Nascent Prostate Cancer Heterogeneity Drives Evolution and Resistance to Intense Hormonal Therapy


SUPPLEMENTARY MATERIAL

- Supplementary Analyses and Discussion
- Supplementary Methods and Materials
- Supplementary Figure Legends
- Supplementary Table Legends
- Supplementary References
SUPPLEMENTARY ANALYSES AND DISCUSSION

Impact of MRI estimates of tumor volume on residual tumor and diversity

Despite concerns that mpMRI can underestimate tumor volumes, we recently reported [1] that MRI baseline estimates of tumor burden prior to NADT, and more specifically the baseline tumor volume adjusted for total prostate size (relative tumor burden), could distinguish exceptional responders from incomplete and nonresponders with an AUC = 0.89 at a cutoff of 0.081. Indeed, positive correlations were observed between baseline tumor burden ($\rho = 0.69$; 95% C.I. 0.47 to 0.84; $p < 0.0001$), relative baseline tumor burden ($\rho = 0.76$; 95% C.I. 0.56 to 0.87; $p < 0.0001$), and final tumor burden ($\rho = 0.72$; 95% C.I. 0.50 to 0.85; $p < 0.0001$) with residual cancer burden (RCB) determined by pathologic examination of surgical specimens (Supplementary Fig. 7A-C). Although measures of genomic diversity correlated with RCB on final pathology, strong correlations of MRI volume tumor estimates were not observed with genomic Shannon diversity index (Supplementary Fig. 7D-F), although its weak positive association with final volume ($\rho = 0.14$; 95% C.I. -0.21 to 0.46; $p = 0.4$) is similar to its correlation with RCB (see Fig. 4F).

Nonetheless, this finding raised the possibility that tumor diversity is distinct from volume in a subset of cases, although both diversity and volume may independently inform on the aggressive properties of a tumor. When RCB was adjusted for baseline tumor volumes, ER and INR cases still exhibited clear separation (Supplementary Fig. 7G) and the adjusted RCB correlated well with RCB across the entire cohort ($\rho = 0.93$; 95% C.I. 0.86 to 0.96; $p < 0.0001$), but this was largely driven by the greater volumes of residual tumor, as ER cases exhibited more baseline volume variability. Thus, while greater initial volume was informative to predict poor response, lower genomic diversity was informative to predict exceptional responses. Tumor heterogeneity, as estimated by the number or MRI-visible lesions in each patient, was independent of outcome (see Supplementary Fig. 5C), and the difference we observed between the number of MRI-visible lesion and high versus low SDI was
not statistically significant (Supplementary Fig. 7H).

We therefore considered whether as an independent factor, baseline tumor volume could improve the fit of the four-factor histogenomic model for predicting response. We considered either the MRI-estimate of absolute baseline tumor burden (Supplementary Fig. 7I) or the prostate volume-normalized (relative) baseline tumor burden (see Fig. 5I) and observed an increase in the AUC to 0.97 and 0.98 from 0.89, respectively (see Fig. 5C). This improvement in model fit could be explained in part if the biological processes driving proliferation (i.e., larger tumors) are distinct but complementary to those processing conferring fitness. Indeed, the notion that heterogeneity (i.e., multiple MRI lesions) is not necessarily linked to tumor volume suggests that the propensity of an independent cell to undergo transformation and develop into a substantive tumor (i.e., initiation) is distinct from growth in the absence of a selective force. Consequently, the probabilities conferred by this model did not improve the association of INR probabilities with genomic SDI using either baseline or relative baseline volumes with coefficients of correlation of 0.22 and 0.18, respectively (Supplementary Fig. 7J-L), suggesting that tumor volume itself is a confounding factor for assessing diversity. However, a limitation of this 5-feature model is that our validation requires a baseline imaging component, which is not available in the prostate TCGA dataset.

Effect of NADT on biochemical recurrence

Although NADT trials have been conducted over the past decade, the long interval between standard-of-care therapy, recurrence, and cancer-specific mortality has made it challenging to determine the precise overall survival benefits of NADT [2, 3]. However, if AR-targeted therapies active in metastatic hormone sensitive prostate cancer are reducing intraprostatic tumor volumes in a subset of patients, a reasonable hypothesis is that these agents also act on occult micrometastatic disease that might otherwise progress later. Indeed, in a meta-analysis of three NADT trials [4-7], patients who exhibited MRD or pCR (defined as < 0.5 cm residual tumor in the largest cross-
sectional dimension) did not experience biochemical recurrence (BCR). However, in contrast to previous studies that had substantial numbers of patients with favorable-intermediate risk, our current study was enriched for unfavorable-intermediate and high-risk patients with greater chances of recurrence. Nonetheless, NADT would be expected to delay or prevent recurrence in at least a subset of the ER patients. Using the high-risk prostate TCGA cohort for which follow-up data was available, we compared the BCR-times by Kaplan-Meier analysis based on their predicted ER or INR status and did not observe a significant difference (Supplementary Fig. 8). This finding suggests that among higher risk patients, the four factors used in our model did not independently affect BCR, and by extension, that NADT had the potential to delay BCR in a subset of those patients.

**Genome-wide genomic analyses**

In addition to a pre-defined set of genomic alterations (see Fig. 1), we also performed an exploratory analysis to identify novel perturbations to the genome that were associated with response to NADT. We first performed a Fisher-Freeman-Halton test of the INR versus ER status of the 20 lesions from the nine patients with multiple lesions in the dataset (seven patients with two lesions each, and two patients with 3 lesions each). This test showed consistency with the null hypothesis of independent within-patient response outcomes in the lesions from these patients ($p = 0.4$), consistent with their sampling as independently contoured tumors on mpMRI. Consequently, all lesion-level analyses were performed considering each lesion independently.

Using segmented copy number values from each focus of tumor in our entire cohort, we identified a series of 180 peaks using GISTIC and called gain or loss values within each peak (Supplementary Table 7), of which 5 were enriched in the INR group ($p < 0.05$, Fisher’s exact test). In addition to peaks identified by this analysis (Supplementary Fig. 9A), which included a portion of chromosome 21 interstitially deleted in a subset of TMPRSS2-ERG fusions (q22.2) and loss at chromosome 10q21 encompassing PTEN, analysis at the lesion level (Supplementary Table 8)
identified a loss at chromosome 5q14.3-15 that was enriched in the ER group (Supplementary Fig. 9B). Interrogating the combined effect of somatic copy number changes and mutations, we examined at the patient- (Supplementary Table 9) and lesion-level (Supplementary Table 10) whether any alteration to a series of 712 cancer-related genes significantly altered in the ER or INR group. As shown in Supplementary Figure 9C, these univariate analyses revealed that mutations and copy number alterations to the hedgehog pathway receptor \(\text{PTCH1}\) were only observed in INR cases \((p = 0.0009, \text{Fisher’s exact test})\) and were also enriched in INR lesions (Supplementary Fig. 9D). Other genes exclusively altered in INR included the tumor suppressor \(\text{FANCC}\) and the oncogene \(\text{MYCN}\). As the majority of these alterations included shallow gains or deletions, we repeated this analysis to identify any gene enriched with biallelic inactivations, either by two-copy deletion or mutation plus loss of heterozygosity. Four genes satisfied these criteria at the patient level, all enriched in ER cases (Supplementary Fig. 9E), which included \(\text{MAP3K7}\) and the oncogene \(\text{FYN}\) \((p = 0.016 \text{ and } 0.042, \text{respectively})\); both are located in a recurrently-deleted region of chromosome 6q15-21, raising the fascinating possibility that alterations which contribute to tumorigenesis may sensitize the tumor to NADT, similar to the protective effects of null oncogene mutations [8]. \(\text{FOXO1}\) deep deletions were also enriched in ER, both at the patient and lesion level (Supplementary Fig. 9F). The lesion-level analysis also identified homozygous deletions or hemizygous loss plus point mutations of \(\text{NOTCH2}\) exclusively in the INR group \((p = 0.046)\).

