

Supplementary Materials and Methods

Nucleic Acid Isolation

Tissue was obtained by performing 3 – 5 punches using a 1mm biopsy needle or scalpel macrodissection using 3-10 x 10um unstained sections. Co-isolation of DNA and RNA from each specimen was performed using the Qiagen Allprep FFPE DNA/RNA kit (Qiagen, Valencia, CA) as described (41).

RNA and DNA Next Generation Sequencing

RNA Ampliseq libraries were performed essentially as described using 1 - 15ng of RNA per sample and the Ion Ampliseq RNA Library kit with barcode incorporation. Templates were prepared using the Ion PI Hi-Q OT2 200 Kit and sequenced using the Ion PI Hi-Q Sequencing 200 Kit (200 base pair reads) on Ion PI Chip v3 as described (18). DNA libraries were generated from 1- 20ng of DNA per sample using the Ion Ampliseq library kit 2.0 (Life Technologies, Foster City, CA) and the DNA component of the Oncomine Comprehensive Assay (OCA) panel or the Comprehensive Cancer Panel (CCP) with barcode incorporation as described (18).

Bioinformatics Analysis

End to end read counts for RNA expression runs were calculated using Torrent Suite's Coverage Analysis plugin v. 5.0.4. All further analyses were conducted using The R Project for Statistical Computing v. 3.2.3 (18, 40). Housekeeping genes from Oncotype DX panel ($n=5$) were considered for normalization, and the 4 with robust expression (*ATP5E*, *ARF1*, *CLTC1*, and *PGK1*) were used for normalization prior to downstream analyses. Non-fusion amplicons were filtered to ensure that all amplicons retained for analysis had ≥ 200 reads in at least two tissue

samples, or > 1000 reads in at least one tissue sample to ensure amplicon performance. Raw read counts were subsequently log₂-transformed, (i.e., log₂(read_count + 1)) and normalized to the geometric mean of expression values for the 4 retained housekeeping genes. For heatmap visualization only, the median amplicon-level expression was calculated across all samples and subtracted from each target amplicon's expression value prior to plotting.

Stringent quality control criteria were employed to ensure only high quality samples were included in subsequent analyses. Sample-level inclusion quality control metrics included at least 500,000 total mapped sequencing reads and at least 60% of all sequenced reads mapping end-to-end. Housekeeping gene read mapping and expression variability were also assessed to filter out low quality samples. For each sample, the proportion of mapped reads mapping end-to-end to each housekeeping gene ('hk_prop_filt') was evaluated in a compendium of 255 samples (including additional samples not described herein), and the following gene-level hk_prop_filt thresholds were applied to our current cohort (based on percentile expression across the compendium) to exclude low-quality samples: (exclude samples if *ATP53* < 0.000133, *AFR1* < .001266, *CLTC* < 0.001894, *PGK1* < 0.000352). Samples with <0.8% of all reads mapping to these four housekeeping genes or standard deviation of log₂-normalized expression values across the four housekeeping genes < 0.0015 were also excluded from our analyses. Samples that did not pass all of the above criteria were not considered for further analysis.

Derivation of Commercially Available Prognostic Scores

After exclusion of low quality amplicons as described above, 29/31 CCP and 12/12 GPS target gene amplicons were retained. For CCP score, we used the previously published formula for calculating CCP scores for non-replicate experiments, except that we used the four GPS

housekeeping genes (identified above) (5). Log₂-normalized expression values for the 29 high quality CCP target transcripts were floored at -5 prior to score derivation to ensure technical artifacts of RNAseq normalization did not impact score derivation. Based on the previously published formula, derived mxCCP scores were then calculated by taking the mean of each CCP retained gene's median-centered expression value to the power of 2, then log₂ transforming the mean (5).

For GPS, scores were derived by adding or multiplying log₂ normalized gene expression values for each score module as previously published (20). The lower bound of expression values for *TPX2* and *SRDA5* were capped at 5 and 5.5 respectively, as previously described (20). However, we omitted multiplying individual expression values by described coefficients, as these were validated using qRT-PCR methodology. As such, each module score was derived as follows:

$$\text{Cellular Organization Module} = FLNC + GSN + TPM2 + GSTM2$$

$$\text{Stromal Module} = BGN + COL1A1 + SFRP4$$

$$\text{Androgen Module} = FAM13C + KLK1 + SRDA5 + AZGP1$$

$$\text{Proliferation Module} = TPX2$$

To derive the full unscaled score, the previously published methodology(20) was used, including the coefficients for adding subcomponents:

$$mxGPS = .735*Stromal - .368*Cellular Organization - .352*Androgen + 0.95*Proliferation.$$

After score derivation, mxCCP scores and mxGPS were converted to percentile distributions, respectively, for ease in downstream interpretation.

Given that Decipher™ GC score was built off a random forest based classifier and not all target transcript locations have been published, direct derivation of the GC score is not feasible. Hence, to generate our derived GC score (mxGC), we first performed median centering of all \log_2 normalized values for all GC target amplicons passing the above QC criteria (24/24). We then took the average of the respective \log_2 normalized *ANO7*, *MYBPC1* and *UBE2C* amplicons (3 each) to generate a single composite expression value for each of these three loci (resulting in 18 target transcripts). As unsupervised, centroid-linkage, hierarchical clustering of normalized expression for these 18 target transcripts demonstrated two clusters of transcripts over- (n=9 targets) and under-expressed (9 targets) with prostate cancer progression (see **Figure S2**), we calculated mxGC scores as: (average of the \log_2 normalized values for the 9 over-expressed targets) – (average of the \log_2 normalized values for the 9 under-expressed targets).

For single gene candidate markers, normalized expression of amplicons targeting *SChLAP1* (AMPLSCHLAP1_1_1.999) and *PRCAT104* (AMPLG053084_T230586_2_1.989) were evaluated as for derived prognostic scores.

Fusion Isoform- and Partner-Level Analyses

For initial validation assessment analyses, fusion isoform-specific amplicons were filtered to only retain those with >1000 reads in at least one sample. Isoform-level (e.g., *TMPRSS2:ERG.E1E4*) \log_2 normalized read counts were calculated as described above. For fusion partner-level (e.g., *TMPRSS2:ERG*) status, read counts for all retained isoforms were then totaled for each sample, and a normalized fusion partner value was calculated by taking the \log_2 of the sum of the all reads over the sum of housekeeping reads for each sample. A sample was determined as *TMPRSS2:ERG* fusion positive if it had more than 500 total reads across isoforms,

and its fusion value was greater than $\log_2(.01)$. Investigation of novel fusion isoforms was carried out by mapping all targeted RNAseq reads to the hg19 reference genome with STAR (v2.3.0) using Gencode v19 splice junction annotation.

AR and AR-splice Variant Analysis

Amplicons were included against full length *AR*, 6 *AR* splice variants (including *ARv7*) and two areas of recurrent hotspot mutations (**Table S2**). The full length *AR* amplicon and 4 *AR* splice variant amplicons passed usual target gene expression filtering above and were considered. Samples passing usual filtering above were further filtered to exclude those with no detectable full length *AR* expression (<-5 normalized \log_2 expression of AMPLAR-FL5-1-2_1_1.762). For each sample, an individual *AR* splice variant was considered detected if >100 reads were present.

