Cancer Research

Utility of Single-Cell Genomics in Diagnostic Evaluation of Prostate Cancer

Joan Alexander[1], Jude Kendall[1], Jean McIndoo[1], Linda Rodgers[1], Robert Aboukhalil[1], Dan Levy[1], Asya Stepansky[1], Guoli Sun[1], Lubomir Chobardjiev[2], Michael Riggs[1], Hilary Cox[1], Inessa Hakker[1], Dawid G. Nowak[1], Juliana Laze[3], Elton Llukani[3], Abhishek Srivastava[4], Siobhan Gruschow[4], Shalini S. Yadav[4], Brian Robinson[5], Gurinder Atwal[1], Lloyd C. Trotman[1], Herbert Lepor[3], James Hicks[1], Michael Wigler[1], and Alexander Krasnitz[1]

Abstract

A distinction between indolent and aggressive disease is a major challenge in diagnostics of prostate cancer. As genetic heterogeneity and complexity may influence clinical outcome, we have initiated studies on single tumor cell genomics. In this study, we demonstrate that sparse DNA sequencing of single-cell nuclei from prostate core biopsies is a rich source of quantitative parameters for evaluating neoplastic growth and aggressiveness. These include the presence of clonal populations, the phylogenetic structure of those populations, the degree of the complexity of copy-number changes in those populations, and measures of the proportion of cells with clonal copy-number signatures. The parameters all showed good correlation to the measure of prostatic malignancy, the Gleason score, derived from individual prostate biopsy tissue cores. Remarkably, a more accurate histopathologic measure of malignancy, the surgical Gleason score, agrees better with these genomic parameters of diagnostic biopsy than it does with the diagnostic Gleason score and related measures of diagnostic histopathology. This is highly relevant because primary treatment decisions are dependent upon the biopsy and not the surgical specimen. Thus, single-cell analysis has the potential to augment traditional core histopathology, improving both the objectivity and accuracy of risk assessment and inform treatment decisions.

Significance: Genomic analysis of multiple individual cells harvested from prostate biopsies provides an indepth view of cell populations comprising a prostate neoplasm, yielding novel genomic measures with the potential to improve the accuracy of diagnosis and prognosis in prostate cancer. *Cancer Res; 78(2); 348–58. ©2017 AACR.*

Introduction

Histopathology of tissue biopsies is a standard method used for evaluating cancer risk. Many decades of experience have led to classification of the histologic types correlated with clinical outcome. Prostate cancer diagnosis is routinely made by obtaining biopsy specimens under ultrasound guidance. The Gleason score, assigned to the prostate biopsy (PB), is a well-established morphologic grading system that predicts adverse pathology of the radical prostatectomy (RP) surgical specimen and biochemical recurrence following local curative treatment of prostate cancers. However, the Gleason score, which is based on changes in glandular architecture, is hampered by multifocality, morphologic heterogeneity of prostatic lesions, sparse stochastic sampling, and inter- and intraobserver variability (1–3). Of the nearly one million men biopsied annually (4), only about one fourth are diagnosed with cancer (5). Half of those diagnosed have a Gleason score of 6 or lower (6), which has very low metastatic potential, and the proper clinical treatment for these men is unclear. Indeed, upon removal of the prostate and subsequent histologic analysis, the Gleason score is often revised, and an upgrade upon surgery is associated with adverse prognosis (7–9). Hence, there is an unmet need for improved diagnostics and risk assessment.

We report here a small pilot study to explore the utility of single nucleus sequencing (SNS) to aid diagnosis. Although the heterogeneity and molecular complexity of prostate tumors have been characterized in several large-scale genomic studies (10–16), none have used multiregional single-cell DNA analysis to examine intraprostatic genomic complexity. The main output of SNS consists of profiles of integer-valued copy-number variation (CNV) in individual cells. Given this output, we can examine intratumor genomic heterogeneity and determine the genealogical relationships among tumor

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. [2]Technological School of Electronic Systems, Technical University of Sofia, Sofia, Bulgaria. [3]Department of Urology, New York University Langone Medical Center, New York, New York. [4]Department of Urology, Weill Medical College of Cornell University, New York, New York. [5]Department of Pathology and Laboratory Medicine, Weill Medical College of Cornell University, New York, New York.

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

Current address for R. Aboukhalil: GenapSys, Redwood City, CA; Current address for G. Sun: Intuit Inc., Mountain View, CA; Current address for A. Srivastava: Department of Urology, Montefiore Medical Center, Albert Einstein College of Medicine, New York, New York; Current address for S. Gruschow: PolicyLab, The Children's Hospital of Philadelphia, Philadelphia, PA; Current address for S.S. Yadav: Department of Urology, Icahn School of Medicine at Mount Sinai, New York, New York; Current address for J. Hicks: Department of Biological Sciences, University of Southern California, Los Angeles, CA.

Corresponding Author: A. Krasnitz, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724. Phone: 516-367-6863; E-mail: krasnitz@cshl.edu

doi: 10.1158/0008-5472.CAN-17-1138

cell subpopulations. As the cells are sampled from a number of anatomically separate locations, we can delineate cell migration patterns within each subpopulation. We can further assess, within each subpopulation, the degree of global chromosomal instability and gain direct insights into molecular mechanisms that may be driving the growth and metastatic potential of malignancy, such as locus-specific amplifications and deletions. Thus, SNS is a source of genomic information complementary to conventional pathology and histology. As very few cells are required, in the hundreds, only minimally invasive procedures are needed.

Here, we describe a small pilot study on 11 patients. In 8 cases, we compare genomic pathology based on SNS to histopathology reports based on standard hematoxylin–eosin (H&E) staining of diagnostic needle core biopsies. Our procedure maintained tissue integrity of cores for downstream microscopic assessment because we used only the cells that exfoliated with gentle washing of the core prior to formalin fixation. By maintaining the association of exfoliated cells with their core of origin, we directly compare those exfoliated cells with histopathology from their anatomic region. Clearly, one distinction in the two procedures is that although histopathology samples core longitudinal sections, analyses of exfoliated cells sample the core surface. For all biopsied patients, we used both standard random cores and MRI-ultrasound fusion–targeted biopsies. The prostate was removed in 5 of these 8 cases, so we also compare single-cell molecular analysis with the final pathologic assessment. In 3 cases (3 of 11) only cores from RP were available for SNS. In the following, we use the terms "core," "sector," or "area" interchangeably to denote an anatomic origin within a prostate of the cells we profile.

As part of our program to evaluate SNS in context with anatomy and histopathology, we have developed new algorithms for statistical inference of clonal structure based on CNV profiles. We also introduce an early version of a "single cell genomics viewer" (SCGV). This is an integrated and interactive visualization platform for CNV profiles in relation to clonal phylogeny, anatomic spread, pathologic score, and genome annotation, among others.

Analysis of several hundred cells per patient provides a detailed evolutionary picture of their prostatic neoplasia. Sectors associated with cancers identified by histopathology (positive Gleason score) typically display sets of cells with statistically significant shared copy-number events, which we interpret as evidence for clonal expansion. With important exceptions, benign sectors typically do not display such clones. In cases with low to intermediate grade disease, we find that SNS has greater sensitivity than core histopathology, as judged by comparison with revised grading following RP. Thus, SNS has the potential to significantly improve tumor staging and grading, and, given the minuscule amount of tissue it requires, could do so with a less invasive procedure such as fine needle aspiration.

Materials and Methods

We performed SNS on a total of 4,021 nuclei from 122 anatomical locations in 11 patients spanning a broad histologic spectrum from benign prostatic epithelium to high-grade prostatic intraepithelial neoplasia (HGPIN) and frank carcinoma (within and beyond the prostate) in both early and advanced stage diseases. The entire workflow of sample and data processing by SNS is depicted in Supplementary Fig. S1.

