Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer

Himisha Beltrán1,2,3, Davide Prandi4,13, Juan Miguel Mosquera1,5, Matteo Benelli4, Loredana Puca1, Joanna Cyrt1, Clarisse Marotz1, Eugenia Giannopoulou6, Balabhadrapatruni V S K Chakravarthi7, Sooryanarayana Varambally7, Scott A Tomlins8, David M Nanus2,3, Scott T Tagawa2,3, Eliezer M Van Allen9,10, Olivier Elemento1,6, Andrea Sboner1,5,11, Levi A Garraway9,10,12,14, Mark A Rubin1,3,5,14 & Francesca Demichelis1,4,11,14

An increasingly recognized resistance mechanism to androgen receptor (AR)-directed therapy in prostate cancer involves epithelial plasticity, in which tumor cells demonstrate low to absent AR expression and often have neuroendocrine features. The etiology and molecular basis for this ‘alternative’ treatment-resistant cell state remain incompletely understood. Here, by analyzing whole-exome sequencing data of metastatic biopsies from patients, we observed significant genomic overlap between castration-resistant tumors that were histologically characterized as prostate adenocarcinomas (CRPC-Adeno) and neuroendocrine prostate cancer (CRPC-NE); analysis of biopsy samples from the same individuals over time points to a model most consistent with divergent clonal evolution. Genome-wide DNA methylation analysis revealed marked epigenetic differences between CRPC-NE tumors and CRPC-Adeno, and also designated samples 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 cancer death in males worldwide1. The mainstay of therapy for patients with metastatic spread, including castration-resistant disease, is hormonal therapy that targets the AR2–4. Enzalutamide and abiraterone are potent AR-targeted therapies approved for the treatment of men with castration-resistant prostate cancer (CRPC)5,6. Although the use of these agents improves the survival and quality of life of individuals with CRPC, most patients ultimately develop resistance to them2. Predictive biomarkers that help distinguish responders from nonresponders before starting the next line of hormonal therapy are 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 therapies11–14. Prognosis of CRPC-NE is poor owing to late recognition, heterogeneous clinical features and lack of effective systemic therapies.

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 CRPC-Adeno and 30 with features of CRPC-NE, as confirmed by pathologic consensus criteria; we studied matched normal cells from all of the patients, multiple tumor biopsies from 17 patients and single tumor biopsies from 64 patients. We hypothesized that CRPC-NE could be distinguished from CRPC-Adeno on the basis of distinct molecular alterations and that this information could improve upon and supplement the current, often challenging diagnostic features that are reliant on morphological characterization. We also hypothesized that CRPC-NE that develops after therapy arises clonally from a CRPC-Adeno precursor, rather than from the selection of pre-existing neuroendocrine clones. Finally, we hypothesized that AR-independent prostate adenocarcinomas that share CRPC-NE-specific molecular alterations may represent tumors that are at high risk for either progression or transition to CRPC-NE.

We evaluated biopsies from a wide range of metastatic sites on the basis of the presence and accessibility of lesions, with more bone biopsies obtained from patients with CRPC-Adeno as compared to those with CRPC-NE (31% CRPC-Adeno versus 2% CRPC-NE;
P < 0.05, binomial test) (Fig. 1a). The clinical and pathologic features of the biopsy samples are summarized in Supplementary Table 1 and Supplementary Figure 1. As expected, we found lower protein expression of the AR, on average, in CRPC-NE samples, using immunohistochemistry (Fig. 1b). We also quantified AR signaling status by measuring the expression of mRNAs that were included in a previously defined AR signature by Hieronymus et al. (Supplementary Table 2) and observed overall lower abundance of these mRNAs in CRPC-NE specimens as compared to CRPC-Adeno specimens (Fig. 1b); however, there was 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 tissue pairs. The mutational landscape of CRPC-NE tumors was similar to that of CRPC-Adeno, but it was also consistent with those in published studies of CRPC-NE tumors (Fig. 1c)—including enrichment for the loss of RB1, which encodes the retinoblastoma tumor suppressor protein (in 70% of CRPC-NE versus 32% of CRPC-Adeno samples; P = 0.003, proportion test), and for the mutation or deletion of TP53, which encodes the p53 tumor suppressor protein (in 66.7% of CRPC-NE versus 31.4% of CRPC-Adeno samples; P = 0.0043, proportion test). Loss of RB1 is common in primary small-cell prostate and lung carcinomas, and it promotes small-cell carcinoma pathogenesis in conjunction with a TP53 mutation16–18; in our series, concurrent loss of RB1 and TP53 was present in 53.3% of CRPC-NE versus 13.7% of CRPC-Adeno samples (P = 0.0004, proportion test).

Another feature distinguishing CRPC-NE tumors 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 AR are commonly observed in CRPC-Adeno and are associated with treatment resistance to AR-directed therapies19,20. This observation was confirmed in our cohort; 29 samples showed a focal amplification or a point mutation in AR, and 21 samples had alterations in genes encoding known AR co-activators (FOXA1, NCOA1, NCOA2 and ZBTB16). In contrast, AR point mutations were notably absent in CRPC-NE samples, and amplifications, when present, were of low level, broad 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 tumors may be due to the clonal selection of non-amplified CRPC-Adeno tumor subpopulations through selective pressure (in the context of AR-directed therapies). The AR splice variant 7 (AR-V7), which is associated with AR-driven resistance in CRPC21,22, was observed in both CRPC-Adeno and CRPC-NE tumors, and although overall AR expression was lower in CRPC-NE tumors, the ratio of AR-V7 to wild-type AR mRNA was significantly lower in CRPC-NE samples than in CRPC-Adeno samples (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 findings imply that distinct biological properties are associated with CRPC-NE.

