Identification of Somatic Structural Variants in Solid Tumors By Optical Genome Mapping

David Y. Goldrich^{1,4}, Brandon LaBarge^{1,4}, Scott Chartrand², Lijun Zhang², Henry B. Sadowski³, Yang Zhang³, Khoa Pham³, Hannah Way³, Chi-Yu Jill Lai³, Andy Wing Chun Pang³, Benjamin Clifford³, Alex R. Hastie³, Mark Oldakowski³, David Goldenberg¹, James R. Broach²

¹ Department of Otolaryngology – Head and Neck Surgery, Pennsylvania State University College of Medicine, Hershey, PA

² Department of Biochemistry and Molecular Biology, Pennsylvania State University College of Medicine, Hershey, PA

³ Bionano Genomics, San Diego, CA

⁴ These authors contributed equally to this work.

Corresponding author:

James R. Broach

Department of Biochemistry and Molecular Biology

Pennsylvania State University College of Medicine

Hershey, Pennsylvania

Email: jbroach@pennstatehealth.psu.edu

Abstract: Genomic structural variants comprise a significant fraction of somatic mutations driving cancer onset and progression. However, such variants are not readily revealed by standard next generation sequencing. Optical genome mapping (OGM) surpasses short read sequencing in detecting large (>500bp) and complex structural variants (SVs) but requires isolation of ultra-high molecular weight DNA from the tissue of interest. We have successfully applied a protocol involving a paramagnetic nanobind disc to a wide range of solid tumors. Using as little as 6.5mg of input tumor tissue, we show successful extraction of high molecular weight genomic DNA that provides a high genomic map rate and effective coverage by optical mapping. We demonstrate the system's utility at identifying somatic SVs affecting functional and cancer-related genes for each sample. Duplicate/triplicate analysis of select samples shows intra-sample reliability but also intra-sample heterogeneity. We also demonstrate that simply filtering SVs based on a GRCh38 human control database provides high positive and negative predictive values for true somatic variants. Our results indicate that the solid tissue DNA extraction protocol, OGM and SV analysis can be applied to a wide variety of solid tumors to capture SVs across the entire genome with functional importance in cancer prognosis and treatment.

Keywords: optical genome mapping, solid tumors, cancer genomics

Introduction

One of the hallmarks of cancer is genomic instability, which often affects genes controlling cell division and genome integrity. The resulting alterations include single nucleotide variant (SNV) point mutations as well as structural variants (SVs), in which larger DNA segments undergo chromosomal perturbations such as deletions, insertions, duplications, inversions, and translocations. For instance, recurrent translocations, such as the Philadelphia chromosome, can activate oncogenes but at the same time reveal avenues for implementing or developing effective targeted drug therapies [1-4]. Likewise, SV identification plays an increasingly important role in cancer diagnosis and prognosis [5,6], and SVs have been shown to play a crucial role in intra-tumoral genetic heterogeneity [7]. Therefore, SV identification and analysis is important to understanding oncogenesis and tumor behavior.

Short read sequencing can readily detect many SNVs, but is less successful in detecting SVs, by either alignment-based or assembly-based methods [8]. Since alignment-based approaches rely on mapping reads to unique positions, repetitive and low-complexity genomic regions can lead to misalignment and false positive SV calls. Additionally, homologous alleles may be incorrectly combined, leading to haploid assembly only representing a single allele or chimeric assemblies mixing alleles. Whole genome and cytogenetic approaches such as whole genome sequencing (WGS), karyotyping, fluorescent in-situ hybridization (FISH) and CNV microarrays also contain significant limitations. Karyotyping provides a comprehensive view of the entire genome but carries limited resolution of ~5Mb and in most cases requires culturing cells before preparing chromosomes. FISH has higher resolution but requires prior knowledge as to which loci to test and has limited throughput. CNV-microarrays offers resolution down to multiple Kb but are insensitive to balanced chromosomal aberrations such as translocations and inversions, are unable to detect low-frequency allelic changes, and cannot distinguish tandem duplications from insertions in trans. Finally, WGS has difficulty with de-novo genome assembly and resolving duplications and repeated sequences [8-10]. Therefore, alternative methods are required to preserve long-range genomic structural information.

Optical genome mapping (OGM) has emerged as a viable option for analyzing large genomes for SVs. OGM preserves long-range information by imaging entire intact molecules of DNA in their native state and, as a result, has contributed to constructing reference genome assemblies, including those for maize, mouse, goat, and humans [11-28]. OGM can detect large (>500bp) and complex SVs that are difficult to detect using traditional short read sequencing alone. OGM preparation and analysis workflow has been successfully applied to liquid-phase tumor and cell culture SV analyses. For instance, Neveling et al. successfully analyzed SV and CNV calls and identified novel potential drivers of leukemia using bone marrow and blood samples [29]. Xu et al utilized OGM to detect previously unrecognized SVs in leukemia patients implicated in oncogenesis and patients' survival [5]. Chan et al used OGM to visualize complex gene fusion maps in previously studied liposarcoma cell line [30]. Pang et al demonstrated the ability of OGM to identify novel somatic SVs in melanoma and other well-studied cancer cell lines [31]. Dixon et al integrated OGM to high-throughput chromosome conformation capture (Hi-C) and WGS to systematically detect SVs in leukemia and other cell lines [32].

Despite its success in visualizing SVs in liquid tumors, OGM has not yet seen widespread application in solid tissue tumors, due primarily to the difficulty of obtaining high quality, high molecular weight DNA from solid tumor samples. Nonetheless, previous work has shown the feasibility of high-quality high molecular weight DNA isolation and analysis using earlier workflow iterations [33], and recent feasibility studies have shown the importance of OGM application to solid tumor analysis [7,34,35]. Peng et al demonstrated large SVs not detected by WGS implicated in metastatic lung squamous cell carcinoma [7], and Jaratlerdiri et al and Crumbaker et al similarly found SVs impacting oncogenic and tumor suppressing genes not identified by NGS or WGS alone in prostate cancer [34,35]. However, these previous methods for extracting gDNA from solid tissue were either prohibitively expensive or yielded low quantities of DNA [36]. We demonstrate here the successful implementation of a workflow to generate ultrahigh molecular weight gDNA and subsequent SV analysis for 20 solid tumor samples comprising a wide variety of solid tissue organ systems.