DNA Variant Calls and Copy Number Analyses

Raw reads were aligned to the reference genome (hg19) using TMAP on Torrent Suite v. 5.0.4 (Thermo Fisher Scientific, Waltham MA). Somatic variants for DNA samples were called using Torrent Variant Caller v. 5.0.4, and annotated and filtered using previously described internal pipelines (18). Normalized, GC-content corrected read counts per amplicon for each sample were divided by those from a pool of normal male genomic DNA samples (FFPE and frozen tissue, individual and pooled samples), yielding a copy number ratio for each amplicon. Gene-level copy number estimates were determined as described(18) by taking the coverage-weighted mean of the per-probe ratios, with expected error determined by the probe-to-probe

variance. Genes with a \log_2 copy number ratio estimate of < -1 or > 0.8 were considered to have high level loss and gain, respectively.

Supplementary Results

mxRNAseq Assay Assessment and Validation

Using our novel ampliseq mxRNAseq on the Ion Torrent Proton sequencer, we applied this panel to a cohort of 198 macro-dissected, routine, clinical FFPE tissue specimens from multiple institutions representing the full spectrum of prostate cancer progression from benign prostatic tissue (both epithelial/stromal mixed and pure stromal samples), GG1-5 localized prostate cancer, hormone sensitive metastatic prostate cancer, and mCRPC. Six of these samples were FFPE CRPC cell lines, benign lymph node tissue or melanoma tissues included for assay controls. In total, 156/198 (79%) samples met high-stringency quality control parameters for both sequencing and housekeeping gene expression metrics (see **Bioinformatics Methods** above) and were considered informative for all subsequent analyses. Of the 294 non-housekeeping target transcript amplicons, 223 (89.7%) were considered informative for gene expression and were included in subsequent analyses after normalization (raw read counts for all amplicons in all samples is provided in **Table S2** and normalized reads from amplicons/samples passing QC are provided in **Table S3**).

The validity of our approach and observed expression profiles are supported by multiple lines of evidence. First, target gene expression profiles were strongly correlated across highly related tumor samples (**Figure S1**). For example, two samples from the same lymph node metastasis (PR-376 and PR-377) showed the most correlated target gene expression across our entire sample compendium ($n=223$ amplicons, $r_p=0.85$, $p<0.0001$, 95% CI 0.80-0.88). Likewise,

the second and third most correlated sample pairs were the two melanoma samples (MO-18 and MO-28, n=223 amplicons, $r_p=0.84$, $p<0.0001$, 95% CI 0.79-0.87) and two immediately adjacent areas of the same high GG tumor focus (PR-412 and PR-413, n=223 amplicons, $r_p=0.82$, $p<0.0001$, 95% CI 0.77-0.86), respectively. Second, we observed highly correlated target gene expression for related amplicons (**Figure S2**). For example, we designed multiple amplicons to assess different regions (e.g. coding, non-coding and antisense) of three loci included in the Decipher™ GC signature (*MYBPC1* and *ANOV7* and *UBE2C*) and observed highly correlated expression (median pairwise r_p of each target gene amplicon at these three loci in all 156 informative samples = 0.85, **Figure S2**), with two amplicons against *MYBPC1* showing the highest correlation of all amplicon pairs on our assay (AMPLMYBPC1_1_1.100 and AMPLMYBPC1_2_1_1.2, n=156 samples, $r_p=0.94$, $p<0.0001$, 95% CI 0.91-0.95). Third, highly correlated transcriptional modules were also observed across prostate cancer progression as expected. For example, as shown in **Figure 1B** and **S2**, over-expression of cell cycle/proliferation, and under-expression of stromal and *AR* signaling modules were observed with prostate cancer progression, consistent with previous reports (5, 20). Fourth, our assay demonstrated highly correlated expression from expected markers in non-cancerous (including non-prostate) tissue components. For example, across the entire cohort of informative samples, expression of the smooth muscle restricted gene *myosin light chain kinase* (*MYLK*) was most correlated to the smooth muscle restricted gene *tropomyosin 2* (*TPM2*; n=156, $r_p=0.90$, $p<0.0001$, 95% CI 0.86-0.92). Likewise, expression of the cytotoxic T lymphocyte marker *CD8A* was most correlated to cytotoxic T lymphocyte/natural killer cell specific protease *granzyme A* (*GZMA*; n=156, $r_p=0.55$, $p<0.0001$, 95% CI 0.43-0.65; **Figure S2**).

Detection of Gene Fusions, Mutations and AR splice variants

ETS gene fusions, present in approximately 50% of prostate cancer from PSA-screened Caucasian cohorts, form the basis of prostate cancer molecular subtyping and can be used for clonality/multifocality assessment (14). While recurrent hotspot mutations are infrequent in prostate cancer, subtypes mutually exclusive with *ETS* gene fusions are defined by hotspot mutations in *SPOP* (often with *SPINK1* over-expression), *BRAF* hotspot mutations or *IDH1* R132H mutations (14). Although the Decipher™ assay has been analytically validated to assess *ETS* gene fusion status indirectly (42), currently available prognostic assays do not directly assess fusion transcript expression nor the *SPOP*, *BRAF* or *IDH1* mutated subtypes. Hence, in addition to amplicons targeting the 3' aspects of *ETS* genes and *BRAF/RAF1* (retained in all reported fusions), our assay included 60 pre-defined gene fusion amplicons. We included five amplicons designed to specifically cover transcribed somatic mutational hotspots in *SPOP*, *IDH1*, *KRAS*, *BRAF*, *NRAS*, as well as two amplicons targeting germline risk variants (in *HOXB13* and *HSD3B1*). Lastly, we included 6 amplicons designed to assess *ARv*'s. Across our cohort, 13/60 pre-specified gene fusion amplicons and 4/6 *ARv* amplicons were informative (detected in at least one specimen). Likewise, 6/7 transcribed hotspot/germline mutation amplicons were informative [all but an amplicon targeting *HSD3B1* p.T367 (rs1047303)].

To validate the performance of these molecular subtyping based amplicons, our cohort included multiple tissue samples and CRPC cell lines with known *ETS* gene fusion/ hotspot mutation/germline variant/*ARv* status. Specifically, for validation of *ETS* gene fusion/*SPOP* mutation subtyping by our approach, we included 24 informative prostate cancer tissue samples and 3 CRPC cell lines with known *ETS* fusion determined by qRT-PCR and *SPOP* mutation status determined by targeted DNA sequencing.(18) As shown in **Figure S3**, our mxRNASeq

results were 100% concordant for known *ETS* gene fusion status (including *T2:ERG* detection in endogenously harboring VCAP cells), and 3' *ETS* gene over-expression was seen nearly exclusively in cases with detectable *ETS* gene fusions (*ERG* > *ETV1* and *ETV1* over-expression in LNCaP as expected) and lacking *SPINK1* outlier over-expression.

While essentially all samples with outlier *ERG* expression (62/64 (96.9%) samples with \log_2 *ERG* expression > 0.25) show detectable expression of fusion isoforms involving *ERG* by mxRNAseq analysis, only 2 of 8 (25%) samples with outlier *ETV1* expression (\log_2 expression >2.0) demonstrate detectable expression of pre-designed targeted fusion isoforms involving *ETV1*. Through unbiased alignment of mxRNAseq sequencing data to the whole transcriptome (see **Supplementary Methods** above), however, we show that all 8 samples with outlier *ETV1* expression show detectable expression of *TMPRSS2-ETV1* or *SLC45A3:ETV1* fusion isoforms (**Figure S4A**), presumably due to combinatorial priming of 5' and 3' partners from fusion isoforms targeted on our panel. These results demonstrate a robust ability to assess diverse *ETS* fusion isoform expression from a single mxRNAseq assay and highlight a unique ability to leverage combinatorial priming for detection of non-targeted gene fusion isoforms.