The overall spectrum of genomic alterations, however, was similar in tumors of both CRPC subtypes (Supplementary Fig. 2) and was maintained after controlling for the site of biopsy and for prior exposure to chemotherapy (Supplementary Fig. 3). There were no significant differences in the distribution of nonsilent point mutations, polyploidy or copy-number genomic burden between the CRPC-Adeno and CRPC-NE tumor subtypes (on average >30% of the genome was aberrant), and tumors of both subtypes showed a significantly higher number of clonal point mutations as compared with those of clinically localized prostate cancer (P < 10−7, Wilcoxon test)23,24. The median number of nonsilent single-nucleotide variants (SNVs) in metastatic samples was 41 (range: 2–729) (Supplementary Tables 3 and 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 studies20,25. Recurrent alterations are shown in Figure 1c and putative gene fusions (as determined by RNA sequencing (RNA-seq) analysis) are in Supplementary Table 5.

The substantial overlap in the overall somatic copy-number landscape between CRPC-Adeno and CRPC-NE tumors (Fig. 1f) was noteworthy in light of the marked genomic differences between adenocarcinomas and small-cell carcinomas observed in other tissue subtypes (for example, lung and gastrointestinal tract tumors). After correcting for admixture of non-tumor cells and for ploidy26, we sought to identify regions of the genome that were differentially altered and noted copy-number alterations that are enriched in either CRPC-NE or CRPC-Adeno samples (Supplementary Tables 6 and 7). Putative cancer genes27 within differentially deleted regions in CRPC-NE with concordant downregulation at the mRNA level (false-discovery rate (FDR) < 10% for both DNA and mRNA) are highlighted in Figure 1f. The top ranked gene was CYLD, which encodes cylindromatosis, a deubiquitinating enzyme reported as a tumor suppressor involved in negative regulation of multiple signaling pathways including those involving nuclear factor (NF)-κB28, transforming growth factor (TGF)-β and Notch29. CYLD was deleted in 51% of CRPC-NE samples and verifed by fluorescence in situ hybridization analysis (Supplementary Fig. 5). We found that genomic loss of CYLD was associated with reduced mRNA expression and a modest decrease in the expression of genes encoding factors that are associated with AR signaling (in this study, in the SU2C/PCF 2015 cohort of CRPC20, and in cell lines; Supplementary Fig. 5), suggesting that CYLD loss alone may be insufficient to promote AR indiffERENCE, and that it might cooperate with other molecular alterations to do so. By extending the computational framework of the algorithm CLONET23,26—which first estimates the purity of somatic aberrations—we assessed for allele-specific copy-number clonality (Supplementary Note) and found both focal and broad copy-neutral or copy-aberrant loss of heterozygosity across our cohort (Supplementary Fig. 5). This included focal allelic imbalance of the DEK proto-oncogene in CRPC-NE tumors as compared to CRPC-Adeno (P = 0.04, binomial test). DEK belongs to a class of DNA topology modulators with a reported oncogenic role in prostate cancer, including in CRPC-NE30,31.

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 or clonal tumor cell populations originating from the primary tumor, with sequential acquisition of genomic alterations (linear)32. Alternatively, independent clones within the primary tumor or the metastasis could give rise to parallel and distinct resistant tumor populations (independent)33. Finally, there may be genetic diversification in the primary tumor or in the metastatic lesion as a mechanism of adaptation, leading to selective pressure and divergent clonal evolution (divergent). Although prior studies have demonstrated concordance of the prostate cancer–specific genomic rearrangement TMPRSS2-ERG3 and other single-gene alterations34,35.
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 to infer clonal-expansion dynamics, we studied serial tumor samples from individuals during the course of their disease. Patient WCMC7520 had a prostatectomy for a clinically localized Gleason 9 prostate adenocarcinoma with local lymph node involvement and was treated initially with adjuvant androgen–deprivation therapy (ADT) followed by castration-resistant adenocarcinoma. However, from these data we cannot definitively rule out but clonal, paths to the lymph node and both distant metastasis from two time points and observed a homozygous deletion of the primary tumor, the local lymph node and the CRPC-NE distant metastasis from two time points and observed a homozygous deletion of the BRCA2 tumor suppressor gene and a mutation in TP53 in biopsy samples from all of the sites, suggesting a common ancestor. Allele-specific DNA analysis highlighted diverse genomic states of other key genes such as MYCN, which encodes the N-myc oncoprotein (Fig. 2b; MYCN has been previously described as oncogenic in CRPC-NE). After comparison of the samples from all of the sites, the patient’s primary prostate specimen had lesions that suggested divergent, but clonal, paths to the lymph node and both distant metastases. However, from these data we cannot definitively rule out...
metastasis-to-metastasis seeding, which was recently proposed as a mechanism of tumor progression\(^3\). Patient WCMC161 showed progression of disease after multiple lines of therapy for CRPC—including the development of new visceral metastases at the time of progression on abiraterone with a liver biopsy showing small-cell carcinoma (Fig. 2c and 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 Shwachman-Bodian-Diamond syndrome (SBDS) gene (in which the allele-specific state of SBDS in metastatic samples was not compatible with linear evolution; Fig. 2c) and that the lymph node, bone and liver metastases arose from a common ancestor. Phylogenetic trees outlining disease evolution from the adenocarcinoma to a neuroendocrine phenotype in individual patients, constructed through analyses of additional multitumor cases (Supplementary Fig. 6), collectively rule against a parallel evolutionary model. On the basis of these temporal case series, the high degree of clonality in CRPC and the overall similarity between the genomic profiles of CRPC-Adeno and CRPC-NE specimens, the most parsimonious model that explains the data is divergent clonal evolution of metastatic CRPC to either an AR-driven or an AR-indifferent state (Fig. 2c, model 5). In other words, CRPC-NE tumors appear clonal in origin with a clonal ancestry traceable back to a CRPC-Adeno precursor.