Identifying ER lesions in INR patients increased our ability to discover somatic alterations that may sensitize tumors to NADT. These lesion-level analyses also identified in ER lesions recurrent deletions to chromosome 5q14, which harbors \(\text{NR2F1}\), a nuclear receptor associated with tumor cell dormancy [9]; tumors with functioning NR2F1 may potentially evade NADT while deletion of that locus is sensitizing. Although these significance levels of these alterations may not be conclusive due to multiple testing, further \textit{in vitro} and \textit{in vivo} functional validation of these findings may shed light on mechanisms of exceptional response.
Germline SNP analyses

To explore potential relationships between individual responses to intense neoadjuvant ADT and background genetics, we performed whole-genome sequencing on either buffy coat- or saliva-derived genomic DNA. Our analysis of inherited loss-of-function alleles to cancer-related genes resulted in the identification of three patients with deleterious point mutations to either BRCA1, BRCA2, or PALB2. The implication of these mutations in distinguishing responders from nonresponders is a component of our primary analyses.

In addition to coding mutations, we profiled a series of 146 risk SNPs previously identified by GWAS projects as having significant associations with prostate cancer (Supplementary Table 11). 6 individual SNPs (rs12621278, rs142436749, rs12785905, rs80130819, rs75823044, and rs11863709) exhibited mutual exclusivity between ER and INR cases, but these did not reach statistical significance by univariate (two-sided Fisher’s exact test) analysis. By two-sided Fisher’s exact test, 3 SNPs exhibited enrichment in ER cases at $p < 0.05$: rs12500426, rs17021918 and rs11568818. Interestingly, no SNPs were individually enriched in INR cases, suggesting that these polymorphisms previously identified by increased prostate cancer risk did not interact with response to intense neoadjuvant ADT.

We next considered minimal multi-factor models, i.e. all possible pairs of SNPs. First, using logistic regression, we identified two pairs of SNPs fitted to models based on presence or absence of the SNP allele in each pair (i.e. 0/0, 0/1, 1/0, or 1/1). The (rs12500426, rs6062509) pair was significant at $p = 0.0005$ by likelihood ratio test and $p = 0.0052$ by the score test. The individual Fisher’s exact test results from these SNPs are $p = 0.028$ and $p = 0.059$, respectively, which were among the lowest of that set. The second pair of SNPs identified by this analysis was (rs2660753, rs6062509) at $p = 0.0018$ by the likelihood ratio test.

Finally, using the Fisher-Freeman-Halton test, we identified a third pair of potentially interesting...
SNPs (rs17599629, rs103294). The null hypothesis of equal probabilities of ER in the four categories was rejected at the $p = 0.00004$ level, but this comes from fractions of 0/12 for the (0/0) combination, 0/3 for the (1/1) combination, 7/8 for the (0/1) combination and 8/14 for the (1/0) combination. Remarkably, all of the ER cases were carriers for either of these SNPs, mutually exclusive of each other. Thus, for this pair of SNPs, having neither SNP nor both SNPs provided an estimated zero probability of ER, but having exactly 1 of the SNPs provides a probability estimated to be greater than 50% for ER.

Length of the AR polyglutamine tract

We investigated whether the length of the AR polyglutamine (PolyQ) tract (CAG repeats) was associated with response to intense neoadjuvant ADT, given its inverse association with risk of developing prostate cancer [10, 11]. Taking advantage of the longer read length from whole-genome sequencing (150 nucleotides paired-end), we employed bioinformatic tools to determine the length of the PolyQ repeat for each patient. As depicted in Supplementary Table 12, there was no significant difference between the PolyQ length between ER and INR cases. For both groups, the median CAG repeat length was 21, ranging from 16-28 for ER and 18-26 for INR. Consequently, the PolyQ length is likely not associated with response to treatment in this study.

Histology for PTEN

As depicted in Figure 1C, loss of function mutations (including copy number losses) to PTEN were enriched in INR cases (11/22) more than ER cases (4/15) but this did not reach statistical significance ($p = 0.19$) at the patient level. Similar trends were observed at the lesion level as well (Fig. 1D). In contrast, per-patient reductions in PTEN by immunohistochemistry (Supplementary Fig. 10A) were enriched in INR versus ER cases at $p = 0.0015$ (Fig. 3A). As shown in Supplementary Figure 10B, among PTEN reduced cases, approximately 54% harbored somatic copy
number losses to PTEN. In contrast, approximately 85% of cases without reductions in PTEN by IHC did not show copy number reductions to PTEN either, resulting in significant enrichment ($p = 0.035$) for PTEN copy number loss when reductions to PTEN were visualized by IHC. Note that we do not attempt to distinguish between hemizygous deletion or epigenetic silencing of PTEN from homozygous deletion by IHC. However, in cases where homozygous deletions of PTEN were determined by somatic copy number analysis, 100% of those cases also exhibited reduced tumor expression of PTEN by IHC (Supplementary Fig. 10B). However, PTEN levels by IHC were unable to distinguish those cases with loss of chromosome 10q, meaning that derivation of this factor for the histogenomic or historadiogenomic model would require a DNA-based approach rather than IHC for PTEN (Supplementary Fig. 10B).

Considering the differences between ER and INR cases, the combination of IHC and/or genomic detection of PTEN alterations did not further improve the identification of INR cases based on IHC (19 out of 22, or 86%), because within the INR group the 3 cases that exhibited intact PTEN by IHC did not harbor alterations to PTEN (Supplementary Fig. 10C). In addition, using concordance for both genomic alterations to PTEN and reductions of PTEN levels by IHC did not improve the sensitivity of identifying INR cases over genomics alone (Supplementary Fig. 10D). Consequently, only IHC for PTEN remains the most sensitive approach for identifying INR cases based on PTEN status despite a 33% false-positive rate if used to identify ER cases (see Fig. 3A).

**Histology for p53**

Immunostaining of pre-treatment biopsies with anti-p53 antibodies was performed in order to examine the relationship between a staining protocol validated for the identification of gain-of-function (hotspot) TP53 mutations and pathologic response after treatment. Slides were scored in a semiquantitative fashion, ranging from 0 to 3 to assess intensity, or using a dichotomous score of 0 or 1 for predicting effects on p53 function based on nuclear accumulation (see Supplementary Table...
As shown in Supplementary Figure 11A-B, the increase in staining intensity did not associate with ER or INR outcome \((p = 0.8)\). Interestingly, ER cases did show a greater proportion of scores equal to 1 (Supplementary Fig. 11B), suggesting modest baseline expression of p53, while extreme scores of 0 (no expression) or 3 (very high expression) were enriched in INR. To determine if these scores were related to particular genomic features, we assessed multiple combinations of scores and genetic alterations. As shown in Supplementary Figure 11C, semiquantitative scores of 1 were neither enriched in the absence of any alterations nor when there were compound mutations that had potentially competing effects (such as a gain-of-function mutation on one allele and deletion of the other allele). In contrast, cases with semiquantitative scores of 0 were modestly enriched, proportionally, for somatic copy number losses and truncating mutations that would otherwise result in loss-of-function (Supplementary Fig. 11D).

Finally, as a dichotomous score of 1 would be calibrated to alterations resulting in gain-of-function to p53 (Supplementary Fig. 11E), we explored the relationship between missense and other in-frame alterations to \(TP53\) with strong nuclear accumulation. Cases scored as 1 were more strongly enriched for nontruncated mutations versus those scored 0, although our multiregion sequencing efforts observed missense mutations, including classical \(TP53\) hotspot SNVs, in cases scored as 0 showing little or no staining \((p = 0.067)\). Scores of 1 did not show enrichment for INR cases (Supplementary Fig. 11F, \(P > 0.9\)). As neither semiquantitative scoring nor dichotomous scoring were strongly associated with hotspot or loss-of-function alterations to \(TP53\) that are considered in the histogenomic or historadiogenomic model, a DNA-based approach rather than IHC for p53 is needed.

Relationships of histologic and molecular features to treatment resistance

By comparing the histologic intensity of IHC in this cohort of patients to pathologic outcome, we found that PSA was inversely proportional, while GR was directly proportional. The fact that PSA
levels tended to be uniformly high yet were heterogeneously reduced in INR cases is consistent with the hypothesis that AR-high tumors are more likely to respond to AR-directed therapy [12, 13], and that lower PSA staining is representative of an AR-low subpopulation. Indeed, in situ PSA levels were reduced by treatment [1], with poor responders a priori less dependent upon AR. This finding is further supported by the observation that baseline GR histologic intensity increased with RCB post-treatment. Nuclear GR expressed in prostate cancer luminal cells is lower than in normal epithelium but is elevated in CRPC as one of several bypass mechanisms to reconstitute AR activity in the absence of bound ligand [14, 15]. In contrast to PSA, GR levels are uniformly low in our cohort but exhibit heterogeneity in INR cases, potentially priming those tumors for resistance to NADT. Interestingly, with both stains, heterogeneity increased proportionally with RCB, and SDI’s calculated based on subpopulations of stained tumor cells also indicated greater in situ diversity. Thus by per-patient analysis, these direct phenotypes of prostate cancer offered visual and microscopic interpretation of subclonal diversity.

