

# Morphology-Predicted Large-Scale Transition Number in Circulating Tumor Cells Identifies a Chromosomal Instability Biomarker Associated with Poor Outcome in Castration-Resistant Prostate Cancer



Joseph D. Schonhoff<sup>1</sup>, Jimmy L. Zhao<sup>2,3</sup>, Adam Jendrisak<sup>1</sup>, Emily A. Carbone<sup>4</sup>, Ethan S. Barnett<sup>4</sup>, Melanie A. Hullings<sup>4,5</sup>, Audrey Gill<sup>1</sup>, Ramsay Sutton<sup>1</sup>, Jerry Lee<sup>1</sup>, Angel E. Dago<sup>1</sup>, Mark Landers<sup>1</sup>, Samuel F. Bakhoun<sup>3,6</sup>, Yipeng Wang<sup>1</sup>, Mithat Gonen<sup>7</sup>, Ryan Dittamore<sup>1</sup>, and Howard I. Scher<sup>2,8</sup>

## ABSTRACT

Chromosomal instability (CIN) increases a tumor cell's ability to acquire chromosomal alterations, a mechanism by which tumor cells evolve, adapt, and resist therapeutics. We sought to develop a biomarker of CIN in circulating tumor cells (CTC) that are more likely to reflect the genetic diversity of patient's disease than a single-site biopsy and be assessed rapidly so as to inform treatment management decisions in real time. Large-scale transitions (LST) are genomic alterations defined as chromosomal breakages that generate chromosomal gains or losses of greater than or equal to 10 Mb. Here we studied the relationship between the number of LST in an individual CTC determined by direct sequencing and morphologic features of the cells. This relationship was then used to develop a computer vision algorithm that utilizes CTC image features to predict the presence of a high (9 or more) versus low (8 or fewer)

LST number in a single cell. As LSTs are a primary functional component of homologous recombination deficient cellular phenotypes, the image-based algorithm was studied prospectively on 10,240 CTCs in 367 blood samples obtained from 294 patients with progressing metastatic castration-resistant prostate cancer taken prior to starting a standard-of-care approved therapy. The resultant computer vision-based biomarker of CIN in CTCs in a pretreatment sample strongly associated with poor overall survival times in patients treated with androgen receptor signaling inhibitors and taxanes.

**Significance:** A rapidly assessable biomarker of chromosomal instability in CTC is associated with poor outcomes when detected in men with progressing mCRPC.

## Introduction

Chromosomal instability (CIN), or more broadly genomic instability, is increasingly being recognized as an essential hallmark of cancer implicated in initiation, progression, metastasis, and therapeutic resistance (1, 2). CIN changes range from whole-chromosomal alterations to structural alterations that include chromosomal rearrangements and complex structural changes (3). Aneuploidy and complex karyotypes, which are direct consequences of CIN, are observed in up to 70% of all cancers including

30% of prostate cancers and 60%–80% of breast, colorectal, and lung cancers (3, 4). They have also been shown to play a causal role in cancer progression and the acquisition of drug resistance in experimental systems (4, 5). Furthermore, analyses of large cohorts of patients from The Cancer Genome Atlas have shown that chromosomal copy-number alteration predicts for an inferior survival and clinical outcomes across a range of tumor types including prostate, breast, endometrial, renal clear cell, thyroid, and colorectal cancers (6, 7).

Given that CIN is by definition a dynamic process that reflects a cancer cell's ongoing ability to acquire chromosomal alterations, direct measurements have been largely limited to the research setting. Proxies or markers of CIN that have been applied in a clinical setting include chromosomal number and structure variation by FISH, comparative genomic hybridization, and next-generation sequencing (8). Each method has limitations, including labor-intensiveness, high cost, and turnaround times that are too long to impact medical decision making. In addition, methods that rely on bulk sequencing of tumors can mask the cellular and clonal heterogeneity arising from CIN. Assessing CIN by sequencing at a single-cell level in a large number of cells in time to inform clinical decisions extends these challenges even further despite the technological feasibility of doing so.

In addition to the heterogeneity of the cell populations within a tumor, it is increasingly recognized that metastases at different sites in a patient often harbor different tumor subclones. The result is that a single-site tissue biopsy is likely to underestimate the genetic diversity of an individual patient's disease. Profiling circulating tumor cells (CTC) at the single-cell level has the potential to capture the genetic diversity of tumor subclones and, through serial sampling, to monitor

<sup>1</sup>Epic Sciences, San Diego, California. <sup>2</sup>Genitourinary Oncology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York. <sup>3</sup>Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, New York. <sup>4</sup>Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York. <sup>5</sup>Current affiliation: University of Texas Southwestern Simmons Comprehensive Cancer Center, Dallas, Texas. <sup>6</sup>Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, New York. <sup>7</sup>Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York. <sup>8</sup>Department of Medicine, Weill Cornell Medical College, New York, New York.

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

**Corresponding Author:** Howard I. Scher, Memorial Sloan Kettering Cancer Center, 1275 York Ave., New York, NY 10065. Phone: 646-888-4878; Fax: 212-988-0851; E-mail: [scherh@mskcc.org](mailto:scherh@mskcc.org)

Cancer Res 2020;80:4892–903

doi: 10.1158/0008-5472.CAN-20-1216

©2020 American Association for Cancer Research.

tumor evolution over time (9–12). This is of particular importance in the metastatic setting where genomic alteration rates can be high and can increase further as the disease progresses, as is the case in tumors with CIN.

Our objective was to develop a biomarker of CIN in CTCs that could be assessed rapidly so that the result could be used to inform a management decision in real time. To do so, we used the Epic Sciences CTC isolation platform, a nonselection-based approach that deposits all nucleated cells from a patient's blood sample onto pathologic test slides, followed by staining and analysis with fluorescent scanners that perform automated high-throughput imaging to detect individual CTCs. In previous work, we used this technology to develop an assay to detect the presence and localization of the androgen receptor (AR)-V7 splice variant in CTCs. The test result was used to inform the choice of a taxane or AR signaling inhibitor (ARSI) for men with progressing metastatic castration-resistant prostate cancer (mCRPC) about to start a second or greater line of therapy. Studied in two independent cohorts, use of the AR-V7 assay to inform the choice of treatment improved patient survival (13, 14), a demonstration of clinical utility that is rarely shown. In our current research, we combined machine learning and digital pathology analyses to evaluate imaging features of individual CTCs, including the sizes and shapes of cytoplasm and nucleus, AR and cytokeratin (CK) expression, and other features. The features were then used to develop and validate, analytically and clinically, a phenotypic classifier of CIN in CTCs.

The focus was the detection and determination of the number of large-scale transitions (LST), a genomic alteration defined as chromosomal breakages that generate chromosomal gains or losses of 10 Mb or more. The LST metric was chosen for its clear cross-platform definition and frequent use as a biomarker of structural CIN arising through homologous recombination deficiency (HRD) associated with the presence of BRCA1/2 mutations in breast cancer (15, 16). To do so, we developed an image-based algorithm that uses the features of individual CTCs as input to predict the number of LSTs (pLST) present in each cell determined by direct single-cell sequencing. A CTC with 9 or more LSTs was scored as having a high LST number. The technical feasibility of determining LST number in single CTCs by next-generation sequencing has been demonstrated (Fig. 1A; ref. 17), and to test performance, the algorithm was applied to a separate analytic validation cohort of blood samples from which the patient-level CIN biomarker was defined as the presence of 3 or more CTCs in a sample that are predicted to have 9 or more LSTs per mL of blood. Finally, we applied the pLST algorithm to a larger clinical validation cohort for the context of use as a prognostic biomarker.

## Materials and Methods

### Patients and clinical specimens

All studies were conducted in accordance with ethical guidelines according to the Declaration of Helsinki. Patients with a diagnosis of mCRPC in need of treatment for progressive disease provided written informed consent to one of two Institutional Review Board–approved biospecimen protocols, MSK 06-107 or MSK 12-245. A total of 391 patients donated one or more pretreatment blood samples for CTC analysis. All patients were being treated at MSK between September 2012 and August 2016, had progressing mCRPC, and were in need of a therapy change. In this study, all cells analyzed were collected from mCRPC patients' blood and analyzed directly. No cultured primary cells or cell lines were used.

### Cohorts

Patients were retrospectively assigned to one of three cohorts: training, analytic validation, or clinical validation (Fig. 1B). The training cohort sample data were used to develop the phenotypic algorithm (pLST) to predict the actual number of LSTs in a given CTC, LSTs being a surrogate measure of CIN (15, 17). The analytic validation cohort sample data were used to independently assess the performance of the phenotypic algorithm's predictions compared with the number of LSTs determined by sequencing. The clinical validation cohort sample data were used to determine the association of the analytically validated pLST-based biomarker to overall survival. All samples in the clinical validation cohort were drawn  $\leq 30$  days prior to the start of the new therapy.