Although 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. To this end, we evaluated CpG-rich methylation on a genome-wide scale by single-cytosine-resolution DNA methylation analysis (using enhanced reduced-representation bisulfite sequencing (eRRBS). In contrast to the mostly similar genomic data, the CRPC-NE and CRPC-Adeno subtypes showed strong epigenetic segregation, using unsupervised analysis of unsellected methylation sites (Fig. 3a, Supplementary Fig. 7 and Supplementary Table 8). In addition, methylation of first exons and gene promoters was associated with...
significant changes in gene expression \( (P < 0.001, \text{Wilcoxon test; Supplementary Fig. 8}) \). 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 the advent of, the CRPC-NE subtype is associated with epigenetic dysregulation. Notably, the epigenetic signal identified three cases with clinical features of AR independence that were binned as adenocarcinoma on the basis of standard pathology but that segregated with CRPC-NE after unsupervised analysis (Fig. 3a and Supplementary Fig. 8). All three of these patients demonstrated radiographic progression in the setting of a stable or low serum amount of the androgen-regulated protein prostate-specific antigen (PSA). These data suggest that clustering predictions based on DNA methylation may provide additional information associated with AR independence and CRPC-NE that potentially improves on those based on tumor morphology.

By using functional enrichment analysis of differentially methylated genes, we identified epigenetically dysregulated pathways, including those involving neuronal, cell-cell adhesion, developmental, epithelial-mesenchymal transition (EMT) and stem cell programs (Fig. 3b), which are thought to be relevant for CRPC-NE progression36. Among our notable findings (Supplementary Fig. 7), we observed hypermethylation and reduced expression of the tumor suppressor gene \( \text{SPDEF} \) (which encodes a prostate-derived Ets factor) is a transcriptional activator and a regulator of cellular differentiation (Fig. 3c). \( \text{SPDEF} \) expression levels estimated in the SU2C/PCF 2015 cohort are shown. The number of samples for each pathway classification is reported inside the square symbols of the legend. Error bars are median absolute deviation. * \( P < 0.05 \), *** \( P < 0.001 \); by Wilcoxon test. (e) Cell viability in the prostate adenocarcinoma cell lines DU145 and LNCaP, as well as in the neuroendocrine prostate cell line NCI-H660, as assessed at 48 h after treatment with escalating doses of the EZH2 inhibitor GSK343 (5.0, 7.5 and 10 \( \mu M \)). Error bars are standard error of the mean.

![Image](https://example.com/image.png)

**Figure 3** Methylation analysis of CRPC-NE and CRPC-Adeno tumors. (a) Hierarchical clustering of eRBB data from 28 samples, using ‘1 – Pearson’s correlation’ as distance measure on unscaled sites. Clinical features of outlier cases are described. CEA, carcinoembryonic antigen; CGA, chromogranin A; NSE, neuron-specific enolase. (b) Left, pie chart showing the number of differentially methylated genes, identified by annotating hyper- and hypomethylated loci (number is reported in parentheses) on GENCODE version 19. Right, selection of functional categories enriched after analysis of differentially methylated genes. (c) Top, genome track of \( \text{SPDEF} \). Hypermethylated loci are reported in the annotation track. Bottom, box plot of \( \text{SPDEF} \) expression levels for samples in this study (left) and in the SU2C/PCF 2015 (SU2C 2015; right) cohort. * \( P < 0.05 \), *** \( P < 0.001 \); by Wilcoxon test. Lower and upper bars correspond to the minimum and maximum non-outlier values of the data distribution. Outliers are defined as values outside of the range \( \text{Q1} – 1.5 \times (\text{Q3} – \text{Q1}) \), \( \text{Q3} + 1.5 \times (\text{Q3} – \text{Q1}) \), where \( \text{Q1} \) and \( \text{Q3} \) are the first and third quartile, respectively. (d) Bar plots showing the effect of EZH2 transcriptional activity across 487 samples of differing pathology classifications. The bars are relative to the median values of changes in mRNA expression, with respect to benign prostate tissue samples, of homeobox genes with reduced expression in CRPC-NE versus CRPC-Adeno samples (FDR < 0.1); expression changes in a selection of mRNAs of EZH2 target genes (\( \text{DDKK1, NKK01, AMID1, HOXA13, HOXA11 and NKK3-I} \)), DNA methyltransferase genes (indicated as DNMTs; \( \text{DNMT1, DNMT3B, DNMT3A and DNMT3L} \)) and \( \text{EZH2} \) are shown. Significance of differences between CRPC-NE and CRPC-Adeno subgroups are shown (maximum \( P = 3 \times 10^{-5} \) for DNMTs). When significant, \( P \) values estimated in the SU2C/PCF 2015 cohort are shown. The number of samples for each pathway classification is reported inside the square symbols of the legend. Error bars are median absolute deviation. * \( P < 0.05 \), *** \( P < 0.001 \); by Wilcoxon test. (e) Cell viability in the prostate adenocarcinoma cell lines DU145 and LNCaP, as well as in the neuroendocrine prostate cell line NCI-H660, as assessed at 48 h after treatment with escalating doses of the EZH2 inhibitor GSK343 (5.0, 7.5 and 10 \( \mu M \)). Error bars are standard error of the mean.
NCl-H660, as compared to that in the prostate adenocarcinoma line LNCaP (Supplementary Fig. 9).