Sample acquisition from RP specimens

A total of 16 tissue biopsies were obtained from 3 patients (COR001.GS9.1, COR002.GS6.1, and COR003.GS9.2) undergoing RP at New York Presbyterian-Weill Cornell Medical Center. The patients provided informed consent, and radical prostatectomy specimens (RPS) were processed and bio-banked according to a protocol previously published (17). The clinical study was conducted following U.S. Common Rule, with approval from Weill Cornell Medical Center Institutional Review Board. The lead study pathologist, after reviewing H&E sections of the banked frozen tissue, selected 5 to 6 sectors of interest from each RPS. A 1 mm diameter core of tissue from each of the sectors of interest was obtained from the frozen tissues. All cores of frozen tissue were placed into sterile tubes and maintained on dry ice for transfer to Cold Spring Harbor Laboratory for SNS. Clinical and pathologic data were collected and maintained in a database curated by the Weill Cornell Medical College Center for Prostate Cancer.

Sample acquisition from PB washings

Under an Institutional Review Board–approved protocol, 8 patients (NYU001-NYU005, NYU007, and NYU009-NYU011) undergoing PB at the Smilow Comprehensive Prostate Cancer Center (SCPCC) at NYU Langone Medical Center participated in the SNS study. Informed consent was obtained independently for the PB and participation in the clinical study, as guided by the principles of the Belmont Report conducted in accordance with the Common Rule. Demographic and clinical information related to risk and aggressiveness of diagnosed cancers was collected and maintained in SCPCC database. All but 1 patient received the standard 12-core TRUS-guided biopsy, and all patients with an MRI lesion underwent MRI-ultrasound fusion–targeted biopsy with 2 to 4 cores obtained from the MRI lesion(s) depending on clinical indication. The biopsy cores were processed separately with the site of origin noted. Individual cores of prostate tissue were placed in site-separated vials filled with 5 mL of sterile wash buffer (1x PBS containing 0.5% BSA and 2 mmol/L EDTA) and gently inverted several times for 60 seconds to enhance exfoliation of prostate cells. After inversion, prostate cores were removed from the wash solution using disposable single-use sterile forceps and transferred to site-separated containers with formalin fixative for histologic processing and pathologic evaluation.

The vials containing the exfoliated cell suspensions were coded in order to identify the site-specific location corresponding to the biopsy template, along with a numerical code to deidentity the patient. The key to the patient ID was maintained by the Study Coordinator at NYU in order to enable correlation of the molecular results with histopathology. The prostate cores were examined by NYU Langone Medical Center pathologists who assigned a Gleason score to all observed prostate cancers along with the total linear dimension of cancer, and the percentage of the tissue core that was Gleason patterns 3, 4, and 5. The presence and extent of HGPIN, perineural invasion, and atypical small acinar proliferation were also noted in the final diagnostic pathology reports. PB washings were kept on wet ice for 1 to 2 hours during transfer to CSHL where the cell suspensions were briefly centrifuged to pellet the cells and lysed using NST-DAPI buffer described in previous studies (18, 19). A total of 5 of these men underwent RP. The linear diameter, Gleason score, and percentage of Gleason 3, 4,

and 5 were reported for all observed tumor(s). The site(s) and extent of extra-prostatic extension, the presence of seminal vesicle invasion, and surgical margins were reported in final pathology report on RPS.

Nuclei isolation from clinical samples, DNA staining, and single-cell FACS

Nuclei were isolated from frozen core biopsies and biopsy washings using NST-DAPI buffer [800 mL of NST (146 mmol/L NaCL, 10 mmol/L Tris base at pH 7.8, 1 mmol/L $CaCl_2$, 21 mmol/L $MgCl_2$, 0.05% BSA, 0.2% NP-40)], 200 mL of 106 mmol/L $MgCl_2$, 10 mL of 500 mmol/L EDTA at pH 8.0, and 10 mg of DAPI. Nuclei were prepared from frozen core biopsy of RPS by finely mincing tissue in 1.0 to 2.0 mL of NST-DAPI buffer as described in a previously published protocol (18, 19). Nuclei were prepared from PB washings by gently centrifuging washings at 1,000 rpm for 5 minutes to pellet the exfoliated cells followed by removal of supernatant and addition of 1.0 mL of NST-DAPI buffer to the cell pellet. All nuclei suspensions were filtered through a 25-μm cell strainer prior to flow sorting. Single nuclei were sorted by FACS using the BD Biosystems SORP flow cytometer by gating cellular distributions based on differences in total DNA content (ploidy) relative to DAPI intensity such that enrichment for aneuploid cells is possible.

Whole-genome amplification and Illumina library construction

Single nuclei were deposited into individual wells in a 96-well plate and amplified using Sigma-Aldrich's GenomePlex WGA4 kit (catalog no. WGA4-50RXN) according to the manufacturer's instructions. Whole-genome amplification (WGA) DNA was sonicated using a Covaris focus acoustics system. The Covaris E210 300± sonication program generated WGA DNA inserts of the desired length of approximately 300 bp (range, 200–400 bp) for library construction. Multiple libraries were combined into pools ranging from 8 to 12 libraries to pools of 96 libraries for 76 bp single-read sequencing on single lanes of Illumina's GAIIx and HiSeq flow cells, respectively. The first 30 bases of each read were trimmed to remove any WGA primer sequence. For the RPS, we profiled about 25 to 100 cells from the 5 to 6 sectors of interest, and for the core washings, we profiled approximately 20 to 25 cells per washing from the standard 12 random and from the MRI-ultrasound fusion–targeted biopsies.

Derivation of integer-valued CN profiles

Whole-genome copy-number profiles for each sample were determined as described (18, 19) with minor modifications. A short summary follows: Illumina single-end reads were aligned end to end without gaps using Bowtie (20) to human genome version GRCh37 with the pseudo-autosomal regions of chromosome Y masked. The genome was partitioned into 20,000 bins with equal expected number of uniquely mapped positions. A read-count vector was formed for each single-nucleus read set, with the numbers of reads mapping to each bin as components. DNA copy-number profiles were derived for each nucleus by first normalizing the corresponding read-count vector to the mean read count of one per bin, then using LOWESS regression to remove bias in the read counts due to variation of bin-wise GC content. A number of regions in the genome, mainly at or adjacent to centromeres, have been found to consistently display anomalously high read depth in both bulk and single-cell sequencing data (21). Bins corresponding to such regions were masked from downstream copy-number analysis. For the remaining bins, a piecewise constant approximation (segmentation) of the copy-number profile (segmentation vector) is computed using circular binary segmentation algorithm as implemented by R language package DNAcopy (22). The result is a segmented profile with a mean value close to 1. The integer position-dependent copy number was estimated by a least-squares fit, under the assumption that the copy number lies in the 1 11 range, and with the cell ploidy as a parameter (23).

Removal of shredded profiles

We find some cell profiles are "shredded," meaning that substantial portions of the genome are homozygously deleted. Although the reason for these deletions is not known, it is unlikely that a cell can sustain such major losses and retain viability, and this widespread damage must therefore have occurred post vivo. Consistent with this assumption, genomic locations of these homozygous deletions do not recur from cell to cell. We remove such incomplete profiles from further consideration if the homozygous losses span over 1% of the genome.

Determination of change points in individual CN profiles

Each segmented CN profile is an integer-valued function of the bin number, which we visualize as proceeding in chromosomal order from chromosome 1 to Y. The integer-valued function is further reduced to a set of change points (CP), also sometimes referred to as a "break points." A CP is specified by its position (bin number) and the sign of CN discontinuity at that position, positive if the function goes up, negative otherwise. To allow for the uncertainty inherent in segmentation, a CP is assumed to be localized in an interval spanning b bins and centered at the most likely CP position as determined by the segmentation algorithm. Our best estimate for b is 3 for the data analyzed here. Thus, a CN profile for the n-th cell is reduced to a set S_n of genomic intervals of length b, one set for each sign, which we also write as $S_n \equiv S^+{}_n \cup S^-{}_n$. For brevity, we will call this form of a CN profile CP-reduced. To further guard against anomalously high read counts in the vicinity of a centromere, we filter out all CP due to copy-number events that lie entirely within the regions spanned by cyto-bands p11 through q11 of each chromosome.

