

# Validation of OncoPanel

## A Targeted Next-Generation Sequencing Assay for the Detection of Somatic Variants in Cancer

Elizabeth P. Garcia, PhD; Alissa Minkovsky, MD, PhD; Yonghui Jia, PhD; Matthew D. Ducar, MS; Priyanka Shivdasani, MS; Xin Gong, MS; Azra H. Ligon, PhD; Lynette M. Sholl, MD; Frank C. Kuo, MD, PhD; Laura E. MacConaill, PhD; Neal I. Lindeman, MD; Fei Dong, MD

• **Context.**—The analysis of somatic mutations across multiple genes in cancer specimens may be used to aid clinical decision making. The analytical validation of targeted next-generation sequencing panels is important to assess accuracy and limitations.

**Objective.**—To report the development and validation of OncoPanel, a custom targeted next-generation sequencing assay for cancer.

**Design.**—OncoPanel was designed for the detection of single-nucleotide variants, insertions and deletions, copy number alterations, and structural variants across 282 genes with evidence as drivers of cancer biology. We implemented a validation strategy using formalin-fixed, paraffin-embedded, fresh or frozen samples compared with results obtained by clinically validated orthogonal technologies.

**Results.**—OncoPanel achieved 98% sensitivity and 100% specificity for the detection of single-nucleotide variants, and 84% sensitivity and 100% specificity for the

detection of insertions and deletions compared with single-gene assays and mass spectrometry-based genotyping. Copy number detection achieved 86% sensitivity and 98% specificity compared with array comparative genomic hybridization. The sensitivity of structural variant detection was 74% compared with karyotype, fluorescence in situ hybridization, and polymerase chain reaction. Sensitivity was affected by inconsistency in the detection of *FLT3* and *NPM1* alterations and *IGH* rearrangements due to design limitations. Limit of detection studies demonstrated 98.4% concordance across triplicate runs for variants with allele fraction greater than 0.1 and at least 50× coverage.

**Conclusions.**—The analytical validation of OncoPanel demonstrates the ability of targeted next-generation sequencing to detect multiple types of genetic alterations across a panel of genes implicated in cancer biology.

(*Arch Pathol Lab Med.* 2017;141:751–758; doi: 10.5858/arpa.2016-0527-OA)

The understanding of cancer as a genomic disease, coupled with the introduction of targeted therapeutic agents to treat cancer, has propelled the genomic profiling of tumor samples to aid in assigning diagnosis and prognosis as well as selecting treatment. The identification of actionable cancer gene mutations in tumor samples may result in improved outcomes for patients.<sup>1</sup>

Somatic mutations in a defined set of oncogenes and tumor suppressor genes have been associated with biologic

significance and clinical actionability.<sup>2,3</sup> Although single-gene testing in specific cancer types has been the standard in molecular diagnostics, the increasing development of novel pathway-specific pharmaceutical agents, along with technologic advances, has made broad screening of actionable mutations across multiple cancer types feasible.<sup>4,5</sup>

Previous methods of mutation detection across multiple genes include multiplex hotspot sequencing, including mass spectrometry-based sequencing methods<sup>6</sup> and multiplex single-base pair extension sequencing.<sup>7</sup> The availability of next-generation sequencing (NGS) technology has further increased the number of genes and types of mutations detectable by a high-throughput assay, and multiple laboratories have successfully validated NGS tests to detect somatic alterations in cancer.<sup>8–10</sup> Considerations for the design and implementation of NGS for cancer have been previously described.<sup>11,12</sup>

In this paper, we report the analytical validation of OncoPanel, a targeted NGS panel of 282 genes selected based on clinical actionability in cancer. Targeted genes are enriched by hybrid capture for sequencing and are designed to detect single-nucleotide variants, insertions and deletions, copy number variations, and a limited number of

Accepted for publication January 13, 2017.

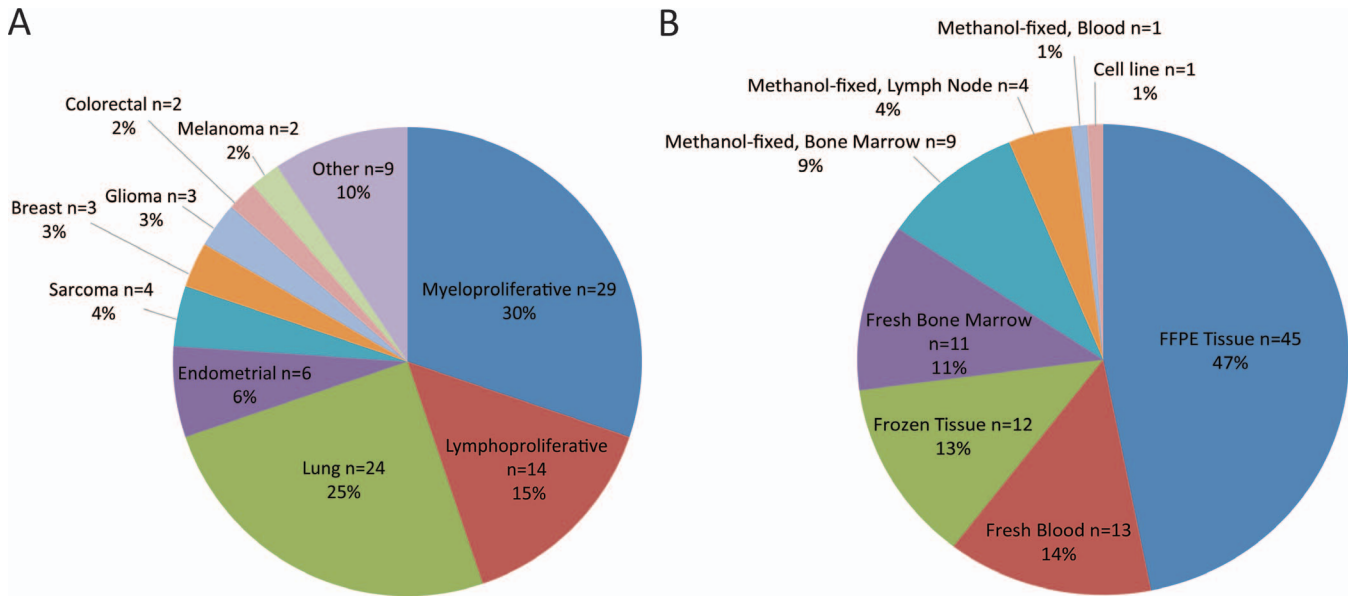
Published as an Early Online Release March 3, 2017.