Expression of mRNA encoding the histone methyltransferase EZH2, which has previously been associated with aggressive disease in prostate cancer38–41, was twofold higher in CRPC-NE than that in CRPC-Adeno (P < 10⁻⁶, Wilcoxon test); EZH2 protein was also more abundant in CRPC-NE (Supplementary Fig. 9). These findings add to its previously reported differential-expression status in CRPC-NE versus localized prostate cancer39,42. Furthermore, EZH2-repressed target genes43,44 were significantly downregulated 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 homeobox factor (HOX)-encoding genes (P = 0.001, Wilcoxon test). Treatment with the EZH2 inhibitor GSK343 resulted in a preferential decrease in the viability of NCI-H660 cells, as compared to that in other non-neuroendocrine prostate cancer cell lines (Fig. 3e), with a substantial downregulation of several CRPC-NE-associated genes after treatment, including NGAM (CD56), MYCN and PEG10 (ref. 45) (Supplementary Fig. 9). Overall these data support a key role of the epigenome in the emergence and/or maintenance of CRPC-NE.

On the basis of the current gaps in the clinical and molecular assessment of CRPC-NE, we used these data to develop a molecular classifier to potentially improve the often-challenging diagnosis of CRPC-NE. This integrated 70-gene neuroendocrine prostate cancer (NEPC) classifier was developed by using expression data of genes that were prioritized by genomic, transcriptomic or epigenomic status (Fig. 4a,b and Supplementary Fig. 7; see Online Methods), and it demonstrated both a precision and a recall of >0.99 in identifying CRPC-NE in our discovery cohort (Fig. 4c,d and Supplementary Table 9). Included within the classifier were also genes we had previously described as being overexpressed in CRPC-NE5, including AURKA (P < 10⁻⁵, Wilcoxon test) and MYCN (P < 10⁻⁴, Wilcoxon test) (Supplementary Fig. 10). Interrogation of transcriptomic data from 683 prostate samples (Supplementary Table 10) using data sets from The Cancer Genome Atlas (TCGA)96, Grasso et al.19 (Michigan 2012 cohort), Robinson et al.20 (SU2C/PCF 2015 cohort) and our group45,47 revealed a positive classifier score in up to 8% of metastatic tumors (n = 191) and none of treatment-naïve prostate adenocarcinomas (n = 460) or of benign prostate (n = 32) (Fig. 4d and Supplementary Figs. 10 and 11). Of those with an elevated classifier score, we reviewed the pathology of those specimens and found that >80% of them had pathologic features of CRPC-NE (the remaining 20% were adenocarcinomas) (Fig. 4d and 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 data sets as compared to other variables, such as conventional neuroendocrine markers (i.e., CGHA, SYP, NSE and CD56 transcripts) plus PSA (Supplementary Fig. 12). AR mRNA expression, AR signaling status and highly ranked differentially expressed genes (such as SPDEF) (Fig. 4c). Although we recognize the influence of other factors, including differences in prior therapies, on gene expression changes, we posit that castration-resistant tumors with a moderate or rising NEPC score may represent tumors with AR-independent features that are either in transition to or at a high risk for CRPC-NE progression during treatment with AR therapies. In fact, a subpopulation of prostate adenocarcinoma cells (LNCaP) that were treated long term with enzalutamide acquired molecular features of CRPC-NE (i.e., methylation of SPDEF) (Supplementary Figs. 8 and 13).

Although we identified cases with low AR signaling and NEPC classifier scores, and cannot exclude a less common alternative distinct subset, our data favors a continuum of progression from an AR-driven state to an AR-independent state that is associated with neuroendocrine molecular features. These findings warrant the clinical evaluation of more specimens to further investigate the potential superiority and applicability of the classifier diagnostic criteria over more conventional diagnostic criteria and to 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 are often impractical. The incorporation of different molecular layers helps apply the classifier to different data sets 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 DNA48). If CRPC-NE alterations could be detected earlier during CRPC-Adeno disease progression, for instance, then such individuals could potentially be selected for CRPC-NE–directed (such as platinum chemotherapy) rather than AR-targeted systemic therapies or for potentially co-targeting therapeutic approaches. Furthermore, 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 support divergent evolution of CRPC-NE from one or more CRPC-Adeno cells (adaptation) rather than linear or independent clonal evolution, with selective pressure of subclonal populations with wild-type AR and the acquisition of new genomic and epigenetic drivers associated with decreased AR signaling and epithelial plasticity. However, there are also other possibilities that cannot be fully excluded, such as the de-differentiation of an adenocarcinoma to a more progenitor-like cell state—with some cells subsequently adopting neuroendocrine features due to local effects.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. All BAM files and associated sample information are described in Supplementary Table 11; data are deposited in dbGap phs000909.v1p1 and are accessible on the cBIO Portal for Cancer Genomics. Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank our patients and their families for participation in this study. We would like to acknowledge R. McBurney, A. Armstrong and A. Bordenave for contributing samples and S.S. Chae for his technical assistance. H.B. is the Damon Runyon-Gordon Family Clinical Investigator and is supported (in part) by the Damon Runyon Cancer Research Foundation (award no. CI-67-13). This work was also supported by the Ann and William Bresnan Foundation (H.B. and D.M.N.), the Department of Defense (grant no. PC121341; H.B.), the US National Institutes of Health (NIH) (grant no. R01 CA116337 (H.B., F.D. and M.A.R.), R01CA157945 (S.V.), R01CA138857 (S.A.T.), 1K08CA138865 (E.M.V.A.), U54 HG003067 (L.A.G.), U01CA162148 (L.A.G.), SU01CA111275-09 (J.M.M., M.A.R. and F.D.), the Starr Cancer Consortium (H.B., L.A.G. and M.A.R.), the Italian Cancer Foundation (L.P.), the Novel Soldati Foundation (I.C.), the A. Alfred Taubman Medical Institute (S.A.T.), a Prostate Cancer Foundation Young Investigator Award (E.M.V.A.), the Associazione Italiana per la Ricerca sul Cancro (AIRC; grant no. IG 13562; F.D.), the European Research Council (consolidator grant (CoG) SPICE (Synthetic lethal phenotypes identification through cancer evolution analysis); F.D.) and the Prostate Cancer Foundation (H.B., S.A.T., M.A.R. and F.D.), H.B., J.M.M., S.A.T., D.M.N., S.T.T., E.M.V.A., O.E., A.S., L.A.G. and 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 that is administered by the American Association for Cancer Research (grant no. SU2C-AACR-DT0712).