Genomically, our approach to sample tumors comprehensively also increased our sensitivity to identify alterations contributing to exceptional response. At the patient level, mutations or copy number changes exclusively found in ER cases included alterations to SPOP, which have been previously associated with response to ADT, are found with less frequency in mCRPC patients, and indicate a more favorable prognosis by conventional therapies [16, 17]. Our analysis also identified recurrent alterations that were enriched in INR patients. These events, which include the TMPRSS2:ERG fusion and loss of PTEN, are common in prostate cancer and have been observed more frequently in NADT-resistant tumors [18, 19]. Interestingly, the co-occurrence of ERG and PTEN heterogeneity (i.e., tumor cells with positive and negative staining for both markers) was exclusively in INR patients. We further selected the histologic feature intraductal carcinoma (IDC-P) for use in our four-factor model also based on its strong enrichment in INR tumors. Recently, IDC-P has become increasingly appreciated for its association with poor overall outcome [20, 21], and
deeper investigations exhibited relationships between IDC-P, hypoxia, and genomic instability [22, 23]. Indeed, PGA was also proportionally increased with both genomic SDI and the volume of treatment-resistant tumors. Incidentally, IDC-P was the only feature distinctly subclonal to INR cases, whereas chromosome 10q loss, TMPRSS2:ERG fusion and TP53 alterations were generally clonal and thus happened early in the tumor’s history, explaining their increased incidence in ER samples by lesion-level analyses. Nonetheless, among cases with any combination of the four factors, they most frequently were observed in the same biopsy block (see Supplementary Table 1), simplifying potential translation of this model into clinical practice.
SUPPLEMENTARY METHODS AND MATERIALS

Patient and biopsy selection

All patients who completed our recent clinical trial [1] of neoadjuvant ADT plus enzalutamide were included in this study. Study participants were predominantly clinical stage T3 and T4 with biopsies in ISUP Grade Groups 4 and 5 (see Table 1), although a subset of patients had biopsies in ISUP Grade Groups of 2 or 3. Collectively, this is a large group of patients who have heterogeneous disease with variable prognoses. This heterogeneity makes development of treatment approaches clinically challenging secondary to the wide range of range of prostate cancer specific mortality, and biochemical or clinical recurrence (ranging from 2% to 70%) following treatment with radical prostatectomy or radiation. As such, ISUP Grade Group 2 and 3 patients, especially those in stage cT3, need further differentiation for tumors that are more indolent versus those that are more aggressive, particularly in the context of mpMRI imaging. Pretreatment clinical staging was determined based on DRE and in some cases, transrectal ultrasound of the prostate, and/or MRI. The TNM categories combined with the Gleason score and serum PSA levels were used to determine the overall stage. Patients with distant metastatic disease beyond N1 (regional) lymph nodes on conventional imaging studies (CT, MRI, or bone scan) were excluded.

From 37 patients, 62 mpMRI-visible lesions were identified and sampled by biopsy (up to 4 lesions per patient). For larger lesions, additional sampling by biopsy of the lesion was performed to account for potential heterogeneity. Of the 62 sampled lesions, 9 did not have any tumor and were excluded from further analysis. Five lesions had less than 5mm of visible tumor on biopsy and were also not considered for molecular analysis. One of these lesions (lesion L11) had sufficient tumor (4.5 mm) for limited histologic analysis. A total of 48 lesions were subjected to combined histologic and molecular analysis. Generally, more than one biopsy was selected for a lesion if it contained large volumes of tumor conducive to laser capture microdissection or if multiple biopsies from that
lesion exhibited varying histologic phenotypes or adverse pathologies, such as Gleason score or perineural invasion. Using maps created from the RP surgical samples, we verified that every major region of residual tumor was in the same anatomical region as a biopsy acquired prior to treatment.

Treatment response

Out of 39 total enrolled patients, two patients did not complete the study and are censored from analysis. One of the remaining 37 patients underwent transurethral resection of the prostate (TURP) due to extension of the prostate tumor to the bladder, and consequently final pathology volumes for that patient are censored despite his classification as a nonresponder to therapy. Thus, at the conclusion of therapy, 36 subjects underwent radical prostatectomy (RP). The RP specimen was grossed and sectioned in the same plane as the mpMRI, using a customized 3-D printed mold designed using the most recent post-treatment mpMRI [24]. Surgical specimens were stained with H&E and additional immunostains to verify the presence of residual tumor, including anti-NKX 3.1, PIN-4 cocktail, CAM5.2, and anti-p63 [25, 26]. Determination of residual disease was performed by three genitourinary pathologists (M.J.M., H.Y., and R.T.L.). Treatment response was determined by the volume of residual cancer burden (RCB), which was calculated by multiplying the number of tissue slices containing residual tumor by the largest cross-sectional width and length of that tissue and by block thickness (0.6 cm). Volume was further corrected by multiplying by 0.4 to account for tumor cellularity.

We considered RCB < 0.05 cm³ as exceptional responders (ER) and we grouped incomplete and nonresponders (INR) into a single category for RCB > 0.05 cm³. Each mpMRI-visible lesion was similarly classified based on the volume of evaluable tumor remaining in the anatomical region of the lesion. If a patient had more than one residual lesion at final pathology, the largest cross-sectional dimension of tumor was used to determine RCB for the purpose of classifying the patient as ER or INR. By definition, an ER case meant that all MRI-visible lesions responded. At the patient
level, a subset of INR tumors exhibited heterogeneity because they included lesions that responded as well as lesions that persisted [1], with variability associated with intrapatient response characteristics. This approach resulted in more ER lesions than ER patients, as non-index responding lesions were harbored by INR patients.

**Histology**

5-micron serial sections of formalin-fixed paraffin-embedded (FFPE) biopsy tissues were stained with H&E, stained with antibodies by immunohistochemistry, or cut onto polyethylene naphthalate membrane slides (MicroDissect GmbH) for laser capture microdissection (LCM). For LCM, additional glass slides were cut after every five membrane slides for additional H&E and IHC stains to serve as references.

For IHC, slides were stained using validated protocols with antibodies against androgen receptor (AR), prostate specific antigen (PSA), glucocorticoid receptor (GR), ERG, PTEN, synaptophysin (SYP), PIN-4, and Ki-67 on an IP FLX autostainer (Biocare Medical). Glass slides containing tissue sections were baked for 30 minutes at 60°C. Following deparaffinization in xylenes and rehydration through graded alcohols, antigen retrieval was performed using a NxGen Decloaker (Biocare Medical) at 110°C for 15 minutes in Tris-EDTA Buffer (Abcam; ab93684), pH 9.0 for PSA, PTEN, and Synaptophysin or Diva Decloaker (BioCare Medical; DV2004MX) for AR, GR, ERG, PIN-4, and Ki-67. Next, a thin border was drawn around the edges of each glass slide using a PAP pen. 300 μL of primary antibody solutions were prepared and incubated at room temperature as follows: anti-AR clone D6F11 (Cell Signaling; 5153) diluted 1:200 into Renoir Red diluent (Biocare; PD904) for 30 m, anti-PSA (DAKO; IR514) ready-to-use (RTU) for 30 m, anti-SYP (Dako; M7315) diluted 1:200 into Renoir Red for 30 m, PIN-4 cocktail (CK5/14 + p63 + P504S) (Biocare Medical; PPM225DSH) RTU for 1 h, anti-PTEN clone D4.3 (Cell Signaling; 9188) diluted 1:100 into Renoir Red for 1 h, anti-Ki-67 clone D2H10 (Cell Signaling; 9027) diluted 1:500 into Cell SignalStain
diluent (Cell Signaling; 8112) for 30 m, anti-GR clone D6H2L (Cell Signaling; 12041) diluted 1:400 into Renoir Red for 30 m, or anti-ERG clone EPR3864 (Abcam; ab92513) diluted 1:200 into SignalStain for 30 m. Secondary detection was achieved with Mach 2 (for PIN-4) (Biocare Medical; MRCT525) or Mach 4 (Biocare Medical; M4U534H) polymer and/or probe (all other antibodies) for 30 minutes. Chromogen development was achieved with either Vulcan Red Fast Chromogen (Biocare Medical; FR805) or Betazoid DAB (Biocare Medical; BDB2004) and counterstained with CAT hematoxylin (Biocare Medical; CATHE) diluted 1:2 into distilled water. PIN-4 stained slides were air dried; all other slides were dehydrated through graded alcohols into xylene. Slides were mounted using Permount (Thermo Fisher).