Supplemental digital content is available for this article at [www.archivesofpathology.org](http://www.archivesofpathology.org) in the June 2017 table of contents.

From the Center for Advanced Molecular Diagnostics, Brigham and Women's Hospital (Drs Garcia, Minkovsky, Jia, Ligon, Sholl, Kuo, MacConaill, Lindeman, and Dong, Messrs Ducar and Gong, and Ms. Shivdasani), and the Center for Cancer Genome Discovery, Dana Farber Cancer Institute (Mr Ducar and Dr MacConaill), Harvard Medical School, Boston, Massachusetts.

The authors have no relevant financial interest in the products or companies described in this article.

Reprints: Fei Dong, MD, Department of Pathology, Brigham and Women's Hospital, 75 Francis St, Amory 3, Boston, MA 02114 (email: [fdong1@partners.org](mailto:fdong1@partners.org)).



**Figure 1.** Distribution of samples included in OncoPanel validation ( $n = 96$ ). Representation of tumor types (A) and sample types (B) included across validation studies. Other tumor types included gastrointestinal stromal tumor, astrocytoma, renal cell carcinoma, sialoblastoma, serous carcinoma, desmoplastic small round cell tumor, basal cell carcinoma, atypical carcinoid, and a College of American Pathologists survey specimen. Abbreviation: FFPE, formalin-fixed, paraffin-embedded.

structural variants. Data analysis is performed with a custom informatics pipeline. The validation of sequence alterations, copy number variations, and structural variants is described here.

## MATERIALS AND METHODS

### Panel Design

A total of 282 genes were targeted based on their designation as oncogenes and tumor suppressor genes and their involvement in cancer-related signaling pathways (Supplemental Table 1; see supplemental digital content at [www.archivesofpathology.org](http://www.archivesofpathology.org) in the June 2017 table of contents). The exons of 275 genes were targeted for detection of single-nucleotide variants, insertions and deletions (indels), and copy number alterations. Selected noncoding regions involved in structural variants, including introns, were targeted for 30 genes. In total, 1.4 Mb of the genome was targeted, including 4450 exons and 103 introns. RNA baits were designed and ordered through Agilent SureSelect (Agilent, Santa Clara, California). Exonic and intronic regions were covered using a 5X and 2X tiling strategy, respectively.

### Sample Selection and Orthogonal Testing

A total of 96 neoplastic samples of various tissues of origin from formalin-fixed, paraffin-embedded (FFPE) specimens (45), blood (13), optimal cutting temperature-embedded frozen tissue (12), bone marrow aspirate (11), methanol-fixed tissue (9 from bone marrow, 4 from lymph node, and 1 from blood), and cell line (1) were used for single-nucleotide variant and indel validation (Figure 1). Single-nucleotide variants and indels had been detected previously using validated assays, including mutation-specific polymerase chain reaction (PCR; *EGFR* p.L858R), Sanger sequencing (*EGFR*, *PIK3CA*, *ERBB2*, *KIT*, *PDGFRA*), pyrosequencing (*KRAS* codons 12 and 13; *NRAS* codons 12, 13, and 61; *BRAF* codons 597 to 600), or mass spectrometry-based genotyping.<sup>6</sup>

Twenty-three FFPE brain tumor specimens previously characterized by array comparative genomic hybridization (aCGH) were used for validation of copy number changes. Array comparative genomic hybridization was performed using the 1X1M Agilent SurePrint G3 Human CGH Microarray chip to identify tumor-specific genomic copy number changes. Genomic DNA isolated

from an FFPE specimen was hybridized with genomic DNA isolated from commercially available pooled reference DNA (Promega, Madison, Wisconsin).<sup>13</sup> The array platform included 963 029 probes spaced across the human genome with a 2.1-kb overall median probe spacing and a 1.8-kb probe spacing in RefSeq genes. A genomic imbalance was reported when a minimum of 8 consecutive probes, which corresponded to approximately 14 to 16 kb, showed an average log<sub>2</sub> ratio above +0.18 or below −0.30. Copy number gains with an average log<sub>2</sub> ratio for a given interval equal to or greater than +2.0 were reported as amplifications.

