Chromosomal Instability in Cell-Free DNA Is a Serum Biomarker for Prostate Cancer

Ekkehard Schütz,1 Mohammad R. Akbari,2,5 Julia Beck,1 Howard Urnovitz,1 William W. Zhang,3 Kirsten Bornemann-Kolatzki,1 William M. Mitchell,4* Robert K. Nam,3 and Steven A. Narod2,5

BACKGROUND: Genomic instability resulting in copy number variation is a hallmark of malignant transformation and may be identified through massive parallel sequencing. Tumor-specific cell free DNA (cfDNA) present in serum and plasma provides a real-time, easily accessible surrogate.

METHODS: DNA was extracted from serum of 204 patients with prostate cancer (Gleason score 2–10), 207 male controls, and patients with benign hyperplasia (n = 10) and prostatitis (n = 10). DNA was amplified by use of random primers, tagged with molecular identifiers, sequenced on a SOLID system, and aligned to the human genome. We evaluated the number of sequence reads of cfDNA in sliding 100-kbp intervals for variation from controls. We used chromosomal regions with significant variations in alignment hits for their ability to segregate patients and matched controls.

RESULTS: Using ROC curves to assess diagnostic performance, we evaluated the number of regions in a first subset (n = 177), with variations in alignment hits alone, provided an area under the curve (AUC) of 0.81 (95% CI 0.7–0.9, P < 0.001). Using 5 rounds of 10-fold cross-validation with the full data set, we established a final model that discriminated prostate cancer from controls with an AUC of 0.92 (0.87–0.95), reaching a diagnostic accuracy of 83%. Both benign prostatic hypertrophy and prostatitis could be distinguished from prostate cancer by use of cfDNA, with an accuracy of 90%.

CONCLUSIONS: Assessment of a limited number of chromosomal structural instabilities by use of massive parallel sequencing of cfDNA was sufficient to distinguish between prostate cancer and controls. This large cohort demonstrates the utility of cfDNA in prostate cancer recently established in other malignant neoplasms.

© 2014 American Association for Clinical Chemistry

The prostate is the most common site of cancer in men, with 240000 new cases diagnosed annually in the US and approximately 28000 yearly deaths. Prostate screening typically relies on digital rectal exam and prostate specific antigen (PSA). Limitations of the PSA test include its relatively low diagnostic sensitivity and specificity and its inability to distinguish between low-grade and high-grade lesions. Recent screening trials suggest that PSA-based screening programs result in small or no reduction in mortality, with significant treatment-related adverse events (1). Better serum/plasma biomarkers are needed to supplement the inexpensive PSA test in the diagnosis and management of a disease with a multiplicity of presentations and clinical outcomes.

The hypothesis that DNA from cell-free plasma or serum can be used for the preclinical detection of human malignancies has been studied for >15 years (2). These studies are all based on the ability to distinguish cancer-specific DNA markers from those of nonmalignant tissues. Our interest in prostate cancer is based on the recent discoveries that almost all cancer types exhibit chromosomal structural instability, through either progressive evolution or catastrophic genomic events (3). A constant amount of routine cell-free DNA (cfDNA) present through apoptosis can be distinguished from neoplastic apoptosis (4). In cancer patients, cfDNA is released from apoptotic cells as nucleosomes from both healthy and diseased tissue that includes tumor cells as well as microbial nucleic acids from systemic infections (5).
clinical utility of cfDNA as a biomarker is contingent on the ability to distinguish between those origins.

The statistical power provided by massive parallel sequencing in next generation sequencing (NGS) platforms is ideally suited for the detection in cfDNA of differences in chromosomal instability as regional DNA ploidy heterogeneity (6). Embryonic cfDNA can be detected in maternal serum even if it represents <0.5% of the total DNA present (7). The number of unique sequence reads is the determining factor in identifying regional DNA ploidy heterogeneity in trace amounts of pathologic cfDNA. By calculating a genome-wide z score, Heitzer et al. (8) could identify tumor-associated aneuploidy by low-coverage cfDNA sequencing and were able to discriminate plasma samples from men with and without prostate cancer at a detection limit of 1% tumor-derived cfDNA. The recent availability of NGS platforms with massive capacity has provided the ability to use cfDNA in blood as the basis for developing cancer biomarkers, either for detection or for monitoring of therapeutic efficacy (9–21). Quantitative detection of cfDNA as a “liquid biopsy” has been demonstrated to accurately reflect the evolving genomic instabilities observed in cancer (9–12, 15–21). Canine mammary carcinomas are a heterogeneous collection of histopathological types. Despite this heterogeneity, paired end massively parallel sequencing by Beck et al. (20) proved that copy-number imbalances of the tumors were reflected by cfDNA. Moreover, minimal residual disease was detected after surgery as a solitary metastasis a year later. Similar genomic rearrangements and mutations in a variety of human cancers can be detected in cfDNA at a high rate without false positives (15–19).

Three recently published analyses of cfDNA sequences focused on the evaluation of the genomic origin of coding and noncoding regions that included repetitive elements (5, 13, 14). In this study, we explored the utility of assessing cfDNA fragments from the entire prostate cancer genome in serum as a genetic signature of apoptotic prostatic cancer cells as a potential clinically useful biomarker.