Expressed Somatic Mutation and Germline Risk Variant Detection

Importantly, for hotspots in *SPOP*, *NRAS*, *BRAF*, *AR* and *HOXB13* assessed by mxRNAseq analysis, mutant allele expression was only observed in samples with corresponding mutations detected in co-isolated DNA ($n=21$ mutations). While a single sample harbored an *SPOP* p.F133V hotspot mutation by targeted DNaseq without detectable mutant transcript expression, highly correlated variant allele frequencies (Pearson correlation, $r=0.78$, $p=1.7 \times 10^{-5}$) were observed overall in detectable variants from corresponding mxRNA/DNaseq (**Figure**

S3B and S4B). In other cohorts, we confirmed the ability of an amplicon targeting known *IDH1* hotspot mutations, seen in 1% of primary prostate cancer (43), to detect mutant allele expression (data not shown). Lastly, the *HOXB13* p.G84E risk allele was robustly detected in a sample from a patient known to harbor this variant included in our cohort (**Figure S3B**).

Full length AR and AR Splice Variant Expression/Detection

ARv7 and other *AR* splice variants have been reported as predictive biomarkers of second-generation anti-androgen (abiraterone and enzalutamide) resistance in men with mCRPC when detected in circulating tumor cells (CTCs) (44). Previous whole transcriptome RNAseq has shown that although *AR* splice variants are increased in men with mCRPC, they are also detectable rarely in both benign prostate tissue samples and localized prostate cancer (45). Given the potential clinical utility of *AR* splice variants, we included nine total amplicons assessing *AR*, including one amplicon designed to assess full length *AR* (*AR FL*), six amplicons designed to assess *AR* splice variants, and two amplicons designed to cover recurrent mutations observed in patients with mCRPC (e.g. p.702H and p.R878L) and associated with poor response to second-generation anti-androgens abiraterone and enzalutamide (46, 47).

As expected, we observed the highest *AR FL* expression in mCRPC, with the highest expression included in three mCRPC samples with *AR* amplifications identified by mxDNAseq in co-isolated DNA. *ARv7* was detectable in 2/6 (33%) mCRPC samples and 1/36 (3%) GG 5 prostate cancer samples. Importantly, both 22RV1 and VCAP CRPC cells showed high *AR FL* and *ARv7* expression (vs. LNCaP cells), as expected (48). Including all detectable variants ($n=4$) in tissue samples with detectable *AR FL* ($n=137$), *AR* splice variants were detected in subsets of benign prostate tissue samples, GG1-5 localized prostate cancer, HN metastases and mCRPC,

respectively (**Figure S5**), consistent with previous reports of detectable AR variants across prostate tissue types (45). By mxRNAseq analysis, we robustly detected known homozygous hotspot AR mutations in LNCaP (p.T878A, RNA variant fraction (VF): 98.4%, 910/925 reads)(49, 50) and 22RV1 (p.H875Y, RNA VF: 97.9%, 642/656 reads)(50) prostate cancer cell lines, and identified somatic p.H875Y (c.2623C>T, chr:66953543) mutations in two separate tissue samples (PR-172: DNA VF: 84.6%, 1387/1648 reads, RNA VF: 58.3%, 961/2250 reads; PR-37: DNA VF: 50.3%, 1064/2111 reads, RNA VF: 72.7%, 1535/5151 reads) previously detected as mutation positive by targeted DNA sequencing (18). Taken together, these results support the ability of our approach to simultaneously assess both target gene, gene fusion, expressed somatic mutations/variant and splice variant detection to enable comprehensive characterization of prostate cancer tissue samples.

Extreme Grade-Discordant Multifocal Prostate Cancer Case Descriptions

In case MF5, a large overall GG5, pT3a tumor with extensive intraductal carcinoma of the prostate (IDC-P) involved the bilateral posterior aspect of the prostate (orange, **Figure S8A**); spatially distinct small GG1 tumors were present in the anterior aspect (green). Two samples were isolated from the overall GG5 tumor (PR-482 [sampled GG4 with extensive IDC-P] and PR-486 [sampled GG3 with extensive IDC-P]; cyan) and one sample each were isolated from distinct GG1 foci (PR-484 and PR-483). As shown in **Figures 3A & S8A**, both samples from the GG5 tumor with extensive IDC-P tumor were *T2:ERG* positive, while the samples from the smaller GG1 tumor foci were *T2:ERG* negative, confirming multiclonality between the GG5 and GG1 foci. As shown in **Figure S7C**, mxCCP score and *SCHLAP1* expression showed marked

differences between the GG1 and GG5 foci samples, while the samples from the GG5 tumors showed more variable mxGPS, mxGC and *PRCAT104* scores/expression.

In MF Case 6, we co-isolated informative samples (cyan) from 6 areas of a large, high grade pT3a tumor focus (overall Gleason score 4+5=9, GG5, orange in **Figure S9**), including 4 areas of GG5 (PR-408, PR-410-412; cyan), one area of GG3 in an area of extra-prostatic extension at the apex (PR-413; cyan), and one area of GG2 at the edge of the focus (PR-409; cyan). Multiple, histologically separate, GG1 tumors were present in the base (green and light green in **Figure S9**), with two samples taken from the largest GG1 focus (PR-406 and PR-407; green) for NGS. Lastly, we also sampled benign prostate tissue (mixed epithelium and stroma) in close proximity to the sampled GG1 focus (gray region). By mxDNAseq, somatic MED12 p.A1222P mutations were present exclusively in the two samples from the GG1 focus (p.V1223G/L and p.L1224F somatic mutations are recurrent in prostate cancer; cBioPortal), and not in the samples from the overall GG5 tumor focus (**Figure S9**). By mxRNAseq, *T2:ERG* fusions were detected in all samples from the GG5 focus, but not in the samples from the GG1 focus **Figures 3A & S9**. Together, these results clearly demonstrate multiclonality of these tumor foci. Importantly, mxRNAseq demonstrated clear differences in mxGPS scores, mxGC scores, *SCHLAPI* expression and *PRCAT104* expression across the GG1 and GG5 tumor foci (**Figure 4C**), with mxCCP scores showing variability in the GG1 focus samples. Also, of note, the GG2 region of the overall GG5 focus (sample PR-409) showed *PRCAT104* expression and mxGPS, mxCCP and mxGC scores more in keeping with a GG2 focus compared to the remaining GG3-5 regions (samples PR-408, PR-410-412).