Materials and Methods

Tumor Samples

Solid tissue was collected following surgical resection for 10 tumors: four squamous cell carcinomas of the tongue, three anaplastic carcinomas of the thyroid, one liver hepatocellular carcinoma, one lung pleomorphic carcinoma, and one bladder tumor. Patients were consented under protocols approved by the Penn State Health Institution Review Board. Tissue was flash frozen and stored at -80°C in the Penn State Institute for Personalized Medicine (IPM). Ten additional fresh frozen solid tumor samples were acquired from BioIVT for the following tumor types: lung adenosquamous carcinoma, liver hepatocellular carcinoma, bladder papillary urothelial carcinoma, kidney renal cell carcinoma, breast ductal carcinoma in-situ, prostate invasive adenocarcinoma, brain anaplastic astrocytoma, ovarian serous carcinoma, colon adenocarcinoma, and papillary thyroid carcinoma.

Bionano Optical Genome Mapping

Ultra-High Molecular Weight gDNA Isolation from Solid Tissue (Figure 1). A 10mg section of frozen tissue was sliced from its parent piece on a sterilized aluminum block over dry ice. The tissue was minced briefly into 1-2mm pieces with a scalpel, and placed into a 15ml conical tube containing 2ml of chilled Bionano Prep SP Animal Tissue Homogenization Buffer (HB) on ice. The tube was then placed in a beaker of ice for effective cooling of the sample during subsequent blending procedure. The tissue was blended at full speed for 10 seconds using the Tissuerruptor II (Qiagen). Following tissue disruption, 6ml of additional chilled HB was added to the sample, and the resultant 8ml was filtered through a 40µm nylon cell filter within a 50ml conical tube.

Homogenate was transferred to a new 15ml tube, inverted, centrifuged at 2000g for 5 minutes at 4°C using a swinging bucket rotor, and decanted. The cell pellet was then resuspended in 300µl of Wash Buffer A (Bionano) and transferred to a 1.5ml Protein LoBind tube. 700µl additional Wash Buffer A was added to the 15ml tube, pipette mixed, and transferred to the same 1.5ml tube, followed by centrifuge at 2000g for 5 minutes at 4°C. The supernatant was then decanted

Figure 1. Workflow for Isolation of High Molecular Weight DNA from Solid Tumors.
See text for detailed description of the workflow.

and the pellet was resuspended in residual volume. 50µl Proteinase K (Bionano) was added to the sample and incubated at room temperature for 3 minutes. 225µl Buffer Lysis and Binding Buffer (LBB, Bionano) was then added, and samples were inverted 15 times to mix. After HulaMixer rotation at 10rpm for 15 minutes at room temperature, 10µl of 100mM Phenylmethylsulfonyl Fluoride Solution (PMSF, Sigma-Aldrich) was added to inactivate Proteinase K, followed by 85µl of Salting Buffer (SB, Bionano).

A single 4mm paramagnetic Nanobind Disc (Bionano) was transferred to the lysate, in addition to 400µl 100% isopropanol. Binding of gDNA to the disc was facilitated by HulaMixer rotation for 15 minutes at room temperature at 10 RPM. Samples were then placed into a clear Dynamag tube rack for quick inversions to further increase gDNA binding. The rack was then inverted and joined to the Dynamag magnetic base. The apparatus was slowly brought upright to allow the disc to adhere to the magnet. The supernatant was removed and 700µl Wash Buffer 1 (Bionano) was dispensed onto the disc, followed by four inversions of the clear rack for washing. This washing step was performed twice and followed by two washes of 500µl Wash Buffer 2 (Bionano). After final decanting, the disc was transferred to a new 1.5 ml Protein LoBind tube using the Magnetic Retriever (Bionano). Elution Buffer (65µl, Bionano) was added and the sample was incubated for 20 minutes at room temperature. gDNA was then transferred to a 2.0ml microfuge tube. The previous tube containing the disc was spun on a benchtop microcentrifuge for 5 seconds and the residual eluate transferred to the 2.0 ml tube. The gDNA was slowly pipetted to homogenize and then rotated via HulaMixer at room temperature for 1 hour at 15 RPM. The sample was then pulse spun and allowed to equilibrate overnight at room temperature.

Ultra-High Molecular Weight gDNA Isolation from Blood. A 650µl aliquot of frozen blood was thawed to room temperature. The sample was inverted 10 times to mix and then pulse spun. 20µl was dispensed onto a Hemocue cuvette and cell density was measured. The amount of blood necessary to obtain 1.5×10^6 cells was calculated (typically 200-600µl). Sample was again inverted 10 times and briefly pulse spun before the calculated volume was transferred to a 1.5ml Protein LoBind tube.

The resultant blood aliquot was spun for 2 minutes at 2,200 x g at room temperature to obtain a WBC pellet. After removing the supernatant, 40µl Stabilizing Buffer and 50µl Proteinase K (Bionano) were added to the WBC pellet. After resuspension, the sample was incubated at room temperature for 3 minutes. 225µl Buffer LBB (Bionano) was added to the sample and mixed by inversion fifteen times. The sample was pulse spun and 10µl of 100mM PMSF (Sigma-Aldrich) was added to the sample, followed by capping and inversion five times and a brief pulse spin. The sample was incubated at room temperature for 10 minutes. Isolation of DNA was then performed essentially as described above for tumor tissue.

Direct Label and Staining (DLS). 750ng of gDNA, determined by Broad Range dsDNA Qubit Assay, was mixed with Direct Label Enzyme-1 (DLE-1) Buffer (Bionano Genomics, Part#20350), DL-Green (Bionano Genomics, Part#20352), and DLE-1 Enzyme (Bionano Genomics, Part#20351) and incubated for 2 hours at 37°C in a thermocycler. 5µl of Proteinase K solution was then added to the reaction and incubated for 30 minutes at 50°C. During Proteinase K digestion, a DLS membrane (Bionano Genomics, Part#20358) was placed hydrated with 25µL

of DLE-1 buffer in one well of a DLS-microplate (Bionano Genomics, Part#20357). The labeled DNA was transferred onto this membrane, incubated at room temperature for one hour, transferred onto another membrane with 25 μ L DLE-1 buffer and incubated for 30 minutes at room temperature. Sample exposure to light was avoided during these steps. Following this drop dialysis cleanup, 5x DTT (Bionano Genomics, Part#20354), Flow Buffer (Bionano Genomics, Part#20353), DNA stain (Bionano Genomics, Part#20356), and nuclease free water were added to the DNA in a 2ml amber tube and the tube was mixed in a hula mixer at 5rpm for 2 hours at room temperature. The sample was then spun down and stored overnight at room temperature protected from light. The DNA concentration of each labeled sample was confirmed within 4-12 ng/ μ l by High Sensitivity dsDNA Qubit Assay and then loaded onto a Bionano Saphyr® Chip (Bionano Genomics, Part#20319) and run on the Bionano Saphyr® instrument, targeting approximately 300x human genome coverage.