### CTC identification and isolation for imaging and LST determination

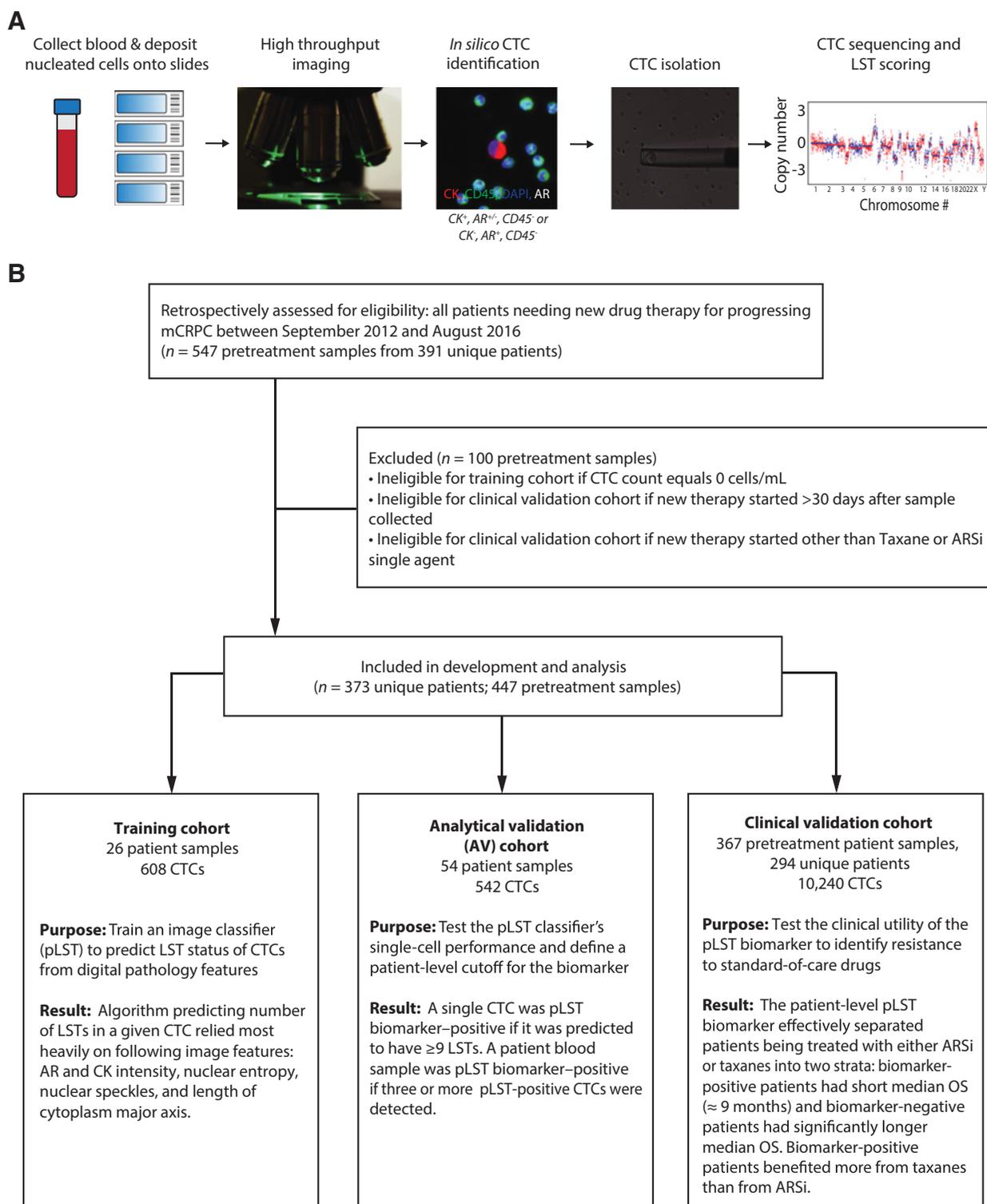
Detection and analysis of CTCs was performed using the Epic Sciences platform as described previously (17). In brief, nucleated cells derived from single blood draws were partitioned onto glass slides, fixed, and then stained for CD45, pan-CK, DAPI, and the N-terminus of the AR as described previously (13, 18). Each cell was imaged ( $\sim 3$  million per slide) and CTCs were identified *in silico* and confirmed by trained human technicians. A CTC was defined as a cell with an intact nucleus that was CK(+), AR( $\pm$ ), and CD45(-); or CK(-), AR(+), and CD45(-). For each sample, at least two slides corresponding to the analysis of approximately 1 mL of blood were analyzed. After imaging, each CTC selected for single-cell sequencing was isolated from the glass slide and the genomic DNA was whole-genome amplified and sequenced on an Illumina NextSeq500 to determine gene region–specific copy-number alterations and the number of LSTs. Additional details on the sequencing workflow are provided in the Supplementary Methods section.

### Image-based pLST algorithm training

The pLST algorithm was developed using the CTCs identified in training cohort samples. All identified CTCs detected in this cohort were digitally imaged and in parallel picked and subjected to single-cell sequencing to determine the true number of LSTs within each CTC. The cell image features [ $x_1, x_2, x_3 \dots x_n$ , (Supplementary Table S1)] were then used to train an algorithm that predicted the ground truth number of LSTs determined by single-cell sequencing (Y). CTC image digital pathology features were extracted from each  $6.8 \times$  CTC image as described previously (18). AR and CK protein expression were measured by fluorescent intensity from secondary antibodies and were utilized as the ratio of expression in a CTC relative to background (cRatio). All CTC image features utilized were  $\log_{10}$  and z-normalized before use.

A pipeline utilizing common supervised classification and regression algorithms was constructed that included Support Vector Machine, Neural Net, Random Forest, and generalized linear model (GLM), among others, to predict the number of LSTs from CTC image features. Ten-fold cross-validation and grid-search parameter tuning were used to train and test performance of each algorithm class. We selected a GLM regression model for its simplicity and relative performance to other models based on the AUC in receiver–operator curves in predicting the categorical LST-high versus LST-low status and by comparing goodness-of-fit using mean squared error for predicting the continuous LST number. Least absolute shrinkage and selection operator (LASSO) regression was used to quantify CTC image feature importance (Supplementary Fig. S1A and S1B). After training and cross-validation, the final model's parameters were locked.

Schonhoff et al.

**Figure 1.**

CTC detection platform and patient flow diagram. **A**, CTC detection, imaging, and LST scoring. Within 24 hours of collection, red blood cells are lysed and nucleated cells are deposited and fixed onto glass slides. Slides are then stained with antibodies against CK, CD45, and the AR N-terminus and DAPI. All cells including leukocytes are imaged. CTCs are defined as CK(+) CD45(−) AR(±), or CK(−) CD45(−) AR(+). Postimaging, single CTCs are isolated using a pipette tip and micromanipulator, then whole-genome amplified, and sequenced to determine the number of LSTs. **B**, Patient flow diagram describing use of the algorithm in the training, analytic validation, and clinical validation cohorts.

A CTC was defined as LST-high or chromosomally unstable if the number of predicted LSTs was greater than a cutpoint determined from the distribution of the number of LSTs in the entire cohort. Using this cutpoint, the number of true positives, true negatives, false positives, and false negatives were calculated and the performance of the classifier assessed using accuracy, sensitivity, and specificity metrics.

#### Analytic validation of the pLST algorithm and biomarker definition

The performance of the pLST classifier to classify a CTC as LST-high or LST-low was then studied in the analytic validation cohort by comparing the predicted LST number and the number determined by direct sequencing for accuracy, sensitivity, and specificity.

Next, to derive a patient sample-level scoring guide, a cutoff of the number of detected predicted LST-high single CTCs was chosen to maximize patient-level accuracy, sensitivity, and specificity in identifying the equivalent number of LST-high CTCs by sequencing within the blood analyzed for each patient. A cutoff expressed as the per mL of blood equivalent was then determined prior to clinical data analysis and served to stratify patients as pLST biomarker-negative or -positive.

#### Clinical validation of the pLST biomarker and statistical analysis

The pLST classifier was applied to all of the CTCs detected in the clinical validation cohort. If the patient had greater than the predefined cutoff of LST-high CTCs/mL, the patient was considered CIN biomarker-positive. Next, biomarker status, positive and negative patient, was studied in relation to overall survival. Overall survival was calculated from the start of therapy; patients lost to follow-up were

right censored. In the case of patients receiving multiple lines of therapy, survival times were right censored from the time of initiation of the new, most recent therapy. HRs were estimated using Cox regression and the probabilities of survival over time were estimated using the Kaplan–Meier method.