Materials and Methods

We studied 204 biopsy-proven cases of prostate cancer and 207 age-matched healthy controls that were asymptomatic and/or had recent negative prostate biopsies obtained from 3 sources under informed consent and institutional review board approval. An additional 10 patients with benign prostate hyperplasia and 10 patients with prostatitis were included in the study. Whole blood was collected and serum was recovered by centrifugation immediately after clotting and stored frozen (−20°C) in aliquots. We analyzed 89 sera from prostate carcinoma with a Gleason score <7 and 84 with a Gleason score ≥7. Gleason score was not recorded for the remaining 31 samples. Sera from patients with prostate carcinoma were obtained from 76 patients diagnosed at age ≤65 years. Immediately before whole genome amplification (WGA), serum samples (≥200 µL) were thawed and centrifuged at 4000g for 20 min to pellet any cellular debris. Total nucleic acids were extracted from 200 µL supernatant with the High Pure Viral Nucleic Acid Kit (Roche Applied Science) according to the manufacturer’s instructions, but without the use of carrier RNA. All samples provided sufficient cfDNA for NGS. Each sample was subjected to commercial WGA in independent duplicates, which was conducted in a LightCycler480 with the addition of 1× EvaGreen (GenomePlex Whole Genome Amplification Kit, WGA, Sigma-Aldrich). To avoid bias caused by overamplification during WGA, the amplifications were monitored in real time and were stopped at the first cycle trending outside the linear amplification range. Massive parallel sequencing used the SOLID-P2 NGS platform. We used ROC analyses to evaluate the performance of the final model in the entire sample and in subgroups defined by Gleason score and age of diagnosis. Specific details of sample preparation for sequencing, sequence alignment, data normalization, analysis of copy number variations (CNVs), and hotspot identification including ROC analysis are provided as Supplemental Fig. 1 and Supplemental Materials and Methods in the Supplemental Data, which accompanies the online version of this article at http://www.clinchem.org/content/vol61/issue1. All samples contained cfDNA sufficient for low cycle number WGA [mean (SD) 16 (4) cycles per sample]. The mean DNA yield was 3.1 µg. Online Supplemental Table 1 provides the mean number of sequences and sequence size per sample. Sequence data has been deposited with the European Nucleotide Archive.

Results

SEQUENCING RESULTS

A total of 1.5 × 10⁹ sequence reads were generated for the 204 patients with prostate cancer and 1.4 × 10⁹ sequence reads were generated for the 207 healthy controls, for a total combined sequence length of 1.1 × 10¹¹ nucleotides. Nucleotides of uniquely human origin were defined by the mapping software. Online Supplemental Table 1 provides a summary of the mean fragment lengths and the nucleotide counts in each group and the proportions of sequences/nucleotides that uniquely map to the human genomic database (HG18).

PREANALYSIS OF COPY NUMBER INSTABILITY IN cfDNA

We performed the initial analysis on a subset of 177 samples to globally test the ability of CNV detection of regions of chromosomal structural instability in cfDNA that distinguish patients with prostate cancer from controls. The copy
number instability (CNI) score, which can be interpreted as a general measure of genomic instability, is directly related (within the technology limits) to the regional chromosomal DNA ploidy heterogeneity, ranging from 0 to 2008 in the 177 samples.; the median in controls was 9 (95% CI 5–12) and 44 (30–67) in prostatic carcinoma. The ROC curve on the basis of this CNI score provided an area under the curve (AUC) of 0.81 (95% CI 0.69–0.90) (see online Supplemental Fig. 2). These preliminary data demonstrated that a higher occurrence of copy number imbalances in cfDNA of patients with malignant prostate tumors could be detected compared to controls. Comparative Circos plots from 5 patient and control samples in which the number of regions significantly deviated from euploidy are illustrated in Fig. 1. Evidence of chromosomal structural instability detected as DNA ploidy heterogeneity was widespread in the prostate cancer cohort.

CROSS-VALIDATION OF cfDNA GENOMIC ALIGNMENT HIT VARIANCE AS A PREDICTOR OF PROSTATE CANCER
The selected consolidated genomic clusters were subjected to 5 independent rounds of 10-fold cross-validation each. The number of regional clusters in the final model was restricted to 20. The mean AUC for the ROC calculated from each of the 50 validation sets was 0.85 (0.06) (P < 10−8).

CHROMOSOMAL REGIONAL CLUSTERS OF DNA PLOIDY HETEROGENEITY
Table 1 illustrates the chromosomal distribution of relative locations and cluster compositions (bin components including closely adjacent genes) with their respective increases or decreases in observed apoptotic DNA serum frequencies, on the basis of the cross-validation results. These 20 clusters in the highest abundance of the cross-validation defined the final model that was used for all subsequent analyses of the regional CNI index. The frequencies of the consolidated regions used for cross-validation and the regions that were in the final model are shown in online Supplemental Fig. 3.