Bionano Access and Solve Pipeline

Genome analysis was performed using Rare Variant Analysis in Bionano Access 1.6 and Bionano Solve 3.6, which captures somatic SVs occurring at low allelic fractions. Briefly, molecules of a given sample dataset were first aligned against the public Genome Reference Consortium GRCh38 human assembly. SVs were identified based on discrepant alignment between sample molecules and GRCh38, with no assumptions about ploidy. Consensus genome maps (*.cmaps) were then assembled from clustered sets of at least three molecules that identify the same variant. Finally, the genome maps were realigned to GRCh38, with SV data confirmed by consensus forming final SV calls. SVs were then annotated with known canonical gene set present in GRCh38, as well as estimated population frequency for each structural variant detected by comparing to a custom control database (n=297) from Bionano Genomics.

Data Comparison

Whole genome imaging data were compared to the human reference genome GRCh38 (hg38) to retain only those SVs not present in the reference genome. SVs were further filtered to eliminate any variant observed in any of the Bionano control samples or, if available, patient-matched blood. Bionano Access-created csv files containing filtered SVs were analyzed to compare SV content across samples. For tissue samples with associated blood samples, control database filtration efficacy was compared to blood-filtering efficacy at identification of somatic mutations. For duplicate/triplicate samples, filtered SVs were compared to determine intra-sample reliability. For identification of cancer related genes, the set of genes affected by SVs in each of the samples was compared to the list of genes causally implicated in cancer available in the Cosmic Cancer Gene Census database (v92) [37] (<https://cancer.sanger.ac.uk/census>).

Results

Patient Clinical Characteristics. Clinical data for the patients from whom tumor samples were acquired are shown in Table 1. 60% (12/20) patients were male, with mean age of 73.5 years at sample acquisition. 45% (9/20) patients identified as Caucasian, 40% (8/20) as Asian, 5% (1/20) as Hispanic and 10% (2/20) not identifying. The majority of IPM-sourced tumor samples were obtained from Caucasian patients (7/10) while the majority of the BioIVT-sourced tumor samples were obtained from patients of Asian ethnicity (8/10). In terms of overall risk factors, 55% (11/20) of patients were self-described current or former tobacco users and 45% (9/20) endorsed some history of alcohol use.

The tumor samples consisted of a variety of stages (Table 1). 75% (3/4) of tongue cancer samples and 100% (3/3) anaplastic thyroid cancers were stage IV cancers, while 100% (2/2) lung and (2/2) bladder cancers were stage II. Limited tumor data was available for the commercially available BioIVT-sourced tumor samples.

DNA quality metrics: All 20 solid tumors yielded high molecular weight gDNA (Table 2). The average concentration across all samples following gDNA isolation was 120 ng/μl by Broad Range dsDNA Qubit Assay. All eluted gDNA were well above the minimal concentration required for DLS labeling (35ng/μl) and the average final DNA yields for each tumor ranged from 1.2-16.4μg/10mg input tissue. Analysis on a Saphyr instrument following DLS labeling revealed that samples achieved an average label density of 14.4/100Kbp, average filtered N50 (>20kbp) DNA size of 242Kbp, average filtered N50 (>150kbp) DNA size of 315Kbp, map rate of 82.62%, effective reference coverage of 320x and average effective DNA throughput (≥150Kbp) of 50Gbp/scan. Rare Variant Pipeline Analysis of the samples yielded an average of 82.4% of molecules aligning to the reference genome. These values are all well above the acceptable range for obtaining high quality data and none of the samples failed any of these quality control metrics.

Identification of somatic structural variants. Rare Variant analysis of the samples revealed a large numbers of variants in each sample, only a fraction of which were likely somatic (Figure 2). The unfiltered analysis yielded an average of 1633 total SVs per sample (range 1241-2000), which include both somatic and germline polymorphic variants. These consisted predominantly of insertions and deletions, with an average of 712 insertions and 604 deletions, a fewer number of inversion (an average of 153) and duplications (an average of 123), and relatively few translocations (an average of 41). Eliminating those SVs found in Bionano's control database of known polymorphic SVs reduced the number of putative somatic structural mutations by 91% to an average of 124 total SVs per sample (Figure 2). Most of the variants eliminated were insertions and deletions, of which on average 97% and 94%, respectively, were removed. On the other hand, less than 0.2% of the translocations were flagged as polymorphic, consistent with the fact that almost no translocations persist in the population as polymorphism.

To determine the efficacy of identifying somatic SVs by filtering against Bionano's database of known polymorphisms, we used as a gold standard the blood samples from four patients from whom we had obtained tongue tumors. That is, we determined the true somatic mutations in each of these four tumors by eliminating those SVs identified in each of the tumors that were also

Table 1. Patient Demographics and Tumor Characteristics

Study ID	Cancer Type*	Age†	M/F	Race	Smoking History	Alcohol History	Pathologic TNM	Cancer Stage
7528	Tongue (SCC)	25	M	White	None	Rare	T3N2bM0	IVa
7052	Tongue (SCC)	35	M	White	None	None	T2N3M0	IVb
7622	Tongue (SCC)	60	F	White	50 pack years	1-2 drinks/ week	T3N0M0	III
7403	Tongue (SCC)	65	M	White	45 pack years	Rare	T2N3bM0	IVb
7518	Thyroid (AP)	70	F	White	20 pack years	2 drinks per day	T4bN1bM1	IVc
7708	Thyroid (AP)	65	M	White	None	None	4aN1bM1	IVc
3717	Thyroid (AP)	80	M	Hispanic	25 pack years	Rare	T4aN1aM1	IV
14369	Lung (pleomorphic carcinoma)	60	M	N/A	60 pack years	None	T2bN1M0	Ila
10974	Liver (metastatic adenocarcinoma of colon)	65	F	N/A	Former	None	T3N2aM1	IVB
3096	Bladder (urothelial carcinoma)	55	M	White	60 pack years	None	T2N0M0	II
73432	Lung (adenosquamous carcinoma)	35	M	Asian	Former (5 pack years)	Former (1 per day, 10 years)	T2aN1M0	IIA
94894	Liver (hepatocellular carcinoma)	70	M	Asian	5 15 pack years	1 per day, 35 years	T1NxM0	I
101558	Bladder (papillary urothelial carcinoma)	65	M	Asian	Former (5 pack years)	1 per day, 20 years	T2NxM0	II
69033	Kidney (renal cell carcinoma)	60	F	Asian	None	None	T2bNxM0	II
79379	Breast (ductal carcinoma in situ)	50	F	Asian	None	None	T3N0M0	IIB
102095	Prostate (invasive)	60	M	White	40 pack years	None	T3bN1M0	IV