Twenty-seven specimens from neoplasms with known rearrangement events from a variety of tissue types were used for structural variant validation (summarized in the Table). Structural variants were validated compared with results obtained through karyotype, fluorescence in situ hybridization, PCR, or reverse transcription PCR.

### DNA Isolation

For FFPE and fresh frozen solid tumors, hematoxylin-eosin-stained slides were prepared and reviewed by a pathologist to identify areas of ≥20% tumor for molecular analysis. For FFPE samples, tumor-enriched areas were macrodissected from ten 5-micron tissue sections. For fresh frozen tissues, tumor-enriched areas were minced. For blood and bone marrow aspirate samples, DNA was extracted directly. In order to obtain sufficient quantities of DNA, several isolations were performed from each sample, pooled to ensure homogeneity, and then aliquoted for use in validation. DNA was isolated using standard extraction methods (Qiagen, Valencia, California) and quantified using PicoGreen dsDNA detection (Life Technologies, Carlsbad, California).

### Library Preparation, Hybrid Capture, and Sequencing

The NGS workflow is summarized in Figure 2. A total of 200 ng of tumor DNA was fragmented to an average of 270 bases using a Covaris LE220 series sonicator (Covaris, Woburn, Massachusetts). Samples were size selected to remove all fragments less than 100 bases. Fragmentation profiles for each sample were then evaluated using a LabChip GX (Perkin Elmer, Hopkinton, Massachusetts). To prepare indexed libraries from each sample, fragmentation was followed by end repair, A-tailing, and adapter ligation using the TruSeq LT library preparation kit (Illumina, San Diego, California).

Structural Variant Detection by Hybrid Capture–Based Next-Generation Sequencing				
Diagnosis	Sample Type	Structural Variant	Orthogonal Method	OncoPanel Detection
Chronic myelogenous leukemia	Cell pellet	<i>FGFR1</i> rearrangement	Karyotype and FISH	+
Chronic myelogenous leukemia	Blood	<i>BCR-ABL1</i>	RT-PCR	+
Chronic myelogenous leukemia	Blood	<i>BCR-ABL1</i>	RT-PCR	+
Chronic myelogenous leukemia	Blood	<i>BCR-ABL1</i>	RT-PCR	+
Chronic myelogenous leukemia	Blood	<i>BCR-ABL1</i>	RT-PCR	+
Chronic myelogenous leukemia	Cell pellet	<i>FIP1L1-PDGFR</i>	FISH	+
Acute lymphoblastic leukemia	Blood	<i>BCR-ABL1</i>	RT-PCR	+
Acute myeloid leukemia	Bone marrow	<i>FUS-ERG</i>	Karyotype and FISH	+
Acute myeloid leukemia	Cell pellet	<i>KMT2A</i> rearrangement	FISH	+
Acute myeloid leukemia	Cell pellet	<i>KMT2A</i> rearrangement	FISH	+
Acute myeloid leukemia	Cell pellet	<i>KMT2A</i> rearrangement	FISH	+
Acute promyelocytic leukemia	Bone marrow	<i>PML-RARA</i>	RT-PCR	+
Acute promyelocytic leukemia	Bone marrow	<i>PML-RARA</i>	RT-PCR	+
Diffuse large B-cell lymphoma	Blood	<i>IGH-MYC</i>	FISH	+
Desmoplastic small round cell tumor	FFPE	<i>EWSR1</i> rearrangement	FISH	+
Desmoplastic small round cell tumor	FFPE	<i>EWSR1</i> rearrangement	FISH	+
Follicular lymphoma	Frozen	<i>IGH-BCL2</i>	PCR	+
CAP survey	DNA	<i>IGH-BCL2</i>	PCR	+
CAP survey	DNA	<i>IGH-BCL2</i>	PCR	–
Lymphoma	Blood	<i>IGH-CCND1</i>	Karyotype	–
Multiple myeloma	Cell pellet	<i>IGH-CCND1</i>	Karyotype and FISH	–
Multiple myeloma	Cell pellet	<i>IGH-CCND1</i>	FISH	–
Multiple myeloma	Cell pellet	<i>IGH-FGFR3</i>	FISH	–
Chronic lymphocytic leukemia	Cell pellet	<i>IGH</i> rearrangement	FISH	–
Lung adenocarcinoma	Frozen	<i>EML4-ALK</i>	FISH	+
Lung adenocarcinoma	FFPE	<i>EML4-ALK</i>	FISH	+
Lung adenocarcinoma	FFPE	<i>ALK</i> rearrangement	FISH	–

Abbreviations: CAP, College of American Pathologists; FFPE, formalin-fixed, paraffin-embedded; FISH, fluorescent in situ hybridization; RT-PCR, reverse transcription–polymerase chain reaction; +, detected; –, not detected.