ROC ANALYSIS
Table 2 summarizes the ROC analysis by use of the CNI index from cfDNA for the study and its component parts. The AUC of the ROC curve for the entire study was 0.92 (95% CI 0.87–0.95) (Fig. 2A). An insignificant improvement in AUC (Fig. 2B) was obtained in patients with a low- or moderate-grade (Gleason score <7) neoplasm vs patients with a high-grade (Gleason score ≥7) neoplasm. Age had little effect on cfDNA distribution. The inclusion of patients with benign prostatic hypertrophy or prostatitis had no effect on the ROC analysis of the entire cohort or as a function of microscopic prostate cancer grade. When using a 10-fold cross-validation, in which 10 random groups of prostate cancer samples were compared with benign samples, the AUC was 0.90 (0.055), similar to the cancer vs control sera. Fig. 3 illustrates the CNI index for all prostate cancer patients and as a function of Gleason score and benign prostatic disease. The latter is significantly different from the cancer cohort (P < 0.00001). The genes that were present in the identified regions are summarized in online Supplemental Table 2.

CORRELATION WITH CONCURRENT PSA CONCENTRATIONS
PSA concentrations were available in a subset of samples (126 cases and 108 controls). The majority of patients with prostate cancer had an increased PSA concentration. Five patients with prostatic carcinoma had a PSA value ≤4 µg/L with clear evidence of chromosomal instability, as measured by cfDNA. Overall there was no correlation between serum PSA concentration and the cfDNA biomarker (r = 0.3).

Discussion
Cancer is characterized by genomic instability (aneuploidy) and is clonal in origin (22). The Mitelman–National Cancer Institute database currently catalogs >63,000 human cancers with individual clonal karyotypes distinct from their cells of origin, including >2,000 gene fusions (23). This massive collection of individual clonal karyotypes is derived from individual progressive mutational events provided by single-nucleotide polymorphisms and chromosomal CNVs in a cellular form of Darwinian speciation (24, 25). Tumor structural chromosomal variations observed as regional DNA ploidy heterogeneity are recapitulated in the cfDNA as a function of apoptosis and the cellular release of nucleosomes that can be distinguished from typical cellular apoptotic DNA.

The clinical variability of prostate cancer is predicated on its immense genomic variability. Massive parallel sequencing on a limited number of primary prostate carcinomas (n = 7) first demonstrated the vast genomic complexity of advanced disease (Gleason score ≥7/stage ≥Tc2). A mean of 3866 somatic mutations (range 3192–5865) per genome was documented with a C-phosphate-G (CpG) mutation rate 10-fold higher than non-CpG mutations (26). Chromosomal fusions were common, with breakpoints clustered in regions of high transcriptional activity (transcriptomes located in euchromatin). A median of 90 rearrangements (range 43–213) per genome were found with multiple intragenic breakpoints predicted to encode truncated proteins of potentially altered biological function. In contrast, we used the statistical power provided by massive parallel sequencing for a population-based analysis of a large cohort of patients with prostate cancer and a wide range of disease severity. Although the NGS platform provided relatively short sequences (40 or 50 bp), they were appropriate for unique assignment to individual chromosomal loci. The advantage of our assignment of 100 kbp and
Fig. 1. Comparative Circos plots of control and prostate cancer cfDNA quantified by massive parallel sequencing.

The chromosome map is located on the external periphery with the kinetochore in red. The relative chromosomal deviations of individual cfDNA samples from combined control average typical cfDNA frequency (green represents gain; red represents loss) are illustrated as inner wheels. (A), Example of 5 healthy controls showing minimal numbers of copy number imbalances. (B), Example of 5 prostate cancer patients showing significant numbers of regional copy number imbalances (DNA ploidy heterogeneity).
greater individual contiguous chromosomal loci is the ability to recognize chromosomal hotspots of regional DNA ploidy heterogeneity. The limitation of this approach is that single or small nucleotide deletion/duplication mutations, individual breakpoints, or breakpoints involving translocations and/or sequence inversions will not be recognized.

We demonstrated variations in the number of cfDNA sequences circulating in the serum of patients with prostate cancer compared with healthy controls. These sequences derive from apoptotic cell death of cancer cells that provide detectable differences from routine cell turnover. The numbers of observed chromosomal fragments in prostate cancer cfDNA can derive from