	adeno-carcinoma)							
80384	Brain (anaplastic astrocytoma)	40	F	White	None	None	NA	NA
81347	Ovarian (serous carcinoma)	75	F	Asian	None	None	T1aN0M0	IA
119664	Colon Cancer (adenocarcinoma)	80	M	Asian	2 pack years	1 per day, 40 years	TXNXMX	UNK
128019	Thyroid (papillary)	35	F	Asian	None	None	T3bNxM0	I

* SCC: squamous cell carcinoma; AP: anaplastic

† ~Age ($\geq$ Age-3 and $\leq$ Age+3)

Table 2. Single Molecule Quality Report Metrics

Tissue	#	Input (mg)	DNA (ng/μl)	DNA Yield (μg/mg)	N50 Kbp (>150Kbp)	Labels/100kbp	Map Rate (%)	Gbp /scan	Effective Coverage
7528 (tongue)	1	17.5	37	0.12	317	12.3	58.8	53	237x
7052 (tongue)	1	17.1	81	0.28	287	15.2	82.4	37	345x
7622 (tongue)	1	18.7	160	0.51	361	13.4	75.8	64	317x
7403 (tongue)	1	18	79	0.26	272	14.8	72.8	33	304x
7518 (thyroid)	1	8.6	28	0.20	265	14.4	76.6	26	312x
7708 (thyroid)	1	10.6	85	0.47	356	13.1	61.2	35	253x
3717 (thyroid)	1	13.2	49	0.22	320	14.5	88.2	58	371x
14369 (lung)	1	11.4	87	0.45	323	14.0	89.6	36	372x
10974 (liver)	1	6.5	82	0.74	289	15.2	87.9	49	360x
3096 (bladder)	1	9.4	59	0.37	319	13.8	78.3	39	325x
73432 (lung)	3	9.6	128	0.86	304	15.0	90.4	51	339x
94894 (liver)	2	9.0	196	1.41	306	14.9	89.3	84	325x
101558 (bladder)	3	9.7	245	1.64	357	15.2	91.8	66	338x
69033 (kidney)	3	10	96	0.63	269	14.6	83.5	41	296x
79379 (breast)	3	13.3	183	1.04	395	14.2	84.1	77	288x
102095 (prostate)	3	10.3	113	0.72	361	14.8	85.1	62	295x
80384 (brain)	2	10.5	168	1.06	292	14.6	90.2	42	306x
81347 (ovary)	2	10.5	168	1.05	292	14.6	90.2	42	330x
119664 (colon)	2	11.3	231	1.33	330	14.9	88.6	42	274x
128019 (thyroid)	4	10	126	0.77	294	14.5	87.6	64	294x
Average	1.9	11.8	120.	0.71	315.	14.4	82.6	50.0	314x

Average values are presented for samples with multiple replicates

Figure 2. Total and Somatic Structural Variants Present in Tumor Samples.

Upper panel: SV counts as determined using the Bionano Rare Variant pipeline, before control database filtration. SV counts are averages for duplicate and triplicate samples. Lower panel: SV counts after filtering total SVs to remove known polymorphic SV found in Bionano's GRCh38 control database. SV counts are averages for duplicate and triplicate samples.

present in the corresponding blood sample. We could then compare those true somatic variants to the list of somatic variants predicted by filtering against the database of polymorphisms. For these four tongue tumor samples, we identified an average of 1474 total SVs per sample. Filtering these SVs using the Rare Variant Analysis pipeline for SVs not found in the Bionano control database yielded an average of 72 total SVs per sample, consisting of 11 insertions (range 9-15), 31 deletions (range 11-47), 3 inversions (range 1-6), 14 duplications (range 2-23), and 14 translocations (6-19). Filtering against the variants found in the corresponding blood samples returned an average of 58 total SVs per sample, consisting of 10 insertions (range 9-10), 20 deletions (range 7-35), 2 inversions (range 0-4), 13 duplications (range 4-24), and 14 translocations (range 6-19) (Figure 3, upper panel).

Comparing the residual SV sets obtained by filtering against Bionano's control database to the sets of true somatic SVs for each sample demonstrated that the control database filtration exhibited strong statistical accuracy (Figure 3, lower panel). Across the four separate samples, the control database exhibited an average sensitivity of 92% (83-96%) and specificity of 98% (range 97-99%). That is, filtering with the control database retained most of the true somatic mutations while eliminating almost all of the polymorphic SVs. Similarly, the average negative predictive value of the filter was 99.6%, demonstrating that an SV identified as germline was indeed a germline variant, while the positive predictive value of 74% (range 60-81%) indicates that a majority, but not all, the variants identified as somatic are in fact somatic. The SVs inaccurately identified as somatic were rare germline variants, predominantly insertions or deletions, essentially private to the patient's genome. As above, we noted that the filtering process did not affect all SV types equally: while most deletions and insertions were flagged as polymorphic and eliminated from the list of somatic mutations, very few duplications and essentially no translocations were identified as polymorphic. This is consistent with observation that few translocations or duplications are stable through meiosis.