Definition of the "feature set" and "incidence table" for a PB sample

A critical step in our analysis is the definition of the "feature set" F for a sample of N cells, $F \equiv \{f_k, 1 \le k \le K\}$. We describe first F^+ by considering the set $S^+ \equiv \cup_n S^+{}_n$ of all positive intervals, present in the CP-reduced profiles of all N cells. The set $F^+ \equiv \{f^+{}_k, 1 \le k \le K^+\}$ is a minimum set of "piercing points" for S^+, as described previously (24). Briefly, the piercing points are a smallest set of points such that each interval in S^+ contains at least one of them. Next, we derive the binary incidence table T^+, with entries $T^+{}_{kn}$, indicating whether $f^+{}_k$ is contained in an interval in $S^+{}_n$. The subset F^- and the incidence table T^- are derived in similar fashion, starting from a set $S^- \equiv \cup_n S^-{}_n$ of all negative intervals. Finally, the table T is obtained by concatenation of T^+ and T^-. We call a feature widely shared by a subset of cells if it is present in at least 85% of the subset.

Computation and significance assessment of pairwise dissimilarity among cells in a biopsy sample

For two cells m and n in a sample, the dissimilarity is derived from the number of overlapping and disjoint features: those shared and those not shared, as derived from the incidence tables T^+ and T^-. For simplicity of presentation, we consider their concatenation T. For the derivation, we use a one-tailed Fisher exact test (in R alternative = "greater") on the respective 2×2 contingency table comprised of the count of features shared and not shared, yielding the P value p_{mn}. We take $\log_{10}(p_{mn})$ as the dissimilarity measure. The resulting $N(N\text{-}1)/2$ dissimilarities for all possible pairs of N cells in the sample are assessed for statistical significance by testing the null hypothesis that T^+ and T^- are random, keeping their row and column sums fixed: That is, the overall counts per feature and overall number of incidences per cell fixed. This is accomplished by creating randomized incidence tables with the preserved margins of the observed incidence matrices, and for each computing the $N(N\text{-}1)/2$ dissimilarities. A total of 500 randomizations were performed for each biopsy patient. Finally, the FDR is computed by comparing the observed dissimilarities to those obtained from the randomizations. An R language implementation of the randomization procedure is provided in the Supplementary Information.

Following this procedure, we found four pairs of single-cell genomes, in a total of thousands of cells compared over 12 samples, which had anomalously low dissimilarity. In these four cases, each cell of a pair originated from two neighboring wells on the 96-well plate, and to rule out the possibility that the DNA of the two neighboring wells were cross-contaminated, we eliminated each pair. As a result of this cautionary tale, we checked the well adjacency of cells considered clones, the quantal nature of their features, and where they reside on the phylogenetic trees. We find no reason to suspect well contamination plays any role in clone identification, and as an extra safeguard, we impose a rule of three (see next section).

Genealogy reconstruction and identification of clones

Genealogical relations among the cells (or more formally, "leaves") in the sample were reconstructed by hierarchical clustering, with the dissimilarity matrix as defined above and with the average linkage. A branch in the resultant tree was termed "cohesive" if, for any pair of its cells, the FDR for the dissimilarity did not exceed a threshold value $t = 0.01$, and if its parent branch did not have this property. A cohesive branch was considered hard clonal if at least four features $f \in F$ could be found, such that each was widely shared by the cells in the branch but not by the cells in the entire tree. In addition, for the reasons stated in the previous section, a branch must contain at least three cells to be designated hard clonal. Among the ancestral branches of a hard-clonal branch, we then identify the one nearest the root for which at least three features are widely shared. Such branches are termed soft clonal. Soft-clonal branches with six cells or more were examined for evidence of subclones as follows. First, the clonal incidence table T_C was derived for a clonal branch C, by reducing the sample-wide incidence table T to a subtable with columns only for cells $c \in C$, followed by removing all the constant rows in that subtable. As a result, the rows of T_C corresponded to the features in the clonal feature set $F_C \subseteq F$. Next, T_C was used to derive within-clone dissimilarities, assess their significance, and reconstruct the within-clone genealogy, following the procedure described for T above. A branch in the resultant tree was deemed hard (soft) subclonal using the above criteria for hard (soft)-clonal branches.

A core was considered clone-harboring if cells originating from the core were found to belong to a hard-clonal branch. We then counted cells originating from the clone-harboring cores and belonging to soft-clonal branches in order to quantify clonal involvement of the cores (Tables 1–3 and Supplementary Table S1).

Determination of clonal features

A distinguishing property of clonal branches is co-occurrence of multiple features across the cells belonging to the branch. We call such "clonal features" and estimate their number in a biopsy sample using the sample-wide incidence table T. Specifically, we compute the covariance matrix among the features (rows of T) and determine, for each feature f, the sum S_f of three largest covariance matrix elements with features other than f. A feature f is declared clonal if the null hypothesis formulated above for T can be rejected at $P = 0.05$ level, Bonferroni-corrected for the number of features, using S_f as a statistic in a right-tailed test. The null hypothesis is tested by permutation as above. The number of clonal features is tabulated for all biopsy samples and may serve as a measure of tumor progression.

Sensitivity of clone detection to the depth of coverage

To examine how the number and cell content of clonal branches vary as a function of the coverage depth in a PB sample, genealogy reconstruction was performed with the number of input sequencing reads per cell reduced by a factor of 2^r, with $r = 1$, 2, 3, and 4, comparing the results with when all R available reads are used. To this end, three random samples of $2^{-r}R$ reads out of the original R were generated and for each of the four values of the reduction factor. For each randomization, the reads were sorted into $20{,}000 \times 2^{-r}$ bins, with the expected number of reads per bin in the normal genome constant across all bins in the genome and for all values of r. The entire processing pipeline as described above was then followed to identify clonal branches. The number of cells within each clonal branch was counted, and we determined whether the clonal cells thus identified in each read reduction simulation matched the clonal cells identified using all reads.

Interactive single-cell genome data viewer

Single-cell genome data can be viewed in a Python Matplotlib interactive application. This application will be referred to in this section as the single-cell viewer SCGV. The SCGV application opens on a heat map–like display, referred to as the heat map view. This heat map view consists of, from top to bottom, a dendrogram representing a clustering of the single-cell data, two tracks indicating, where appropriate, the clonal and the subclonal identity of cells, a heat map representing copy number, and, finally, four annotation tracks. The cells from a single tumor sector can be viewed in a heat map view where the dendrogram is a subtree of the full tree rather than a reclustering of the cells in that sector. Detailed data from a set of cells can be displayed in a genome view.