Supplementary Figures

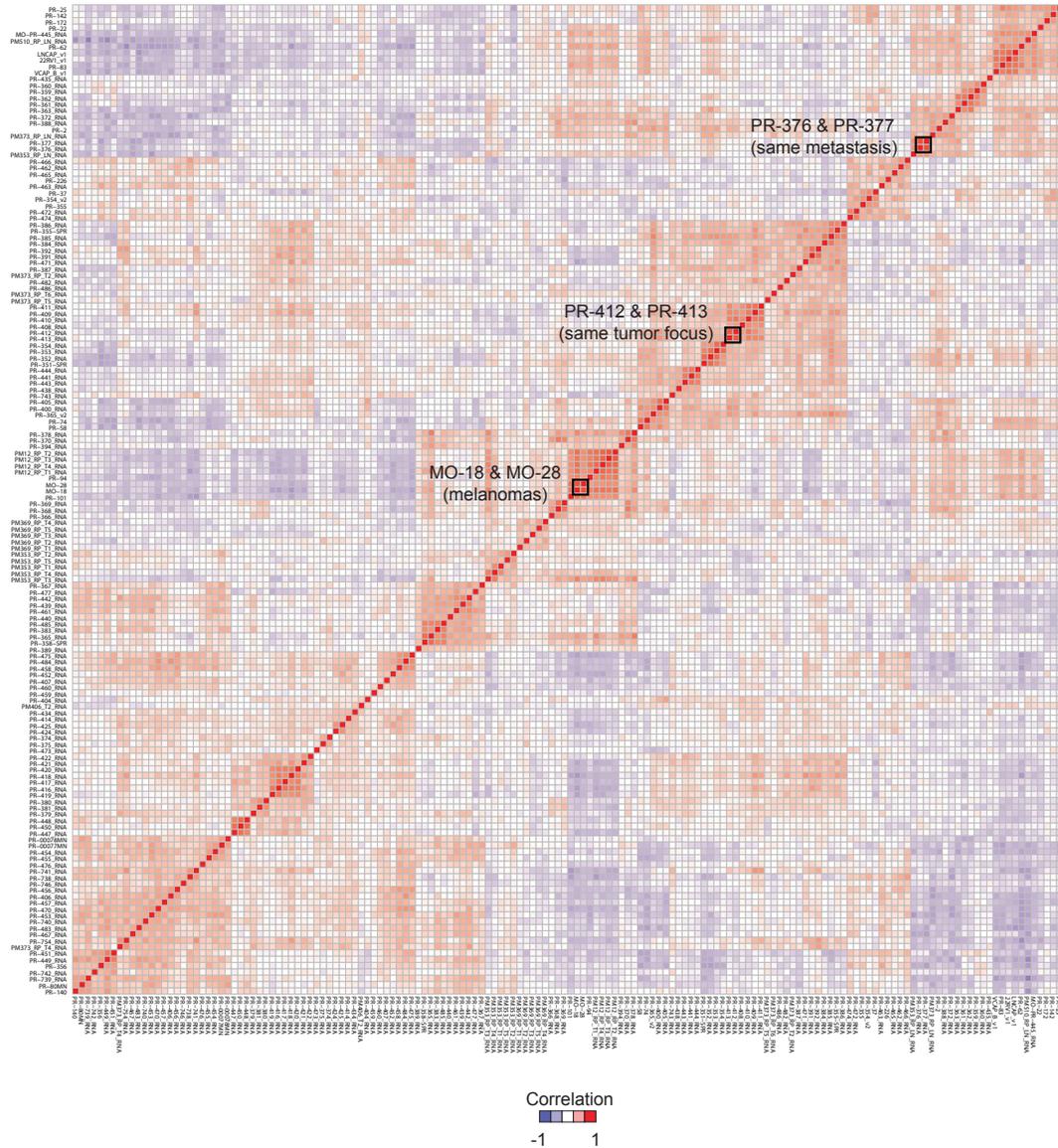


Figure S1. mxRNAseq sample-level target-gene correlation matrix. Heatmap displaying Pearson correlation coefficients (r) for 156 informative tissue samples profiled by mxRNAseq. Correlation coefficients were calculated from \log_2 normalized expression values for 223 high-quality target amplicons passing quality control filters. Samples are ordered from left to right (and bottom to top) by hierarchical clustering distance, with displayed correlation values colored

according to the legend at bottom (range: -1 to 1). The top three most highly correlated pairs of samples are highlighted, namely PR-376 and PR-377 ($r=0.85$, two samples from the same lymph node metastasis); MO-18 and MO-28 ($r=0.84$, two melanoma samples); and PR-412 and PR-413 ($r=0.82$, two samples from the same high grade [GG5] tumor focus).

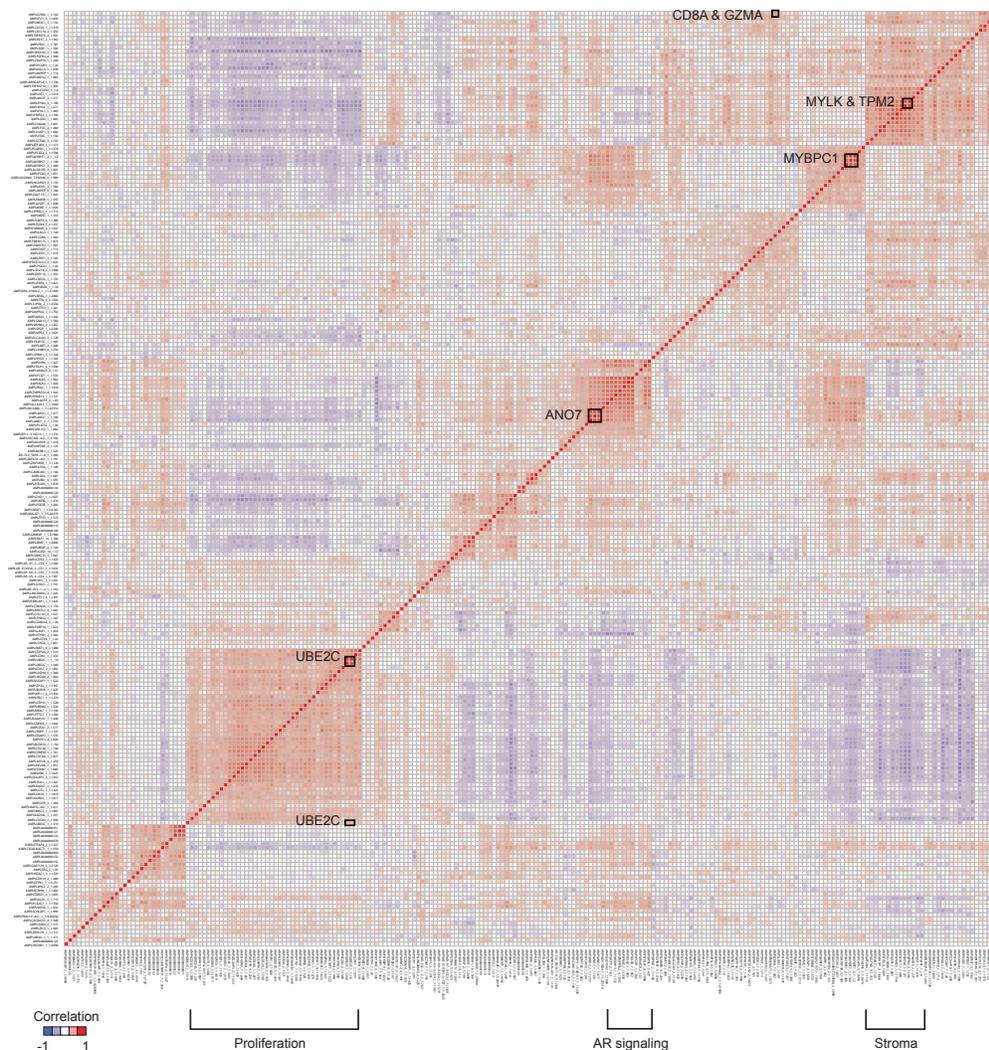


Figure S2. mxRNAseq amplicon-level correlation matrix. Heatmap displaying Pearson correlation coefficients (r) for 223 high-quality mxRNAseq target amplicons passing quality control filters. Correlation coefficients are calculated from \log_2 normalized expression values across 156 informative tissue samples profiled by mxRNAseq. Amplicons are ordered from left to right (and bottom to top) by hierarchical clustering distance, with displayed correlation values colored according to the legend at the bottom (range: -1 to 1). Highly inter-correlated targets from proliferation, androgen receptor (AR) signaling, and stromal transcriptional modules are

highlighted. For several transcripts included in the Decipher GC prognostic signature (*MYBPCI*, *ANO7*, and *UBE2C*), multiple amplicons targeting coding, non-coding and/or antisense transcripts demonstrated highly correlated expression across our full cohort. Several sets of highly correlated target amplicons for non-cancer tissue components are also highlighted, including smooth muscle restricted gene *TPM2* (most highly correlated with smooth muscle restricted gene *MYLK* ($r=0.90$)), and cytotoxic T lymphocyte marker *CD8A* [most highly correlated with cytotoxic T lymphocyte/natural killer cell specific protease *GZMA* ($r=0.55$)].