Adapters included unique 6-nucleotide bar codes or indexes as well as sequences required for PCR enrichment. Using adapter-specific primers, libraries were then PCR enriched for 10 cycles, purified, and quantified by qPCR (KAPA SYBR Fast qPCR kit, Wilmington, MA). Equimolar concentrations of each library were pooled into sets of 24 and hybridized at 65°C for 24 hours to OncoPanel, a custom RNA bait set (Agilent SureSelect), in order to enrich for the targeted 282 genes. Pools were then washed, PCR enriched, and purified. Hybrid-captured pools were then quantified using qPCR (KAPA SYBR Fast qPCR kit) and resuspended to 13 pM for sequencing using the Illumina HiSeq2500 in fast mode with 100 × 100 paired-end reads per the manufacturer's instructions (Illumina).

### Informatics Pipeline

Pooled sample reads were deconvoluted and sorted using the Picard tools (for details, see <https://broadinstitute.github.io/picard/command-line-overview.html#Overview>; accessed May 2013). Reads were aligned to the reference sequence b37 edition from the Human Genome Reference Consortium, using bwa (<http://bio-bwa.sourceforge.net/bwa.shtml>, version 0.5.9; accessed May 1, 2013). Duplicate reads were identified and removed using Picard (version 1.90). The alignments were further refined using the Genome Analysis Toolkit (GATK, version 1.6-5-g557da77) for localized realignment around indel sites and recalibration of the quality scores (<https://software.broadinstitute.org/gatk/>; accessed May 2013).

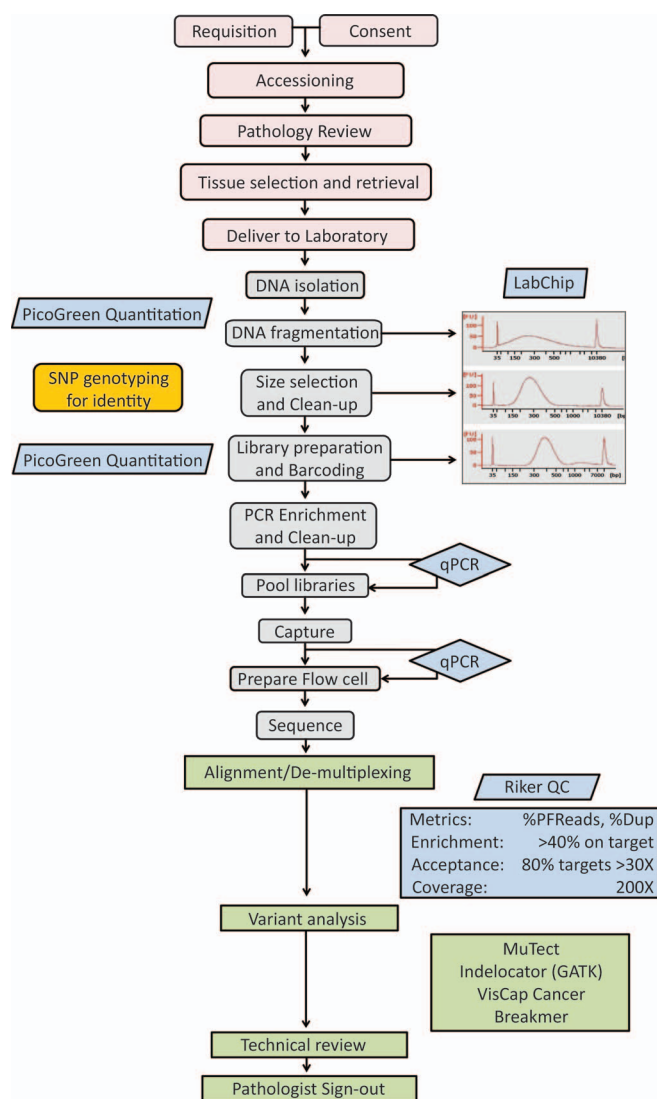
Mutation analysis for single-nucleotide variants was performed using MuTect v. 1.0.27200 (<https://confluence.broadinstitute.org/display/CGATools/MuTect>; accessed May 2013) and annotated by Oncotator (<http://www.broadinstitute.org/oncotator>; accessed May 2013), developed by the Cancer Biology Group at the Broad Institute. MuTect was made available through the generosity of Kristian Cibulskis and the Cancer Genome Analysis Program at the Broad Institute.<sup>14</sup> Insertions and deletions were called using Indelocator (<https://confluence.broadinstitute.org/display/CGATools/Indelocator>; accessed May 2013). Integrative Ge-

nomics Viewer (IGV; version 2.0.16 or later)<sup>15,16</sup> and our internally developed application were used for visualization and interpretation. Variants were filtered to exclude those that occur at a populational frequency of greater than 0.1% in the Exome Sequencing Project database (<http://evs.gs.washington.edu/EVS/>; accessed May 2013).<sup>17</sup> For each sequencing run, nonneoplastic FFPE liver and blood samples were included as controls. Variants identified in these control samples due to sequencing artifacts were filtered. Any filtered variants that were reported in COSMIC more than twice were rescued and presented for manual review.