A zoom-in feature enables the user to select and view in detail any rectangular portion of the heat map, retaining the subtree for that portion. Further, the columns of the heat map can be

Table 1. Case descriptions

Case	Age	Sample[a]	Sectors	Gleason score biopsy	Gleason score final[b]	Proportion of sectors with pathology[c]	Proportion of sectors with clonality[d]	Highest involvement of cancer[e]	Mean involvement of cancer[f]	Multiple clones and/or subclones	Clonal heterogeneity[g]	Number of clonal features[h]	Proportion of clonal cells (clonal/total)	Clonal spread[i]
NYU003.Benign.1	47	PBXW	13	Benign	NA	0/13	0/13	0	0	No	0	0	0/310	0
NYU002.Pin.1	72	PBXW	13	HGPIN	NA	0/13	0/13	0	0	No	0	1	0/579[j]	0
COR002.GS6.1	62	TCRP	5	6 (3+3)	6 (3+3)	2/5	2/5			No	1	34	4/403[j]	0.01
NYU005.GS6.2	64	PBXW	14	7 (3+4)	6 (3+3)[k]	4/14	1/14	30	5	No	1	0	8/309	0.03
NYU001.GS7.1	63	PBXW	14	7 (4+3)	7 (3+4)	8/14	8/14	100	40	Yes	2	54	143/615[j]	0.23
NYU007.GS7.2	65	PBXW	13	6 (3+3)	7 (3+4)[l]	1/13	4/13	30	2	Yes	3	31	42/279	0.09
NYU010.GS7.3	79	PBXW	15	7 (3+4)	NA	6/15	2/15	90	11	Yes	3	25	20/341	0.04
NYU004.GS7.4	75	PBXW	14	8 (4+4)	7 (4+3)[k]	6/14	5/14	100	23	Yes	2	41	51/314	0.14
NYU011.GS7.5	63	PBXW	10	7 (4+3)	7 (4+3)	5/10	4/10	60	14	Yes	2	50	21/221	0.08
COR001.GS9.1	77	TCRP	6	9 (5+4)	9 (5+4)	4/6	3/6			No	1	285	86/261	0.29
COR003.GS9.2	80	TCRP	5	8 (4+4)	9 (4+5)[l]	3/5	3/5			Yes	2	69	117/389	0.28
Median Age	65	Total	122	—–	—–	39/122	32/122	—–	—–	—–	—–	—–	492/4021	—–

NOTE: The 11 clinical cases span a broad histologic spectrum from benign prostatic epithelium to HGPIN and frank carcinoma in both early and advance stage disease.

[a]Nuclei were analyzed from locations within and beyond the prostate from either fresh washings of prostate needle core biopsies (PBXW) or frozen tissue cores from radical prostatectomy specimens (TCRP).

[b]The Gleason score of the RP specimen.

[c]The proportion of sectors (cores) with a Gleason score ≥ 6.

[d]The proportion of sectors with clonality is computed using the definition of clone-harboring cores in Materials and Methods.

[e]The highest percentage of core involvement by cancer in the most extensively involved core.

[f]The average involvement of cancer over all random cores and single most extensively involved core for each MRI-targeted biopsy.

[g]Clonal heterogeneity is defined as $= \sum_{\text{clones}} \max(1, \text{number of subclones})$.

[h]The number of clonal features determined using the definition of a clonal feature in Materials and Methods.

[i]Clonal spread is defined as the average proportion of cells in a sector from a clone affecting the highest number of sectors.

[j]Total cells exclude cells of bulk and/or urine samples.

[k]Downgrading in the Gleason score of the RP specimen.

[l]Upgrading in the Gleason score of the RP specimen.

Table 2. Clonality and Gleason status

A. Core biopsies			
Clonality	**<Gleason 6[a]**	**≥Gleason 6**	**Total**
No	77	13	90
Yes	6	26	32
Total	83	39	122
B. Patients			
Clonality	**<Gleason 6**	**>Gleason 6**	**Total**
No	2	0	2
Yes	0	9	9
Total	2	9	11

NOTE: Summary of clonal properties: clonality of cell populations sampled from the prostate correlates with a Gleason score equal to or greater than 6 judged by diagnostic biopsy (**A**) and patient's Gleason status (**B**).
[a]Includes cores with HGPIN and benign prostatic epithelium.

reordered by the sector. The latter feature is useful, for example, for identifying sectors harboring clonal cell populations.

The color coding on the heat map is assigned by copy number for each cell individually. The median copy number is white, median minus 1 is light blue, median minus 2 is dark blue, median plus 1 is light red, and median plus 2 is dark red. Copy number 0 regions are colored yellow regardless of difference from the median. The annotation tracks are sector, ploidy, multiplier, and error. The sector track is color coded to indicate which sector of the tumor the cell came from. The other tracks are encoded in gray scale. The ploidy track indicates the ploidy gate the cells were sorted from. The multiplier is the computed ploidy as described in the Materials and Methods section. The error is the sum of squares of the normalized bin count values minus the segmented values. This is an indication of data quality. If clonal branches are found, cells comprising each clonal branch are indicated in color in the clone track. In the same fashion, cells comprising soft subclonal branches are indicated in the subclone track.

The genome view has one panel per cell. The genome is represented left to right from chromosome 1 through chromosome Y in chromosome position coordinates. There is one data point for each genome bin. The gray line is the normalized bin count. The blue line is the segmented value for each bin. The γ axis is copy number. The cell ID, sector, ploidy, segment error, quantal error, and percent shredded are indicated in the title of each panel. Segment error is the sum of squares of the normalized bin count minus segmented value for each bin. The quantal error is the sum of squares of the segment value minus the rounded segment value (copy number) for each bin. The shredded value is the percentage of bins in the autosomes at copy-number zero.

Results

We applied methods for genome analysis on isolated nuclei obtained from either diagnostic needle core biopsies or surgical specimens. In the former case, the tissue availability for SNS is constrained by the requirements of histopathology: in order to render a diagnosis based on tissue morphology, each needle core must be preserved in its entirety for multisectional microscopic examination, and the core integrity must not be disrupted by sacrificing any portion of it for other purposes. We therefore only used cells exfoliated from the cores. In addition, we isolated and analyzed a small number of cells from urine, collected prior to invasive procedures in patient cases NYU001.GS7.1, NYU002.Pin.1, and COR002.GS6.1. Mapping sparse sequence data yielded copy-number profiles, and these profiles placed cells into phylogenetic trees. We created an integrated view of the relevant data with pathologic assessment and anatomy in an SCGV. The flow diagram of data processing is depicted in Supplementary Fig. S1.

Processing, viewing, and interpretation

We first describe one case (NYU007.GS7.2) in detail to illustrate our methods and their interpretation. A graphical summary of results for this case is presented in Fig. 1A–G. Similar case reports for all are found in the Supplementary Information, where they are illustrated by Supplementary Figs. S2–S22.

A 65-year-old man referred for PB underwent a standard 12-core biopsy procedure with additional tissue cores directed into an MRI lesion highly suspicious for cancer using MRI-ultrasound fusion–targeted biopsy. The pathology report of the prostate biopsies identified only 1 core out of 13 cores with a positive but low Gleason score (GS 6).

We analyzed on average 20 to 25 cells from each core or sector (Fig. 1A). In single-nucleus sequencing, we aim for about two million reads per nucleus. Sequence reads with high-confidence maps to the reference genome were enumerated in 20,000 consecutive genomic "bins," and bin counts used to make a segmented copy-number profile, as previously described (18, 19). No segments shorter than 5 bins were allowed. Thus, assuming a genome size of 3×10^9 base pairs, on average the genomic resolution of our copy-number analysis was 7.5×10^5 base pairs. In the global view of the SCGV (Fig. 1B), the segmented profile of each cell is represented as a separate column in a red-blue heat map, with bins arranged in the genome order as rows. This cell information is integrated with sector and ploidy encoded as color

Table 3. Correlation of genomic and histopathologic measures of malignancy with the diagnostic and revised Gleason scores

	SCORE			
	Diagnostic Gleason		Revised Gleason	
Measure	**Correlation**	***P* value**	**Correlation**	***P* value**
Clonal heterogeneity	0.36 (−0.43)	0.26 (0.35)	0.86 (0.76)	0.01 (0.11)
Proportion of clonal cells	0.46 (−0.12)	0.14 (0.8)	0.79 (0.63)	0.01 (0.16)
Proportion of clonal features	0.55 (0.12)	0.08 (0.8)	0.79 (0.63)	0.01 (0.16)
Proportion of sectors with clonality	0.55 (0.12)	0.08 (0.8)	0.79 (0.63)	0.01 (0.16)
Clonal spread	0.55 (0.12)	0.08 (0.8)	0.79 (0.63)	0.01 (0.16)
Proportion of sectors with pathology	0.71 (0.36)	0.02 (0.4)	0.7 (0.32)	0.03 (0.5)
Highest involvement of cancer	0.83 (0.67)	0.01 (0.14)	0.78 (0.53)	0.02 (0.3)
Mean involvement of cancer	0.8 (0.6)	0.01 (0.17)	0.7 (0.32)	0.03 (0.5)
Diagnostic Gleason score	1 (1)	0.002 (0.03)	0.64 (0)	0.06 (0.1)

NOTE: Summary of the correlations of genomic and histopathologic measures of malignancy with the diagnostic and revised Gleason scores. Tabulated are Kendall rank correlations and the corresponding *P* values for each measure and each Gleason score, computed for the eight diagnostic biopsy cases in Table 1. The values in parentheses correspond to the five diagnostic biopsy cases for which histopathologic evaluation of a resected prostate is available.