AUTHOR CONTRIBUTIONS

H.B., M.A.R., L.A.G. and F.D. initiated and designed the study; H.B., S.A.T., D.M.N. and S.T.T. enrolled subjects and contributed samples and clinical data; J.M.M., L.P., J.C., C.M., B.V.S.K.C. and S.V. performed experiments; D.P., M.B., E.G., E.M.V.A., O.E., A.S. and F.D. did the statistical and bioinformatics analyses; H.B., M.A.R., L.A.G. and F.D. supervised the research; H.B., M.A.R., L.A.G. and F.D. wrote the first draft of the manuscript; and all authors contributed to the writing and editing of the revised manuscript, and approved the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics, 2015. CA Cancer J. Clin. 65, 5–29 (2015).
2. Huggins, C. & Hodges, C.V. 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. J. Urol. 168, 9–12 (2002).
3. Chen, C.D. et al. Neuroendocrine prostate cancer and Synergy of p53 and Rb deficiency in a conditional mouse model identifies a novel class of HSP90 pathway modulators. Cancer Cell 10, 23–34 (2006).
4. Epstein, J.I. et al. Proposed morphologic classification of prostate cancer with neuroendocrine differentiation. Am. J. Surg. Pathol. 38, 756–767 (2014).
5. Watson, P.A. et al. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. Nat. Rev. Cancer 15, 701–711 (2015).
6. Zhang, Z. et al. Synergy of p53 and Rb deficiency in a conditional mouse model for metastatic prostate cancer. Cancer Res. 66, 7889–7898 (2006).
31. Lin, L. et al. Mechanisms underlying cancer growth and apoptosis by DEK overexpression in colorectal cancer. PLoS One 9, e111260 (2014).
32. Nowell, PC. The clonal evolution of tumor cell populations. Science 194, 23–28 (1976).
33. Maley, C.C. et al. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. Nat. Genet. 38, 468–473 (2006).
34. Hansel, D.E. et al. Shared TP53 gene mutation in morphologically and phenotypically distinct concurrent primary small-cell neuroendocrine carcinoma and adenocarcinoma of the prostate. Prostate 69, 603–609 (2009).
35. Gundem, G. et al. The evolutionary history of lethal metastatic prostate cancer. Nature 520, 353–357 (2015).
36. Smith, B.A. et al. A basal stem cell signature identifies aggressive prostate cancer phenotypes. Proc. Natl. Acad. Sci. USA 112, E6544–E6552 (2015).
37. Steffan, J.J. et al. The transcription factor SPDEF suppresses prostate tumor metastasis. J. Biol. Chem. 287, 29968–29978 (2012).
38. Börro, S.T. et al. Genome-wide DNA methylation events in TMPRSS2-ERG fusion-negative prostate cancers implicate an EZH2-dependent mechanism with miR-26a hypermethylation. Cancer Discov. 2, 1024–1035 (2012).
39. Varambally, S. et al. Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. Science 322, 1695–1699 (2008).
40. Yu, J. et al. A polycomb repression signature in metastatic prostate cancer predicts cancer outcome. Cancer Res. 67, 10657–10663 (2007).
41. Kanduri, M. et al. The placental gene PEG10 promotes progression of neuroendocrine prostate cancer. Cell Rep. 12, 922–936 (2015).
42. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. Cell 163, 1011–1025 (2015).
43. Carreira, S. et al. The estrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. Nat. Commun. 5, 5383 (2014).
44. Cheng, A.S. et al. EZH2-mediated concordant repression of Wnt antagonists promotes β-catenin-dependent hepatocarcinogenesis. Cancer Res. 71, 4028–4039 (2011).
45. Akamatsu, S. et al. The transcription factor SPDEF suppresses prostate tumor metastasis. J. Biol. Chem. 287, 29968–29978 (2012).
46. Maley, C.C. et al. AR-V7 and resistance to enzalutamide and abiraterone in advanced prostate cancer. Sci. Transl. Med. 6, 254ra125 (2014).
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 #1210013164) or retrospectively (IRB #40905010441), and germline (normal) DNA was 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 blinded or randomization was used. Tissues were processed as previously described⁴⁷,⁴⁹. All hematoxylin and eosin–stained slides were reviewed by board-certified pathologists (J.M.M. and M.A.R.). Tumors were classified based on histomorphology as adenocarcinoma (A) or CRPC-NE (B–E) based on a published pathologic classification system⁸ (Supplementary Fig. 1).

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 and IRB #1210013164) or retrospectively (IRB #40905010441), and germline (normal) DNA was 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 blinded or randomization was used. Tissues were processed as previously described⁴⁷,⁴⁹. All hematoxylin and eosin–stained slides were reviewed by board-certified pathologists (J.M.M. and M.A.R.). Tumors were classified based on histomorphology as adenocarcinoma (A) or CRPC-NE (B–E) based on a published pathologic classification system⁸ (Supplementary Table 1 and Supplementary Fig. 1).