To assess the independent prognostic value of the CIN biomarker in relation to the total number of CTCs and to account for other potential confounding factors, a Cox proportional hazards model was constructed as described previously (13). The models included a combination of pretreatment clinical variables, a combined factor derived from CTC/mL  $\geq 3$  (median) and CIN biomarker status, and a two-way interaction term between therapy and biomarker status. All analyses were performed using R software (v3.6) along with the “survival” and “survminer” packages.

## Results

### Patient characteristics

Patients with progressing mCRPC needing to start a new systemic therapy contributed one or more pretreatment blood samples from September 2012 until August 2016. Among the distinct training, analytic validation, and clinical validation cohorts, a total of 447 blood samples were collected from 366 unique patients (Table 1; Fig. 1B). All samples were analyzed using the Epic Sciences platform (Fig. 1A). Samples from the training cohort were used to develop the phenotypic pLST (predicted Large-Scale Transition) scoring algorithm. Those from the analytic validation cohort were used to test the performance of the algorithm to determine the cell-level cutoff that defines the pLST biomarker, and those from the clinical validation cohort to associate the CIN biomarker result with outcomes following treatment with

**Table 1.** Patient and sample demographics by cohort and CTC detection frequency.

	Training cohort	Analytic validation cohort	Clinical validation cohort
Cohort size, age, and survival			
No. of samples	26	54	367
No. of unique patients	26	46	294
Age at baseline, median (range), years	70 (48–91)	70 (49–86)	69 (40–89)
No. of patients with unique death events (%)	N/A	N/A	175 (59.5)
Next therapy—no. of samples (%)			
ARSi	14 (54)	18 (33)	245 (67)
Taxane	8 (31)	11 (20)	122 (33)
Radium <sup>a</sup>	2 (8)	0 (0)	0 (0)
Platinum <sup>a</sup>	2 (8)	4 (8)	0 (0)
Other	0 (0)	21 (39)	0 (0)
Therapy line—no. of samples (%)			
First-line treatment	6 (23)	21 (39)	129 (35)
Second-line treatment	7 (27)	10 (19)	101 (28)
Third-line or greater treatment	13 (50)	23 (43)	137 (37)
Pretreatment baseline lab values			
Albumin, median (range), g/dL	4.1 (3.6–4.8)	4.1 (3.2–4.8)	4.2 (2.4–4.6)
Hemoglobin, median (range), g/dL	10.7 (7.0–14.3)	12.2 (9.2–14.6)	12.3 (7.1–151)
Lactate dehydrogenase, median (range), U/L	257 (123–976)	228 (124–890)	218 (52–2115)
PSA, median (range), ng/mL	88.2 (0.1–947.0)	15.1 (0.05–3022)	38.33 (0.06–16275)
Alkaline phosphatase, median (range), U/L	128.5 (51–1043)	104 (28–511)	108 (40–2170)
Presence of liver and/or lung metastases, no. of samples (%)	3/26 (11.5)	12/54 (22.2)	58/367 (15.80)
CTC Detection			
Total CTC/mL, median (range)	20.7 (3.6, 934.3)	7.75 (1.1, 595)	3 (0, 1149.9)

Abbreviations: ARSi, androgen receptor signaling inhibitor; N/A, not applicable; PSA, prostate-specific antigen.

<sup>a</sup>Single agent or combination.

Schonhoff et al.

standard of care–approved drugs, including ARSi and taxane-based chemotherapy. All patients gave signed informed consent to an IRB-approved protocol before starting study procedures.

#### CTC detection and LST determination in the training cohort

From 26 patients, 768 CTCs representing a range of sizes, shapes, and protein expression levels of the N-terminal AR and CK were identified, isolated, and submitted for DNA whole-genome amplification and low-pass sequencing to determine the number of LSTs within each cell based on the number of contiguous regions of chromosomal breakage of at least 10 Mb. The overall success rate was 79% (608 of the 768 cells submitted). The method utilizes commercially available reagents and was previously developed using LNCaP, PC3, and VCaP prostate cancer cell lines along with mCRPC patient samples (17). Contiguous regions of chromosomal breakage of at least 10 Mb were used to calculate the number of LSTs in CTCs. The imaging features were then used to train a pLST algorithm to predict the number of LSTs detected by single-cell genomic sequencing (Fig. 2). Representative images of LST-low and LST-high CTCs are shown in Fig. 2A. The results from the 608 CTCs sequenced in the training cohort showed an approximate bimodal distribution of LSTs comprising an LST-low group (0–8 LSTs) and an LST-high group (9 or more LSTs) per cell (Fig. 2B). The observed bimodal distribution of LSTs in CTCs is consistent with the distribution found in tumor tissues obtained from patients with HRD-associated cancers (breast, ovarian, and prostate) sequenced with the FDA-approved MSK-IMPACT assay. Notable is that HRD-associated tumors with *BRCA1/2* loss of function predominantly had 9 or more LSTs, while tumors from other lineages and those with intact *BRCA1/2* status predominantly had LST numbers near or at zero (19).

#### pLST Algorithm training

Next, 19 digital pathology features were extracted from each segmented and cropped 4-channel cell image (CD45, CK protein, DAPI, AR) that describe the size, shape, image texture, and protein expression within the cytoplasm and nucleus (Fig. 2B). Significant correlations were observed between the CTC image features and number of LSTs determined by single-cell whole-genome sequencing (Fig. 2B; Supplementary Fig. S2). CTCs with a high number of LSTs generally, although not exclusively, trended to smaller, more circular, and to express higher levels of AR and CK proteins, and have a higher nuclear texture or heterogeneity as measured through the entropy metric within the DAPI channel. Entropy is an image texture feature that measures pixel intensity randomness, where a high entropy broadly indicates a higher heterogeneity in image texture (Supplementary Fig. S3).

The observed relationship of CTC image features and number of LSTs in a cell made it possible to develop a computer vision algorithm using the image features to predict the number of LSTs (Fig. 2C). To do so, a GLM was selected because of its simplicity and ability to directly determine the relative contribution of each feature to the final model. The GLM model uses multivariate linear regression and provided comparable performance to other linear and nonlinear models tested including a Neural Net, Support Vector Machine, and Random Forest (20). Once training and cross-validation were completed, LASSO regression was used to rank CTC image feature importance (Supplementary Fig. S1). The top five CTC image features in the GLM-based classifier were AR intensity, CK intensity, nuclear entropy, nuclear speckles, and the major axis length of the cytoplasm. While these features were the most important contributors, it should be noted

that other features such as nuclear area or circularity were predictors but dropped out of the model during LASSO regression once the top ranked contributors were incorporated. The final locked algorithm predicted the number of LSTs on a continuous basis with an AUC of 0.83. Furthermore, when the cutoff of 9 or more was applied to define an LST-high CTC biomarker for potential clinical application, a single-cell accuracy of 77%, sensitivity of 68%, specificity of 84%, and positive predictive value (PPV) of 78% was observed (Fig. 2D).