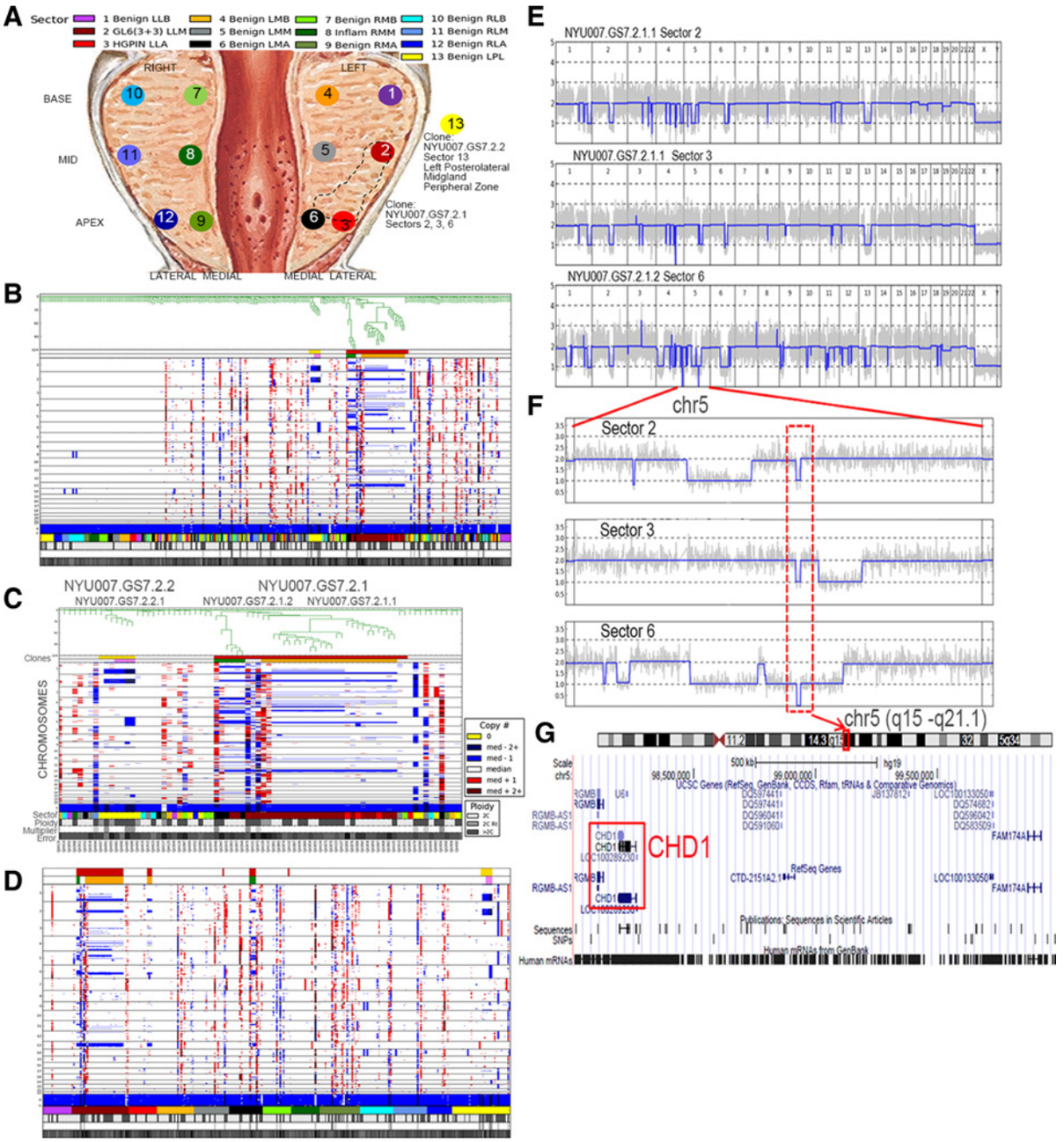

Figure 1.

SCGV images for the case NYU007.GS7.2. SCGV is an interactive and integrative tool for data visualization built in Python. **A,** Illustration of the prostate, a walnut-sized organ, in which a dozen or more biopsies are taken, and single isolated nuclei prepared from each location. The copy-number profiles were determined from low coverage sequence and arranged in a phylogenetic tree. **B,** Plot is one level of the viewer, showing the profiles for each of several hundred nuclei as columns, integrated with information about the sector location, sector pathology, ploidy, and noise. From this level, one can call up at various scales portions of the populations (**C**), or reorder the heat map by sector (**D**). **E-F,** One can view groups of profiles in greater detail and at any scale. **G,** From here, one can open the UCSC Genome Browser to view the genetic loci with annotation, which in this case illustrates that an early event in the ontogeny of this cancer has been a homozygous deletion of the *CHD1* gene.

bars and a gray scale, respectively, indicated in tracks beneath the heat map. We provide a key from sector to Gleason score at the right (FIg. 1C) and place a phylogenetic tree above the copy-number heat map, as we now describe.

We assume throughout that a population with an aberrant and shared copy-number profile consists of cancer cells. The justification for this assumption is presented later. Our primary computational task, therefore, is to determine clonal structure.

It is based on capturing the intuitive notion of shared copy-number events, and we give that an objective and quantitative meaning. We then use hierarchical clustering to reconstruct the phylogeny of the cells in the sample. The viewer software arranges the cells as leaves of a phylogenetic tree, above the heat map. We next use statistical criteria to determine which of the branches of the tree qualify as clones and subclones, and indicate the results in two tracks, beneath the tree and above the heat map.

From this global view, we can zoom and examine in greater detail any aspect of the global view in a separate interface. Zooming (Fig. 1C), we see that the cancer has spread from the main sector 2 to adjacent sectors 3 (called "HGPIN") and 6 (called "benign"). To see this more clearly, we can reorder the heat map by sector (Fig. 1D). The bin counts and segmented profiles for any subset of cells can be reached from any interface displaying part of or the entire heat map, creating the profile interface (Fig. 1E), which can be rescaled over any chromosomal region (Fig. 1F and G). These profile views are interfaces from which genome annotation can be read (Fig. 1H). For example, a magnified plot of chromosome 5 shows a narrow loss of 5q21 in individual cells, outlined in red, a region that encompasses the chromodomain helicase DNA binding protein-1 gene (*CHD1*). This gene functions as a chromatin remodeler and is a known tumor-suppressor gene involved in prostate cancer biology (25). This region undergoes further homozygous loss in sector 6 (bottom profile, Fig. 1E).