For each stain, internal or external controls were employed to validate staining. For anti-AR, nuclear staining of luminal cells in normal prostatic tissue were positive controls, while ADT-treated tissue was used as a negative control. For anti-PSA, normal prostatic tissue was the positive control and non-prostate tissue was negative control. For anti-PTEN, normal tissue was used as the positive control and tissue previously confirmed to be genomically deleted for PTEN was the negative control. For anti-GR, endogenous staining in normal prostatic tissue was the positive control and FFPE-embedded LREX GR-knockout cells (kind gift of Charles Sawyers and David Wise) were negative controls for GR. For anti-SYP, pancreatic tissue (sourced from the NCI Laboratory of Pathology) was used as the positive control and PSA-positive prostate adenocarcinoma was the negative control. For anti-Ki-67, lymph node tissue invaded by MYC-amplified prostate cancer was used as the positive control and normal prostate luminal cells were the negative controls. For anti-ERG, normal prostate luminal cells were the negative control, and endogenous staining by endothelial cells was used as the positive control.

Calls of PTEN reduction at the focus or patient level were made on the basis of least 5% of cancer cells showing reduced levels of PTEN intensity relative to normal glands or stromal cells in the same tissue section. Calls of nuclear ERG expression at the focus or patient level were made on
the basis of any cancer cell exhibiting positive nuclear ERG staining.

After mounted slides were fully dried, residual mounting media was removed using xylene. Slides were then digitized on Carl Zeiss AxioScan.Z1 microscope slide scanner equipped with a Plan-Apochromat 20× NA 0.8 objective, 266% LED intensity, 200 μs exposure time. Tissue images were acquired using ZEN Blue 2012 (Zeiss) with objective/magnification and pixel:distance calibrations recorded within the scanned CZI file. Digitized slides have been deposited in the Cancer Imaging Archive (https://www.cancerimagingarchive.net/).

*Laser capture microdissection*

Areas of interest on one or more biopsy tissues were identified by two genitourinary pathologists (R.T.L. and H.Y.) based on concordance or discordance with diagnostic pathology reports from each patient, the size of the tumor lesion estimated by mpMRI, adverse pathologic features identified on H&E tissues, or differential staining identified by IHC. Up to four areas per biopsy slide were annotated digitally on scanned H&E or IHC slides using ZEN Browser (Zeiss) and displayed on a second monitor for reference during LCM.

PEN-membrane slides were briefly baked, deparaffinized, rehydrated and stained with Paradise stain (Thermo Fisher) according to the manufacturer’s protocol. Approximately 10,000-50,000 tumor cells per ROI were captured from serial sections using an ArcturusXT Ti microscope onto CapSure Macro LCM Caps (Thermo Fisher) using the infrared capture and ultraviolet cutting lasers. Adjacent stromal tissue that was incidentally captured was ablated using the UV laser. Micrographs of each cap were taken after each LCM session and cross-referenced against reference slides to verify the regions captured. Normal DNA was acquired from benign areas not involved with cancer, either from biopsies previously determined to be tumor-free, or from the post-treatment (surgical) specimen. After LCM, each captured tumor area was re-reviewed by two blinded genitourinary pathologists (H.A.S. and R.T.L.) to determine the highest Gleason pattern in the captured region, the
percentage of the captured region corresponding to Gleason patterns 4 and 5, and whether the captured area contained ductal morphology, cribriform architecture, intraductal carcinoma of the prostate, diminished PTEN staining, or positive nuclear ERG staining.

**DNA extraction**

Whole blood collected in K2EDTA tubes was centrifuged at 22°C for 20 minutes at 300 × g. After removal of the plasma layer, the buffy coat interface was transferred to a new tube and frozen at -80°C. gDNA was extracted from 10-100 μL of buffy coat using the DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer’s protocol with two elutions of buffer AE (100 μL per elution). gDNA from 500 μL of preserved saliva was extracted using the prepIT L2P Kit (DNA Genotek) following the manufacturer’s instructions.

DNA was extracted from LCM tissues by using a clean scalpel to excise the LCM cap polymer and immerse it in Buffer ATL from the QIAamp DNA FFPE Tissue Kit (Qiagen). The initial proteinase K digestion step was performed overnight, and the remainder of the extraction protocol was followed according to the manufacturer’s instructions. DNA from blood, saliva and tissues were quantified using Picogreen reagent (Thermo Fisher).

**Sequencing and data processing**

20-200 ng of double-stranded genomic DNA from tumor or normal regions was sheared (Covaris) and prepared into libraries using the SureSelect Human All Exon V7 Low Input Exome kit (Agilent) for whole-exome sequencing (WES). Pooled libraries were sequenced with 150 cycles paired-end plus 8 indexing cycles on either a HiSeq 4000 (Illumina) or a NovaSeq 6000 (Illumina). 200 ng of double-stranded genomic DNA from blood or saliva was sheared (Covaris) and prepared into dual-barcoded libraries using the Illumina TruSeq Nano kit and sequenced with 150 cycles paired-end plus 16 dual-barcoding indexing cycles on a NovaSeq 6000 for whole-genome
sequencing (WGS). WES of microdissected foci was performed to a mean on-target depth of 52.1× unique coverage. Raw sequencing data has been deposited in the Database of Genotypes and Phenotypes (dbGaP) accession number phs001938.v2.p1.

FASTQ files from WES were filtered to remove read pairs flagged as failed by sequencer. Pass-filter reads were then trimmed using SureCall Trimmer version 4.0.1 (Agilent) and aligned at the lane level with the Burrows Wheeler Aligner BWA-MEM version 0.7.17 [27] to version hg19 of the human genome (b37 with decoy chromosomes). The SAM alignment files were coordinate-sorted and duplicate-marked using PICARD version 2.18.27 SortSam and MarkDuplicates then quality score recalibrated using version 4.1.3.0 of the Genome Analysis Toolkit (GATK) BaseRecalibrator and ApplyBQSR. Lane-level BAM files were merged using PICARD MergeSamFiles and duplicate-marked at the sample level again using PICARD. On-bait capture efficiencies and library complexity were determined using PICARD CalculateHsMetrics.

FASTQ files from WGS were filtered to remove read pairs flagged as failed by the sequencer. Pass-filter reads were then trimmed using Trimmomatic version 0.39 and aligned at the lane level with the Burrows Wheeler Aligner BWA-MEM version 0.7.17 to version hg19 of the human genome (b37 with decoy chromosomes). All post alignment steps were performed using the Spark hyperthreading implementation of GATK 4.1.3.0. The SAM alignment files were coordinate-sorted using GATK MarkDuplicatesSpark and tag-set with PICARD SetNmMdAndUqTags. Base quality score recalibration was performed using GATK BQSRPipelineSpark, and then all lane-level BAM files were merged using PICARD MergeSamFiles. GATK MarkDuplicatesSpark and PICARD SetNmMdAndUqTags was run a second time on the merged, sample-level BAM file. WGS coverage metrics and library complexity were determined using PICARD CalculateWgsMetrics. WGS was performed at a mean depth of 46.4× unique coverage.

Somatic variant calling
Somatic point mutations and indels were called on intervals from the Agilent bait design BED files using MuTect2 (part of the GATK4 package), first by running MuTect2 in tumor-only mode on all of the normal tissue BAM files individually, with the parameters disable-read-filter set to MateOnSameContigOrNoMappedMateReadFilter and max-mnp-distance set to 0. The “padded” bait BED file was used as the interval file of covered regions for all MuTect2 and filtering steps. Each of the output VCF files from this analysis of normal BAMs was merged into a database using GATK GenomicsDBImport with the setting merge-input-intervals set to true, and then generating a panel of normal from the database using GATK CreateSomaticPanelOfNormals. The WGS BAM files from buffy coat or saliva gDNA was subsetted to serve as a second matched normal using PICARD FilterSamReads with the Agilent bait design BED intervals file. MuTect2 was run in somatic mode on each tumor BAM paired with the normal WGS-subsetted BAM from the patient and the somatic panel of normals with af-of-alleles-not-in-resource set to 0.0000025 to exclude sites present in gnomAD and disable-read-filter set to MateOnSameContigOrNoMappedMateReadFilter. GetPileupSummaries and CalculateContamination were used on each tumor BAM file and the resultant contamination table was used to filter somatic mutations using FilterMutectCalls. CollectSequencingArtifactMetrics and FilterByOrientationBias were used to further filter mutations for 8-oxoG artifacts using the settings -AM G/T -AM C/T. These pass-filter mutations were then functionally annotated using Oncotator version 1.9.70 (database version April052016), while a second set of mutations that failed any of the filters were functionally annotated into a separate file.