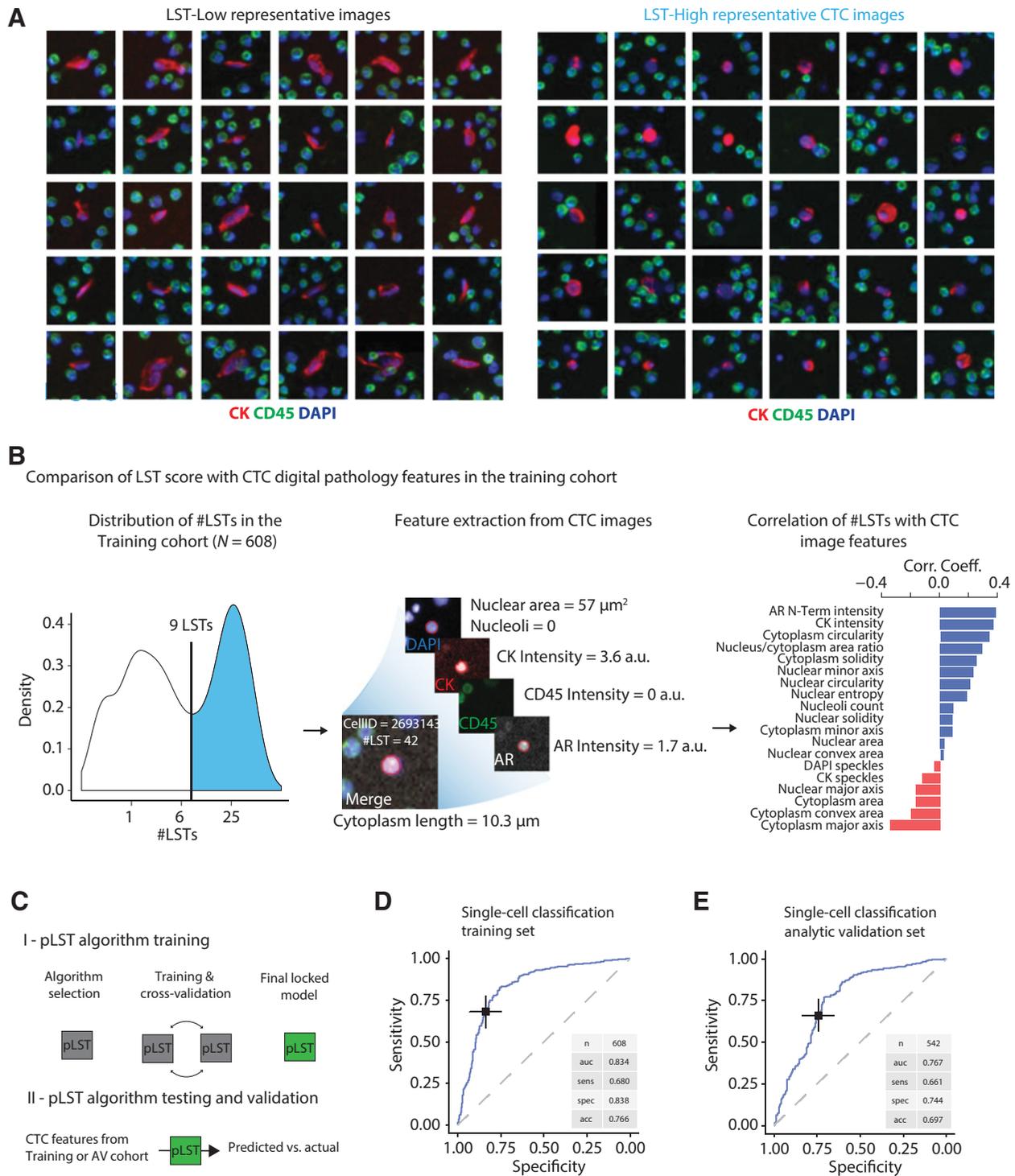
#### Analytic validation and performance of the pLST algorithm

The pLST classifier's performance was then evaluated in a separate analytic validation cohort that included 54 unique blood samples not used in the training cohort from 46 patients prior to starting a new line of therapy from which 542 CTCs were successfully sequenced (success rate = 87% of the 625 CTCs submitted; Table 1). Here, an AUC of 0.77 was observed and the locked pLST algorithm classified CTCs into the LST-high or LST-low groups with an accuracy of 70%, sensitivity of 66%, specificity of 74%, and PPV of 77% at the single-cell level (Fig. 2E). Summary results for both the AV and training cohort are presented in Supplementary Table S2.

#### Copy-number alterations in pLST-high and pLST-low CTCs

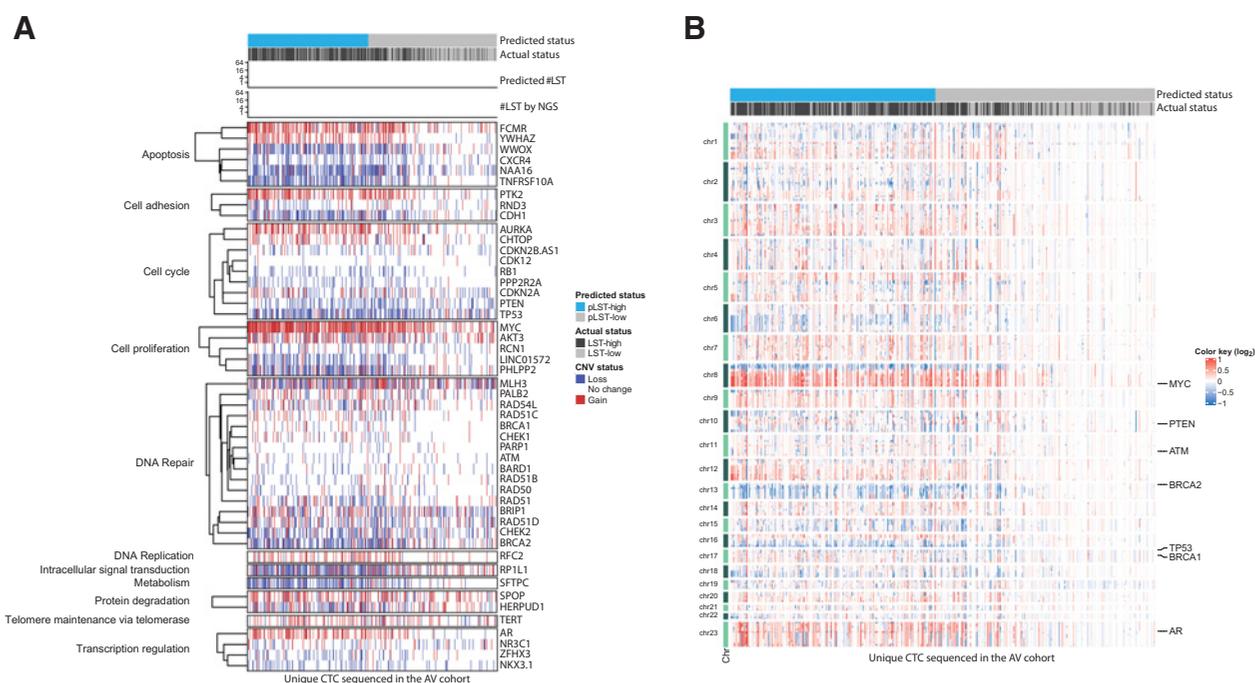
CTCs classified as pLST-high or -low in the analytic validation cohort were then studied for the relationship between the copy-number alterations found in gene regions involved in prostate cancer oncogenesis by direct sequencing. A total of 104 prostate cancer–related genes were qualitatively selected for analysis based on literature and database searches of known CNV altered genes (17). The percentage of CTCs harboring each gene region alteration is shown in Supplementary Fig. S4. With the limitations of low-pass sequencing in mind, gains or losses in prostate cancer–related gene regions were determined using the previously published analysis pipeline developed in prostate cancer cell lines and mCRPC patient samples (17). For this analysis, 91.5% (496 of the 542) successfully sequenced CTCs were found to have at least one copy-number alteration in a prostate cancer–related gene region and Fig. 3A presents a heat map of the top 30 altered gene regions along with other prostate cancer gene ROIs, including those that regulate DNA replication, cell-cycle, proliferation, apoptosis, DNA repair, and prostate lineage specification, and their relation to actual and predicted LST numbers. Full genome CNV plots are also shown in Fig. 3B. That the copy-number losses and gains were in regions where prostate cancer–related genes are typically found supports the validity of our finding, and were unlikely to be due to random sequencing dropout or polymerase overamplification, a limitation of single-cell whole-genome sequencing technology. The result showed that, in the analytic validation cohort, 44% of the CTCs (115/262) classified as pLST-high and 23% (63/280) of the pLST-low had *BRCA2*, *BRCA1*, or *ATM* copy loss ( $P < 0.0001$ ). Whether this observation is causal and representative of functional HRD or a consequence of increased CIN in LST-high cells is not clear. Furthermore, 29% (76/262) of the pLST-high CTCs and 8% (21/280) of the pLST-low CTCs had AR gene amplification ( $P < 0.0001$ ), aligning with the observation that the intensity of AR protein expression is also a significant predictor of high LSTs. A boxplot comparing LSTs, predicted LSTs, AR copy-number status, and protein expression intensity is presented in Supplementary Fig. S5. Ongoing in separate work is the investigation of the relationship between AR amplification in tissue and ctDNA to that of CTCs. Other genes of interest that showed association with

## Chromosomal Instability in Circulating Tumor Cells

**Figure 2.**

Training and analytic validation of the pLST classifier. **A**, Image gallery of representative CTC images in the LST-low ( $\leq 8$  LSTs) and LST-high ( $\geq 9$  LSTs) groups. **B**, Extraction of CTC image features and comparison of those features with the number of LSTs in each CTC. Pearson correlation coefficients are shown. **C**, Schematic of the development of phenotypic LST-high classifier. **D**, Performance of the pLST algorithm in the training cohort. **E**, Performance of the pLST algorithm in the analytic validation cohort. The cross-hair in each ROC curve denotes the pLST-high/low cut-off point from which sensitivity, specificity, and accuracy were calculated.

Schonhofs et al.

