Supporting Information

Hieronymus et al. 10.1073/pnas.1411446111

SI Materials and Methods

Contemporary Cohort Sample Collection and Clinical Annotation. A total of 104 tumor samples and matched normal samples were obtained from patients treated by prostatectomy at Memorial Sloan–Kettering Cancer Center, and the study was conducted with Institutional Review Board approval (HBUC HBS2011066, IRB WA0369-11). Following radical prostatectomy, patients were followed with history, physical examination, and serum prostate-specific antigen (PSA) testing every 3 mo for the first year, 6 mo for the second year, and annually thereafter. Biochemical recurrence (BCR) was defined as an increase in PSA of ≥0.2 ng/mL on two occasions. The 2005 Stephenson nomogram for postoperative risk of BCR at 5 y was used (1). Patient follow-up data used here was updated through March 2013 for both the initial and contemporary cohorts.

Tumor Dissection, DNA Extraction, and Array Hybridization. Radical prostatectomy specimens were collected during surgery, snap-frozen in liquid nitrogen, and stored at −80 °C. Specimens were dissected from the frozen block for tumor samples containing greater than 70% tumor cell content and adjacent normal. In the majority of cases only one focus was dissected, but where there was insufficient tumor, tumor was pooled from two to three foci. The dominant focus or foci with highest Gleason sum score (8–10) or stage determining (pT3 and above) was used for selection. DNA was extracted (using phenol/chloroform/isoamyl ethanol after proteinase K digestion), digested, random primed (Bioprime; Invitrogen), labeled, and hybridized to the Agilent 1M-feature array comparative genomic hybridization (aCGH) platform according to the manufacturer's instructions, using pooled reference DNA (Promega) as the normal control. The only exception was the four formalin-fixed, paraffin-embedded (FFPE) biopsy cases, which were hybridized to the 244 K aCGH using the Agilent FFPE protocol.

DNA Copy Number Analysis. All copy number array data from patients in the contemporary cohort were quantified, normalized, segmented, and analyzed with RAE, all as previously described (2). To assess the percentage of autosomal genome affected by copy number alteration (CNA) in both the initial and contemporary cohorts, the per-tumor parameterization by RAE was used (3). Briefly, in every tumor RAE independently and adaptively infers the diploid component of copy number based on the total distribution of segmentation genome-wide, allowing for different profiles of signal-to-noise, purity, and clonality in heterogeneous cohorts of tumors. RAE then builds in each tumor four discriminators of discrete classes of CNA (gain, amplification, loss, and deletion: A_0, A_1, D_0, and D_1). For the analyses described here, CNAs are those segments whose value of A_0 or D_0 (gains and losses or more, respectively) were greater than or equal to 0.75. The total genomic territory spanned by these contiguous segments (in the event of focal alterations embedded among larger lower-level CNAs) was summed and a percentage was generated using the size of the autosomal human genome (chromosomes 1–22). The resulting CNA burden in either the initial or contemporary cohorts were similar when putative germ-line copy number polymorphisms were excluded, because of their minority contribution to overall CNA burden in the presence of large-scale (arm-length) somatic CNAs in prostate cancer genomes. For outcome analyses performed using a stratification of cases on the basis of their CNA burden, we grouped cases with greater or less than 5.41% of their autosomal genome affected by CNA, which is the mean CNA burden observed across the prostate cancer genomes that defined the intermediate alteration/risk clusters (1, 3, and 4) we previously inferred in the initial cohort data (2).

Statistical Analyses. For cohort characteristics, P values were determined by Wilcoxon rank sum for continuous variables and by Fisher's exact test for categorical variables. We tested for a univariate association between CNA and BCR or metastasis in the initial and contemporary cohorts separately. CNA was then added to a multivariable Cox proportional hazards regression model, including a postoperative nomogram for the risk of BCR (preoperative total PSA, pathologic grade, year of surgery, extracapsular extension, seminal vesicle invasion, lymph node invasion, and positive surgical margins). The postoperative risk was estimated based on the 2005 Stephenson nomogram for postoperative risk of BCR at 5 y (1). We repeated this process for the subset of patients with a pathologic Gleason grade of 7. The proporitional hazards model was not checked, as our view is that formal testing of the proportional hazards assumption is of limited utility. It is known that for outcomes such as BCR or metastases after treatment that the proportional hazards assumption is not met with long follow-up time. We see no reason to call into question the proportional hazards assumption; we find no large deviations of the assumption in the survival curves generated. For Kaplan–Meier analyses, follow-up time was censored to 14.5 y and the Mantel–Cox log rank significance is reported. We also repeated this process based on a CNA burden stratified at cut-points of 5.41% and 1.34%, which represent the mean CNA burden from patient genomes that defined the intermediate and low alteration clusters reported previously (2) (clusters 1, 3, and 4 vs. 2, respectively). To assess the discriminative accuracy of the models developed, we calculated Harrell's C concordance statistic after 10-fold cross-validation. Analyses were conducted in SPSS, Stata 12, and R.