---

### Table 1. Regions of genomic instability basis for cfDNA prostate cancer biomarkers.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Region, kbp</th>
<th>Bin size, kbp</th>
<th>Abundance vs controls</th>
<th>Global/local</th>
<th>Weight factor</th>
<th>Bin components</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS1</td>
<td>39 200-39 400</td>
<td>200</td>
<td>Loss</td>
<td>Global</td>
<td>1.0</td>
<td>MACF1, HSPE1, 11 transcribed loci; [RRAGC, MYCBP, GJA9, RHBDL2]</td>
</tr>
<tr>
<td></td>
<td>149 250-149 450</td>
<td>200</td>
<td>Loss</td>
<td>Global</td>
<td>1.0</td>
<td>SEC22B (l), PDE4DIP, NOTCH2, NBP1, 2 uncharacterized gene loci, 21 transcribed loci</td>
</tr>
<tr>
<td></td>
<td>212 200-212 600</td>
<td>400</td>
<td>Loss</td>
<td>Global</td>
<td>1.0</td>
<td>PPP2R5A, SNORA14B, TMEM206, NENF, ATF3, HNRNPH1, 1 uncharacterized gene locus, 17 transcribed loci; [INTS7, DTL, MIR372]</td>
</tr>
<tr>
<td>HS3</td>
<td>110 000-110 600</td>
<td>600</td>
<td>Gain</td>
<td>Global</td>
<td>1.0</td>
<td>6 transcribed loci</td>
</tr>
<tr>
<td>HS4</td>
<td>186 600-187 500</td>
<td>900</td>
<td>Loss</td>
<td>Local</td>
<td>1.0</td>
<td>FAT1, MRPS36P2, 3 uncharacterized gene loci, 21 transcribed loci; [TLR3, FAM149A, ORAOV1P1, CYP4V2, KLKB1, F11, SLC25ASP6*, MTNR1A]</td>
</tr>
<tr>
<td>HS6</td>
<td>46 750-47 150</td>
<td>400</td>
<td>Gain</td>
<td>Global</td>
<td>1.0</td>
<td>ANKRD66, MEP1A, GPR116, GPR110, 1 uncharacterized gene locus, 9 transcribed loci</td>
</tr>
<tr>
<td>HS8</td>
<td>43 150-43 450</td>
<td>300</td>
<td>Gain</td>
<td>Local</td>
<td>1.0</td>
<td>HGSNAT, RNY5P6, VN1R45/46P, AFG3L2, PO1EA, 3 poorly characterized loci, 3 transcribed loci</td>
</tr>
<tr>
<td></td>
<td>43 050-43 550</td>
<td>500</td>
<td>Gain</td>
<td>Global</td>
<td>0.5</td>
<td>FNTA, POMK, HGSNAT, RNY5P6, VN1R45/46P, AFG3L2P1, PO1EA, SNX1P27*, 10 transcribed loci</td>
</tr>
<tr>
<td>HS9</td>
<td>120 750-121 150</td>
<td>400</td>
<td>Gain</td>
<td>Local</td>
<td>0.5</td>
<td>SNTB1, 1 uncharacterized gene locus, 7 transcribed loci; [TAP2, DSCC1, COL14A1]</td>
</tr>
<tr>
<td>HS10</td>
<td>17 900-18 300</td>
<td>400</td>
<td>Gain</td>
<td>Local</td>
<td>1.5</td>
<td>MiR4289, 1 uncharacterized gene locus, 3 transcribed loci; [GOLM1]</td>
</tr>
<tr>
<td></td>
<td>27 600-28 000</td>
<td>400</td>
<td>Gain</td>
<td>Local</td>
<td>1.0</td>
<td>MiX, ARMC4, RPL36AP55, 1 uncharacterized gene locus, 4 transcribed loci; [PTCHD3, RAB18]</td>
</tr>
<tr>
<td>HS12</td>
<td>46 750-47 050</td>
<td>300</td>
<td>Gain</td>
<td>Local</td>
<td>1.0</td>
<td>SLC38A2, 1 uncharacterized gene locus, 2 transcribed loci</td>
</tr>
<tr>
<td></td>
<td>109 500-109 900</td>
<td>400</td>
<td>Gain</td>
<td>Global</td>
<td>0.5</td>
<td>UBE3B, MMAB, MKV, FAM222A, TRPV4, GLTP, TCHP, 1 uncharacterized gene locus, 25 transcribed loci; [USP30, ALKBH2, UNG, ACACB, FOXN4, MLOH, KCTD10]</td>
</tr>
<tr>
<td>HS13</td>
<td>21 200-22 100</td>
<td>900</td>
<td>Loss</td>
<td>Local</td>
<td>1.5</td>
<td>ESRRAP2, MIPEP2, GRK6P1, GAPDHPS2, RNASSPS2, ZDHHC20, HST112BPS3, FICU2, FNTAP2, RN64-39P, RPS5P10, FGPF9, LINC00424, NME1P1, 3 uncharacterized gene loci (+/-), 32 transcribed loci; [FF88, IL17D, N6AMT2, XPO4, PRPAP2*, LAT52, SAP18, SKA3, MRPL57]</td>
</tr>
<tr>
<td>HS15</td>
<td>19 450-19 750</td>
<td>300</td>
<td>Gain</td>
<td>Global</td>
<td>1.0</td>
<td>No genes or transcribed loci, high density of repetitive elements</td>
</tr>
<tr>
<td></td>
<td>52 900-53 200</td>
<td>300</td>
<td>Gain</td>
<td>Local</td>
<td>1.5</td>
<td>EEF1A1, 1 transcribed loci; [FAM214A, ONECUT1, RPSAP55]</td>
</tr>
<tr>
<td></td>
<td>61 500-62 000</td>
<td>500</td>
<td>Gain</td>
<td>Global</td>
<td>0.5</td>
<td>VP513C, 3 uncharacterized gene loci, 12 transcribed loci (ROA)</td>
</tr>
</tbody>
</table>