**Figure 3.**

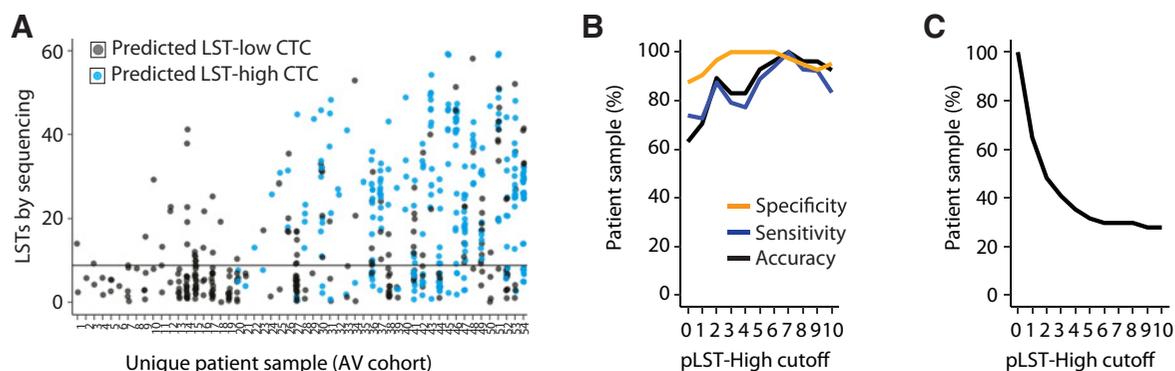
Copy-number variations in CTCs single cell sequenced in the AV cohort. **A**, Heatmap of prostate cancer-related gene copy-number variations in single CTCs sequenced in the analytic validation cohort and comparison with predicted and actual LST number. **B**, Heatmap of alterations across all chromosomes.

pLST-high classification were *AURKA* and *MYC* gene amplification. A heatmap identical to **Fig. 3** using CTCs in the training cohort is presented in Supplementary Fig. S6.

#### A patient-level scoring guide to define the CIN biomarker-positive patient strata

We then went on to determine a cutpoint for the number of pLST-high CTCs in an individual sample needed to score a sample as pLST biomarker-positive and pLST biomarker-negative (CIN-positive or CIN-negative). For each patient at a given pLST-high

CTC cutpoint, a sample was reported as true positive when both the predicted number of pLST-high CTCs and the actual number of LST-high CTCs by single-cell DNA sequencing were greater than the cutpoint. This two-component CIN biomarker was needed because the PPV of the algorithm to identify true LST-high CTCs was only 76% in the analytic validation cohort. However, when a specific cutoff for the number of LST-high CTCs present in a sample was included, the PPV increased to over 95% at the patient level after more than one pLST-high CTC was detected (**Fig. 4A** and **B**).

**Figure 4.**

Development of a patient-level scoring guide to define the CTC-CIN biomarker. **A**, Comparison of the number of predicted LSTs with the actual number of LSTs by patient sample. An LST-high CTC is defined as a CTC having  $\geq 9$  LSTs, while an LST-low CTC has  $\leq 8$  LSTs. CTCs that were predicted to be in the LST-high group are colored in cyan, while those predicted to be in the LST-low group are colored in gray. Transparent colors are shown to make all data points visible. **B**, Patient-level accuracy, sensitivity, and specificity by the number of predicted LST-high CTCs used as the sample-level cutoff for biomarker positivity. **C**, Prevalence of biomarker positivity by the number of predicted LST-high CTCs/mL used as the sample-level cutoff. Using the cutoff of  $\geq 3$  predicted LST-high CTCs/mL, prevalence of pLST biomarker-positive samples was 41%.

Patient-level accuracy, sensitivity, specificity, and PPV were all found to be optimal when 3 or more LST-high CTCs were detected in a patient sample from the analytic validation cohort. Here it is important to note that a total of 6 million nucleated cells are analyzed for each patient (see Materials and Methods), corresponding to about 1 mL of blood. Requiring multiple observations to define the LST-high CTC phenotype improved performance metrics to the level needed for a biomarker assay result that could be used in a clinical practice setting to predict sensitivity to a drug class and inform treatment selection. At a cutoff of 3 or more LST-high CTCs per mL of blood, the patient sample-level accuracy was 86%, sensitivity 83%, specificity 100%, and PPV of 100% (Fig. 4B) and the overall frequency of patients predicted to have  $\geq 3$  LST-high CTCs per mL of blood exceeded 40% (Fig. 4C).

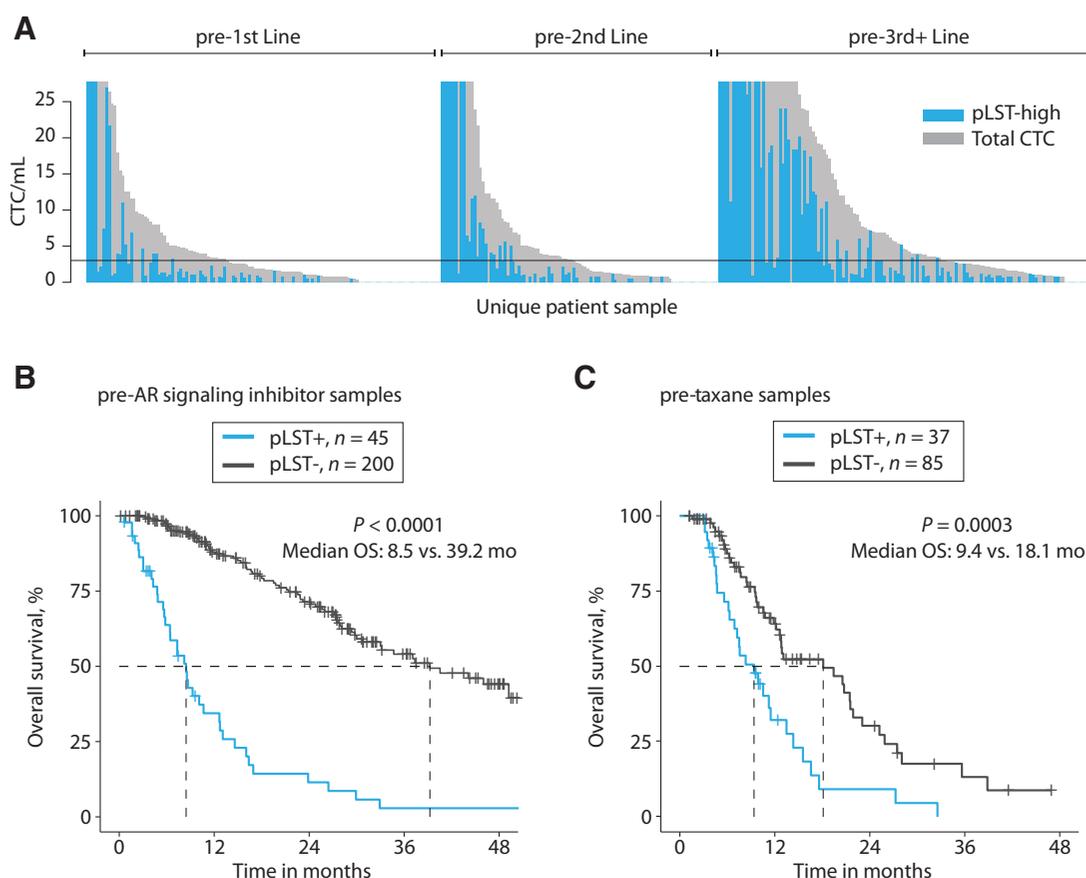
#### CTC-CIN biomarker positivity prior to therapy initiation associates with early drug resistance and poor survival

In the clinical validation cohort, a total of 10,240 CTCs were analyzed in the 367 samples from 294 unique patients with mCRPC. Overall, using the cutoff of  $\geq 3$  CTCs predicted to have  $\geq 9$  LSTs per mL of blood, 22.3% of the samples (82/367) were CIN biomarker-positive, a rate that increased in frequency by line of therapy (Supplementary Table S3). A bar plot comparing total CTCs and pLST-high CTCs by line of therapy is presented in Fig. 5A and a scatter plot of pLST-high

CTCs versus total CTCs is presented in Supplementary Fig. S7. All were pretreatment samples obtained a median of 3 days (range, 0–30) prior to starting either an ARSi such as enzalutamide and abiraterone or a taxane-based chemotherapy, such as docetaxel or cabazitaxel. Supplementary Table S4 compares pretreatment clinical features of patients between therapy classes, Supplementary Table S5 compares CIN biomarker positivity by therapy class, and Supplementary Table S6 compares clinical features between CIN biomarker-positive and -negative strata.

To further examine the CIN biomarker for the context of use of prognosis, we assessed the relationship of CIN biomarker status with overall survival. By Kaplan–Meier analysis, patients who were CIN biomarker-positive pretreatment had a significantly shorter median overall survival than those who were CIN biomarker-negative following treatment with both an ARSi (8.5 vs. 39.2 months,  $P < 0.0001$ ) or taxane (9.4 vs. 18.1 months,  $P = 0.0003$ ; Fig. 5B and C).