The nonproprietary cell cycle progression (CCP) signature was defined as the average of the log gene-expression level (Affymetrix platform) of the 30 CCP genes (*ASF1B, ASPM, BIRC5, BUB1B, CDCA3, CDCA8, CDK1, CDK10, CDKN3, CENPF, CENPM, CEP55, DLGAP5, DTL, FOXM1, KIAA0101, KIF11, KIF20A, MCM10, NUSAP1, PBK, PLK1, PRC1, PTTG1, RAD51, RAD54L, RRM2, SKA1, TK1, TOP2A*) with available expression data in our cohort; the CCP signature score used for multivariate cox proportional hazards regression analyses was defined as the percentile CCP signature value, using the outcome of BCR. Although the current cohort could not be analyzed with the commercial CCP test because of consent and other limitations, the gene expression represented by the CCP signature could be compared with CNA burden in the initial cohort for which RNA expression data exists.

Low-Input Whole-Genome Sequencing. FFPE biopsy tumor and matched adjacent normal DNA was obtained from four patients with prostate cancer. Whole-genome sequencing libraries were prepared using 100-ng or 250-ng input dsDNA (qubit/picogreen quantified). Paired-end 100-bp libraries were generated using the KAPA LTP Library Preparation Kit for Illumina sequencing platforms (Kappa Biosystems) and sequenced on HiSEq. 2000 instrumentation according to manufacturer's instructions. After sample demultiplexing, raw sequencing reads were aligned to the hg19 build of the human reference genome with BWA (4). Libraries were de-duplicated, SAM (5) files merged, read groups

enforced, and alignment metrics determined with the Picard suite. For DNA copy number analysis, the aligned first read in the pair was retained and coverage calculated in 100-kb nonoverlapping windows of the autosomal genome in which G+C% content and mappability were also determined. In each genome, a locally weighted polynomial regression was fit between read coverage and G+C% content (30–70%) in alignable regions not overlapping known gaps in the reference assembly. Because of the variability of G+C% profile among the individual genomes sequenced here, we performed a G+C%-content normalization using the loess fit to normalize read counts in each independent tumor and matched normal genome before segmentation. A ratio of DNA copy number was inferred in tumor and matched normal genomes of identical input analyte in all cases except for two 100-ng matched normal libraries that failed QC and sequencing. In this case, DNA copy number ratios were inferred for both the 100-ng and 250-ng tumor libraries from the 250-ng libraries, respectively. A pseudocount was added to read counts genome-wide in both samples, the ratio was normalized by library size, and segmented with circular binary segmentation (α = 0.01, nperm = 10,000, undo.SD = 3). Individual samples were then analyzed with RAE to determine regions of copy number alteration, as described above.

Data Access. Study array data were deposited in National Center for Biotechnology Information Gene Expression Omnibus under accession number GSE54691. The segmented and normalized DNA copy number data as well as all clinical annotation for both cohorts are also available via the Memorial Sloan–Kettering Cancer Center Prostate Cancer Genomics Data Portal: http://cbio.mskcc.org/prostate-portal.

1. Stephenson AJ, et al. (2005) Postoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy. *J Clin Oncol* 23(28):7005–7012.
2. Taylor BS, et al. (2010) Integrative genomic profiling of human prostate cancer. *Cancer Cell* 18(1):11–22.
3. Taylor BS, et al. (2008) Functional copy-number alterations in cancer. *PLoS ONE* 3(9): e3179.
4. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.
5. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

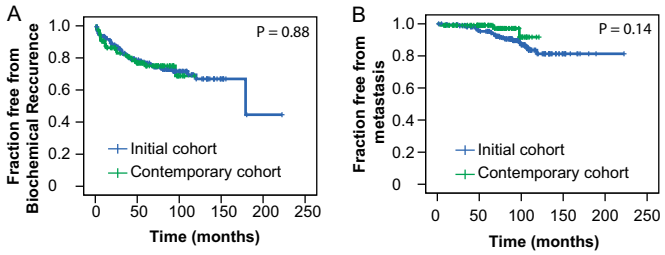

Fig. S1. Initial and contemporary cohorts do not differ in their times to BCR. (*A*) Kaplan–Meier plot for BCR in the initial and contemporary cohorts. (*B*) Kaplan–Meier plot for metastasis in the initial and contemporary cohorts. Mantel–Cox log-rank significance is shown.