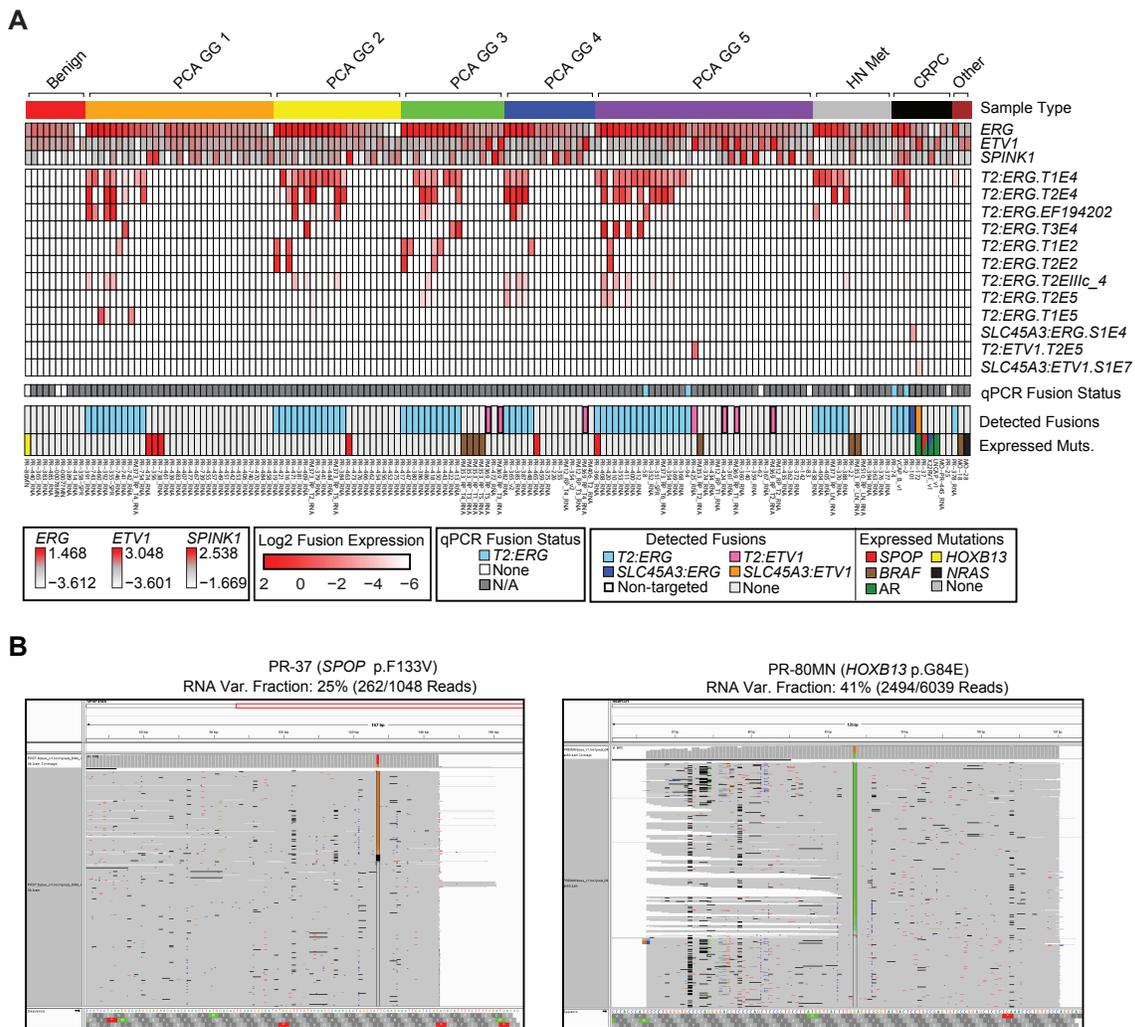


Figure S3: mxRNAseq detection of prostate cancer relevant gene fusions, expressed somatic mutations and germline risk variants. A) An integrative heatmap of mxRNAseq detected *ETS* gene fusions and expressed somatic/germline variants. Samples classes are as in **Figure 1**, with expression of key subtyping genes (*ERG*, *ETV1*, *SPINK1*) and individual gene fusion isoform expression (sorted vertically by descending total expression abundance). Below the heatmap, known fusion status obtained by prior qRT-PCR (or targeted RNAseq profiling) is indicated, along with mxRNAseq determined gene fusion status (bolded boxes indicate non-

targeted fusion isoforms detected by unbiased realignment of mxRNAseq reads). Somatic and germline mutations (Muts.) identified by targeted next-generation sequencing of co-isolated DNA are also indicated, with 20/21 (95.2%) detected by mxRNAseq profiling. **B)** Integrative Genomics Viewer (IGV) display of two representative expressed DNA alterations detected by mxRNAseq. On the left, a somatic *SPOP* F133V hotspot mutation seen at 25% variant fraction (262/1,048 total reads) by mxRNAseq profiling of sample PR-37 (CRPC) is shown, and on the right, a germline *HOXB13* G84E variant seen at 41% variant fraction (2,494/6,039 total reads) by mxRNAseq profiling of sample PR-80MN (normal prostate tissue).

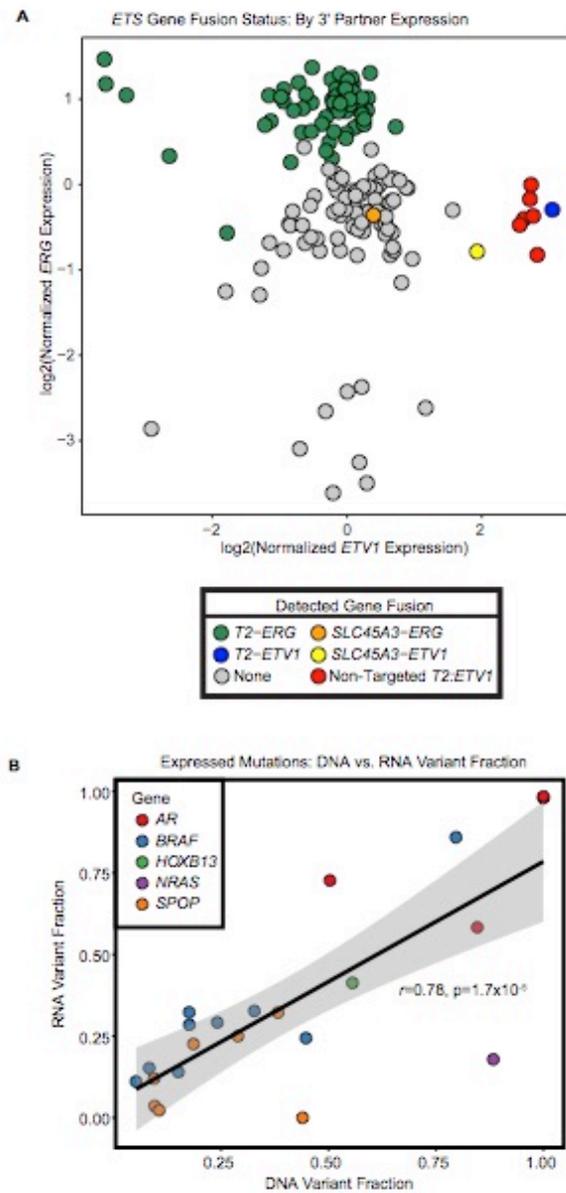


Figure S4. *ETS* 3' fusion expression and DNA and RNA variant fraction concordance for germline risk variant and somatic hotspot mutations. A. Scatter plot of log₂ normalized 3' amplicon *ETV1* vs. *ERG* expression is shown for all informative samples, with samples colored by detected *ETS* fusion isoform. Non-prespecified *T2:ETV1* fusions identified by realignment of

read-level data are shown in red. **B.** A dot plot summarizing observed DNA (x-axis) and RNA (y-axis) variant fractions for prioritized germline and/or somatic hotspot mutations originally detected by targeted DNA sequencing and assessed by mxRNAseq profiling. Overall correlation between observed DNA and RNA variant fractions ($r=0.78$, 1.7×10^{-5}), and linear trend line with 95% confidence interval is displayed. Points are colored by corresponding gene, with 20/21 (95.2%) mutations revealed by DNA profiling detected by mxRNAseq assessment.