We see already that single-cell analysis finds subpopulation structure and detects more cancer than histopathology. Three standard biopsy cores, not one, have cancer cells, and these can be organized as a single clone (NYU007.GS7.2.1) with two subclones (NYU007.GS7.2.1.1 and NYU007.GS7.2.1.2). One of these subclones (NYU007.GS7.2.1.1) resides mainly in sector 2, with some in 3, and the other (NYU007.GS7.2.1.2) mainly in sectors 2 and 6. Subclone NYU007.GS7.2.1.1 is by far the largest population and fills entirely core 2, the only core that pathology gave a positive score. There is also a second apparently unrelated clone (NYU007.GS7.2.2) in sector 13, which comprised three same-site MRI-targeted cores that were called benign. From the data, we readily see that the profiles of clone NYU007.GS7.2.1 are far more complex than those of clone NYU007.GS7.2.2, whereas subclones NYU007.GS7.2.1.1 and NYU007.GS7.2.1.2 are roughly comparable. Finally, we note that following RP, the GS was upgraded to a GS 7 (3+4) with identification of a single tumor focus localized to the left posterior–lateral apex (see Fig. 1A).

Finally, we note the presence of sporadic copy-number events not generally shared between cells. Such events are present in some cells in all cases, and in most sectors, whether clonal or not, and whether cancer is present or not. Noted previously, we considered them artifactual, perhaps arising from nuclear degradation or cleavage occurring during the biopsy procedure.

Correlating histopathology and genomic pathology

Single-cell genome features consist of clonal and subclonal population structure, relative proportion of populations, complexity of genomic profiles, the number and location of invaded sectors, and locus-specific information. The latter may be useful for subtype analysis (15), but we do not use it further in this study. We tabulate genome features and histopathology of the core biopsies in Table 1 and Supplementary Table S1. The latter is interactive, with links leading to the relevant image files either from the SCGV or histopathology. We also present in Fig. 2 a graphical summary of the clonal structure, placed in its anatomic context, for all six diagnostic-biopsy cases in which clones were observed.

The data indicate that clonality is strongly associated with a positive Gleason score, by core (Table 2A and Supplementary Table S2) from biopsy and from postsurgical specimens (122 samples). The association, although strong and highly significant [Fisher exact test odds ratio 24.6, 95% confidence interval (CI), 9.33 to infinity; P value $= 1.26 \times 10^{-11}$], is not perfect, as we have already seen in case NYU007.GS7.2, and discordance was observed in both directions. However, histopathology examines longitudinal sections of the biopsy core, whereas SNS samples exfoliated cells from the perimeter of the core, so the two methods do not examine the same cells. Moreover, the Gleason score assesses architecture and cell morphology, whereas cancer cells may have migrated singly and not yet established architectural features. Per patient, overall correlation between the presence of clonal cells and a positive Gleason score is shown in Table 2B (11 samples, P value $= 0.018$, FE test). These data are the main justification for the assumption stated earlier that we treat cells as cancer when they share abnormal genomes. We note that, by this criterion, none of the cells isolated from urine represented malignant populations.

Next, we considered a relation between histopathologic findings and clonal heterogeneity at diagnosis, defined here as the number of subclones plus the number of clones without subclones. We computed a rank-based (Kendall) correlation of this quantity with the overall Gleason score before and, where available, after regrading following RP, for the 8 patients who underwent diagnostic biopsy. This computation was repeated for other useful genomic descriptors, namely, the proportion of cores containing clonal cells; the proportion of cells judged clonal; the genome complexity defined here as the number of clonal features (*cf* Materials and Methods: Subsection 11); and the clonal spread, defined as the average proportion of cells in a sector from a clone affecting the highest number of sectors. For comparison, we computed the correlation of the original and the revised overall Gleason scores with four measures of malignancy derived from histopathologic evaluation: the original overall Gleason score itself; the proportion of cores called Gleason positive; the percentage of a core involved in cancer, averaged over all cores; and the maximal percentage of a core involved in cancer among all cores. The results are found in Table 3, together with P values. Also shown in Table 3 in parentheses are correlations and P values for a subset of 5 patients who went on to RP with subsequent regrading. The correlation of the original overall Gleason score with itself is one by definition. Not surprisingly, the other three measures derived from the diagnostic histopathologic evaluation also are strongly correlated with the diagnostic Gleason score. However, and more importantly, the five genome-derived descriptors improve in both correlation and P value following regrading. Upon regrading, heterogeneity is the best performing of all nine parameters, followed closely by the other four genomic measures. All five genomic measures better correlate with the revised Gleason score than the four measures derived from conventional pathology.

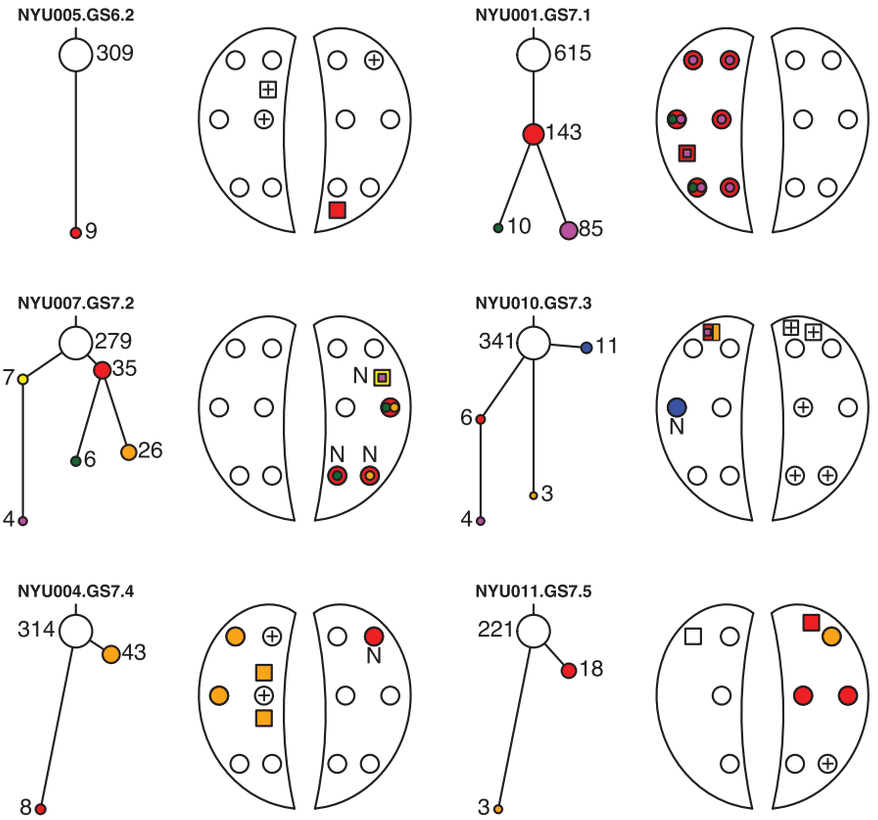

Figure 2.
Clonal structure and spread. Clonal structure is represented as a tree, alongside with a schematic depiction of the diagnostic PB, for the six diagnostic-biopsy cases in which clones were identified. For each tree, the number of cells analyzed by sparse sequencing is represented at each node, with all cells at the root. Clones and subclones are shown as colored nodes of the tree, with the number of cells sampled from each indicated. To the right of each tree, a schematic cross-section of the prostate is shown, with the locations of origin for the standard 12-core biopsy scheme depicted as circles, and the locations of the additional MRI-guided biopsies depicted as squares. The fill colors at each location correspond to the clones and subclones found therein. Cores with pathologic finding of malignancy but no clonal populations detected are indicated by "+." Cores with clonal populations detected but no pathologic finding of malignancy are indicated by "N" nearby.