For copy number analysis, a custom R-based tool (VisCap-Cancer) was used to calculate the fractional coverage of specified genomic intervals compared with the median fractional coverage obtained in a panel of 67 FFPE nonneoplastic samples. Coverage across each interval captured was calculated using the "DepthOfCoverage" program of GATK.<sup>18</sup> The final outputs of the tool were log<sub>2</sub> ratio values, which were plotted by relative genome order. All VisCap copy number plots were manually reviewed and interpreted with consideration of tumor percentage from initial pathologist review.

Structural variant analysis was performed using Breakmer,<sup>19</sup> followed by review of reads in IGV. Breakmer used a split-read, assembly, and realignment approach for calling structural variants at a target gene level. Based on sequence reads mapped to a reference genome, the algorithm initiated by extracting non-duplicate, misaligned reads, including reads that have been partially aligned to the reference and unmapped reads with a mapped paired end. For the extracted reads mapped to any given targeted genomic region, a "kmer" subtraction from reference was performed. The kmers that only exist within the sample set were retained for assembling contigs that may contain structural variants. The iterative assembly process began with a seed kmer that retrieved associated reads containing the kmer. Each contig assembled with the minimum number of 2 reads was aligned against the target gene reference sequence.





**Figure 2.** OncoPanel workflow through accessioning, laboratory processing, and software analysis. Clinician order and accessioning: Once the clinician has ordered OncoPanel and consent has been received, the case is reviewed by pathologists to select an appropriate specimen for testing, and, for a formalin-fixed paraffin-embedded specimen, the block is retrieved and cut. Hematoxylin-eosin-stained slides are reviewed, and areas of sufficient tumor density are circled for macrodissection of unstained slides and delivered to the laboratory. Laboratory processing: DNA is extracted from the indicated areas, fragmented, and used for library preparation. Quality controls include fragment analysis before and after library preparation, and parallel genotyping by mass spectrometry to assess sample identity. Libraries are polymerase chain reaction (PCR) enriched, undergo hybrid capture using the OncoPanel bait set, and again PCR-enriched prior to sequencing. Data analysis: Sequencing data are demultiplexed, are aligned, and duplicates are removed prior to analysis for variants. Single-nucleotide variants, insertions and deletions, copy number alterations, and structural variants are identified as described in Materials and Methods.

## RESULTS

### DNA Input

Seven samples with known alterations as well as FFPE liver and blood controls were tested with 200, 150, 100, and 75 ng of starting DNA. The starting DNA input was correlated with the amount of DNA recovered after

fragmentation and size selection. The median recovered DNA quantity was 56.1 ng (range, 34.9–90.4 ng) for samples with greater than or equal to 150 ng of starting DNA, compared with 29.6 ng (range, 16.7–49.8 ng) for samples with less than or equal to 100 ng of starting DNA (Figure 3, A). Analysis of the mean target coverage, normalized to pass-filter reads, achieved with each starting condition showed a trend of increasing mean target coverage with increasing input DNA. These trends were attributed to greater complexity as supported by an increase in percent selected bases accompanied by a concomitant decrease in duplication rate (Figure 3, B).

Low starting DNA input did not adversely affect variant detection. A total of 10 of 10 known alterations (5 indels, 3 single-nucleotide variants, 1 copy number variant, and 1 structural variant) were detected in the tumor tissues irrespective of initial DNA input. In addition, we found similar allele fractions of variants in tests originating from various DNA concentrations (Figure 3, C). Because of the correlation between DNA input and quality metrics, 200 ng of DNA was recommended for NGS analysis. However, we concluded that at least 75 ng of DNA was sufficient as the starting input for library construction and sequencing.

### Sequencing Metrics

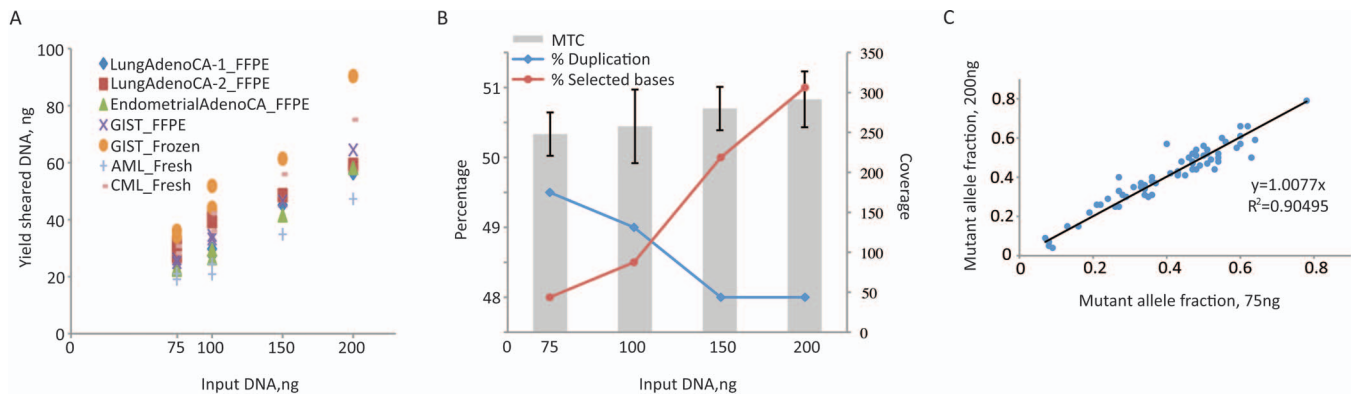
Across validation runs the failure rate was 3.4% (10 of 293 samples) as defined by a mean target coverage of less than or equal to 50× or if 80% or fewer bases were covered at less than or equal to 30×. Analysis of the 283 passing samples across 5 sequencing runs demonstrated that the assay achieved mean target coverage of 208× (range, 81–425), with 24.57 million passing filter reads per sample. The mean duplication rate was 68% (range, 52%–89%), whereas the mean selected bases was 50.3% (range, 35%–56%), with an average of 96.9% of bases achieving 30× or greater coverage.