Continued on page 244
duplications, deletions, and association between euchromatin and heterochromatin. More condensed heterochromatin is relatively protected from DNA degradation and would be expected to yield increased DNA sequence similarity 149, member A; one cut homeobox 1; TMEFV1, transmembrane protein 206; NME1P1, nucleoside diphosphate kinase 1 pseudogene 1; PPIAP21, phosphatase 2, 3-aminotransferase 2, mitochondrial intermediate; LINC00424, mitochon- 

dian chromosomal protein 536 pseudogene 2; MACF1, microtubule-actin croslinking factor 1; HSF3P18, heat-shock 10-kDa protein 1 pseudogene 8; RRAGC, Ras-related GTP binding C; NMYC, MYC binding protein; GAP, gap junction protein, alpha 9, 59 kDa; RBBPL2, ribosomal protein L2 pseudogene; GPD2, glyceraldehyde 3-phosphate dehydrogenase pseudogene 52, RNA; NME1P1, nucleoside diphosphate kinase 1 pseudogene 1; TMEFV1, transmembrane protein 206; NME1P1, nucleoside diphosphate kinase 1 pseudogene 1; PPIAP21, phosphatase 2, 3-aminotransferase 2, mitochondrial intermediate; LINC00424, mitochon- 

dian chromosomal protein 536 pseudogene 2; MACF1, microtubule-actin croslinking factor 1; HSF3P18, heat-shock 10-kDa protein 1 pseudogene 8; RRAGC, Ras-related GTP binding C; NMYC, MYC binding protein; GAP, gap junction protein, alpha 9, 59 kDa; RBBPL2, ribosomal protein L2 pseudogene; GPD2, glyceraldehyde 3-phosphate dehydrogenase pseudogene 52, RNA; NME1P1, nucleoside diphosphate kinase 1 pseudogene 1; PPIAP21, phosphatase 2, 3-aminotransferase 2, mitochondrial intermediate; LINC00424, mitochon- 

dian chromosomal protein 536 pseudogene 2; MACF1, microtubule-actin croslinking factor 1; HSF3P18, heat-shock 10-kDa protein 1 pseudogene 8; RRAGC, Ras-related GTP binding C; NMYC, MYC binding protein; GAP, gap junction protein, alpha 9, 59 kDa; RBBPL2, ribosomal protein L2 pseudogene; GPD2, glyceraldehyde 3-phosphate dehydrogenase pseudogene 52, RNA; NME1P1, nucleoside diphosphate kinase 1 pseudogene 1; PPIAP21, phosphatase 2, 3-aminotransferase 2, mitochondrial intermediate; LINC00424, mitochon- 

dian chromosomal protein 536 pseudogene 2; MACF1, microtubule-actin croslinking factor 1; HSF3P18, heat-shock 10-kDa protein 1 pseudogene 8; RRAGC, Ras-related GTP binding C; NMYC, MYC binding protein; GAP, gap junction protein, alpha 9, 59 kDa; RBBPL2, ribosomal protein L2 pseudogene; GPD2, glyceraldehyde 3-phosphate dehydrogenase pseudogene 52, RNA; NME1P1, nucleoside diphosphate kinase 1 pseudogene 1; PPIAP21, phosphatase 2, 3-aminotransferase 2, mitochondrial intermediate; LINC00424, mitochon- 

dian chromosomal protein 536 pseudogene 2; MACF1, microtubule-actin croslinking factor 1; HSF3P18, heat-shock 10-kDa protein 1 pseudogene 8; RRAGC, Ras-related GTP binding C; NMYC, MYC binding protein; GAP, gap junction protein, alpha 9, 59 kDa; RBBPL2, ribosomal protein L2 pseudogene; GPD2, glyceraldehyde 3-phosphate dehydrogenase pseudogene 52, RNA; NME1P1, nucleoside diphosphate kinase 1 pseudogene 1; PPIAP21, phosphatase 2, 3-aminotransferase 2, mitochondrial intermediate; LINC00424, mitochon- 

dian chromosomal protein 536 pseudogene 2; MACF1, microtubule-actin croslinking factor 1; HSF3P18, heat-shock 10-kDa protein 1 pseudogene 8; RRAGC, Ras-related GTP binding C; NMYC, MYC binding protein; GAP, gap junction protein, alpha 9, 59 kDa; RBBPL2, ribosomal protein L2 pseudogene; GPD2, glyceraldehyde 3-phosphate dehydrogenase pseudogene 52, RNA; NME1P1, nucleoside diphosphate kinase 1 pseudogene 1; PPIAP21, phosphatase 2, 3-aminotransferase 2, mitochondrial intermediate; LINC00424, mitochon- 

dian chromosomal protein 536 pseudogene 2; MACF1, microtubule-actin croslinking factor 1; HSF3P18, heat-shock 10-kDa protein 1 pseudogene 8; RRAGC, Ras-related GTP binding C; NMYC, MYC binding protein; GAP, gap junction protein, alpha 9, 59 kDa; RBBPL2, ribosomal protein L2 pseudogene; GPD2, glyceraldehyde 3-phosphate dehydrogenase pseudogene 52, RNA; NME1P1, nucleoside diphosphate kinase 1 pseudogene 1; PPIAP21, phosphatase 2, 3-aminotransferase 2, mitochondrial intermediate; LINC00424, mitochon- 