To determine the prognostic value of the CIN biomarker in the context of total CTC counts and other previously established prognostic factors, a multivariate Cox proportional hazard models were constructed. A base model containing only clinical and laboratory variables showed that higher line of therapy, LDH and PSA, and the presence of liver and/or lung metastases was associated as expected with shortened survival times (Table 2).



**Figure 5.**

Association of the CTC-CIN biomarker ( $\geq 3$  pLST-high CTCs in a blood sample) pre-ARSi and pre-taxane with overall survival. **A**, Bar plot of total CTC/mL and pLST-high CTC/mL by line of therapy. **B** and **C**, Kaplan–Meier analysis of overall survival in the clinical validation cohort by therapy class.

Schonhofs et al.

**Table 2.** Clinical validation cohort: multivariable Cox proportional hazard analyses of overall survival.

	Base model		CIN/CTC Model		Therapy interaction model	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
Treatment (taxane, ARSi)	1.2 (0.8-1.7)	0.426	1.1 (0.8-1.6)	0.587	1.8 (1.1-2.9)	0.018
Line of therapy > 2 (Y, N)	2.4 (1.7-3.5)	<0.001	2.3 (1.6-3.3)	<0.001	2.3 (1.6-3.2)	<0.001
Lactate dehydrogenase > 250 U/L (Y, N)	3.1 (2.2-4.2)	<0.001	2.6 (1.8-3.6)	<0.001	2.4 (1.7-3.4)	<0.001
PSA > 20 ng/mL (Y, N)	1.8 (1.2-2.6)	0.002	1.7 (1.2-2.5)	0.005	1.8 (1.2-2.6)	0.003
Age > 65 years (Y, N)	1.1 (0.8-1.4)	0.645	1.3 (0.8-1.5)	0.420	1.2 (0.9-1.6)	0.330
Liver and/or lung mets pretreatment (Y, N)	1.5 (1-2.1)	0.043	1.6 (1.1-2.3)	0.011	1.6 (1.1-2.3)	0.012
CIN/CTC status (pos. or neg./high or low)						
Negative/high	—		0.9 (0.6-1.3)	0.510	1.0 (0.6-1.6)	0.902
Positive/high	—		2.9 (2.1-4.2)	<0.001	5.0 (3.1-7.9)	<0.001
Therapy interaction (taxane, ARSi)						
Negative/high, taxane	—		—		0.8 (0.4-1.6)	0.470
Positive/high, taxane	—		—		0.3 (0.2-0.6)	0.001

Abbreviations: ARSi, androgen receptor signaling inhibitor; CIN, chromosomal instability; CTC, circulating tumor cell; mets, metastases; PSA, prostate-specific antigen.

The CIN biomarker (positive/negative) and CTC count (high/low by median dichotomization; median = 3/mL) were then combined into a single factor with levels CIN-negative/CTC-low (50% of patients), CIN-positive/CTC-high (22% of patients), and CIN-negative/CTC-high (28% of patients). For comparison, Kaplan-Meier curves for CTC-high and CTC-low by therapy class are presented in Supplementary Fig. S8. Added to the base multivariate model, the CIN biomarker and total CTC covariate CIN-positive/CTC-high status remained significant (HR = 2.9; 95% CI, 2.1-4.2;  $P < 0.001$ ). However, CIN-negative/CTC-high status applied in a similar fashion was not (HR = 0.9; 95% CI, 0.6-1.3;  $P = 0.5$ ). Similar results were shown in a third model that included a two-way interaction of therapy class and CIN/CTC factors where CIN-positive/CTC-high status was strongly prognostic (HR = 5.0; 95% CI, 3.1-7.9;  $P < 0.001$ ) while survival was improved for CIN-positive/CTC-high patients treated with a taxane relative to an ARSi (HR = 0.3; 95% CI, 0.2-0.6;  $P = 0.001$ ). In contrast, no significant prognostic difference was found for CIN-negative/CTC-high patients on taxane compared with ARSi (HR = 0.8; 95% CI, 0.4-1.6;  $P = 0.47$ ; **Table 2**). Finally, a fourth model was constructed using factors with total CTC counts dichotomized at the top 20th percentile, identifying a similar number of patients compared with a CTC-CIN biomarker to create a new set of levels. In this model, CTC-CIN was again found to be significantly associated with poor OS (CIN-positive CTC-high HR = 3.1; 95% CI, 2.2-4.5;  $P < 0.001$ ) and (CIN-positive CTC-low HR = 3.0; 95% CI, 1.7-5.1;  $P < 0.001$ ; Supplementary Table S7).

Taken together, the results confirm that the CIN biomarker is a strong prognostic marker of overall survival, independent of other known clinical features and total CTC number. It also suggests that CIN biomarker positivity, while prognostic of a shorter survival, may predict for a more favorable outcome with taxane-based chemotherapy relative to an ARSi, based on the significant statistical interaction observed between CIN biomarker positivity and therapy class (**Table 2**); however, additional studies will be needed to determine the utility of CTC-CIN as a predictive biomarker for drug response and treatment selection.

## Discussion

A critical unmet need in the management and treatment of patients with progressing mCRPC is for biomarkers that can be

assessed in a timeframe to inform a management decision. Here we developed an image-based algorithm to create a CTC biomarker of chromosome instability in a phlebotomy sample that assesses morphologic features and protein expression levels by fluorescent staining to determine the number of LSTs in a single CTC. The biomarker was developed using a training cohort consisting of blood samples from patients with progressing mCRPC about to start a new line of therapy and validated in a separate analytic validation cohort to determine sensitivity, specificity, PPV, and a patient-level cutoff of positivity.

The pLST algorithm classifies individual CTCs as LST-high (9 or more predicted LSTs in a cell) or LST-low (8 or fewer) with an accuracy in the range of 70%-80% based on features extracted from CTC images. Morphologically, LST-high CTCs identified by the algorithm are typically, but not exclusively, smaller, rounder, have higher CK and AR expression, and greater heterogeneity in nuclear texture as measured through DAPI intensity (e.g., higher nuclear entropy). The CTC CIN biomarker is reported as positive if 3 or more cells in a sample are predicted to be LST-high, shown with a PPV of >95% at the individual patient level in both the analytic and clinical validation cohorts based on confirmation by direct sequencing, and strongly associates with a shortened survival. The requirement for detection of multiple cells overcomes the observed performance at the single-cell level that precludes making a call on a single cell (AUC of 0.77 in ROC analysis and specificity of 74% in the independent AV cohort).

Clinical validation was achieved in a larger independent cohort in which pretreatment CIN biomarker-positivity was again shown to be a strong independent prognostic marker for inferior overall survival independent of the line of therapy or the class of therapy received. The association of the CIN biomarker with overall survival was stronger than total CTC number, the presence of visceral metastasis, or other routinely used prognostic factors in the progressing mCRPC setting. Overall, the results are consistent with previous reports showing the relationship of prostate cancer cell and nuclear morphology to overall genomic and/or chromosomal instability and poor outcome (21-23), including a recent analysis of a large cohort of tissue hematoxylin and eosin sections relating nuclear texture heterogeneity to poor outcome in prostate cancer (24).

Many links between phenotype and genomic and/or chromosomal instability have been established. Aneuploidy in a single cell estimated

by quantifying nuclear size is one example, where greater nuclear area correlates with greater total DNA content (25, 26). Increased AR signaling has been shown to enhance TOP2B recruitment to sites of TMPRSS2-ERG genomic breakpoints to promote TOP2B-mediated dsDNA breaks (27), and to be involved in regulating DNA damage checkpoint signaling in prostate cancer (28, 29). Similarly, high expression of the most abundant CKs, CK8 and CK18, has been shown to associate with resistance to multiple cancer therapies (30–32), and may be a marker of increased PI3K/Akt pathway activation in tumor cells (33). On the other hand, it should be noted that reduced CK marker and/or AR expression is also observed in a subset of CRPCs that are resistant to next-generation AR-targeted therapies by a transition to an epithelial-to-mesenchymal (EMT) like state or lineage plasticity (33). One likely explanation is that the definition of a CTC used here required CK or AR expression, possibly excluding CTCs that have undergone EMT/lineage plasticity (34–36). Future studies that include additional CTC markers, such as synaptophysin, chromogranin A, and vimentin, may help capture this group of CTCs (34–36). Given the weak association between CK expression and LSTs in CTCs seen here, more study is needed to understand the complex relationship between epithelial marker and AR expression and the status of genomic instability.