Immunohistochemistry. Immunohistochemistry (IHC) was performed using a Bond III automated immunostainer (Leica Microsystems, IL, USA) with the following antibodies and dilutions: anti-AR (MU256-UC, clone F39.4.1, BioGenex, CA, USA; dilution 1:800 with casein), anti-synaptophysin (RM-9111-S, clone SP11, Thermo Scientific; dilution 1:100), anti–chromogranin A (MU126-UC, clone LK2H10, BioGenex, CA, USA; dilution 1:400), anti–CD56 (NCL-SDS56-504, clone CD56a, Leica Biosystems, IL, USA; dilution 1:50), anti–PSA (MU014-UC, clone ErPr-8, BioGenex, CA, USA; dilution 1:400), anti–PSA (MU013-UC, clone BGX013A, BioGenex, CA, USA; dilution 1:250), anti–PSA (M3620, clone 3E6, Dako, CA, USA; dilution 1:100), anti–Ki67 (M7240, clone MIB-1, Dako, CA, USA; dilution 1:50), anti–ERG (ab29513, clone EPR8384, Abcam, MA, USA; dilution 1:100), anti–EZH2 (612667, clone 11/EZH2, BD Biosciences, CA, USA; dilution 1:20), anti–MLH1 (554073, clone G168-728, BD Biosciences, CA, USA; dilution 1:400), anti–PMS2 (556415, clone A16-4, BD Biosciences; dilution 1:100), anti–MSH2 (NA27, clone FE11, Calbiochem, CA, USA; dilution 1:100) and anti–MSH6 (610919, clone 44/MSH6, 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 pH 6 retrieval for anti-ERG, anti–PSA and anti–PSMA; pressure cooker pH 6 retrieval for anti–EZH2; no retrieval for anti–chromogranin A and anti–PSAP; and heat-mediated pH 9 retrieval for all of the other antibodies. Study pathologists performed semi-quantitative evaluation for protein expression in nuclear (AR, ERG, MLH1, MSH2, MSH6 and PMS2), cytoplasmic (synaptophysin and 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 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 fresh-frozen paraffin-embedded (FFPE) tissue blocks and examined by the study pathologists to select high-density cancer foci and ensure high purity of cancer DNA. We used previously developed protocols⁴⁷ that were successfully used in our earlier genomic studies.²⁰,²³,²⁴,⁴⁹ Following this protocol, each case was quantified for tumor purity and annotated for discrete areas of macromissection, avoiding regions of necrosis or high stromal content. All cases were also quantified for tumor purity using an algorithm called CLONET.²³,²⁶ 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 single-nucleotide polymorphism (SNP) genotype data from whole-exome sequence coverage to quantify the percentage of reads supporting the considered aberration. The resultant tumor purity values were used to adjust the genomic data for downstream processing and analysis.

DNA extraction, tumor purity, and exome sequencing. Slides were cut from frozen or fresh-frozen paraffin-embedded (FFPE) tissue blocks and examined by the study pathologists to select high-density cancer foci and ensure high purity of cancer DNA. We used previously developed protocols⁴⁷ that were successfully used in our earlier genomic studies.²⁰,²³,²⁴,⁴⁹ Following this protocol, each case was quantified for tumor purity and annotated for discrete areas of macromissection, avoiding regions of necrosis or high stromal content. All cases were also quantified for tumor purity using an algorithm called CLONET.²³,²⁶ 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 single-nucleotide polymorphism (SNP) genotype data from whole-exome sequence coverage to quantify the percentage of reads supporting the considered aberration. The resultant tumor purity values were used to adjust the genomic data for downstream processing and analysis.

Extraction and sequencing were performed as previously described.⁴⁸,⁵¹ Briefly, DNA was extracted using Promega Maxwell 16 MDx. DNA was stored at −20 °C. Whole-exome capture libraries were constructed from tumor and normal tissue after sample shearing, end repair, and phosphorylation and ligation to barcoded sequencing adapters. Ligated DNA was size-selected for lengths between 200 and 350 bp and subjected to either exonic hybrid capture using SureSelect v2/v4 Exome bait (Agilent) or HaloPlex Exome (Agilent) (Supplementary Tables 11 and 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 100× for the tumor and germline samples.

Sequence data processing and quality control. A fraction of study samples (n = 78) were preprocessed with the analytical pipeline of the Englander Institute for Precision Medicine at Weill Cornell–New York Presbyterian Hospital (IPM-Exome-pipeline v0.9)⁴⁹. FastQC was run on the raw reads to assess the quality of the raw reads as previously described⁴⁸. The remaining 36 samples were processed using the Broad Institute Firehose infrastructure (http://www.broadinstitute.org/cancer/cga/Firehose).⁴⁹ Cross-contamination between samples from other individual subjects sequenced in the same flow cell was monitored with the ContEst algorithm.²³,²⁶ Normal-tumor pairs were checked for consistency using SNP panel identification assay (SPIA)⁵³. Processing pipelines returned segmented files for somatic copy-number aberrations (SCNAs) (Supplementary Table 6). No differences were observed in the results from 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, based on Bayesian statistical analysis to nominate putative SNVs upon coverage, allelic fraction, and base-qualities extraction. Next, we used Oncotator to annotate point mutations with variant- and gene-centric information relevant to cancer, and lastly we used MutSig 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-values 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.²⁶ 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 SCNA’s log2 ratio distribution, which is reflective of an aneuploid genome, and is then used to 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)⁵⁷ 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 less than −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 allelic fractions (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 is 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 also provides a space transformation from the beta versus log2 ratio to the cnB versus cnA (Supplementary Note and Supplementary Fig. 14), where each segment is visualized at coordinates representing the number of copies of allele A (cnA) and allele B (cnB). In the absence of parental allelic information, we consider cnA ≥ cnB. Noninteger 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 versus 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 described46,47. 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 2 × 75-bp 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; http://hgdownload.cse.ucsc.edu/downloads.html#human; major haplotypes only) using STAR v2.3.0e (ref. 58), 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 analysis59. 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) were estimated by counting all nucleotides mapped to each gene and were normalized by the total number of mapped nucleotides (per million) and the gene length (per kb). Differential expression analysis was performed using the Mann-Whitney Wilcoxon test after transforming the FPKMs via log2(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 + s.d.) < 1 FPKM were discarded.