The number of small nucleotide variants (SNVs) reported per tumor focus is the number of pass-filter, coding, non-synonymous mutations (i.e. missense, nonsense, frameshift, splice-site, in-frame deletion/insertion, start/stop-codon, and de novo start mutations) that are further filtered to allow for no more than 10% strand bias at each site.

With limited power of this cohort to discover new mutations, variants considered in unbiased univariate analyses were curated first by only selecting the pass-filter mutations that overlapped with
those deposited in COSMIC, and then further limiting those genes that also overlapped with a series of 727 cancer-related genes frequently mutated in prostate cancer and other human tumors [28, 29]. SNVs reported in large prostate cancer cohort studies [30, 31] were backfilled if not yet in COSMIC, and novel frameshifts were manually included if they were detected in >1 focus from cases where multiple foci were microdissected. SNVs and SCNAs considered in curated and pathway analyses were further classified as gain-of-function, loss-of-function, or hotspot mutations by manual interrogation in the Precision Oncology Knowledge Base (OncoKB, http://www.oncokb.org).

Underlying processes of mutational signatures were estimated from WES data using the deconstructSigs [32] module for R version 3.5, using the Nature 2013 [33] signatures as the reference. All pass-filter synonymous and non-synonymous mutations were considered for this analysis, provided that they had less than 10% strand bias, were covered by the tumor and normal sample by at least 16× coverage, and were greater than 2% variant allele fraction in the tumor.

Somatic copy number alterations (SCNAs) were called across genomic intervals specified by the Agilent library design BED file with variable resolution depending on bait spacing. The design BED file was preprocessed with GATK PreprocessIntervals, with bin-length set to 0 and interval-merging-rule set to OVERLAPPING_ONLY. The BED file was also annotated with GC content using AnnotateIntervals. These interval files were used in all copy number calling steps. Read counts were first obtained from all tumor and normal BAM files using GATK CollectReadCounts. The normal read count files were compiled into a panel of normal using GATK CreateReadCountPanelOfNormals, excluding cases that had read depth coverage outside the 95% confidence interval for the dataset. The panel of normals was used to smooth read counts across all samples using GATK DenoiseReadCounts. WGS matched-normal BAM files subsetted for the region of coverage of the WES bait library were then processed with GATK CollectAllelicCounts to identify regions of potential LOH. GATK CollectAllelicCounts was also applied to each tumor BAM file. GATK ModelSegments used smoothed read counts from each tumor BAM along with the
paired normal/tumor allelic counts for generating copy number estimates that were then called using GATK CallCopyRatioSegments. Finally, potential noise from each WES-derived SCNAs SEG file was removed setting small segments to zero if the size of the segment divided by the number of segments was less than an *a priori* determined cutoff value of 750. All zero-value segments were merged using bedtools version 2.29.0.

Determination of recurrent regions of genomic gains or losses, as well as discrete gene-level somatic copy number changes, were performed using GISTIC 2.0.23 on the GenePattern platform (module version 6.15.28). The cleaned focus-level SEG file was used as input with the following parameters: focal length cutoff set to 0.50, gene GISTIC set to yes, confidence level set to 0.90, cap value set to infinite, broad analysis set to on, max sample segments set to 10,000, arm peel set to no, and gene collapse method set to extreme. Post-GISTIC log₂-copy number ratio threshold values were used for calling discrete gene-level calls as follows: > 1.3 = 2-copy gain; 0.1 to 1.3 = 1-copy gain; -1.3 to -0.1 = 1-copy loss; < -1.3 = 2-copy loss. Chromosome arm-level changes followed the same threshold scheme per tumor focus, using the convention that at least 50% of the chromosome arm (noncontiguously) must be affected. Focal regions of significant gains or losses (peaks) were considered both by the magnitude of the log₂-copy number change and the *p* value of the 2-sided Fisher’s exact test when comparing ER to INR cases. Percent genome altered (PGA) was calculated for each focus by filtering the whole-genome estimated copy number SEG file for only those log₂-copy number ratios greater or less than ±1, adding the size of each segment, and dividing the sum by 3,095,677,412, the size in nucleotides of the reference genome. Graphical depictions of mutations for each focus/lesion/case (Oncoprints) were generated using Oncoprinter [34].

For consolidating genomic data from individual foci to lesions, a pre-determined threshold was set for a minimum number of foci to harbor that alteration to call a particular copy number event for the entire lesion. For lesions sampled by 7 or 8 foci, the threshold was 3; for 3, 4, 5 or 6 foci the threshold was 2, and for 1 or 2 foci the threshold was 1. For each focus, per-gene somatic copy
number calls from GISTIC (i.e., -1, -2, 0, 1 or 2) were averaged across all foci from each lesion. If the absolute value of that average was greater than the pre-determined threshold, the averaged value was rounded to the nearest integer. For point mutations, the lesion was considered to harbor the mutation if any number of foci harbored it. This approach allowed us to conservatively call copy number changes and mutations with similar sensitivities as bulk sequencing of non-microdissected tumor tissue. Univariate affected gene analyses considered any past-threshold somatic copy number alteration or mutation once consolidated at the lesion level. Similarly, genes with biallelic events were re-aggregated at the lesion level based on the same foci consolidation scheme. The log2-copy number changes of significantly altered focal peaks (gains or losses) were averaged across foci per lesion and univariate analyses performed on the aggregated data. For SNVs and PGA, values were averaged across all foci consolidated into a lesion. For mutational signatures, the proportion of each signature was averaged across foci in the same lesion. After all aggregation, lesion level was used directly for patient-level data except in cases where there were multiple lesions per patient; the index lesion served as the representative lesion for those patients.

For determining affected pathways at the patient or lesion level, the patient or lesion consensus effect for each gene (either mutation or copy number change) was considered individually as a gain-of-function and/or loss-of-function event based on the effect predicted in OncoKB. Genes were organized into a subset of cancer-related pathways [35, 36], and a pathway was reported affected in a lesion or case based on whether any genes in each group exhibited the predicted effect to perturb that pathway. For the cell cycle, epigenetic, MYC, NOTCH and WNT pathways, only 1 gene alteration counted towards the pathway being altered. For the DNA repair, PI-3K, and RAS-MAPK pathways, two genes had to be affected.

Germline variant calling

Germline variants were identified from buffy coat or saliva WGS BAM files individually by
running GATK HaplotypeCaller with ERC set to GVCF and referencing dbSNP version 138. All VCF files were joined into a cohort by running GATK CombineGVCFs and genotyped jointly with GATK GenotypeGVCFs. All discovered variants were subjected to sequential recalibration. SNVs were first recalibrated using GATK VariantRecalibrator with the following parameters:

```
--resource:hapmap,known=false,training=true,truth=true,prior=15.0 hapmap_3.3.b37.vcf
--resource:omni,known=false,training=true,truth=false,prior=12.0 1000G_omni2.5.b37.vcf
--resource:1000G,known=false,training=true,truth=false,prior=10.0 1000G_phase1.snps.high_confidence.b37.vcf
--resource:dbsnp,known=true,training=false,truth=false,prior=2.0 dbsnp_138.b37.vcf
```

SNPs with outputs applied using GATK ApplyVSQR with the following parameters:

```
-tS-filter-level 99.5 -mode SNP
```

Similarly, indels were then recalibrated using GATK VariantRecalibrator with the following parameters:

```
--resource:mills,known=false,training=true,truth=true,prior=12.0 Mills_and_1000G_gold_standard.indels.b37.vcf
--resource:dbsnp,known=true,training=false,truth=false,prior=2.0 dbsnp_138.b37.vcf
```

and outputs applied using GATK ApplyVSQR with the following parameters:

```
-tS-filter-level 99.0 -mode INDEL
```

The VCF file of recalibrated SNPs was annotated in the “id” column with the rsid of SNPs from dbSNP using bcftools (Samtools version 1.3.1).