The CIN and nuclear-localized AR-V7 protein biomarkers (13, 14, 37) assessed in individual CTCs along with a previously reported biomarker of CTC heterogeneity (18) are all associated with resistance to ARSi and a poor prognosis. Future work will focus on the combined and independent frequency of the AR-V7, CIN, and other biomarkers to determine the added value of using two or all three measures to determine which patients will do poorly with ARSi and other classes of drug (38, 39). Separately, others have reported that the presence of CIN in breast cancer, as determined by several independent methods, may be a marker for resistance to taxane treatment as well (40, 41), and prospective efforts are ongoing to test whether targeting CIN with DNA damage-directed therapies such as a PARP inhibitor (NCT03712930) or platinum-based agents would be beneficial.

It should also be recognized that LST number is only one functional measure of HRD, and other measures of genomic instability exist including telomere allele imbalance and loss of heterozygosity, as well as other genomic consequences of deficiency in DNA repair mechanisms. Also, it is likely that CTCs with a high number of LSTs include diverse genomic aberrations including aneuploidy or other mechanisms relating to chromosomal instability regardless of HRD gene status. Accordingly, models of BRCAness or HRD typically include HRD-associated mutational signatures in addition to metrics such as LST. As a result, a high number of LSTs here may not always relate to functional HRD. On the other end of the spectrum, the precise origin of cells with only a few alterations is not always clear and reflects the technical limitations in CTC characterization. In bulk sequencing of localized and some metastatic prostate cancers low or absent copy-number burden in prostate cancer have been observed, possibly due to the lower coverage, and the overall burden was shown to be associated with disease extent (6, 7). Although more recent deep whole-genome and whole-transcriptome sequencing has revealed a wide diversity of alterations in the vast majority of metastatic biopsies sequenced that would be potentially missed with the single CTC sequencing technology used here (42). Furthermore, data on the single-cell nature of copy number or LST burden in single tumor cells isolated from biopsy in a large cohort is lacking and technical

limitations require that the data generated to date be interpreted with caution.

Collectively, we observe a relationship between CTC morphology and CIN, and demonstrate the feasibility of developing and clinically validating a robust image-based CTC-CIN biomarker that can be applied with rapid turnaround time (estimated at 2–3 days) to individual patients.

### Disclosure of Potential Conflicts of Interest

J.D. Schonhoft reports other compensation from Epic Sciences (salary and stock options) during the conduct of the study and other compensation from Epic Sciences (salary and stock options) outside the submitted work. A. Jendrisak reports other compensation from Epic Sciences (salary and stock options) during the conduct of the study and other compensation from Epic Sciences (salary and stock options) outside the submitted work. A. Gill reports other compensation from Epic Sciences (salary and stock options) during the conduct of the study and other compensation from Epic Sciences (salary and stock options) outside the submitted work. R. Sutton reports personal fees from Epic Sciences during the conduct of the study and personal fees from Epic Sciences outside the submitted work. A.E. Dago reports other compensation from Epic Sciences (salary and stock options) during the conduct of the study and other compensation from Epic Sciences (salary and stock options) outside the submitted work. S.F. Bakhoun reports personal fees and other compensation from Volastra Therapeutics Inc. (SAB, board of directors, consulting fees) and personal fees from Sanofi (consulting) outside the submitted work; in addition, S.F. Bakhoun has a patent for Targeting chromosomal instability issued (related to chromosomal instability and associated pathways in cancer). Y. Wang reports other compensation from Epic Sciences (salary and stock options) during the conduct of the study and other compensation from Epic Sciences (salary and stock options) outside the submitted work. R. Dittamore reports other compensation from Epic Sciences (was an employee of Epic Sciences) outside the submitted work, and was an employee at Epic Sciences, which developed the CTC platform utilized in the study. H.I. Scher reports grants from Epic Sciences (support to institution—Memorial Sloan Kettering Cancer Center) and nonfinancial support from Epic Sciences (travel paid by Epic Sciences) during the conduct of the study; personal fees and nonfinancial support from Astera Biotherapeutics (compensated—board of director and member), Ambray Genetics Corporation, Konica Minolta Inc. (compensated—consultant), Pfizer Inc. (compensated—consultant), WCG Oncology (compensated—consultant); nonfinancial support from Amgen (uncompensated—consultant), Bayer (uncompensated—consultant), ESSA Pharma (uncompensated—consultant), Janssen Research & Development LLC (uncompensated—consultant), Janssen Biotech Inc. (uncompensated—consultant), Menarini Silicon Biosystems (uncompensated—consultant), Sanofi Aventis (uncompensated—consultant), Phosphatin Therapeutics (travel expenses), and Prostate Cancer Foundation (travel expenses); grants from Illumina, Inc. (support to institution—Memorial Sloan Kettering Cancer Center), Janssen (support to institution—Memorial Sloan Kettering Cancer Center), Prostate Cancer Foundation (support to institution—Memorial Sloan Kettering Cancer Center), Menarini Silicon Biosystems (support to institution—Memorial Sloan Kettering Cancer Center), Thermo Fisher Scientific (support to institution—Memorial Sloan Kettering Cancer Center), outside the submitted work. No potential conflicts of interest were disclosed by the other authors.

### Authors' Contributions

J.D. Schonhoft: Conceptualization, data curation, formal analysis, writing-original draft, writing-review and editing. J.L. Zhao: Formal analysis, writing-original draft, writing-review and editing. A. Jendrisak: Conceptualization, formal analysis, methodology. E.A. Carbone: Acquisition of data. E.S. Barnett: Project administration, writing-review and editing, acquisition of data. M.A. Hullings: Project administration, acquisition of data. A. Gill: Formal analysis, methodology, writing-review and editing. R. Sutton: Formal analysis, methodology, writing-review and editing, acquisition of data. J. Lee: Formal analysis, methodology, writing-review and editing, acquisition of data. A.E. Dago: Formal analysis, methodology, writing-review and editing. M. Landers: Conceptualization, formal analysis, methodology, writing-review and editing. S.F. Bakhoun: Formal analysis, writing-review and editing. Y. Wang: Conceptualization, formal analysis, methodology, writing-review and editing. M. Gonen: Formal analysis, writing-review and editing. R. Dittamore: Conceptualization, formal analysis, methodology, writing-review

## Schonhoff et al.

and editing. H.I. Scher: Conceptualization, formal analysis, supervision, funding acquisition, methodology, writing-original draft, writing-review and editing.

### Acknowledgments

We thank the patients taking part in this study, and the clinical and laboratory staff at Memorial Sloan Kettering Cancer Center and Epic Sciences.

Epic Sciences provided facilities, materials, and services to perform isolation and sequencing of circulating tumor cells. Work at Memorial Sloan Kettering Cancer Center (MSK) was supported by the Sidney Kimmel Center for Prostate and Urologic Cancers, and funded in part by the NIH/NCI Cancer Center Support Grant to MSK (P30 CA008748), the NIH/NCI SPORE in Prostate Cancer grant to MSK (P50 CA092629), the Prostate Cancer Foundation, and the Prostate

Cancer Clinical Trials Consortium. J.L. Zhao is supported by an NIH/NCI NRSA training award to MSK (T32 CA009207), an ASCO Conquer Cancer Young Investigator Award, and a K12 Paul Calabresi Career Development Award for Clinical Oncology.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 13, 2020; revised June 29, 2020; accepted August 14, 2020; published first August 19, 2020.