### Reproducibility

To assess for reproducibility, 17 samples with known pathogenic alterations were tested in triplicate across 3 independent experiments. Across 153 total tests, 1 variant, *NPM1* p.W288fs\*12, was not detected in 1 of 9 replicates. Comparison of variant frequencies across replicate samples showed high reproducibility for all variants except *NPM1*, where a large range of 0 to 0.7 variant allele fractions was observed (Figure 4). Review of the region in IGV showed evidence of the expected 4-nucleotide insertion in the discordant sample and also demonstrated that *NPM1* was poorly covered across all samples tested. The intrarun reproducibility was found to be 100%, 100%, and 98% across the 3 experiments, and overall interrater reproducibility was 99.5%.

### Accuracy of Single-Nucleotide Variant and Indel Detection

Test accuracy was assessed by comparison of results with known genomic alterations detected by other analytic methods. A total of 78 known alterations (46 single-nucleotide variants and 32 indels) in 119 samples were evaluated. Alterations had been previously assessed by clinically validated assays based on mass spectrometry genotyping,<sup>6</sup> mutation-specific PCR, pyrosequencing, or Sanger sequencing. Sequencing identified 45 of 46 known single-nucleotide variants, with a sensitivity of 97.8% (95% confidence interval [CI], 87%–99.9%). One false-negative result was seen in *JAK2* p.P132T, which was filtered by the informatics pipeline because of a minor allele frequency of



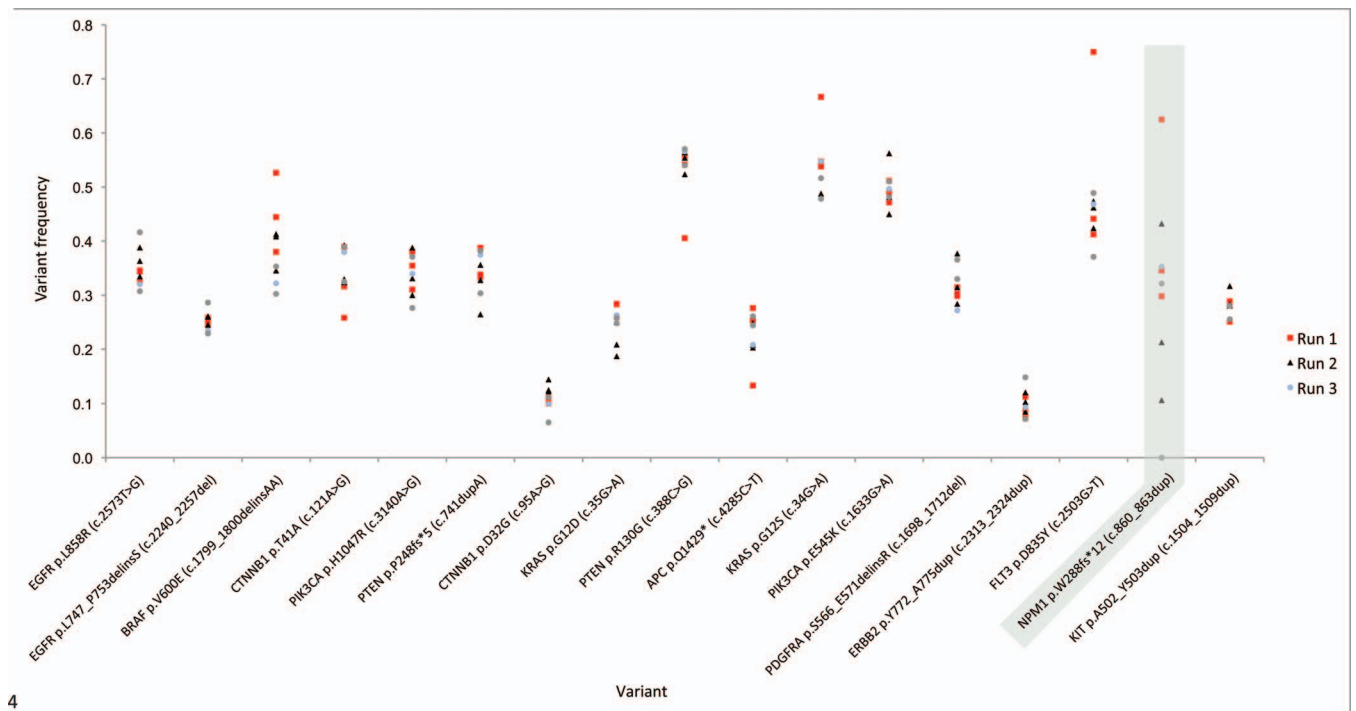
**Figure 3.** Minimal required DNA input for sequencing. A minimum of 75 ng of DNA was required for OncoPanel testing. Seven samples were assessed in OncoPanel at 75, 100, 150, and 200 ng of starting DNA with respect to sequencing metrics and variant detection. A, Recovery of size-selected fragment DNA decreased with decreasing input DNA. B, Higher input DNA amount was associated with higher percent selected bases and lower duplication rate. C, Comparison of variant allele fractions observed in samples sequenced with 75 or 200 ng of starting DNA.