dian chromosomal protein 536 pseudogene 2; MACF1, microtubule-actin croslinking factor 1; HSF3P18, heat-shock 10-kDa protein 1 pseudogene 8; RRAGC, Ras-related GTP binding C; NMYC, MYC binding protein; GAP, gap junction protein, alpha 9, 59 kDa; RBBPL2, ribosomal protein L2 pseudogene; GPD2, glyceraldehyde 3-phosphate dehydrogenase pseudogene 52, RNA; NME1P1, nucleoside diphosphate kinase 1 pseudogene 1; PPIAP21, phosphatase 2, 3-aminotransferase 2, mitochondrial intermediate; LINC00424, mitochon- 

dian chromosomal protein 536 pseudogene 2; MACF1, microtubule-actin croslinking factor 1; HSF3P18, heat-shock 10-kDa protein 1 pseudogene 8; RRAGC, Ras-related GTP binding C; NMYC, MYC binding protein; GAP, gap junction protein, alpha 9, 59 kDa; RBBPL2, ribosomal protein L2 pseudogene; GPD2, glyceraldehyde 3-phosphate dehydrogenase pseudogene 52, RNA; NME1P1, nucleoside diphosphate kinase 1 pseudogene 1; PPIAP21, phosphatase 2, 3-aminotransferase 2, mitochondrial intermediate; LINC00424, mitochon- 

dian chromosomal protein 536 pseudogene 2; MACF1, microtubule-actin croslinking factor 1; HSF3P18, heat-shock 10-kDa protein 1 pseudogene 8; RRAGC, Ras-related GTP binding C; NMYC, MYC binding protein; GAP, gap junction protein, alpha 9, 59 kDa; RBBPL2, ribosomal protein L2 pseudogene; GPD2, glyceraldehyde 3-phosphate dehydrogenase pseudogene 52, RNA; NME1P1, nucleoside diphosphate kinase 1 pseudogene 1; PPIAP21, phosphatase 2, 3-aminotransferase 2, mitochondrial intermediate; LINC00424, mitochon- 

dian chromosomal protein 536 pseudogene 2; MACF1, microtubule-actin croslinking factor 1; HSF3P18, heat-shock 10-kDa protein 1 pseudogene 8; RRAGC, Ras-related GTP binding C; NMYC, MYC binding protein; GAP, gap junction protein, alpha 9, 59 kDa; RBBPL2, ribosomal protein L2 pseudogene; GPD2, glyceraldehyde 3-phosphate dehydrogenase pseudogene 52, RNA; NME1P1, nucleoside diphosphate kinase 1 pseudogene 1; PPIAP21, phosphatase 2, 3-aminotransferase 2, mitochondrial intermediate; LINC00424, mitochon- 

dian chromosomal protein 536 pseudogene 2; MACF1, microtubule-actin croslinking factor 1; HSF3P18, heat-shock 10-kDa protein 1 pseudogene 8; RRAGC, Ras-related GTP binding C; NMYC, MYC binding protein; GAP, gap junction protein, alpha 9, 59 kDa; RBBPL2, ribosomal protein L2 pseudogene; GPD2, glyceraldehyde 3-phosphate dehydrogenase pseudogene 52, RNA; NME1P1, nucleoside diphosphate kinase 1 pseudogene 1; PPIAP21, phosphatase 2, 3-aminotransferase 2, mitochondrial intermediate; LINC00424, mitochon- 

dian chromosomal protein 536 pseudogene 2; MACF1, microtubule-actin croslinking factor 1; HSF3P18, heat-shock 10-kDa protein 1 pseudogene 8; RRAGC, Ras-related GTP binding C; NMYC, MYC binding protein; GAP, gap junction protein, alpha 9, 59 kDa; RBBPL2, ribosomal protein L2 pseudogene; GPD2, glyceraldehyde 3-phosphate dehydrogenase pseudogene 52, RNA; NME1P1, nucleoside diphosphate kinase 1 pseudogene 1; PPIAP21, phosphatase 2, 3-aminotransferase 2, mitochondrial intermediate; LINC00424, mitochon- 

dian chromosomal protein 536 pseudogene 2; MACF1, microtubule-actin croslinking factor 1; HSF3P18, heat-shock 10-kDa protein 1 pseudogene 8; RRAGC, Ras-related GTP binding C; NMYC, MYC binding protein; GAP, gap junction protein, alpha 9, 59 kDa; RBBPL2, ribosomal protein L2 pseudogene; GPD2, glyceraldehyde 3-phosphate dehydrogenase pseudogene 52, RNA; NME1P1, nucleoside diphosphate kinase 1 pseudogene 1; PPIAP21, phosphatase 2, 3-aminotransferase 2, mitochondrial intermediate; LINC00424, mitochon-
although TMPRSS2-ERG and related fusions are associated with aggressive disease. TMPRSS2-ERG fusion can be detected by low-coverage high-throughput sequencing of cfDNA from advanced-stage prostate cancer patients preselected for high tumor cfDNA concentrations (8). ERG was not recognized by our group analysis without any patient preselection.