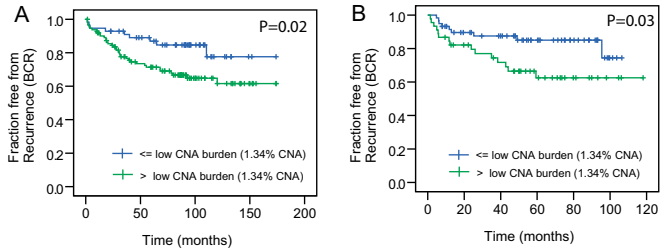

Fig. S2. CNA burden is associated with recurrence even at low alteration levels. Kaplan–Meier plots for BCR in the (*A*) initial cohort and (*B*) contemporary cohort are shown. Cases are stratified at low CNA burden (1.34% CNA, the mean of low CNA clusters in ref. 2). Strata with CNA burden greater than (green) or less than or equal to (blue) low CNA burden are shown. The Mantel–Cox log-rank significance value is shown for each.

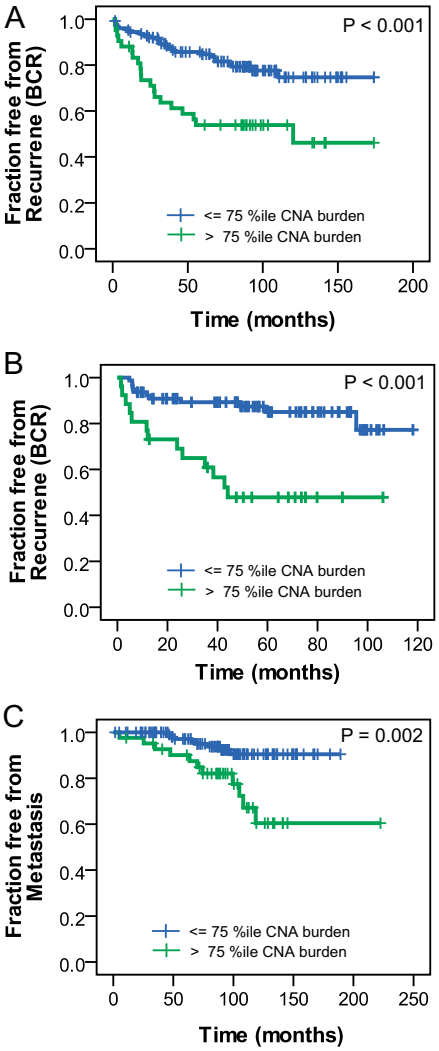

Fig. S3. CNA burden is associated with BCR and metastasis at 75th percentile CNA burden. Kaplan–Meier plots for BCR in the (*A*) initial cohort and (*B*) contemporary cohort are shown. (*C*) Kaplan–Meier plot for metastasis in the initial cohort is shown. Strata with CNA burden greater than (green) or less than or equal to (blue) 75th percentile of the CNA burden in the cohort are shown. The Mantel–Cox log-rank significance value is shown for each.

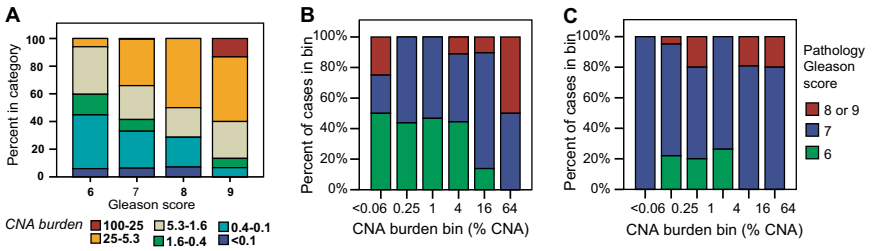

Fig. S4. A broad range of Gleason scores are seen across CNA burden. (*A*) Histogram of CNA burden across Gleason score for the initial and contemporary cohorts combined. (*B*) Histogram of Gleason scores is shown over a range of copy number alteration bins (CNA burden). Histogram for the initial cohort. (*C*) Histogram for the contemporary cohort. Gleason 6 (green), Gleason 7 (blue), and Gleason 8 or 9 (red) are shown.