greater than 0.1% in the ESP database. OncoPanel achieved specificity of 100% (22 227 of 22 228; 95% CI, 100%–100%) compared with all methods. In 1 case, *PIK3CA* p.N345K was detected by NGS but not by mass spectrometry genotyping. Upon review, there was evidence for a low-allele fraction variant by mass spectrometry, consistent with an allele fraction of 10% observed by NGS. The discordance was attributed to improved sensitivity by NGS compared with mass spectrometry.

Of the 32 indels evaluated, 27 were identified by OncoPanel, with a sensitivity of 84.3% (95% CI, 66.5%–94.1%). All discordant results were observed in myeloid neoplasms with in tandem duplication mutations in *FLT3* and indels in *NPM1*. Coverage was poor for *NPM1* exon 12, averaging only 50× (range, 18× to 148×) across validation

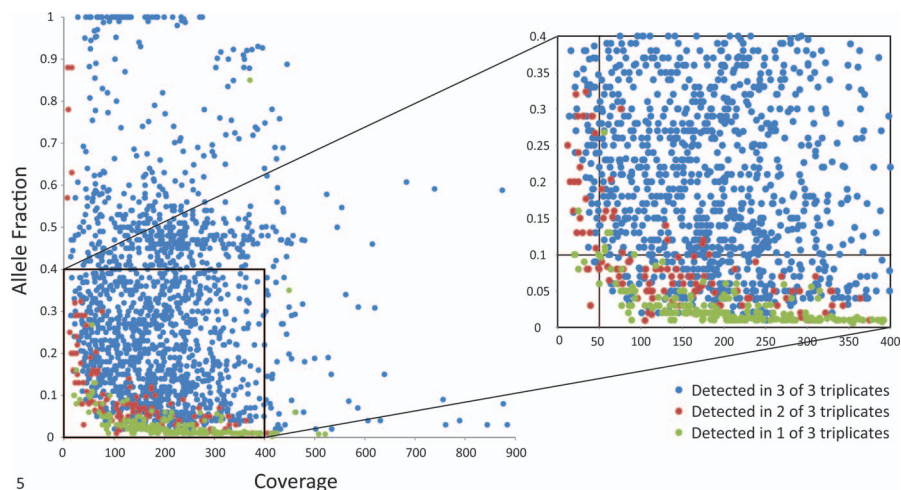
runs. This exon only contains 13 codons and was covered at 2.3 baits per nucleotide (targeted for 5× tiling). Inspection of IGV showed that the variant had been detected, but the call was filtered because of low frequency. In addition, GATK Indelocator did not efficiently detect insertion and deletion events greater than 20 nucleotides. All 3 *FLT3* in tandem duplications included for validation were not identified by the informatics pipeline, but there was evidence for the events upon review in IGV. Indel detection by NGS achieved 100% specificity (95% CI, 99.9%–100%), with no false positives observed.

We recognized limitations of the existing assay design and bioinformatics pipeline to reliably detect *FLT3* in tandem duplication and *NPM1* mutations. We addressed these



**Figure 4.** Reproducibility. Allele fractions of known single-nucleotide variant, and insertion and deletion alterations were consistent across reproducibility runs. Ten samples were assessed for reproducibility in triplicate across 3 runs for 17 known genetic variants. The *NPM1* variant that was inconsistently identified is highlighted in gray.

**Figure 5.** Limit of detection. Single-nucleotide variants, and insertions and deletions were reproducibly detected at a minimum of 50× coverage and 0.1 variant allele fraction. DNA extracted from 12 tumor samples was diluted in nonneoplastic DNA to 100%, 50%, and 20% of the original sample and tested by OncoPanel. Analysis of coverage and variant allele fractions was performed for all variants identified.



issues by requiring manual review of *FLT3* and *NPM1* for mutations in IGV for all samples.

### Lower Limit of Detection

Lower limit of detection was based on analysis of reproducibility for variants detected across 12 neoplastic samples with tumor purity between 80% and 100%. Samples were diluted with nonneoplastic DNA to final dilutions of 100% (no addition of diluent), 50%, and 20% of DNA isolated from the neoplastic sample. Each sample was tested in triplicate at each dilution. DNA isolated from FFPE tumor samples was diluted with DNA from FFPE nonneoplastic liver, whereas DNA isolated from fresh frozen and blood samples was diluted with DNA from normal blood.