An independent method of assessing chromosomal imbalances in cancer is the clinically useful tool of cytogenetics. Although the Progenetix cytogenetics database of 702 prostate carcinomas (http://www.progenetix.org) cannot be directly compared with our more discrete top 20 loci of gains and losses, comparison of gains and losses is of interest. Our data span the Gleason score spectrum, whereas the cytogenetic clinical data will be skewed toward more clinically advanced disease. The cytogenetics recorded chromosomal imbalances associated with gains and losses over large regions of all 22 somatic chromosomes, whereas we focused on discrete chromosome regions (Table 1 bins) with the greatest number of gains or losses. Nevertheless, chromosome 8 was recorded as having the greatest number of gains by cytogenetics, in agreement with our identification of 3 gain loci. Similarly, the largest cytogenetic recorded losses occurred in chromosome 13. We found a large region of chromosome 13 with a loss. The major disparity was observed on chromosomes 6 and 16, with minor cytogenetic p-arm gains overlapping our observed gains, but was not consistent with the losses observed by cytogenetics in the q arm, which were not observed in the top 20 loci of our cfDNA analysis.

Analysis of the chromosomal genetic content identified in Table 1 and relevant functional content in online Supplemental Table 2 provides insight into the dynamic real-time elements of chromosomal instability in prostate cancer. These 20 regions of chromosomal instability represent <0.5% of the total chromosomal DNA distributed over 12 chromosomes. Five regions were associated with significant reductions in copies of cfDNA in 3 chromosomes (HS1, HS4, and HS13) (Table 1) vs healthy controls. The remainder of these chromosomal hotspots were associated with significant increases in copies of cfDNA. Among the 99 functional genes and 25 pseudogenes contained within these loci of chromosomal instability (see online Supplemental Table 2), 49 of the functional genes (49%) have functions directly relevant to chromosomal instability and cancer. These include genes involved in cell division, transcription/translation, altered membrane structure function, differentiation, and apoptosis. Target functions include kinetochore and sister chromatid stabilization; S-phase signaling and G1-S regulation; transcription factors controlling cell cycle progression, epigenetic regulation, membrane oncogenes, and release of metabolic growth factors; and morphogenesis/cytoskeleton organization. One notable gene of interest is a hormone gene [RORA (RAR-related orphan receptor A)] enhancer within a fragile region of the genome (see online Supplemental Table 2). Genes over-expressed in a variety of cancers include FAM149A (family with sequence similarity 149A, member A), ONECUT1 (one cut homeobox 1), TOX2 (tox high mobility group box family member 2), and GDAPIL1 (ganglioside induced differentiation associated protein 1-like 1). Additionally, within these foci of instability are 9 genes of unknown function and a large number of transcribed but untranslated RNA that are potential sources of regulatory miRNAs. The large number of cancer-related genes found in the abnormal cfDNA distribution are consistent

### Table 2. AUC based on frequencies of apoptotic DNA in serum of patients with prostatic cancer.

<table>
<thead>
<tr>
<th>Sample set size (n)</th>
<th>Prostate cancer</th>
<th>Control</th>
<th>Prostate cancer</th>
<th>Control</th>
<th>AUC (95% CI)</th>
<th>Accuracy (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>204</td>
<td>207</td>
<td>All</td>
<td>Healthy</td>
<td></td>
<td>0.92 (0.87-0.95)</td>
<td>0.84 (0.80-0.88)</td>
</tr>
<tr>
<td>89</td>
<td>207</td>
<td>Gleason score &lt;7</td>
<td>Healthy</td>
<td></td>
<td>0.94 (0.89-0.97)</td>
<td>0.89 (0.84-0.92)</td>
</tr>
<tr>
<td>84</td>
<td>207</td>
<td>Gleason score ≥7</td>
<td>Healthy</td>
<td></td>
<td>0.91 (0.85-0.96)</td>
<td>0.87 (0.83-0.91)</td>
</tr>
<tr>
<td>204</td>
<td>227</td>
<td>All</td>
<td>Healthy and OMCa</td>
<td></td>
<td>0.92 (0.87-0.95)</td>
<td>0.84 (0.79-0.88)</td>
</tr>
<tr>
<td>89</td>
<td>20</td>
<td>Gleason score &lt;7</td>
<td>OMC</td>
<td></td>
<td>0.93 (0.83-0.98)</td>
<td>0.91 (0.86-0.96)</td>
</tr>
<tr>
<td>84</td>
<td>20</td>
<td>Gleason score ≥7</td>
<td>OMC</td>
<td></td>
<td>0.90 (0.78-0.97)</td>
<td>0.88 (0.81-0.94)</td>
</tr>
<tr>
<td>192</td>
<td>201</td>
<td>41 &lt; age &lt; 81b</td>
<td>41 &lt; age &lt; 81</td>
<td></td>
<td>0.91 (0.86-0.95)</td>
<td>0.84 (0.79-0.88)</td>
</tr>
<tr>
<td>76</td>
<td>174</td>
<td>41 &lt; age &lt; 65</td>
<td>41 &lt; age &lt; 65</td>
<td></td>
<td>0.92 (0.85-0.97)</td>
<td>0.86 (0.80-0.90)</td>
</tr>
</tbody>
</table>

*a Age range between youngest prostate cancer (n = 41) and oldest control (n = 81).
*b Other medical conditions (OMC) samples were 10 benign prostate hypertrophy and 10 prostatitis. These were not included in the original ROC AUC and cross-validation analyses. When added as additional controls for confirmation, the ROC AUC did not deteriorate the specificity or sensitivity of the original set, which serves as additional rationale for the use of the selected regions.
with chromosomal instability that is reflected in the cfDNA.