### References

- Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* 2010;11:220–8.
- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144:646–74.
- Sansregret L, Vanhaesebroeck B, Swanton C. Determinants and clinical implications of chromosomal instability in cancer. *Nat Rev Clin Oncol* 2018;15:139–50.
- Targa A, Rancati G. Cancer: a CINful evolution. *Curr Opin Cell Biol* 2018;52:136–44.
- Vargas-Rondón N, Villegas VE, Rondón-Lagos M. The role of chromosomal instability in cancer and therapeutic responses. *Cancers* 2017;10:4.
- Hieronimus H, Murali R, Tin A, Yadav K, Abida W, Moller H, et al. Tumor copy number alteration burden is a pan-cancer prognostic factor associated with recurrence and death. *Elife* 2018;7:e37294.
- Hieronimus H, Schultz N, Gopalan A, Carver BS, Chang MT, Xiao Y, et al. Copy number alteration burden predicts prostate cancer relapse. *Proc Natl Acad Sci U S A* 2014;111:11139–44.
- McGranahan N, Burrell RA, Endesfelder D, Novelli MR, Swanton C. Cancer chromosomal instability: therapeutic and diagnostic challenges. *EMBO Rep* 2012;13:528–38.
- Juric D, Castel P, Griffith M, Griffith OL, Won HH, Ellis H, et al. Convergent loss of PTEN leads to clinical resistance to a PI(3)K $\alpha$  inhibitor. *Nature* 2015;518:240–4.
- McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell* 2017;168:613–28.
- Perdigones N, Murtaza M. Capturing tumor heterogeneity and clonal evolution in solid cancers using circulating tumor DNA analysis. *Pharmacol Ther* 2017;174:22–6.
- Lambros MB, Seed G, Sumanasuriya S, Gil V, Crespo M, Fontes M, et al. Single-cell analyses of prostate cancer liquid biopsies acquired by apheresis. *Clin Cancer Res* 2018;24:5635–44.
- Scher HI, Lu D, Schreiber NA, Louw J, Graf RP, Vargas HA, et al. Association of AR-V7 on circulating tumor cells as a treatment-specific biomarker with outcomes and survival in castration-resistant prostate cancer. *JAMA Oncol* 2016;2:1441–9.
- Scher HI, Graf RP, Schreiber NA, Jayaram A, Winquist E, McLaughlin B, et al. Assessment of the validity of nuclear-localized androgen receptor splice variant 7 in circulating tumor cells as a predictive biomarker for castration-resistant prostate cancer. *JAMA Oncol* 2018;4:1179–86.
- Popova T, Manié E, Rieunier G, Caux-Moncoutier V, Tirapo C, Dubois T, et al. Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer Res* 2012;72:5454–62.
- Marquard AM, Eklund AC, Joshi T, Krzystanek M, Favero F, Wang ZC, et al. Pan-cancer analysis of genomic scar signatures associated with homologous recombination deficiency suggests novel indications for existing cancer drugs. *Biomark Res* 2015;3:9.
- Greene SB, Dago AE, Leitz LJ, Wang Y, Lee J, Werner SL, et al. Chromosomal instability estimation based on next generation sequencing and single cell genome wide copy number variation analysis. *PLoS One* 2016;11:e0165089.
- Scher HI, Graf RP, Schreiber NA, McLaughlin B, Jendrisak A, Wang Y, et al. Phenotypic heterogeneity of circulating tumor cells informs clinical decisions between AR signaling inhibitors and taxanes in metastatic prostate cancer. *Cancer Res* 2017;77:5687–98.
- Jonsson P, Bandlamudi C, Cheng ML, Srinivasan P, Chavan SS, Friedman ND, et al. Tumour lineage shapes BRCA-mediated phenotypes. *Nature* 2019;571:576–9.
- Sommer C, Gerlich DW. Machine learning in cell biology—teaching computers to recognize phenotypes. *J Cell Sci* 2013;126:5529–39.
- Miller C, Partin W, Epstein I, Veltri RW, Miller MC, Partin AW, et al. Ability to predict biochemical progression using gleason score and a computer-generated quantitative nuclear grade derived from cancer cell nuclei. *Urology* 1996;48:685–91.
- Pienta KJ, Partin AW, Coffey DS. Cancer as a disease of DNA organization and dynamic cell structure. *Cancer Res* 1989;49:2525–32.
- Uhler C, Shivashankar GV. Nuclear mechanopathology and cancer diagnosis. *Trends Cancer* 2018;4:320–31.
- Kleppe A, Albrechtsen F, Vlatkovic L, Pradhan M, Nielsen B, Hveem TS, et al. Chromatin organisation and cancer prognosis: a pan-cancer study. *Lancet Oncol* 2018;19:356–69.
- Doležel J, Greilhuber J, Suda J. Estimation of nuclear DNA content in plants using flow cytometry. *Nat Protoc* 2007;2:2233–44.
- Zeimet AG, Fiegl H, Goebel G, Kopp F, Allasia C, Reimer D, et al. DNA ploidy, nuclear size, proliferation index and DNA-hypomethylation in ovarian cancer. *Gynecol Oncol* 2011;121:24–31.
- Hafner MC, Aryee MJ, Toubaji A, Esopi DM, Albadine R, Gurel B, et al. Androgen-induced TOP2B-mediated double-strand breaks and prostate cancer gene rearrangements. *Nat Genet* 2010;42:668–75.
- Polkinghorn WR, Parker JS, Lee MX, Kass EM, Spratt DE, Iaquinta PJ, et al. Androgen receptor signaling regulates DNA repair in prostate cancers. *Cancer Discov* 2013;3:1245–53.
- Goodwin JF, Schiewer MJ, Dean JL, Schrecengost RS, de Leeuw R, Han S, et al. A hormone-DNA repair circuit governs the response to genotoxic insult. *Cancer Discov* 2013;3:1254–71.
- Karantza V. Keratins in health and cancer: more than mere epithelial cell markers. *Oncogene* 2011;30:127–38.
- Fortier AM, Van Themsche C, Asselin É, Cadrin M. Akt isoforms regulate intermediate filament protein levels in epithelial carcinoma cells. *FEBS Lett* 2010;584:984–8.
- Liu F, Chen Z, Wang J, Shao X, Cui Z, Yang C, et al. Overexpression of cell surface cytokeratin 8 in multidrug-resistant MCF-7/MX cells enhances cell adhesion to the extracellular matrix. *Neoplasia* 2008;10:1275–84.
- Bauman PA, Dalton WS, Anderson JM, Cress AE. Expression of cytokeratin confers multiple drug resistance. *Proc Natl Acad Sci U S A* 1994;91:5311–4.
- Beltran H, Hruszkewycz A, Scher HI, Hildesheim J, Isaacs J, Yu EY, et al. The role of lineage plasticity in prostate cancer therapy resistance. *Clin Cancer Res* 2019;25:6916–24.
- Davies AH, Beltran H, Zoubeidi A. Cellular plasticity and the neuroendocrine phenotype in prostate cancer. *Nat Rev Urol* 2018;15:271–86.
- Quintanal-Villalonga Á, Chan JM, Yu HA, Pe'er D, Sawyers CL, Sen T, et al. Lineage plasticity in cancer: a shared pathway of therapeutic resistance. *Nat Rev Clin Oncol* 2020;17:360–71.
- Armstrong AJ, Halabi S, Luo J, Nanus DM, Giannakakou P, Szmulewitz RZ, et al. Prospective multicenter validation of androgen receptor splice variant 7 and

**Chromosomal Instability in Circulating Tumor Cells**

- hormone therapy resistance in high-risk castration-resistant prostate cancer: The PROPHECY study. *J Clin Oncol* 2019;37:1120-9.
38. Scher HI, Schonhoft J, Graf RP, Jendrisak A, Barnett E, Jayaram A, et al. Examination of the additive value of CTC biomarkers of heterogeneity (Het) and chromosomal instability to nuclear-localized (nl) AR-V7+ CTCs in prediction of poor outcomes to androgen receptor signaling inhibitor (ARSi) in metastatic castration resistant. *J Clin Oncol* 37:15s, 2019(suppl; abstr 5075).
  39. Brown LC, Halabi S, Schonhoft J, Luo J, Nanus DM, Giannakakou P, et al. Association of circulating tumor cell chromosomal instability with worse outcomes in men with mCRPC treated with abiraterone or enzalutamide. *J Clin Oncol* 38:6s, 2020(suppl; abstr 183).
  40. A'Hern RP, Jamal-Hanjani M, Szász AM, Johnston SRD, Reis-Filho JS, Roylance R, et al. Taxane benefit in breast cancer - a role for grade and chromosomal stability. *Nat Rev Clin Oncol* 2013;10:357-64.
  41. Swanton C, Nicke B, Schuett M, Eklund AC, Ng C, Li Q, et al. Chromosomal instability determines taxane response. *Proc Natl Acad Sci U S A* 2009;106:8671-6.
  42. Quigley DA, Dang HX, Zhao SG, Lloyd P, Aggarwal R. Genomic hallmarks and structural variation in metastatic prostate cancer. *Cell* 2018;174, 758-69.

# Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

## Morphology-Predicted Large-Scale Transition Number in Circulating Tumor Cells Identifies a Chromosomal Instability Biomarker Associated with Poor Outcome in Castration-Resistant Prostate Cancer

Joseph D. Schonhoff, Jimmy L. Zhao, Adam Jendrisak, et al.

*Cancer Res* 2020;80:4892-4903. Published OnlineFirst August 19, 2020.

**Updated version** Access the most recent version of this article at:  
doi:[10.1158/0008-5472.CAN-20-1216](https://doi.org/10.1158/0008-5472.CAN-20-1216)

**Supplementary Material** Access the most recent supplemental material at:  
<http://cancerres.aacrjournals.org/content/suppl/2020/08/19/0008-5472.CAN-20-1216.DC1>

**Cited articles** This article cites 42 articles, 12 of which you can access for free at:  
<http://cancerres.aacrjournals.org/content/80/22/4892.full#ref-list-1>

**Citing articles** This article has been cited by 1 HighWire-hosted articles. Access the articles at:  
<http://cancerres.aacrjournals.org/content/80/22/4892.full#related-urls>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, use this link  
<http://cancerres.aacrjournals.org/content/80/22/4892>.  
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.