Germline mutations were then annotated using Oncotator version 1.9.70 (database version April052016), considering only obvious loss-of-function mutations (e.g., frameshift, nonsense) identified by filtering for mutations overlapping with the cancer gene census germline mutation database. For a mutation to be considered as a germline mutation it had to be considered heterozygous by HaplotypeCaller and be covered at greater than 15× coverage.

For PolyQ length estimation, WGS BAM files were parsed to extract reads corresponding to the
AR locus, reindexed, and lifted over from b37 to hg19 using samtools. CAG repeats corresponding to the PolyQ tract were identified using the TREDPARSE HLI short tandem repeat caller[37] for Python version 2.7. The tred.py script was run in haploid mode with the hg19 genome.

For SNP analysis, a curated list of 163 prostate cancer risk SNPs was compiled based on a review of the literature [38]. RSIDs for each SNP were annotated for each heterozygous or homozygous call made by GATK HaplotypeCaller. A case was considered a carrier if either heterozygous or homozygous for the SNP. Comparing INR to ER cases, all possible pairs of SNPs were tested using the likelihood ratio test of the logistic regression model. Pairs of SNPs were also tested as model-free 2x4 tables using the Fisher-Freeman-Halton test.

Tumor phylogeny estimation

To estimate the evolutionary path of each tumor and identify tumor subpopulations, we aggregated gene-level somatic mutation and somatic copy number data by patient. For somatic mutations, we compiled a list of all high-quality (no strand bias), high-depth (at least 100× coverage), pass-filter coding mutations for each patient. In cases where the number of independent mutations across all foci was less than 50, stringency restrictions were lessened, starting with including noncoding mutations, then relaxing strand bias, then by decreasing the depth requirements to half of the average depth of each sample but no less than 16× coverage. The rationale for this conservative selection process was to start with the mutations that would be the most informative for inferring clonality from variant allele frequency. Then for each independent mutation within each patient, the same chromosomal coordinate was assessed against the raw list of variants and backfilled for the remainder of the tumor foci. This Bayesian process allowed previously filtered but shared mutations to be used as evidence for supporting common ancestry in conjunction with high quality mutations in other samples from the same patient, while also allowing unshared high quality mutations as evidence for subclonal populations.
For somatic copy number alterations, per-sample determination of gain or loss segments by GATK resulted in close (but not exact) boundaries for the same event across multiple foci from the same sample. To prevent this noise from supporting spurious subclones during analysis, the cleaned SEG file processed with GISTIC 2.0 identified several hundred broad or focal peaks with harmonized boundaries for each sample. Copy number gains and losses were then re-called based on the unified segment boundaries within each case, allowing for both shared and unshared SCNA events across foci. A histogram of variant allele frequencies for each sample was used to estimate the clonal and subclonal SNVs for determining tumor purity and inferring major alleles, minor alleles, and cellular prevalence of SCNA segment.

Prior to running PhyloWGS [39] version 1.0, the create_phylowgs_inputs tool was used to parse informative SCNAs (CNVs) and mutations (SSMs) from multiple VCF and SEG files into a single set of inputs. The SSM file was then parsed to re-annotate each mutation’s coordinate with the functional annotation from Oncotator. Finally, the multievolve script was used to build phylogenetic trees. For each case, 40 MCMC chains were run simultaneously with 1,000 burn-in samples and 2,500 MCMC-samples, for a total of 100,000 possible trees considered for each patient. After tree generation, chains were merged using the write_results tool, and the mutation and tree JSON files were parsed to select the tree with the most negative log likelihood score. Although multiprimary trees were allowed in the output, only 1 out of 37 trees generated a polytumor tree, suggesting that 100,000 potential trees allowed for sufficient estimation of subclonal reconstruction without defaulting to multiple independent clones.

The best scoring tree was then pruned using a predefined set of rules [40] to conservatively decrease the number of potential subclones. If any given node did not have at least 5 SNVs or SSMs assigned to it, it was merged with its sibling node with the greatest number of events. If that had no siblings, it was merged with its most immediate ancestral node, unless it was a direct descendant of the germ/normal node with no descendants, in which case it was eliminated. The only nodes with
fewer than 5 SNVs/SSMs were the parent cancer node with at least one descendant that could not be merged. The one tree that suggested a polyploid evolutionary path before pruning remained a polyploid after pruning, suggesting sufficient evidence supported those node assignments.

The mean value of the cellular prevalences of all alterations assigned to the node served as the cellular prevalence for each node in the tree. Graphical depictions of node sizes on tree figures are proportional to the cellular prevalence. Shannon diversity indices (SDI) were calculated using the conventional formula for the SDI ($H$), in which

$$H = - \sum p_i \ln p_i$$

where $p_i$ is the proportion of each node’s cellular prevalence in the entire population. The script for processing PhyloWGS output has been deposited at https://github.com/CBIIT/lgcp.

**Automated image analysis**

Whole biopsy slide CZI files of slides stained with antibodies against AR, PSA, Ki-67, SYP, and GR were imported into Definiens Developer XD 64. The magnification for each analysis was set to 20×, with 0.21 μm/pixel for all solutions. IHC stain was identified as brown chromogen. The tumor cells were counted using their nuclear stain. Composer magnification was set to 6× with 12 training subsets, with training segmentation at level 9 for all stains except for SYP, which was segmented at level 8. Segments were classified as tumor or stroma, with normal glands excluded. Within the tumor pattern, cellular analysis magnification was set to 10× with 12 training subsets, and within nuclear detection, the hematoxylin threshold was set at 0.05-0.2 and the brown chromogen was set between 0.15 and 0.3, with typical nuclear size set to 30 μm. For PSA and SYP, nuclei were used to simulate cells, where cytoplasm was estimated at a maximum distance of 4-8 μm around the nucleus.

For Ki-67, SYP and GR, an additional filter was used to exclude nuclei between 15-25μm when non-specific tumor calling was present. Nuclear classification for AR was low versus medium at 0.7 and medium versus high at 0.95. Cellular classification for PSA was no expression versus low at 0.2,
low versus medium at 0.4, and medium versus high at 0.5. Nucleus classification for Ki-67 was low versus medium at 0.65 and medium versus high at 0.86. Cellular classification for SYP was no expression versus low at 0.2, low versus medium at 0.4 and medium versus high at 0.5. Nuclear classification for GR was low versus medium at 0.4 and medium versus high at 0.6.

The total number of positively stained nuclei were reported, along with distribution of low-, medium-, and high-intensity stained nuclei for each tumor focus. A percent positive histology index (HI) score was calculated using a weighted average divided by the total number of nuclei, where

\[
\text{index} = \left( \frac{1 \times \text{nuclei/cell stained low} + 2 \times \text{nuclei/cell stained medium} + 3 \times \text{nuclei/cell stained high}}{3 \times \text{total nuclei}} \right)
\]

The Shannon diversity index (SDI) was calculated as described above. The nuclei or cells quantified for each case were aggregated across blocks prior to calculating HI or SDI.

*The Cancer Genome Atlas (TCGA) analysis*

Whole exome sequencing, clinical, pathological, and recurrence data from 188 tumors harboring Gleason patterns 4 and/or 5 (*i.e.*, Gleason scores of 4+4=8, 4+5=9, 5+4=9 and 5+5=10) were accessed and downloaded from the National Cancer Institute Genomic Data Commons (http://gdc.cancer.gov) via authorized access project #9940 under dbGaP accession ID phs000178. These ISUP GG4-5 cases were selected based on their better representation of the greater risk composition of our trial cohort for the purposes of comparing between TCGA and our trial cohort.

Single sample whole-exome sequencing data was aligned and processed nearly identically to the study cohort data for the purposes of identifying the frequencies of somatic mutations, germline mutations, somatic copy number alterations, SNV counts and PGA. All frameshift mutations called by MuTect2 in TCGA cases were included for the purposes of estimating the frequency of mutational events, irrespective of their overlap in COSMIC. ERG-fusion status and pre-operative PSA levels were included if available. Phylogeny estimation followed the same procedure as single-
focus study samples, considering 100,000 possible tree structures for each tumor. The proportions of somatic mutations and somatic copy number alterations detected in the Gleason score 8-10 prostate TCGA cases was used to validate the detection of similar events in the current study cohort.