NanoString assay. We employed a custom NanoString assay for cases without sufficient material for RNA-seq to evaluate the expression of AR-signaling genes, TPMRSS2-ERG fusion transcript, and neuroendocrine-associated genes. FFPE samples were cut into 10-μm-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 the total RNA concentration and the percentage of RNA >300 nucleotides (nt) in length. For samples with more than 50% of total RNA >300 nt, 100 ng input RNA was used; for samples with less than 50% of total RNA >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 manufacturer’s directions. Briefly, total RNA was hybridized overnight at 65 °C, then run on the Prep Station at maximum 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 six 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 were defined using the thresholds on the log2 ratio as described above. The relative frequencies of deletions and amplifications in CRPC-NE and CRPC-Adeno specimens are compared by means of binomial test adjusted for multiple-hypotheses testing with the 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 P value and FDR correction for 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) that were previously reported as defining the pathway52. For each specimen with expression levels available, either from RNA-seq or Nanostring analysis, 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 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. To detect ERG rearrangement status we used several assays, including fluorescence in situ hybridization (FISH) break-apart assay and/or immunohistochemistry (IHC) as previously described46. 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 the DNA, mRNA or protein level, the sample was considered positive. For other noncanonical and canonical fusion transcript detection, FusionSeq was used as previously described61.

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 (RP11-3272F22) plus a reference probe spanning a stable region of the chromosome (RP11-488120). All clones were tested on metaphase spreads62,63. CYLD deletion was defined by the absence at least one copy, on average, per nucleus, as 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 Medicine Epigenomics Core Facility as previously described63 and included 18 CRPC-Adeno samples from 10 individuals and 10 CRPC-NE samples from 8 individuals. In brief, the preparation steps included: 1) MspI enzyme digestion; 2) end repair of digested DNA; 3) adenylation; 4) adaptor ligation, with pre-anneled 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 the 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 (log2; ratio ≤−1), as inferred from the corresponding WES data (Supplementary Table 6), differentially methylated sites (CRPC-NE versus CRPC-Adeno) were identified by methylKit65. Differentially methylated genes were identified by annotating, using BedTools65, differentially methylated regions on gene promoters (defined as ≥2-kb genomic regions upstream of the set of 5′ gene coordinates), first exons, gene bodies, CpG islands (using the University of California Santa Cruz (UCSC) table browser) and CpG shores (2-kb genomic regions upstream and downstream of CpG islands). GENCODE v19 was used as a gene set. ToppFun65 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 SDPEF was performed using OneStep qMethyl PCR Kit per the manufacturer’s protocol (Zymo Research, Irvine, CA). 20 ng of genomic DNA from cell lines was used for qMethyl PCR. Primers used for qMethyl PCR were: Primer1 forward: 5′-CCGGTACATGCGTGTTC-3′, Primer1 reverse: 5′-ATTGCGGCTACACTCTCTG-3′, Primer2 forward: 5′-GATCTCTGCTTCCGCACTCTCTC-3′, Primer2 reverse: 5′-CCAGCCGCTCTCAGGCA-3′. Amplification parameters were: 45 cycles (denaturation: 95 °C, 30 s; annealing: 64 °C, 30 s; extension: 72 °C, 30 s).

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 the CRPC-NE vector based on a set of 70 genes (Supplementary Table 9 and Supplementary Figs. 10 and 15), using normalized FPKM values of the test sample. The gene set stems from the integration of differentially deleted or amplified and/or expressed and/or methylated genes in CRPC-NE and CRPC-Adeno. Specifically, 16 differentially deleted genes were selected from among putative cancer genes27 (see section ‘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 ≤ 0.1, resulting in a total of 2,425 genes, which correspond to 1,301 overexpressed and 1,124 underexpressed genes. For each gene, we performed a ‘receiver-operator curve’ (ROC) analysis using the normalized FPKMs as a threshold parameter and calculated the area under the curve (AUC). ROCs were built by considering only samples sequenced, excluding two samples (7,520 and 4,240) that were previously published6, leaving 34 CRPC-Adeno and 13 CRPC-NE samples. Only those differentially expressed genes with an AUC ≥ 0.95 and with a ≤ 0.05 change (± 0.05) were included in the classifier, resulting in a list of 49 genes (25 overexpressed and 24 underexpressed in CRPC-NE versus CRPC-Adeno), 21 of which were found as differentially methylated between CRPC-NE and CRPC-Adeno samples. Concordant information between RNA and methylation was found for 11 genes (see Supplementary Table 9).

In addition, we considered two genes (MYCN and AURKA) that we previously described as associated with a CRPC-NE phenotype5, EZH2 (FDR = 7.9 × 10−10) and DNM1L (FDR = 6.9 × 10−9) for their role in controlling DNA methylation67, and RB1 (FDR = 0.056), which was reported as a key driver in the pathogenesis of CRPC-NE16–18. As a result of each of the 70 genes, we corrected 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 data sets (Supplementary Table 10). RNA-seq data were processed as described above. Processed SU2C/PCF 2015 (ref. 20) and Grasso et al.19 (Michigan 2012) data were downloaded from cBioPortal68. Because data for four genes (ARHGAP8, BRINP1, C70orf76 and MAP10) were not available from cBioPortal, for Michigan 2012 we used a reduced version of the integrated NEPC score (indicated as ‘integrated NEPC score’). Samples with integrated NEPC scores ≥ 0.40 (elevated integrated NEPC score in the main text) were nominated as putative CRPC-NE tumors (Fig. 4c and Supplementary Table 14). In order to take into account the lower signal-to-noise ratio and the reduced version of the integrated NEPC score in the Michigan 2012 microarray data, in Figure 4d we consider those samples with an integrated NEPC Score ≥ 0.25 (significant integrated NEPC score in Fig. 4 legend) to be ‘CRPC-NE-like’. AR signaling and integrated NEPC score values for each sample are reported in Supplementary Table 15.