Single-cell versus bulk analysis of cores

We asked whether bulk sequence analysis would achieve results comparable with those obtained by single-cell analysis. To this end, we examined in detail in one specimen, NYU001.GS7.1, with 12 ultrasound-guided cores and 2 MRI-ultrasound-fusion–guided cores. Eight cores showed both histopathology and genomic clonality, and six showed neither. On all but one core, we sequenced WGA from a hundred to a thousand nuclei and obtained sparse (~2 million mapped reads per core) sequence copy-number profiles (Supplementary Table S3, and associated hyperlinks to profiles). We obtain flat profiles from the bulk analysis of the six cores without cancer. In only four of the seven cores with cancer do we observe an abnormal profile from bulk analysis. Signal is not apparent in three, undoubtedly because so few cells from those cores are from the cancer (see Supplementary Table S3). In three of the four cores where we do see signal from bulk analysis, we see distinctive copy-number features in single-cell profiles (see hyperlinks within Supplementary Table S3), but those distinctive features are absent in the corresponding bulk profile.

To extend the comparison of bulk to single-cell copy-number profiles of core biopsy tissue to additional patient cases, we performed, for all 6 patients with clonal populations detected in core biopsies, an *in silico* pooling of sequence reads from single cells, followed by copy-number profile derivation from the resulting pooled set of reads. This analysis was carried out for all cores with clonal populations and, additionally, for a small number of clone-free cores with Gleason score of 6 or higher. In the NYU001.GS7.1 case, where both *in vitro* and *in silico* bulk copy-number profiles of cores are available, these profiles are in good agreement, with the exception of a single core, for which only ten cells were used to derive an *in vitro* profile. In the remaining five patient cases, the results of the *in silico* bulk analysis are consistent with our findings for the NYU001.GS7.1 case, namely that large-scale copy-number lesions are not apparent unless clonal cells predominate in the specimen.

Clonal specificity of lesions harboring driver genes

We repeatedly observed that genomic lesions affecting genes with a known role in prostate cancer may be clone- and/or subclone-specific. Examples include deletions of 10q23, a region that encompasses PTEN (a tumor-suppressor gene) and is frequently lost in prostate cancer, in cases NYU001.GS7.1 and NYU010.GS7.3, but only in their respective clone and subclone NYU001.GS7.1.1 and NYU010.GS7.3.1.1. In case NYU010.GS7.3, we have one out of three independent clones showing loss of 8p (NKX3.1) and gain of 8q, a region containing the c-MYC oncogene. In addition, case NYU007.GS7.2 has a subclone (NYU007.GS7.2.1.2) with a narrow deletion of 18q21, a region demonstrating frequent allelic losses and implicated in prostate cancer progression.

Discussion

Clinical correlations

We have completed a pilot study of the utility of sparse sequencing of single cells in the evaluation of prostate cancer risk. Our major observations are summarized in Tables 2 and 3. With coverage at about two million sequence reads per cell, we are able to observe clonal genomic CNV patterns and tumor heterogeneity; the complexity of genomic alterations; and amplifications and deletions of specific loci, such as PTEN and RB. We can

infer anatomic spread and clonal expansion between cores, and estimate the proportion of neoplastic cells per core washing.

The single most salient observation in the single-cell data is the statistically robust correlation of "clonal" patterns of copy-number events with a Gleason score of 6 or greater (Table 2). There are two-way discordances between observed clonality and histopathology: instances of cores with morphologic malignancy but without observed clonal copy-number changes, and the reverse. These may arise from the method of sampling: for genome analysis, we sample exfoliated cells from the washings of a core, and therefore its periphery, whereas histopathology is determined from longitudinal cross-sections of the core. Such diverse sampling methods will expose different sets of cells for analysis. In our opinion, sampling is the source of the discrepancy. However, we cannot exclude other possibilities. First, malignancy may manifest first in one of two different ways, either by morphology or by genomic change. Second, early malignant change in genomes may involve mechanisms that we cannot presently observe with sparse sequencing: point mutations, copy-neutral rearrangements, quasi-stable epigenetic changes in gene expression, for examples.

Ideally, we would like to correlate genome profiles with clinical outcome, as has been done for metastatic prostate cancer (26). However, pathologists typically assess the Gleason score following RP, and often that changes the score, and hence risk (7–9). Thus, we do have some component of this study, which can be correlated to outcome: do the genome measures of cores predict the improved assessment afforded by examination of the excised prostate? Of the eight cases for which we had core histopathology, five also underwent RP. Of those five, one Gleason score was upgraded and two were downgraded. Five of five measures of genome pathology correlated better to the revised Gleason than four of four measures obtained from core histopathology (Table 3).

The best genomic parameter for predicting the revised Gleason score was genomic heterogeneity. We note, however, one outstanding case, examined only after surgical excision, "COR001. GS9.1." This cancer had a very high Gleason score, had invaded outside the capsule with extension into the periprostatic soft tissue, and had extensive genomic rearrangements, but it was not heterogeneous. Based on this one example, we expect that heterogeneity is a high-risk predictor except in cases where one dominant cancer subclone with extensive genome alterations has finally emerged that overtakes the other clones. Given that the genomic scores are somewhat independent, an algorithm based on parameters, both histologic and genomic, but trained on many more cases might greatly enhance assessment of risk and decisions about treatment.

Technical considerations

We have used existing laboratory protocols for obtaining single-cell DNA sequence. Building upon previous binning, read counting, and integer-valued segmentation, we added new statistical methods for inferring phylogeny from "clonal" and "subclonal" patterns. We still handle a few steps manually. Among these are elimination of genome patterns from shredded nuclei, and choices about how to handle chromosome ends and centromeres, all discussed in the Materials and Methods section. Work remains to achieve a fully automated procedure that could be implemented in a clinical setting.

New with this report is our SCGV tool that allows us to visualize all the information with a graphical user interface. SCGV integrates DNA profile information with data quality, ploidy, subpopulation structure, sector anatomy, histopathology, and the genetic content of loci specified by the user. In this respect, SCGV is distinct from the published visualization software, which is more narrowly focused on genomics (21). Importantly, SCGV enables seamless transitions among multiple complementary views of genomic, histopathologic, and anatomic data, including clonal structure, as it results from the analysis described here. A portable version of the viewer, with additional advanced features, is available at https://github.com/KrasnitzLab/SCGV and will be described in detail in future publication, along with a portable version of the SNS computational workflow, currently under development.

We do not currently use our present protocols to observe single-nucleotide variation in single cells, but we are designing and testing newer single-nuclei sequencing methods that will enable us to do so. Such information may facilitate diagnostic risk assessment. There is no technical obstacle to pooling libraries made from single cells of the same clonal expansion to obtain deeper sequence and more genomic information.

With that approach in mind, and knowing that larger clinical studies demand affordability, we have explored modifications to the experimental procedure described here, with a view of reducing cost per cell from its present approximate value of $40. As the cost per cell is partly driven by sequencing, we examined the efficacy of using much lower coverage to identify tumor clones. Our preliminary results, as shown in Supplementary Table S4, suggest that 8- to 16-fold reduction of coverage per cell would not significantly affect our ability to identify clones of cancer cells in PB samples. Still further work in progress (27, 28) suggests that we can do this with even lower coverage, potentially as low as 50,000 reads per cell, making possible sparse analysis of nearly 5,000 cells on a single lane of an Illumina HiSeq 2000, with a more focused (and less expensive) follow-up on a subset of the interesting cells. Partly, the costs reflect reagent costs and labor, which can be reduced with microfluidics and automation. We expect that by combining these modalities, costs can be reduced to less than a dollar per nucleus.