Whole slide images of frozen tumor sections used for TCGA molecular analysis were accessed on the NCI Genomic Data Commons and cross referenced against a prior publication [41] for the presence of invasive cribriform carcinoma or intraductal carcinoma (IDC-P), as they were not distinguished. Cases previously annotated as ICC were re-reviewed by a genitourinary pathologist (R.T.L.) and further classified as IDC-P based on the presence of nuclear atypia, packed lumina, ducts with florid luminal carcinoma, and the number of glands involved with carcinoma, while excluding atrophic glands with flattened epithelia. For each call of IDC-P presence or absence, a confidence score was calculated based on these factors, the presence of adjacent invasive adenocarcinoma, and whether IDC-P was clearly visible on formalin sections from the same patient. Only cases with >99% confidence of IDC-P or no IDC-P were used in the validation cohort. This restriction for calling IDC-P was the major factor in reducing the number of evaluable TCGA validation cases from 188 to 58.

p53 histology

Staining for p53 was performed as previously described using a protocol validated for detection of TP53 point mutations [42]. Briefly, using a Ventana Benchmark autostaining system, antigen retrieval was performed in CC1 buffer and slides were subsequently incubated with anti-p53 antibody clone BP53-11 (Roche), followed by detection with the iView HRP system (Roche/Ventana Medical Systems, Oro Valley, AZ).

Scoring of p53 staining was performed independently by three genitourinary pathologists (D.C.S., H.B.K., and T.L.L.). Using reference slides as internal controls for each batch, cases were scored using either a dichotomous scoring method (0 or 1, where 1 indicates TP53 overexpressed
likely due to a TP53 missense mutation), or a semiquantitative score (of 0, 1, 2 or 3 for nuclear staining intensity, where 0 = no staining, 1 = < 5% positive staining, 2 = 5-10% positive staining, and 3 = >10% positive staining).

Aggregation of p53 staining scores was performed only at the patient level. If the dichotomous score was 1 for any biopsy within a case, the score of 1 was ascribed to the entire case. Similarly, if a semiquantitative score was 3 for any biopsy in a case, the score of 3 was assigned to the entire case. Similarly, if any biopsy was a 0, the entire case was called 0. In discordant cases where scores of 3 and 0 were in the same case, 3 was used for the entire case. For the remainder of cases, scores of 1 or 2 were used, going with the higher score when heterogeneous for multiple biopsies from the same case.

When comparing p53 staining to TP53 point mutations, we considered all missense mutations that were identified by MuTect2, not limiting calls just to those mutations also overlapping in COSMIC. If a mutation was found in any LCM focus, it was ascribed to the entire case for the purposes of comparing staining to sequencing.
Supplementary Fig. 1. Genomic landscape of each laser capture microdissected foci from baseline biopsies prior to treatment with neoadjuvant androgen deprivation therapy plus enzalutamide. Genomic data are shown per LCM-dissected focus (N = 141), and clinical data are shown per patient (N = 37). Tumor foci are grouped by patient, color-coded to the left of the SCNA heatmap. One patient classified clinically as a non-responder underwent TURP instead of RP such that no calculation of RCB was possible. RCB: Residual cancer burden (cm³); PGA: percentage of the genome altered (%); SNVs: small nucleotide variants (N).
Supplementary Fig. 2. Genomic landscape of mpMRI-visible prostate tumors dissected from baseline biopsies prior to treatment with neoadjuvant androgen deprivation therapy plus enzalutamide. Genomic and imaging data are shown per mpMRI-visible lesion \( (N = 48) \), and clinical data are shown per patient \( (N = 37) \). One patient classified clinically as a non-responder underwent TURP instead of RP such that no calculation of RCB was possible. RCB: Residual cancer burden \( (cm^3) \); BTV: baseline MRI tumor volume \( (cm^3) \); PGA: percentage of the genome altered \( (%) \); SNVs: small nucleotide variants \( (N) \).
Supplementary Fig. 3. Genomic landscape of 188 whole-exome cases from TCGA-PRAD. (A) Data was processed through the same bioinformatic pathway as the current study, with clinical and genomic data displayed per person. DSS: months disease-specific survival; PGA: percentage of the genome altered (%); SNVs: small nucleotide variants (N). (B) Comparison of proportions for 28 curated molecular features across the current study cohort and the prostate cancer TCGA cases with Gleason scores 8-10 (N = 188).
Supplementary Fig. 4. Mutational signatures. The catalog of somatic mutations from each laser capture microdissected tumor focus was parsed for single-based mutation type (A-B) and flanking-base-contextualized mutational signature (C-D). Data were summarized at the patient (A,C) and lesion (B,D) level. Average proportions for each mutation type were compared between ER and INR cases/lesions, and p values were determined by the method of Wei and Johnson [43] for performing a nonparametric test between two groups of a correlated multivariate outcome using a random sample from the permutation distribution.
Supplementary Fig. 5. Comparison of clinicopathologic features between responder and incomplete and nonresponder mpMRI-visible lesions. (A) Baseline lesion-level MRI tumor volumes comparing ER to INR. \( p = 0.0071 \) by Mann-Whitney U test. Effect size \( d \) is determined as the Hodges-Lehmann median percent differences, based on the \( \log_{10} \)-transformed value. (B) The highest Gleason Grade Grouping (GG) of all pre-treatment biopsies for each mpMRI lesion. \( p = 1 \) by the Cochran-Armitage test for trend for each GG between ER and INR. Effect size is given by the estimated odds ratio (95% confidence interval), which is for each increase in Gleason grade group. (C) The number of mpMRI-visible lesions per case, comparing ER to INR. \( p = 0.4 \) by Mann-Whitney U test. Effect size \( d \) determined as the absolute difference of medians of each group. (D) Frequency and odds ratio for five histologic and pathologic features, comparing INR cases to ER cases. Odds ratio > 1 favors INR. *26 mpMRI-visible ER lesions were sampled on biopsy, but only 25 were used for the majority of the analyses. Only Gleason pattern and ERG were measured in 26 lesions.
Supplementary Figure 6

See separate PDF file.

Supplementary Fig. 6. Phylogenetic clone trees from each case in the clinical study. After processing with PhyloWGS, 100,000 different potential structures for each case were considered and the tree with the most negative log likelihood score was selected. After pruning each tree according to a priori rules for a minimum number of events per node, the mean cellular prevalence per node was determined and the node size graphically scaled proportionally to that prevalence.
Supplementary Fig. 7. Analyses of MRI-estimated tumor volumes. (A-C) Scatter plot of the log_{10}-transformed residual cancer burden versus the log_{10}-transformed baseline (A), relative baseline (B) or final (F) MRI-estimated tumor burden. (D-F) Scatter plot of the genomic Shannon diversity index (SDI) for each case versus the corresponding log_{10}-transformed baseline (D), relative baseline (E), or final (F) MRI-estimated tumor burden. For (A-D), Values of zero were transformed to nominal values to preserve their rank within the dataset, but omitted from the graph. \( N = 37 \) for all plots except for plots involving final tumor volumes, where one patient underwent simultaneous TURP and TURBT. For (A-G), nonparametric Spearman correlation analyses rho (\( \rho \)) values are shown with their respective p values. Lines and bands represent linear regression lines and 95% confidence intervals.
confidence intervals, respectively. (H) The number of mpMRI-visible lesions per case, comparing cases with genomic Shannon diversity indices (SDI) less than the median SDI or greater than or equal to the median SDI. \( p = 0.3 \) by Mann-Whitney U test. Effect size \( d \) determined as the absolute difference of medians of each group. (I) Receiver operator characteristic (ROC) curve for the INR probabilities in distinguishing ER from INR in the trial cohort based on the four-factor model plus baseline tumor burden. (J-K) Spearman correlations of the genomic SDI and INR probabilities for the trial cohort using the 5-factor model based on either baseline (J) or relative baseline (K) tumor burden.
Supplementary Fig. 8. Kaplan-Meier survival analysis of the TCGA validation cohort.