By massive parallel sequencing of cfDNA, we identified a small number of foci (<0.5% of total DNA) of chromosomal instability, a recognized common phenomenon of malignant neoplasia. These foci are populated with a variety of genes relevant to the process of neoplasia that accurately predict prostatic carcinoma without dependence on the stage of disease or tumor aggressive potential. These core genes may provide clues to the genomic core for prostate cancer distinct from the progressive genomic instabilities driven by the multiplicity of fusions with their individual related gene dysfunctions responsible for the wide variance in clinical phenotypes.

cfDNA is currently in an exponential clinical development growth phase. Cancer-derived DNA present in blood was first reported in 1948 (27), but the field remained in a dormant state for >50 years. Measurement of absolute concentrations of cfDNA has been suggested for the diagnosis (28) and prognosis (29) of breast and lung cancer (30) but has little clinical utility. The rapid development of NGS platforms with their massive sequence capacities, however, has provided the statistical power to drive clinical applications. Massive parallel sequencing to detect genomic alterations in chromosomal copy number in blood was reported initially for the detection of fetal trisomy 21 in 2008 (31). The method has been validated for trisomies 13, 18, and 21 as a clinical laboratory procedure with a remarkable diagnostic accuracy >99% (32). The statistical power provided by NGS platforms through massive parallel sequencing coupled with mass assembly [mass sequence and assembly (MSA)] allowed us to report algorithms predictive of variant Creutzfeldt-Jakob in ruminantia (33, 34) and multiple sclerosis (35) and breast cancer (14, 36) in humans with-

---

**Fig. 2.** ROC curve of validated data set of prostate cancer vs controls (black line). Dashed line depicts the borders of the 95% CI. (A), Prostate cancer patients (n = 202) compared with controls (n = 207). (B), Prostate cancer patients with Gleason score <7 (n = 89) compared with controls (n = 207).

---

**Fig. 3.** CNI index determined in cfDNA by sequence ploidy heterogeneity for prostate carcinoma, benign prostatic hypertrophy, and prostatitis compared with controls. Boxes (interquartile ranges) and whiskers (5th and 95th percentiles) are shown together with the median (black horizontal line) of the investigated groups.
out genomic knowledge, as is the case in this study. Despite the potential for bias introduced by whole-genome sequencing, cancer-derived cfDNA has been demonstrated to recapitulate genomic tumor DNA (11, 20, 21, 37). Individual tumor genomic analysis brings significant advantages to the power of cfDNA diagnostics. Preliminary data recently reported at the 2014 American Society of Clinical Oncology meeting on small numbers of head/neck and colorectal cancers indicates that patient responses to DNA targeted therapy (radiation, 5-fluorouracil, and/or cis-platinum) is most effective ($P \leq 0.01$) in patients with chromosomal number imbalance (38, 39).

Although tumor tissue is still the gold standard for clinical molecular diagnostics, major disadvantages are involved in acquiring tissue samples (i.e., biopsies may be complicated and are invasive procedures). On the other hand, liquid biopsies, owing to their minimal invasiveness, can be scheduled more frequently and may be the only option for some patients. Furthermore, a tissue biopsy is often obtained from only 1 tumor site, and the genetics of distant metastases cannot be assessed (8, 21). It is arguable that cfDNA sequencing better reflects the genomes of all cancer subclones present in a patient. If validated with larger numbers of patients, massive parallel sequencing will provide substantial value to the selection of therapeutic regimens by li-

References

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures or Potential Conflicts of Interest: Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:

Employment or Leadership: E. Schütz, Chronix Biomedical; J. Beck, Chronix Biomedical; H. Urnovitz, Chronix Biomedical; K. Bornemann-Kolatzi, Chronix Biomedical; W. M. Mitchell, Chronix Biomedical, Vanderbilt University.

Consultant or Advisory Role: W. M. Mitchell, Chronix Biomedical.

Stock Ownership: E. Schütz, Chronix Biomedical; J. Beck, Chronix Biomedical; W. M. Mitchell, Vanderbilt University.

Honoraria: E. Schütz, Chronix Biomedical.

Expert Testimony: None declared.


Other Remuneration: E. Schütz, Chronix Biomedical.

Role of Sponsor: The funding organizations played a direct role in the design of study, review and interpretation of data, and preparation and final approval of manuscript.


26. Berger MF, Lawrence MS, Demichelis F, Dier Y,