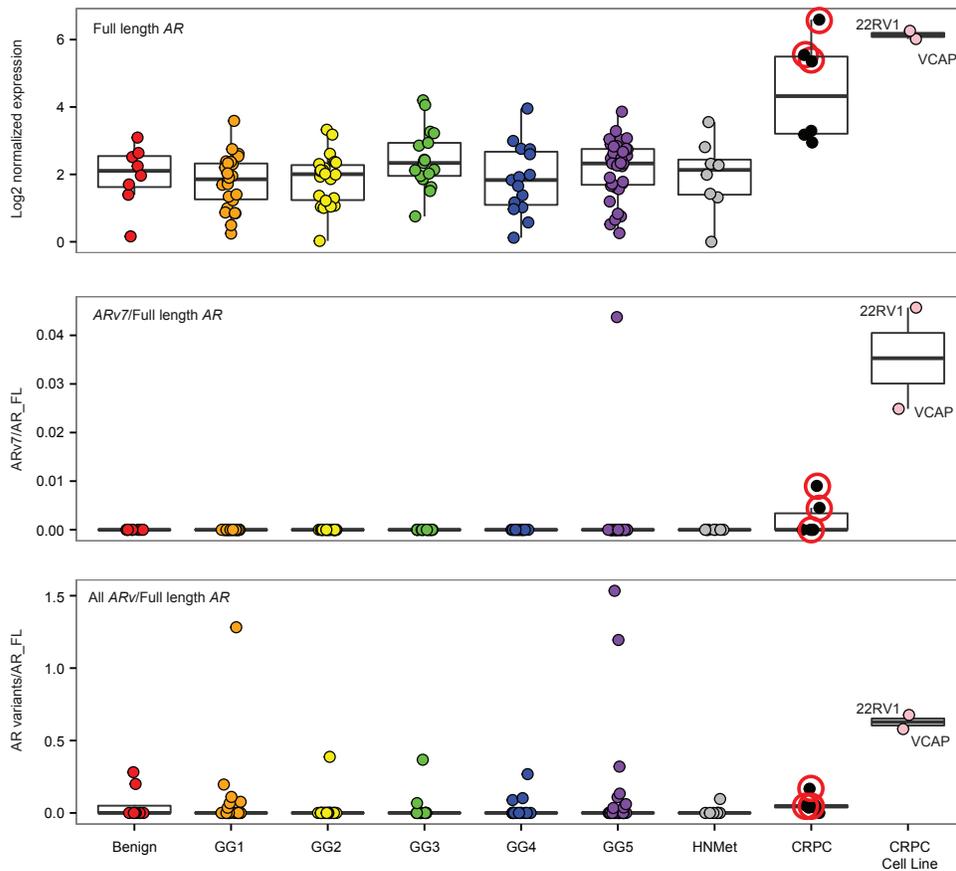


Figure S5. Androgen receptor full-length and splice variant expression. Summarized (from top to bottom) full length androgen receptor (*AR*), androgen receptor splice variant (*ARv7*), and aggregate *AR* splice variant expression across 137 profiled prostate tissue samples [including benign, grade group (GG) 1-5, hormonal therapy naïve metastases (HN met) and castrate resistant prostate cancer (CRPC)] and CRPC cell lines with detectable full-length *AR* expression. Each panel stratifies expression across GG/sample type, and samples with focal *AR* amplification by mxDNAseq are circled in red.

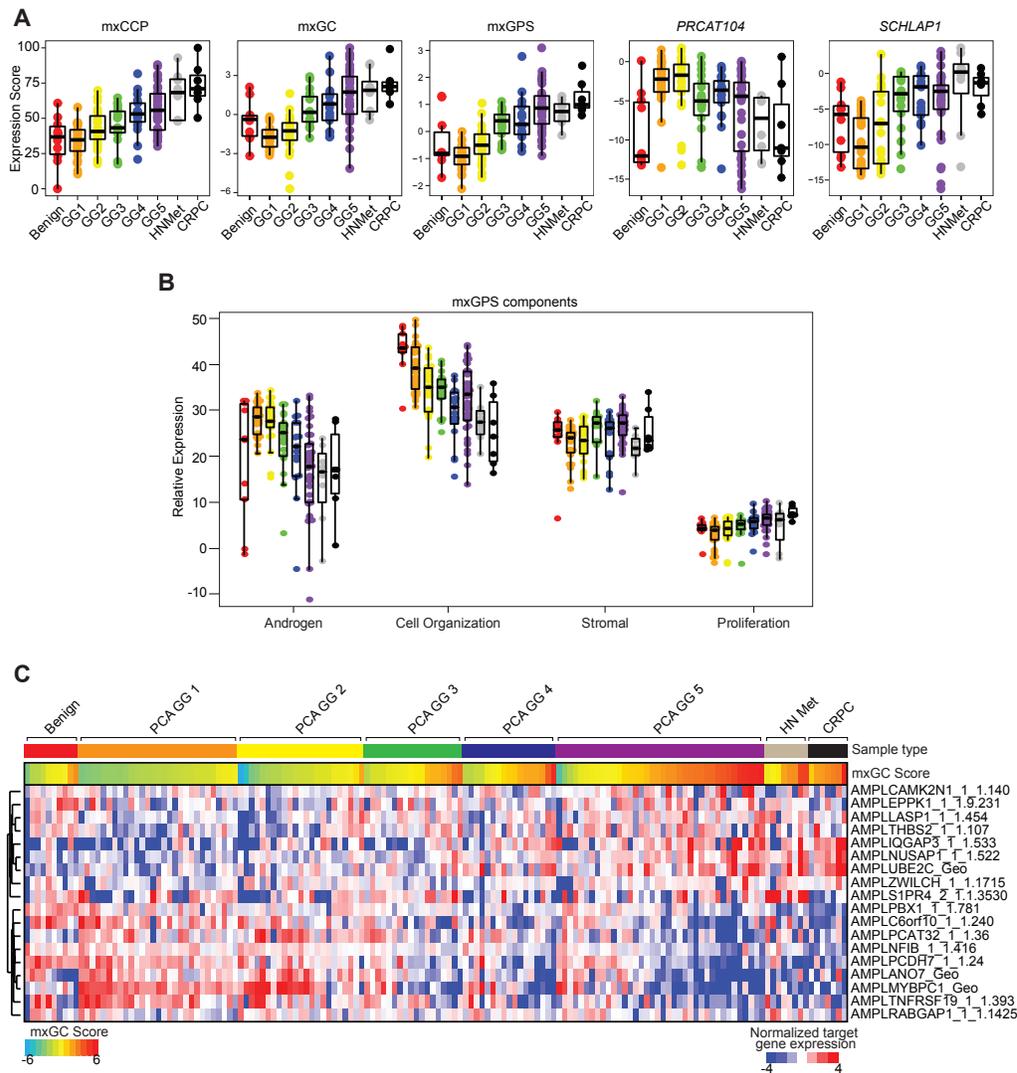


Figure S6. Stratified analysis of derived prognostic scores, including component modules and selected long non-coding RNA biomarkers. A) Derived Prolaris™ Cell Cycle Progression (mxCCP) score, GenomeDx Decipher™ Genomic Classifier (mxGC) score, and OncotypeDX™ Genomic Prostate Score (mxGPS) and expression levels of selected long non-coding RNA (lncRNA) biomarkers associated with aggressive prostate cancer (prostate cancer) are stratified by Grade Group (GG)/sample type for 156 informative samples [including benign, GG1-5,