Biochemical recurrence-free survival data was downloaded from the NCI Genomic Data Commons. TCGA cases scored based on the four-factor model were segregated based on the 60% cutpoint into cases predicted to be ER or INR. Their relative survival curves were compared using the Log-rank (Mantel-Cox) test. The $p$ value of 0.9 and Chi-square of 0.018 indicates that the curves are not significantly different. Effect size is given as log-rank hazard ratio (95% confidence interval).
Supplementary Fig. 9. Differential enrichment of significant somatic events. (A-B) GISTIC was run on the entire cohort, identifying 180 distinct peaks at the patient level (A) and 184 peaks at the lesion level (B). Data is displayed as a volcano plot, with amplifications in red and deletions in blue, plotting statistical significance on the Y axis. Alterations enriched in INR cases/lesions are plotted right of zero, and alterations enriched in ER cases/lesions are plotted left of zero. The X axis
indicates the number of cases or lesions (either ER or INR) harboring each alteration. (C-D) Oncoprint of genomic alterations comparing INR to ER cases (C) or lesions (D), with $p < 0.05$. (E-F) Oncoprint of genomic alterations comparing INR to ER cases (E) or lesions (F) with $p < 0.05$. For A-F, all analyses represent the results of 2-sided Fisher’s exact tests. Effect sizes, given as odds ratios, are provided in Supplementary Tables 9 and 10. As these are exploratory analyses and no adjustments for multiple testing were applied, they may be underpowered for per-cytoband and per-gene analyses.
Supplementary Fig. 10. Comparison of PTEN immunohistochemistry and PTEN genomics. (A) Representative micrographs of PTEN IHC considered intact or unaltered (top) and reduced or absent (bottom). Bar: 100 µm. (B) Two-sided Fisher’s exact test comparing either any PTEN copy number loss, homozygous deletion only, or loss of chromosome 10q to the observations of PTEN reduction by IHC. Effect size given as odds ratio, where OR > 1 favors IHC reduction association with the
somatic copy loss event. (C) Two-sided Fisher’s exact test comparing PTEN reduction by IHC between ER and INR cases. (D) Two-sided Fisher’s exact test comparing the concordance between PTEN alterations and PTEN reduction by IHC between ER and INR cases. Effect size given as odds ratio, where OR > 1 favors PTEN reduction with INR. For (B-D), the comparison of proportions is based on the IHC staining in each group indicated, but p values and odds ratios based on actual counts of cases.
Supplementary Fig. 11. Comparison of p53 immunohistochemistry and TP53 genomics. (A) Representative micrographs of p53 IHC scored using the semiquantitative method, where 0 = no staining, 1 = <5% staining, 2 = 5-10% staining, and 3 = >10% staining. (B) Distribution of ER and INR cases based on semiquantitative score. $p = 0.8$ by the Cochran-Armitage test for trend comparing the semiquantitative scores between ER and INR cases. Due to the semiquantitative nature of the data, no effect size statistic is reported. (C-D) Two-sided Fisher’s exact test comparing the proportions of any TP53 alteration or compound TP53 alterations (C) or truncating or loss of function (LOF) TP53 alterations (D) within all cases based on the p53 semiquantitative score. Effect size given as odds ratio, where OR > 1 favors a score of 1 in the absence of alterations. (E) Representative micrographs of p53 IHC scored using the dichotomous method, where 1 = predicted to have gain-of-function alterations resulting in p53 accumulation, such as a missense mutation, and 0 = no mutation. For (A) and (E), bar = 100 µm. (F) Two-sided Fisher’s exact test comparing the proportions of either ER or INR cases, or the observations of nontruncating TP53 alteration across all cases, with a p53 dichotomous score of 0 or 1. Effect size given as odds ratio, where OR > 1 favors a score of 1 in the INR or where nontruncating mutations were observed. For (C, D, and F), the comparison of proportions is based on the IHC staining in each group indicated, but $p$ values and odds ratios based on actual counts of cases.
SUPPLEMENTARY TABLE LEGENDS

**Supplementary Table 1:** See Microsoft Excel file. **Clinical, genomic and histologic data for each laser capture microdissected focus.** Related to Supplementary Figure 1.

**Supplementary Table 2:** See Microsoft Excel file. **Clinical, genomic and histologic data for each MRI-contoured tumor lesion.** Related to Supplementary Figure 2A.

**Supplementary Table 3:** See Microsoft Excel file. **Clinical, genomic and histologic data for each patient in the NADT clinical trial.** Related to Figure 1A.

**Supplementary Table 4:** See Microsoft Excel file. **Clinical, genomic and histologic data for each patient in the prostate cancer TCGA validation cohort.** Related to Supplementary Figure 3.

**Supplementary Table 5:** See Microsoft Excel file. **Individual cell or nucleus counts for each stained IHC slide used for automated image analysis and manual scoring of p53.** Related to Figure 3J. Negative, low, medium, and high counts are shown for each of anti-AR, anti-PSA, anti-Ki-67, anti-SYP and anti-GR immunostains. Aggregate histology indices (HI) and Shannon diversity indices (SDI) on a per-patient basis are also shown. Dichotomous and semi-quantitative scores are shown for anti-p53 immunostain.

**Supplementary Table 6:** See Microsoft Excel file. **Multivariate logistic regression of the four-factor model and its comparison to each 5-factor model.** For each of the four factors selected for the multivariate histogenomic model, the individual two-sided Fisher’s exact test unadjusted p value and odds ratio is shown, along with each factor’s logistic regression score, odds ratio after adjustment for the other 3 factors, and adjusted logistic regression score. For each potential combination of 4 factors (16 possibilities) the predicted INR probability and prevalence of INR in the cohort is shown. Comparing the associations of each factor in our model to genomic SDI and prediction of INR outcome using a t test, IDC-P was the second most important feature (uncorrected p = 0.088), with only ERG showing the greatest importance (uncorrected p = 0.3). When TP53 was omitted from our logistic regression model, ERG remained tightly associated with INR probability but no longer carried associations with genomic diversity, demonstrating the importance of the coincidence of these alterations for driving evolutionary processes. The importance of each of the four factors in this correlation was assessed by fitting a new logistic regression model with the factor removed and comparing the probabilities to genomic SDI with the factor’s mean effect subtracted. Addition of the log10-transformed MRI estimate of baseline tumor burden or baseline relative tumor burden resulted in adjustment towards 0 or 1, and the individual patient probabilities of INR based on each of the three models is shown.

**Supplementary Table 7:** See Microsoft Excel file. **GISTIC peaks at the person level.** Wide and focal peaks were identified by GISTIC from the combined focus-level SEG file. The presence or absence of each peak was consolidated by lesion, and then subsetted by patient based on the index lesion determined a priori for each case. For each peak, the number of ER and INR cases are indicated, along with the two-sided Fisher’s exact test unadjusted p value.
Supplementary Table 8: See Microsoft Excel file.
GISTIC peaks at the lesion level. Wide and focal peaks were identified by GISTIC from the combined focus-level SEG file. The presence or absence of each peak was consolidated by lesion. For each peak, the number of ER and INR lesions are indicated, along with the two-sided Fisher’s exact test unadjusted $p$ value.

Supplementary Table 9: See Microsoft Excel file.
Affected genes at the person level. Germline mutations, somatic mutations, and somatic copy number alterations were aggregated across foci by lesion and then subsettled by patient based on the index lesion determined a priori for each case. For each gene, the two-sided Fisher’s exact test unadjusted $p$ value is shown along with the odds ratio with its 95% confidence interval ($<1$ indicates association with INR; $>1$ indicates association with ER). The $\phi$ value indicates the directionality of enrichment. Negative $\phi$ values indicate enrichment in INR cases.

Supplementary Table 10: See Microsoft Excel file.
Affected genes at the lesion level. Germline mutations, somatic mutations, and somatic copy number alterations were aggregated across foci by lesion. For each gene, the two-sided Fisher’s exact test unadjusted $p$ value is shown along with the odds ratio with its 95% confidence interval ($<1$ indicates association with INR; $>1$ indicates association with ER). The $\phi$ value indicates the directionality of enrichment. Negative $\phi$ values indicate enrichment in INR lesions.

Supplementary Table 11: See Microsoft Excel file.
Risk SNP status by case. The curated list of 147 risk SNPs is shown by rsid for each case. The individual calls by GATK HaplotypeCaller were summarized as 0 for non-carrier calls of 0/0 or ./ (no call), and as 1 for carrier calls of 0/1, 1/1, 0|1, or 1|1.

Supplementary Table 12: See Microsoft Excel file.
Polyglutamine (PolyQ) repeat length. For each case, the CAG repeat length estimated by TREDPARSE is shown.
SUPPLEMENTARY REFERENCES