The number and types of somatic variants in a tumor varied substantially across the collection of samples (Figure 4). Several tumor samples, including those from colon, bladder, kidney and all four from thyroid, contained relatively few somatic SVs whereas others, including those from prostate, ovaries, lung and brain, carried a large number of somatic SVs. Since these samples for the most part serve as single representatives of each tumor type, we cannot extrapolate to the tumor types as a whole the contribution of SVs to cancer onset and development for each class of tumor. However, it is noteworthy that the SNV mutational burden in thyroid cancers is among the lowest among all tumor types and that measure of genome instability is mirrored in the low number of somatic SVs in all four of the samples examined [38]. Similarly, the SNV mutational burden in lung cancers is among the highest across all tumor types and both of the lung tumors examined here also carry a high level of somatic SV.

Duplicate Sample Analysis. We compared SV calls from separate isolates of the same sample to assess consistency and reproducibility of the method, albeit without knowing the extent of tumor heterogeneity of the individual samples. Six samples underwent triplicate analysis, and four samples underwent duplicate analysis (Table 2). After identifying SVs using the Rare Variant Analysis pipeline and filtering them against the Bionano control database of known polymorphisms, we recovered an average of 116 somatic SVs shared among the separate isolates of the same tumor. These comprised on average of 23 insertions, 29 deletions, 10 inversions, 11

Figure 3. Efficacy of the Somatic Variant Identification Using a Control Database of Known Polymorphisms.

Upper Panel: Number and distribution of somatic structural variant in four tongue tumors as determined by filtering against SVs in the patient's genome from peripheral blood (left) or against Bionano's control database of known polymorphisms.

Lower Panel: Values for sensitivity, specificity and positive and negative predictive values for identification of somatic structural variants obtained by filtering total identified SVs to remove those present in a control database of know human polymorphisms. Data obtained by this filtering method were compared to those obtained by filtering the total SVs to remove those present in the genomes obtained from peripheral blood from the each of the patients from whom the tumors were removed.

Figure 4. Global View of Structural Variants in Solid Tumor Samples

Diagrams of somatic structural variants in all the solid tumor genomes, filtered to remove known polymorphisms, showing translocations and inversions in the center, copy number on the inner ring and insertions (green), deletions (orange), inversions (light blue) and duplications (violet) on the next to most outer ring. Chromosomes are ordered sequentially in the outer ring on which are indicated cytological banding patterns and the centromere (red bar).

Table 3. Duplicate Sample Analysis

Shown are the number of somatic structural variants shared among the multiple isolates of the same sample and the percentage of those relative the total number of somatic variants found in all the isolates of the same sample.

	Total	%	Insertion	%	Deletion	%	Inversion	%	Duplication	%	Translocation	%
Brain*	134	70	5	63	21	78	9	69	13	72	86	69
Colon*	63	93	15	83	36	97	2	100	9	90	1	100
Liver*	45	70	14	74	21	81	1	17	4	67	5	71
Ovary‡	338	86	136	82	59	87	4	80	40	85	99	91
Bladder‡	30	88	11	79	18	100	1	100	0	100	0	0
Breast‡	221	92	9	82	33	85	23	88	14	88	142	95
Kidney‡	19	76	6	67	11	85	2	100	0	0	0	100
Lung‡	221	66	18	69	53	75	59	73	26	50	65	63
Prostate‡	69	48	8	47	22	61	3	38	1	25	35	44
Thyroid‡	19	86	7	88	10	91	0	100	2	100	0	0
Average (all)	116	78	23	73	28	84	10	76	11	68	43	63
Duplicate Average	145	80	43	75	34	86	4	66	17	78	48	83
Triplicate Average	97	76	10	72	25	83	15	83	7	60	40	50

* = duplicate sample % = % of SV calls shared among duplicate/triplicate samples
‡ = triplicate sample

duplications and 43 translocations (Table 3). As noted above, the number of SVs identified in a tumor varied widely across the different tumors examined, with lung, breast, brain and ovarian tumors showing a high level of somatic SVs while the others containing a relative low number of SVs. Moreover, the percentage of SVs shared among different isolates of the same tumor also varied among the different tumor types. However, the percentage of shared SVs and the total number of SVs were uncorrelated. Assuming that the higher values for shared SVs reflect the reproducibility of the method, then we might postulate that the lower shared values represent both the reproducibility and the tumor heterogeneity. This would suggest that these brain, liver, lung and prostate tumors had a relatively high level of tumor heterogeneity.

Identification of cancer gene mutations. While as noted above we cannot generalize regarding the role of structural variants in onset and progression of different tumor types, our results

indicate that we can extract from the structural variant list clinically relevant data on individual tumors that might inform prognosis or treatment options. We examined the somatic structural variants in each tumor sample for those that affected genes previously associated with cancer. In particular, we annotated those genes altered by a structural variant, either by disruption, duplication, deletion or fusion, and intersected that list with the set of cancer-related genes in the Cosmic database (v92) [37]. The resultant list by tumor type is provided in Table 4 and subdivided into oncogenes, tumor suppressor genes and gene fusions. We included only those oncogenes that were potentially activated by duplication or gene fusion and only those tumor suppressor genes that were potentially inactivated by deletion, insertion or fusion. As evident, every tumor sample carried at least one such cancer gene mutation and most contained multiple hits. Several of these genes offer the opportunity for targeted therapies, focused either directly on the oncogene, as would be the case for CDK6 and ERBB2, or at the pathway downstream of the affected gene, as would be the case for BRAF and CDKN2A. Other affected genes, such as MSH2, RAD51B, RAD21 and RAD18, suggest that the potential of therapy based on possible ensuing genome instability, such as immunotherapy or PARP inhibitors. Many of these variants would not be readily identified by targeted gene panels generally used for clinical assessment of tumor genomes. Thus, the identification of somatic structural variants by OGM could provide useful clinical insights not readily available through standard next generation sequencing.

In addition to identifying individual cancer related genes in tumor types, our results provide a panoramic view of the entire tumor genome and reveal large scale genomic features not readily available from standard sequencing techniques. As evident in the results in Figure 4, our data provide a rapid snapshot of the extent of genomic instability in each of the tumors. Such images present an integrated picture of the aneuploidies, translocations, inversions, deletions and insertions, which offers a readily digestible impression of the extent of genetic instability underlying a tumor. Moreover, several large-scale features are evident in these data. For instance, Figure 5 details a chromothripsis event on a portion of chromosome 5 in one of the lung tumor samples. The detection and mapping of such a feature are difficult to achieve by short read sequencing but can indicate poor prognosis and the corresponding need for aggressive therapy.