hormonal therapy naïve metastases (HN met) and castrate resistant prostate cancer (CRPC)] profiled by mxRNAseq. **B)** Relative expression for individual mxGPS component modules from the 156 informative samples is shown across sample types. **C)** Unsupervised hierarchical clustering of mxRNAseq log₂ normalized expression values for 18 published components of GenomeDx Decipher™ GC assay. Samples are organized from left to right by ascending GG/sample type, and by increasing mxGC score (derived by average expression of over-expressed in prognosis targets [top cluster, *n*=9 targets] – average expression of under-expressed in prognosis targets [bottom cluster, *n*=9 targets]) within each sample category. Expression values for 15 of 18 components corresponds to values from single target amplicons, while expression values for *ANO7*, *MYPBC1*, and *UBE2C* represent average expression across multiple corresponding amplicons targeting coding, non-coding, and/or antisense regions.

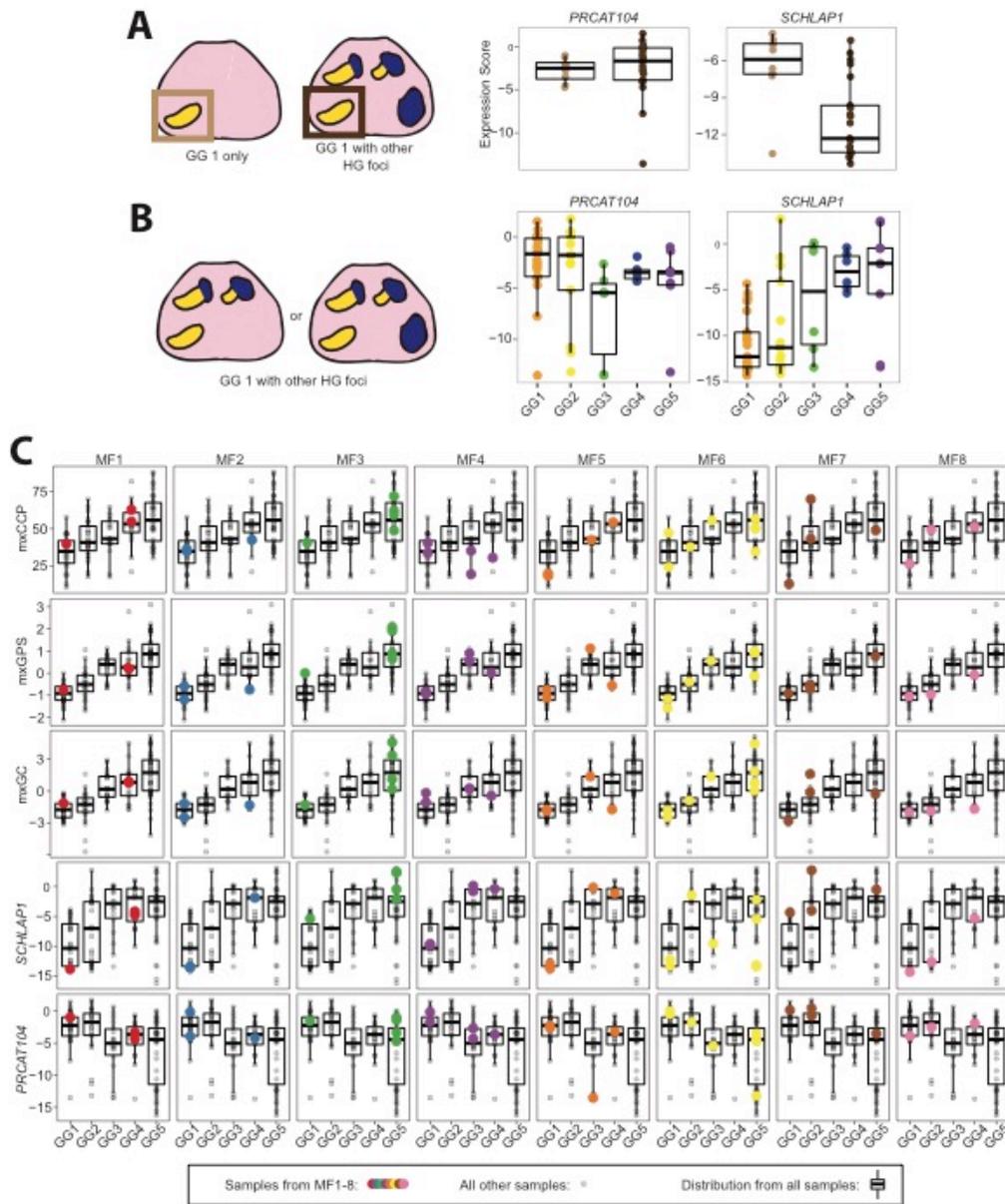


Figure S7. Derived prognostic scores are not robust to multifocal prostate cancer with extreme grade differences. A) Comparison of *PRCAT104* and *SCHLAP1* expression by mRNAseq from samples of pT2 GG1 prostate cancer without HG foci (clinically indolent, light brown, n=8 from 6 cases) vs. scores from GG1 prostate cancer foci (n=21 from 15 cases) where other HG foci were present. Two-sample, unpaired two-sided t-test p values are shown. **B)** *PRCAT104* and *SCHLAP1* expression are plotted for all samples from the 15 cases from **A** where

samples were taken from both GG1 prostate cancer foci as well as other concurrent higher grade foci. Samples were stratified by the GG of the profiled component. Spearman rank (r_s) correlation and p values are shown. C) Derived prognostic scores (mxCCP, mxGPS, mxGC) and expression levels for *PRCAT104* and *SCHLAPI* are stratified by GG and plotted for 125 primary prostate cancer specimens, with individual samples from 8 multifocal cases with extremes of grade differences (cases MF1-8; at least one sample from a focus of GG1 and at least one sample from a focus of \geq GG4) highlighted. For each case, profiled samples are indicated by colored points overlying the overall cohort distribution as in **Figure 3C**.

