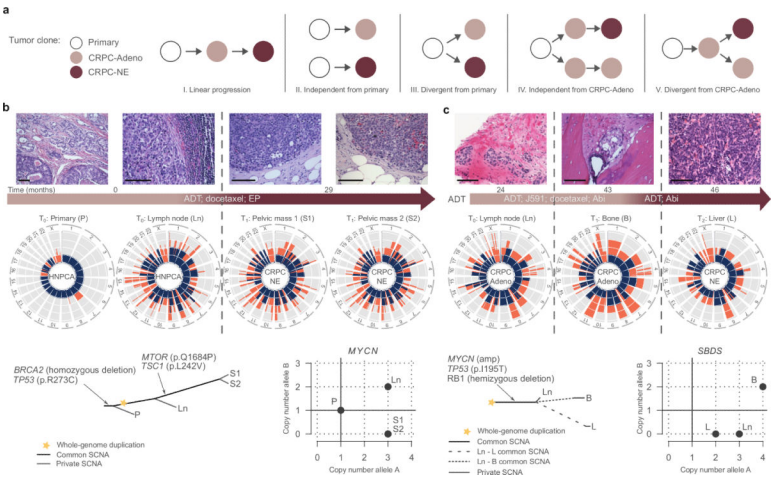

Figure 2. Tracing CRPC-NE emergence through allele specific analysis

(**a**) Potential models of evolution that occur during prostate cancer progression towards the neuroendocrine phenotype**:** linear progression from primary untreated adenocarcinoma to CRPC-Adeno to CRPC-NE; independent progression of two distinct clonal populations within the primary or metastatic CRPC-Adeno towards either CRPC-Adeno or CRPC-NE; divergent clonal evolution of CRPC-NE from either primary adenocarcinoma or CRPC-Adeno. *indicates favored model. (**b**) Allele specific analysis of primary prostate adenocarcinoma and local lymph node metastasis removed at time of radical prostatectomy (RP) and two metastatic CRPC-NE (treated) tumors (3 years after RP) from subject WCMC7520. H&E pathology images and intervening therapies are shown in the timeline. Scale bars, 100 μm. (**c**) Allele specific analysis of tumors at three time points from patient subject WCMC161 during castration resistance: lymph node (CRPC-Adeno), bone biopsy (CRPC-Adeno), and liver biopsy (small cell CRPC-NE). H&E pathology images and intervening therapies are shown in the timeline. ADT= androgen deprivation therapy; EP= etoposide and cisplatin chemotherapy; Abi= abiraterone acetate with prednisone. Circos plots summarize genome-wide allele specific DNA quantity in tumor cells. Individual's tumor phylogeny sketched upon allele-specific analysis including genome-wide amplification and ploidy assessment. Scale bars, 100 μm.