Cell line studies. LNCaP clone FGC, NCI-H660 cells and medium were purchased from the American Type Culture Collection (ATCC), and the cells were 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 medium containing 1 µM enzalutamide for approximately 6 months. A dose-response curve was made by plating 2.5 × 105 cells in 10 µl Matrigel (Corning) onto an Ibidi 96-well microplate and covering them with 40 µl of medium containing increasing doses of enzalutamide for 10 d (medium changed every 4 d). Viability was analyzed using the CellTiter-Blue assay (Promega); western blot analysis was performed with 50 µg protein, and staining was done with 1:1,000 anti-NCAM (Abcam, ab137086), 1:500 anti-SDPEF (Biorbyt orb13642), and 1:10,000 anti-GAPDH (Millipore, AB2302) antibodies. Site-directed CpG methylation of SDPEF was performed as described above. Cells were blocked in 5% BSA in PBS for 1 h on ice, then stained with 5 µg/ml Brilliant Violet 421–conjugated anti–human CD56 (Biologend) for 1 h on ice in the dark. Cells were washed twice with 8 ml PBS (8 min × 500g) and resuspended in FACS sorting buffer (1× PBS, 1 mM EDTA, 25 mM HEPES pH 7.0, 1% FCS (heat inactivated), 1% pluronic). 0.5 µg/ml DAPI was added immediately before sorting to stain dead cells. Cells were sorted on a BD FACSCalibur under the guidance of the Well Cornell Flow Cytometry core and analyzed with FloJo vX.0.7.

CLYD expression was silenced using ON-TARGET plus CYLD-specific short interfering RNA (siRNA) (Dharmacon L-004609-00-0005). ON-TARGET plus non-targeting pool (Dharmacon, D-001810-10-05) was used as a 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′-TTTGGGTTGGTGAAGAGATCAAT-3′ (forward), 5′-TCTCCGGTCAGACTCCTG-3′ (reverse). Human ACT7B (encoding β-actin): 5′-TCCCCTGAAGAGACTGACTG-3′ (forward) 5′-GATGTTTGGTGGAATGCCACA-3′ (reverse). CYLD relative expression was normalized to expression of ACT7B.

LNCaP, NCI-H660, and DU145 cells and medium were purchased from ATCC and used for GSK343 drug-treatment studies using escalating doses (5, 7.5, 10 µM). Cells were plated in Ibidi 96-well microplates embedded in 5 µl Matrigel. Matrigel droplets were then covered with 40 µl medium. After 48 h, medium was replaced with fresh medium containing increasing doses of GSK343 for 14 d (medium changed every 4 d) 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 analysis was performed for quantitative analysis of mRNA expression of the reported genes.

Statistical analysis. For statistical tests, we used the two-sided Mann-Whitney-Wilcoxon test (also referred to as Wilcoxon test in the main text) to check for significant shifts 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 Figure 5c uses a Student’s 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 the Benjamini-Hochberg procedure. No statistical methods were used to predetermine sample size.

NATURE MEDICINE doi:10.1038/nm.4045
49. Beltran, H. et al. Whole-exome sequencing of metastatic cancer and biomarkers of treatment response. *JAMA Oncol.* 1, 466–474 (2015).
50. Esgueva, R. et al. Next-generation prostate cancer biobanking: toward a processing protocol amenable for the International Cancer Genome Consortium. *Diagn. Mol. Pathol.* 21, 61–68 (2012).
51. Van Allen, E.M. et al. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat. Med.* 20, 682–688 (2014).
52. Cibulskis, K. et al. ContEst: estimating cross-contamination of human samples in next-generation sequencing data. *Bioinformatics* 27, 2601–2602 (2011).
53. Demichelis, F. et al. SNP panel identification assay (SPIA): a genetic-based assay for the identification of cell lines. *Nucleic Acids Res.* 36, 2446–2456 (2008).
54. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31, 213–219 (2013).
55. Lawrence, M.S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214–218 (2013).
56. Mosquera, J.M. et al. Concurrent AURKA and MYCN gene amplifications are harbingers of lethal treatment-related neuroendocrine prostate cancer. *Neoplasia* 15, 1–10 (2013).
57. Lin, P.C. et al. Epigenomic alterations in localized and advanced prostate cancer. *Neoplasia* 15, 373–383 (2013).
58. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21 (2013).
59. Habegger, L. et al. RSEQtools: a modular framework to analyze RNA-seq data using compact, anonymized data summaries. *Bioinformatics* 27, 281–283 (2011).
60. Park, K. et al. Antibody-based detection of ERG rearrangement–positive prostate cancer. *Neoplasia* 12, 590–598 (2010).
61. Sboner, A. et al. FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. *Genome Biol.* 11, R104 (2010).
62. Akalin, A. et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Nucleic Acids Res.* 13, R87 (2012).
63. Dale, R.K., Pedersen, B.S. & Quinlan, A.R. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics* 27, 3423–3424 (2011).
64. Chen, J., Bardes, E.E., Aronow, B.J. & Jegga, A.G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* 37, W305–W311 (2009).
65. Viré, E. et al. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 439, 871–874 (2006).
66. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2, 401–404 (2012).