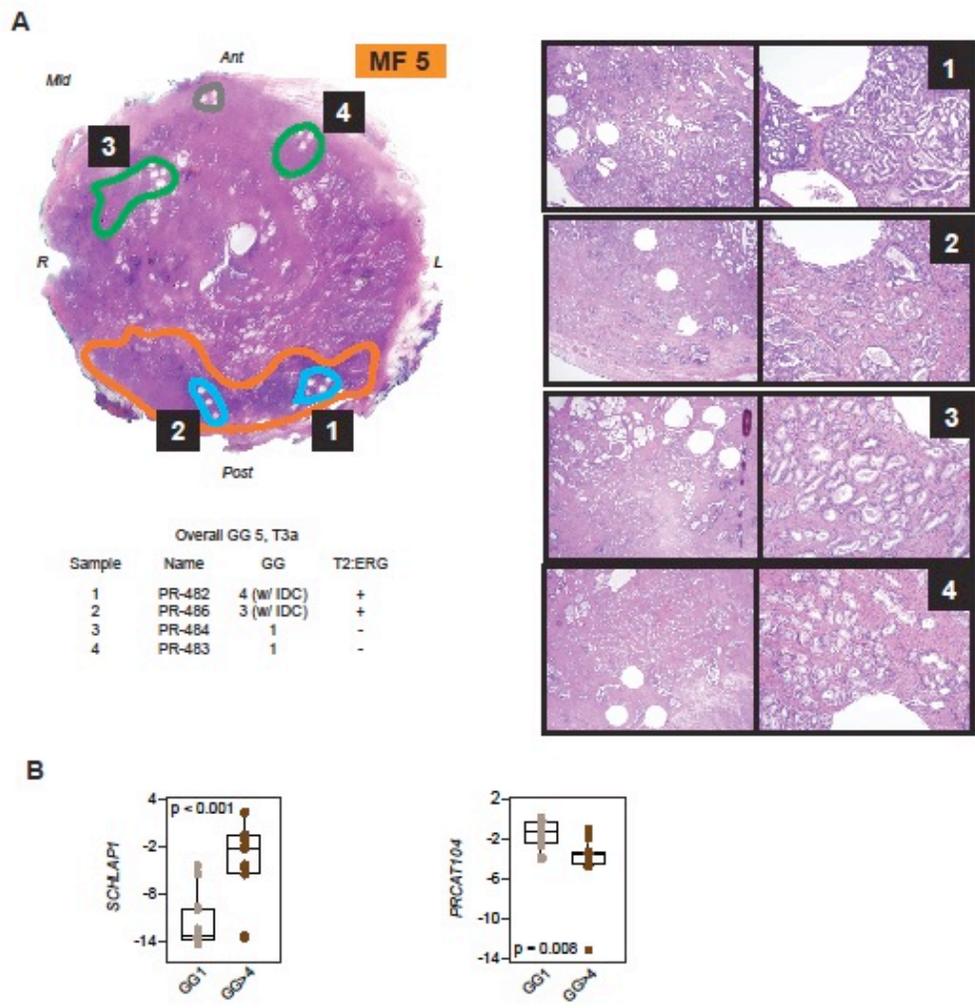


Figure S8. Histology and molecular subtyping of an extreme case of prostate cancer multifocality (MF 5). Histology and mxRNAseq support extreme multifocality in case MF 5,

which had a large posterior pT3a GG5 index tumor focus (orange) with prominent intraductal carcinoma (IDC) of the prostate, and two spatially distinct small anterior GG1 foci (green). Informative samples from the GG5 (cyan) and GG1 (green) foci are indicated (as well as a sampled area of normal prostate stroma in gray), with the table showing sample name, profiled morphologic GG, and *T2:ERG* fusion status by mxRNAseq. Histopathology of the indicated samples (original magnification 2x left; 10x right) are shown to the right. The whole mount section was taken after isolation of samples (by punching) to confirm sampling of the appropriate focus.

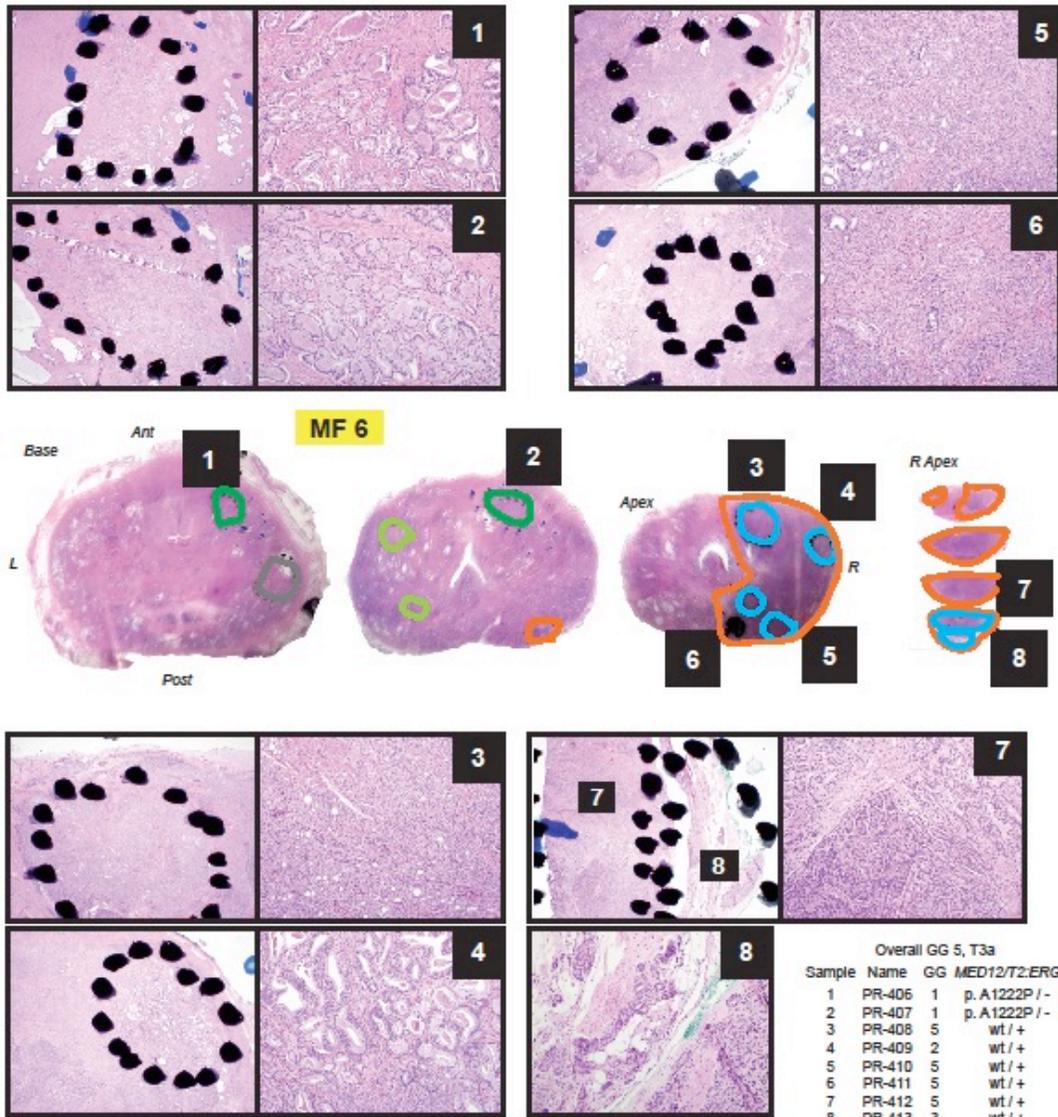


Figure S9. Histology and molecular subtyping of an extreme case of prostate cancer

multifocality (MF 6). Histology, mxDNaseq and mxRNAseq support extreme multifocality in case MF 6, which had a right sided pT3a GG5 index tumor focus (orange), and multiple spatially

distinct small anterior GG1 foci (green and light green). Informative samples from the GG5 focus (cyan, including an area of extraprostatic extension in the apex [Samples 7 and 8; PR-412 and PR-413) and the largest GG1 focus (green) are indicated (as well as a sampled area of normal prostate epithelium and stroma in gray). The table shows sample name, profiled morphologic GG, and *MED12* mutation status by mxDNAseq/*T2:ERG* fusion status by mxRNAseq. Histopathology of the indicated samples (original magnification 2x left; 10x right) are shown. **B)** Expression levels of single prognostic biomarkers are not robust to multifocal prostate cancer with extreme grade differences. Expression of *SCHLAP1* and *PRCAT104* from all GG1 samples vs. GG \geq 4 tumor foci from the 8 multifocal cases with extreme grade group differences shown in Figure 3C and 4A are plotted (two-sample, unpaired two-sided t-test p values are shown).