Discussion

In this report we described the application of optical genome mapping to solid tumors, which we suggest can significantly augment the genomic analysis of such tumors obtained by next generation sequencing. Genomic analysis of tumors has stimulated major advances in cancer diagnosis, prognosis and treatment, shifting the focus from morphological and histochemical characterization to consideration of the landscape of driver mutations in the tumor [39-41]. Somatic driver events in a tumor – point mutations and structural variants (SVs) including insertions, deletions, inversions, translocations and copy number changes – are currently identified in solid tumors by some combination of RNA sequencing and genome sequencing of either targeted gene panels, whole exomes or whole genomes. As noted in this report, OGM can provide a pervasive view of the structural variants in a tumor and the cancer related genes on which they impinge, thus identifying affected genes agnostically, without prior bias imposed by gene panels.

Table 4. Structural Variants Affecting Cancer Relevant Genes

Sample	Oncogene	Tumor Suppressor	Gene Fusion
Prostate	ERBB2 (Dup)	PTEN (Δ)	PTEN-LINC01374
	GATA2 (T)	NF1 (Δ)	DHX30-GATA2
	NUP98 (T)		CASC15-NUP98
			PRKAR1A-FRMPD4
			ERG-TMPRSS2
			FREM1-MYH9
Ovarian		NUMA1 (T)	NBEA-ZFHX3
		NF1 (I)	HMG2P46-BLOC1S6
		SMARCA4 (I)	LPP-PIEZO1
Kidney	PRKAR1A (T)	CDCN2A (Δ)	PRKAR1A-FRMPD4
	ERBB2 (Dup)	ZFHX3 (Δ)	
Colon		FHIT (Δ)	
Breast	ERBB4 (Dup)	USP8 (T)	USP8-PRPSAP
	ERBB2 (Dup)	PRKAR1A (T)	PRKAR1A-FRMPD4
		RAD51B (Δ)	LINC01476-BRIP1
		CDKN2A (Δ)	SYK-CFAP77
Brain	SETBP1	LARP4B (T)	CCDC158-LARP4B
		CSMD3 (T)	CA13-CSMD3
		LRP1B (Δ)	DPYD-SETBP1
		RAD21 (Δ)	CNBD1-AC083836.1
Bladder		DDX10 (Δ)	
Tongue	BRAF (T)	CDKN2A (Δ)	EPHB1-BRAF
	CDC6 (T)	PTPRD (Δ)	CDC6--AC091551.1
	CCND2 (T)	RAD51B (T)	PCLO-RAD51B
	CCND1 (Dup)	LRP1B (Δ)	
		CDKN1B (Δ)	
Thyroid		YWHAE (T)	ABR-YWHAE
		PTPRD (Δ)	CDK12-CSF3
			RAD18-SRGAP3
			SHROOM3-AFF1
Liver	VTI1A (T)	RMI2 (T)	VTI1A-NHLRC2

	MAP3K13 (T)	NCOR (T)	C3orf70-MAP3K13
	MACC1 (T)	CBLC (T)	AC005062.1-MACC1
	NSD3 (T)	MSH2 (T)	NSD3-AC087623.2
			RASGEF1B-VTI1A
			RMI2-TOX3
			NCOR1-LRRC75A
			MSH2-CYP3A43
Lung	CTNND2 (T)	PTPRD (Δ)	CTNND2-TRIO
	IKBKB (T)	RAD51B	DUSP10-CTNND2
		FUS	IKBKB-FAM91A1
		LRP1B	FUS-CNOT1
			PDE6D-RAD51B
			PRKCH-HIF1A
			GAS7-LYRM9
			EHBP1-LRP1B

T, translocation; Dup, duplication; I, insertion; Δ , deletion.

Figure 5. Chromothripsis of Chromosome 5p in a Lung Tumor

Shown is a truncated Circos plot of the lung tumor, focused on the region of chromosome 5, highlighting the chromothripsis event that occurred on its p arm. The organization of the circos plot is as indicated in the legend to Figure 4.

In liquid tumors such as leukemia, genomic analysis is augmented by karyotyping, which gives a panoramic, albeit low resolution, view of the entire genome. Despite the low resolution, the genome wide view of the structural changes afforded by karyotyping reveals diagnostic features of the tumor that have strong prognostic value. Given the consistent correlation of clinical outcomes with specific mutation classes, the World Health Organization (WHO), National Comprehensive Cancer Network (NCCN) and European Leukemia Net (ELN) agencies developed recommendations for diagnosis and management of acute myeloid leukemia in adults based on the spectrum of somatic point mutations and SVs generally revealed by karyotyping. SVs, particularly translocations and inversions, are major considerations in this diagnosis. Since karyotyping is a very challenging technique to apply to solid tumors, the clinician does not have access to a comparable global view of a solid tumor's genome and the role of SVs in prognosis has been possibly underappreciated. Applying OGM broadly to cancer types and correlating SVs revealed by that analysis could provide new genomics markers for prognosis and treatment selection.

Some prior studies have begun to demonstrate the utility of Bionano UHMW DNA isolation protocols in solid tissue tumor analysis. These include studies of lung squamous cell carcinoma and metastatic prostate carcinoma [7,34,35]. This current report demonstrates the utility of the a DNA isolation protocol and SV analysis in a wide variety of solid tissue types, and expands the feasibility of such analysis for previously unused human tissue types. The high DNA yield, high effective coverage, map rate and other molecular quality metrics shown across tumor types confirm how our extraction and analysis workflow can be effectively applied to many solid tissue tumors. This high-resolution genome imaging for many types of tumors will allow increased detection of cancer-associated SV events, potentially leading to new insights and improved personalized treatment options.

This current DNA isolation protocol carries a number of advantages. Tissue handling can be performed at room temperature. The current protocol showed successful DNA isolation in solid tissue samples of <20mg, and even as low as 6mg. The low tissue input requirement carries important applications for rare cancer samples, human tissue biopsy testing and other low-quantity specimen acquisition. Additionally, utilizing the novel paramagnetic Nanobind disks rather than prior agarose gel plugs greatly decreases time needed to complete DNA isolation to only 5 hours. The ability to isolate DNA from up to 8 simultaneous samples using the current protocol greatly amplifies throughput and reduces tissue-to-data processing time, increasing both laboratory convenience as well as expanding potential for clinical utility where rapid data turnaround is paramount. Furthermore, the strong inter-sample SV correspondence shown by most tissue types in duplicate/triplicate sample analysis demonstrates the reproducibility of this technique; intra-sample heterogeneity of select samples may be attributed to non-tumor normal tissue within some tissue fragments, or attributed to specific cancer subtype, and merits further investigation. Although the isolation protocol described here affords many advantages, there are some limitations to this protocol. While high quality DNA isolation and OGM SV analysis was obtained for a wide variety of tumor types that were tested, it may not be generalizable to every additional untested solid tumor type. Future directions include continuing to validate this protocol in additional tissue types, and assessing additional tumor samples to assess broader trends in the role of specific OGM-identified SVs in individual cancer subtypes.