The present method should be considered in relationship to other methods for sequence analysis. In a small study, we performed sparse sequencing from many nuclei culled from cores, and there was, not surprisingly, less information (Supplementary Table S3). Clearly, whole-genome deep sequencing of bulk DNA holds the promise of identifying critical alterations driving cancer. But this method is far too expensive for determining if a cancer lesion is present or not in multiple cores. Moreover, the presence of normal cells, which often greatly outnumber the cancer cells, dilutes the signal of copy-number changes and makes very deep coverage needed to detect point mutations. Worse, we have repeatedly observed in this study clonal and subclonal specificity of genomic lesions harboring genes implicated in prostate cancer. Such subpopulation specificity likely is not limited to copy-number gains and losses and extends to other types of genomic variation, including point mutations. Clonal specificity of genomic lesions can be examined by resorting to single-cell analysis using sparse sequencing. The latter is sufficient to detect copy-number changes and that is sufficient to deduce clonal population structure. Once a clonal identity is determined, the libraries from cells with that identity can be cherry-picked and pooled, and a complete sequence obtained for identifying critical point mutations. Thus, inexpensive sparse single-cell sequencing may be a gateway to more comprehensive deep-sequencing methods.

Disclosure of Potential Conflicts of Interest

S.S. Yadav reports receiving commercial research grant from Prostate Cancer Foundation. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: J. Alexander, B. Robinson, H. Lepor, J. Hicks, M. Wigler
Development of methodology: J. Alexander, D.G. Nowak, J. Hicks, M. Wigler, A. Krasnitz
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J. Alexander, L. Rodgers, H. Cox, D.G. Nowak, J. Laze, E. Llukani, A. Srivastava, S. Gruschow, S.S. Yadav, B. Robinson, H. Lepor, M. Wigler
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): J. Alexander, J. Kendall, R. Aboukhalil, D. Levy, G. Sun, G. Atwal, L.C. Trotman, H. Lepor, M. Wigler, A. Krasnitz
Writing, review, and/or revision of the manuscript: J. Alexander, J. Kendall, R. Aboukhalil, D.G. Nowak, S.S. Yadav, B. Robinson, L.C. Trotman, H. Lepor, M. Wigler, A. Krasnitz
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): J. Kendall, J. McIndoo, A. Stepansky, M. Riggs, I. Hakker, J. Laze, E. Llukani, A. Srivastava, S. Gruschow, S.S. Yadav, H. Lepor, M. Wigler, A. Krasnitz
Study supervision: S. Gruschow, J. Hicks, M. Wigler, A. Krasnitz
Other (development of software tools for visualization of computational analysis results): L. Chobardjiev
Other (performed SNS methods): I. Hakker

Acknowledgments

We thank Ashutosh Tewari for clinical insight and contributions to study design. We thank Eric Antoniou, Elena Ghiban, and the CSHL DNA Sequencing Core for next-generation sequencing and Pamela Moody and Sean D'Italia of the CSHL Flow Cytometry Shared Resource, which are supported by Cancer Center Support Grant 5P30CA045508. We thank Anthony Leotta and Peter Andrews for informatics support. We thank Marlene Sosa for clinical assistance at SCPCC at NYU Langone Medical Center. The sequencing data are deposited at NCBI with accession number SRA (pending). This work was supported by grants from the Simons Foundation (SFARI 235988 to M. Wigler), the National Cancer Institute (5U01CA188590 to A. Krasnitz and M. Wigler), the Department of the Army (DOD W81XWH-12-1-0455 to J. Hicks), Global Prostate Cancer Research Foundation (to M. Wigler), and Long Island Cruizin' for a Cure (to J. Hicks and L.C. Trotman).

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 May 1, 2017; revised August 23, 2017; accepted November 10, 2017; published OnlineFirst November 27, 2017.

References

1. Bolenz C, Gierth M, Grobholz R, Köpke T, Semjonow A, Weiss C, et al. Clinical staging error in prostate cancer: localization and relevance of undetected tumour areas. BJU Int 2009;103:1184–9.
2. Goodman M, Ward KC, Osunkoya AO, Datta MW, Luthringer D, Young AN, et al. Frequency and determinants of disagreement and error in gleason scores: a population-based study of prostate cancer. Prostate 2012;72: 1389–98.
3. King CR, Long JP. Prostate biopsy grading errors: a sampling problem? Int J Cancer 2000;90:326–30.
4. Gann PH, Fought A, Deaton R, Catalona WJ, Vonesh E. Risk factors for prostate cancer detection after a negative biopsy: a novel multivariable longitudinal approach. J Clin Oncol 2010;28:1714–20.
5. Siegel R, Ward E, Brawley O, Jemal A. Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. CA Cancer J Clin 2011;61:212–36.
6. Carter HB, Partin AW, Walsh PC, Trock BJ, Veltri RW, Nelson WG, et al. Gleason score 6 adenocarcinoma: should it be labeled as cancer? J Clin Oncol 2012;30:4294–6.
7. Freedland SJ, Kane CJ, Amling CL, Aronson WJ, Terris MK, Presti JC Jr, et al. Upgrading and downgrading of prostate needle biopsy specimens: risk factors and clinical implications. Urology 2007;69:495–9.
8. Sved PD, Gomez P, Manoharan M, Kim SS, Soloway MS. Limitations of biopsy Gleason grade: implications for counseling patients with biopsy Gleason score 6 prostate cancer. J Urol 2004;172:98–102.
9. Mufarrij P, Sankin A, Godoy G, Lepor H. Pathologic outcomes of candidates for active surveillance undergoing radical prostatectomy. Urology 2010;76:689–92.
10. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, et al. Punctuated evolution of prostate cancer genomes. Cell 2013;153:666–77.
11. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet 2012;44:685–9.
12. Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, et al. The genomic complexity of primary human prostate cancer. Nature 2011;470:214–20.
13. Boutros PC, Fraser M, Harding NJ, de Borja R, Trudel D, Lalonde E, et al. Spatial genomic heterogeneity within localized, multifocal prostate cancer. Nat Genet 2015;47:736–45.
14. Cooper CS, Eeles R, Wedge DC, Van Loo P, Gundem G, Alexandrov LB, et al. Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. Nat Genet 2015;47:367–72.
15. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative genomic profiling of human prostate cancer. Cancer cell 2010;18:11–22.
16. Tomlins SA, Mehra R, Rhodes DR, Cao X, Wang L, Dhanasekaran SM, et al. Integrative molecular concept modeling of prostate cancer progression. Nat Genet 2007;39:41–51.
17. Dev H, Rickman D, Sooriakumaran P, Srivastava A, Grover S, Leung R, et al. Biobanking after robotic-assisted radical prostatectomy: a quality assessment of providing prostate tissue for RNA studies. J Transl Med 2011;9:121.
18. Baslan T, Kendall J, Rodgers L, Cox H, Riggs M, Stepansky A, et al. Genome-wide copy number analysis of single cells. Nat Protoc 2012;7:1024–41.
19. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. Nature 2011;472:90–4.
20. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009;10:R25.
21. Garvin T, Aboukhalil R, Kendall J, Baslan T, Atwal GS, Hicks J, et al. Interactive analysis and quality assessment of single-cell copy-number variations. Nat Methods 2015;12:1058–60.
22. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 2004;5:557–72.
23. Kendall J, Krasnitz A. Computational methods for DNA copy-number analysis of tumors. Methods Mol Biol 2014;1176:243–59.
24. Krasnitz A, Sun G, Andrews P, Wigler M. Target inference from collections of genomic intervals. Proc Natl Acad Sci U S A 2013;110:E2271–8.
25. Burkhardt L, Fuchs S, Krohn A, Masser S, Mader M, Kluth M, et al. CHD1 is a 5q21 tumor suppressor required for ERG rearrangement in prostate cancer. Cancer Res 2013;73:2795–805.
26. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, et al. Integrative clinical genomics of advanced prostate cancer. Cell 2015; 162:454.
27. Aboukhalil R. "Elucidating Cancer Evolution Using Single-Cell Sequencing and Comparative Genomics." PhD thesis, Cold Spring Harbor Laboratory, 2016.
28. Wang Z, Andrews P, Kendall J, Ma B, Hakker I, Rodgers L, et al. SMASH, a fragmentation and sequencing method for genomic copy number analysis. Genome Res 2016;26:844–51.