Table S1. Cohort characteristics

Characteristic	Initial cohort (n = 168)	Contemporary cohort (n = 104)	P
Age (y)	58 (53, 63)	58 (53, 63)	0.9
Surgery type			
LP	48 (29%)	22 (21%)	0.037
RALP	0 (0%)	3 (2.9%)	
RP	120 (71%)	79 (76%)	
Pre-Treatment tPSA (ng/mL)	6.0 (4.5, 9.3)	5.3 (4.3, 8.0)	0.12
Adjuvant chemotherapy	6 (3.6%)	2 (1.9%)	0.7
Adjuvant hormonal therapy	20 (12%)	8 (7.7%)	0.3
Adjuvant radiation	16 (10%)	15 (14%)	0.2
Extracapsular extension	63 (38%)	52 (50%)	0.045
Lymph node invasion	11 (6.5%)	7 (6.7%)	1
Seminal vesicle invasion	18 (11%)	14 (13%)	0.6
Positive surgical margin	46 (27%)	17 (16%)	0.039
Gleason score at pathology			
6	51 (30%)	16 (15%)	0.022
7	98 (58%)	78 (75%)	
8	10 (6.0%)	4 (3.8%)	
9	9 (5.4%)	6 (5.8%)	
Stephenson nomogram	0.95 (0.86, 0.97)	0.94 (0.87, 0.97)	0.8

The median follow-up time is 94.8 and 70.3 mo for the initial and contemporary cohorts, respectively. Values are displayed as median (IQR) and frequency (percentage). P values were determined by Wilcoxon rank sum test for continuous variables and Fisher's exact test for categorical variables. LP, laproscopic prostatectomy; RALP, robotic-assisted laproscopic prostatectomy; RP, radical prostatectomy.

Table S2. The association of CNA burden by quartile with the risk of BCR and metastasis in univariate Cox proportional hazards regression models

		Initial cohort*				Contemporary cohort†			
Event	Variable	P	Hazard ratio	95% CI		P	Hazard ratio	95% CI	
BCR	CNA burden								
	≤75th percentile	Ref.	—	—	—	Ref.	—	—	—
	>75th percentile	0.001	2.65	1.48	4.75	<0.001	4.43	1.97	9.94
BCR	CNA burden					0.004			
	≤25th percentile	Ref.	—	—	—	—	—	—	—
	25–50th percentile	0.440	1.56	0.50	4.85	0.947	0.95	0.24	3.82
	50–75th percentile	0.208	1.85	0.71	4.81	0.574	0.65	0.15	2.91
	>75th percentile	0.003	3.95	1.59	9.85	0.020	3.81	1.24	11.72
Metastasis	CNA burden					Insufficient events			
	≤75th percentile	Ref.	—	—	—				
	>75th percentile	0.004	3.81	1.53	9.48				
Metastasis	CNA burden					Insufficient events			
	≤25th percentile	Ref.	—	—	—				
	25-50th percentile	0.847	0.79	0.07	8.72				
	50-75th percentile	0.443	1.90	0.37	9.80				
	>75th percentile	0.033	5.17	1.15	23.35				

Ref., reference category.

*Initial cohort total n = 168, BCR n = 46 (43 per 1,000 person-years), metastatic events n = 19 (15 per 1,000 person-years).

†Contemporary cohort total n = 104, BCR n = 24 (53 per 1,000 person-years), metastatic events n = 3 (5 per 1,000 person-years).

Table S3. Multivariate models of BCR association with CNA burden and RNA-based CCP gene expression signature

Cohort	Variable	model P	P	hazard ratio	95% CI		*n* events	*n* total
Initial (full)	CCP signature (percentile)	<0.001	0.05	1.124	1.000	1.263	32	127
	CNA burden (percentile)		<0.001	1.097	1.049	1.148		
Initial cohort Gleason 7	CCP signature (percentile)	0.003	NS				17	72
	CNA burden (percentile)		0.004	1.099	1.030	1.172		

Dataset S1. Clinical data for the contemporary cohort

Dataset S1

Dataset S2. Updated clinical data for the initial cohort, nonneoadjuvant treated primary tumors used for survival analyses

Dataset S2

Dataset S3. Updated clinical data for the initial cohort, all primary tumors (not to be used for survival analyses)

Dataset S3

Dataset S4. Updated clinical data for the initial cohort, metastatic tumors (not to be used for recurrence analyses)

Dataset S4

Dataset S5. Whole-genome sequencing metrics for low-input FFPE biopsy

Dataset S5