Conclusion

We demonstrate the utility of a DNA isolation protocol for high molecular weight DNA extraction and OGM SV analysis of a wide variety of solid human tumor types on the Bionano Saphyr system, including breast, colon, liver, brain, bladder, kidney, lung, ovary, prostate and thyroid cancer tissue. The system can be used to accurately detect genetic mutation hallmarks in cancer tissue samples, including rearrangements such as translocations, gene fusions and copy number alterations. Somatic SVs can be determined by comparison filtering with the Bionano control sample database, or against a matched pair sample. Importantly, Bionano SV pipelines can detect SVs with complex breakpoint structures that are difficult to detect with other technologies. Our results indicate that the solid tissue DNA extraction protocol can be applied to a wide variety of solid tumors, and the Saphyr system can capture in a streamlined workflow a broad spectrum of SVs with functional importance and provide great utility in cancer study.

Author Contributions: Conceptualization: DG, ARH, JRB; Methodology, HBS, HW, KP MO, YZ; Software, LZ; Validation, CYJL; Formal analysis, AWCP, BC, CYJL, DYG, LZ; Investigation, DYG, BL, SC; Writing - Initial Draft: DG, ARH; Writing – Editing: DG, AWCP, ARH; Writing - Final Draft: JRB; Visualization, DYG, ARH, BC; Supervision, ARH, DG, JRB, HBS, MO; Project administration, ARH, JRB; Funding acquisition, DG. All authors have read and agreed to the published version of the manuscript.

Funding: Funding for this study was provided in part by a grant from the George Lavery Foundation (DG).

Institutional Review Board: Patients were consented under protocol 40532 approved by the Penn State Health Institution Review Board

Data Availability: Primary Bionano Saphyr data are available on request (jbroach@pennstatehealth.psu.edu)

Conflict of Interest: HBS, YZ, KP, HW, CYJL, AWCP, BC and ARH are employees of Bionano Genomics. The authors have no other conflicts of interest.

References

1. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR: **A census of human cancer genes.** *Nat Rev Cancer* 2004, **4**:177-183.
2. Kantarjian H, Sawyers C, Hochhaus A, Guilhot F, Schiffer C, Gambacorti-Passerini C, Niederwieser D, Resta D, Capdeville R, Zoellner U, et al.: **Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia.** *N Engl J Med* 2002, **346**:645-652.
3. Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, Ou SH, Dezube BJ, Janne PA, Costa DB, et al.: **Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer.** *N Engl J Med* 2010, **363**:1693-1703.
4. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, et al.: **Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer.** *Nature* 2007, **448**:561-566.
5. Xu J, Song F, Schleicher E, Pool C, Bann D, Hennessy M, Sheldon K, Batchelder E, Annageldiyev C, Sharma A, et al.: **An Integrated Framework for Genome Analysis Reveals Numerous Previously Unrecognizable Structural Variants in Leukemia Patients' Samples.** *bioRxiv* 2019:563270.
6. Zhu Y, Brown HN, Zhang Y, Stevens RG, Zheng T: **Period3 structural variation: a circadian biomarker associated with breast cancer in young women.** *Cancer Epidemiol Biomarkers Prev* 2005, **14**:268-270.
7. Peng Y, Yuan C, Tao X, Zhao Y, Yao X, Zhuge L, Huang J, Zheng Q, Zhang Y, Hong H, et al.: **Integrated analysis of optical mapping and whole-genome sequencing reveals intratumoral genetic heterogeneity in metastatic lung squamous cell carcinoma.** *Transl Lung Cancer Res* 2020, **9**:670-681.
8. Alkan C, Sajjadian S, Eichler EE: **Limitations of next-generation genome sequence assembly.** *Nat Methods* 2011, **8**:61-65.
9. Liu S, Song L, Cram DS, Xiong L, Wang K, Wu R, Liu J, Deng K, Jia B, Zhong M, et al.: **Traditional karyotyping vs copy number variation sequencing for detection of chromosomal abnormalities associated with spontaneous miscarriage.** *Ultrasound Obstet Gynecol* 2015, **46**:472-477.
10. Trask BJ: **Human cytogenetics: 46 chromosomes, 46 years and counting.** *Nat Rev Genet* 2002, **3**:769-778.
11. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, Lee J, Lam ET, Liachko I, Sullivan ST, et al.: **Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome.** *Nat Genet* 2017, **49**:643-650.
12. Cao H, Hastie AR, Cao D, Lam ET, Sun Y, Huang H, Liu X, Lin L, Andrews W, Chan S, et al.: **Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology.** *Gigascience* 2014, **3**:34.
13. Cao H, Wu H, Luo R, Huang S, Sun Y, Tong X, Xie Y, Liu B, Yang H, Zheng H, et al.: **De novo assembly of a haplotype-resolved human genome.** *Nat Biotechnol* 2015, **33**:617-622.
14. Chaisson MJP, Sanders AD, Zhao X, Malhotra A, Porubsky D, Rausch T, Gardner EJ, Rodriguez OL, Guo L, Collins RL, et al.: **Multi-platform discovery of haplotype-resolved structural variation in human genomes.** *Nat Commun* 2019, **10**:1784.