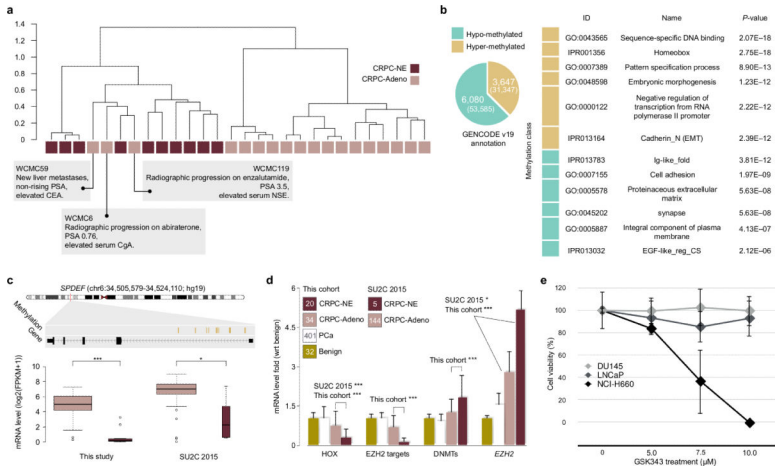

Figure 3. Methylation analysis of CRPC-NE and CRPC-Adeno

(**a**) Hierarchical clustering of 28 eRRBS samples data using (1 - Pearson's correlation) as distance measure on unselected sites. Clinical features of outlier cases are described. (**b**) Left, pie chart showing the number of differentially methylated genes, identified by annotating hyper- and hypo- methylated loci (number is reported between parentheses) on GENCODE version 19. Right, table shows a selection of terms enriched by differentially methylated genes. (**c**) Top, genome track of *SPDEF*. Hyper-methylated loci are reported in the annotation track. Bottom, box plot of expression levels of *SPDEF* samples for This Study (left) and SU2C/PCF 2015 (right) cohorts. (**d**) Bar plots highlight the effect of *EZH2* transcription activity across 487 samples with different pathology classification. The bars are relative to the mRNA level fold (with respect to benign prostate tissue samples) of homeobox genes under-expressed in CRPC-NE versus CRPC-Adeno (FDR < 0.1); a selection of *EZH2* target genes (*DKK1*, *NKD1*, *AMD1*, *HOXA13*, *HOXA11*, *NKX3-1*); DNA methyltransferase genes - indicated as DNMTs (*DNMT1*, *DNMT3B*, *DNMT3A*, *DNMT3L*); *EZH2*. Significance of differences between CRPC-NE and CRPC-Adeno subgroups are shown (max $P = 3*10^{-5}$ for DNMTs). When significant, p-values in SU2C/PCF cohort are shown. The number of samples for each pathology classification is reported inside the square symbols of the legend. (**e**) Cell viability in prostate adenocarcinoma cell lines (DU145, LNCaP) the neuroendocrine prostate cell line NCI-H660 assessed at 48 hours after treatment with escalating does of the EZH2 inhibitor GSK343 (5, 7.5, 10uM).

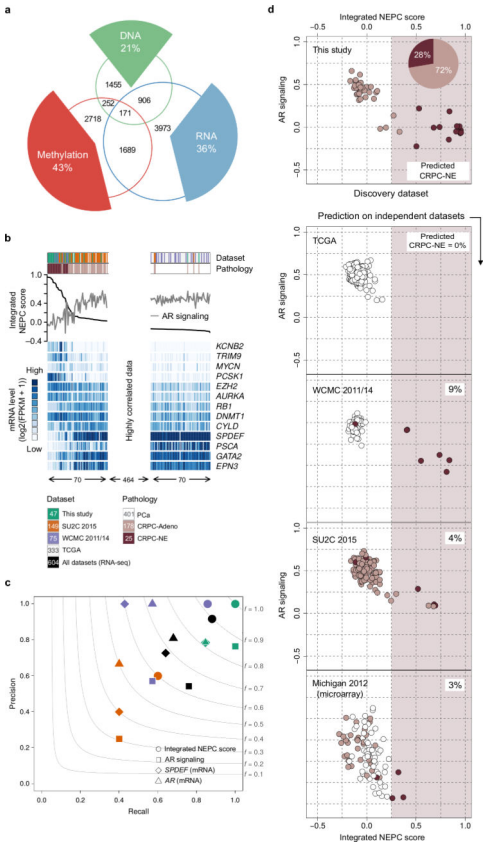

Figure 4. Integrative DNA, RNA and Methylation analysis

(**a**) Weighted Venn diagram with the number of protein-coding genes significantly differentially observed in the three data layers. The superimposed pie chart reports the estimation of the impact of each layer upon the following priority rule: methylation overall and DNA over RNA. (**b**) Integrated NEPC score analysis across 604 samples from four different RNA-Seq prostate cancer datasets (This Study, SU2C/PCF 2015, WCMC 2011/2014 and TCGA). Samples are ordered by decreasing values of Integrated NEPC score (only a fraction of data is shown, entire data are reported in **Supplementary Fig. 10**). Top, annotation tracks report original dataset and pathology classification. Middle, plot reports Integrated NEPC score (black line) and AR signaling (grey line) across samples. Bottom, heat map of normalized FPKMs for a selection of the 70 genes (in rows) across samples (in columns). (**c**) Prediction accuracy of CRPC-NE samples by precision and recall statistics for Integrated NEPC Score (circles), AR signaling (squares), mRNA level of *SPDEF* (diamonds), *AR* (triangles) in RNA-seq datasets: This Study (green), SU2C/PCF 2015 (orange), WCMC 2011/14 (violet) and all datasets (black). Grey curves represent F-measure levels, defined as the harmonic mean of precision and recall. Due to the absence of CRPC-NE samples (positive events), TCGA data were not reported here. (**d**) AR signaling versus Integrated NEPC score across 730 samples from five independent prostate datasets using transcriptome data as proxy. The old-rose shaded area refers to significant values of Integrated NEPC Score. Predicted CRPC-NE percentages calculated by excluding benign samples.