15. Cockburn IA, Mackinnon MJ, O'Donnell A, Allen SJ, Moulds JM, Baisor M, Bockarie M, Reeder JC, Rowe JA: **A human complement receptor 1 polymorphism that reduces Plasmodium falciparum rosetting confers protection against severe malaria.** *Proc Natl Acad Sci U S A* 2004, **101**:272-277.
16. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, et al.: **Origins and functional impact of copy number variation in the human genome.** *Nature* 2010, **464**:704-712.
17. Cooper GM, Zerr T, Kidd JM, Eichler EE, Nickerson DA: **Systematic assessment of copy number variant detection via genome-wide SNP genotyping.** *Nat Genet* 2008, **40**:1199-1203.
18. de Cid R, Riveira-Munoz E, Zeeuwen PL, Robarge J, Liao W, Dannhauser EN, Giardina E, Stuart PE, Nair R, Helms C, et al.: **Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis.** *Nat Genet* 2009, **41**:211-215.
19. English AC, Salerno WJ, Hampton OA, Gonzaga-Jauregui C, Ambreth S, Ritter DI, Beck CR, Davis CF, Dahdouli M, Ma S, et al.: **Assessing structural variation in a personal genome-towards a human reference diploid genome.** *BMC Genomics* 2015, **16**:286.
20. Hastie AR, Dong L, Smith A, Finklestein J, Lam ET, Huo N, Cao H, Kwok PY, Deal KR, Dvorak J, et al.: **Rapid genome mapping in nanochannel arrays for highly complete and accurate de novo sequence assembly of the complex Aegilops tauschii genome.** *PLoS One* 2013, **8**:e55864.
21. Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS, et al.: **Improved maize reference genome with single-molecule technologies.** *Nature* 2017, **546**:524-527.
22. Mak AC, Lai YY, Lam ET, Kwok TP, Leung AK, Poon A, Mostovoy Y, Hastie AR, Stedman W, Anantharaman T, et al.: **Genome-Wide Structural Variation Detection by Genome Mapping on Nanochannel Arrays.** *Genetics* 2016, **202**:351-362.
23. Pendleton M, Sebra R, Pang AW, Ummat A, Franzen O, Rausch T, Stutz AM, Stedman W, Anantharaman T, Hastie A, et al.: **Assembly and diploid architecture of an individual human genome via single-molecule technologies.** *Nat Methods* 2015, **12**:780-786.
24. Sarsani VK, Raghupathy N, Fiddes IT, Armstrong J, Thibaud-Nissen F, Zinder O, Bolisetty M, Howe K, Hinerfeld D, Ruan X, et al.: **The Genome of C57BL/6J "Eve", the Mother of the Laboratory Mouse Genome Reference Strain.** *G3 (Bethesda)* 2019, **9**:1795-1805.
25. Seo JS, Rhie A, Kim J, Lee S, Sohn MH, Kim CU, Hastie A, Cao H, Yun JY, Kim J, et al.: **De novo assembly and phasing of a Korean human genome.** *Nature* 2016, **538**:243-247.
26. Shi L, Guo Y, Dong C, Huddleston J, Yang H, Han X, Fu A, Li Q, Li N, Gong S, et al.: **Long-read sequencing and de novo assembly of a Chinese genome.** *Nat Commun* 2016, **7**:12065.
27. Usher CL, Handsaker RE, Esko T, Tuke MA, Weedon MN, Hastie AR, Cao H, Moon JE, Kashin S, Fuchsberger C, et al.: **Structural forms of the human amylase locus and their relationships to SNPs, haplotypes and obesity.** *Nat Genet* 2015, **47**:921-925.
28. Zook JM, Hansen NF, Olson ND, Chapman L, Mullikin JC, Xiao C, Sherry S, Koren S, Phillippy AM, Boutros PC, et al.: **A robust benchmark for detection of germline large deletions and insertions.** *Nat Biotechnol* 2020.

29. Neveling K, Mantere T, Vermeulen S, Oorsprong M, van Beek R, Kater-Baats E, Pauper M, van der Zande G, Smeets D, Weghuis DO, et al.: **Next generation cytogenetics: comprehensive assessment of 48 leukemia genomes by genome imaging.** *bioRxiv* 2020:2020.2002.2006.935742.
30. Chan S, Lam E, Saghbini M, Bocklandt S, Hastie A, Cao H, Holmlin E, Borodkin M: **Structural Variation Detection and Analysis Using Bionano Optical Mapping.** *Methods Mol Biol* 2018, **1833**:193-203.
31. Pang AWC, Lee J, Anantharaman T, Lam ET, Hastie A, Borodkin M: **Comprehensive Detection of Germline and Somatic Structural Mutation in Cancer Genomes by Bionano Genomics Optical Mapping.** *Journal of Biomolecular Techniques : JBT* 2019, **30**:S9-S9.
32. Dixon JR, Xu J, Dileep V, Zhan Y, Song F, Le VT, Yardimci GG, Chakraborty A, Bann DV, Wang Y, et al.: **Integrative detection and analysis of structural variation in cancer genomes.** *Nat Genet* 2018, **50**:1388-1398.
33. Zhang Y, Broach J: **Abstract 5125: A novel method for isolating high-quality UHMW DNA from 10 mg of freshly frozen or liquid-preserved animal and human tissue including solid tumors.** *Cancer Research* 2019, **79**:5125-5125.
34. Crumbaker M, Chan EKF, Gong T, Corcoran N, Jaratlerdsiri W, Lyons RJ, Haynes AM, Kulidjian AA, Kalsbeek AMF, Petersen DC, et al.: **The Impact of Whole Genome Data on Therapeutic Decision-Making in Metastatic Prostate Cancer: A Retrospective Analysis.** *Cancers (Basel)* 2020, **12**.
35. Jaratlerdsiri W, Chan EKF, Petersen DC, Yang C, Croucher PI, Bornman MSR, Sheth P, Hayes VM: **Next generation mapping reveals novel large genomic rearrangements in prostate cancer.** *Oncotarget* 2017, **8**:23588-23602.
36. Whitlock R, Hipperson H, Mannarelli M, Burke T: **A high-throughput protocol for extracting high-purity genomic DNA from plants and animals.** *Mol Ecol Resour* 2008, **8**:736-741.
37. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA: **The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers.** *Nat Rev Cancer* 2018, **18**:696-705.
38. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, et al.: **Mutational heterogeneity in cancer and the search for new cancer-associated genes.** *Nature* 2013, **499**:214-218.
39. Berger MF, Mardis ER: **The emerging clinical relevance of genomics in cancer medicine.** *Nat Rev Clin Oncol* 2018, **15**:353-365.
40. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW: **Cancer genome landscapes.** *Science* 2013, **339**:1546-1558.
41. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhsng CZ, Wala J, Mermel CH, et al.: **Pan-cancer patterns of somatic copy number alteration.** *Nat Genet* 2013, **45**:1134-1140.