A total of 2011 alterations, as detected by MuTect and GATK, were analyzed across all samples (Figure 5). Most of the alterations represented single-nucleotide polymorphisms identified at varying allele fractions from dilution. For variants with coverage 50× or greater and allele fraction 10% or greater, alterations were detected across all 3 triplicate runs for 1309 of 1330 alterations (98.4% concordance). If the coverage was less than 50× or allele fraction was less than 10%, alterations were detected across all 3 triplicate runs for 348 of 681 variants (51.1% concordance). These findings established our limit of detection at 10% allele fraction and 50× coverage for reproducible results.

### Copy Number Analysis

Copy number analysis by sequencing was compared to analysis by aCGH for 23 brain tumors, including gliomas, medulloblastomas, and meningiomas. For each case, 41 select genomic regions, including single genes and chromosome-level alterations, were evaluated based on known pathogenicity in neurologic neoplasms. Interpretations of copy number alterations were based on review of copy number plots from aCGH and NGS by 2 independent and blinded reviewers.

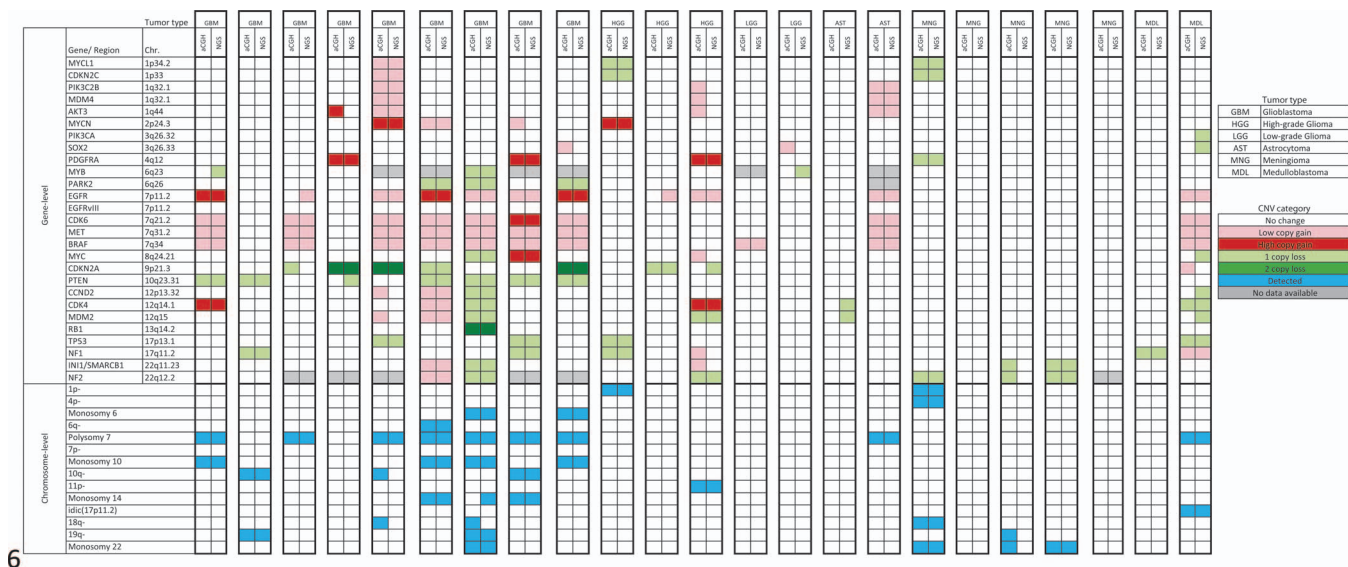
Thirteen amplifications and four 2-copy deletions were reported by aCGH across the 23 samples included in this comparison. A total of 16 of 17 changes were concordant in OncoPanel, with the exception of an *AKT3* high-copy gain in 1 glioblastoma. Further inspection revealed that aCGH detected a gain of 8 consecutive probes that spanned intron 2. Because *AKT3* intron 2 was not targeted in OncoPanel, this result was attributable to differences in assay design. Of 68 low gains and 67 single-copy losses reported by aCGH,

sensitivity by OncoPanel was 84% (95% CI, 72%–91%) and 87% (95% CI, 76%–93%), respectively. Factors that contributed to discordant results observed for low gains and losses with OncoPanel included assay-specific differences as well as differences in interpretation of the copy-neutral baseline by independent reviewers. Because copy number analysis by NGS did not establish ploidy, low-level gains and losses could not be determined with confidence in samples without a clearly inferable diploid baseline. Overall copy number gains or losses were observed in 152 of 930 regions by aCGH and 145 of 930 regions by NGS, with a concordance of 96% (Figure 6). OncoPanel achieved 86% overall sensitivity (131 of 152; 95% CI, 79%–91%) and 98% overall specificity (764 of 778; 95% CI, 97%–99%) for copy number detection compared with that of aCGH.

### Structural Variant Analysis

Structural variant detection was compared for 27 samples with known rearrangements as determined by reverse transcription-PCR (*BCR-ABL1*, *PML-RARA*), karyotype, or fluorescence in situ hybridization (*ALK*, *EWSR1*, *FGFR1*, *FGFR3*, *FIP1L1-PDGFRα*, *FUS-ERG*, *IGH-MYC*, *IGH-BCL2*, *IGH-CCND1*, and *KMT2D*). As shown in the Table, the correct structural variant was identified by sequencing in 20 of the 27 unique samples. Of the 7 discordant results, 6 were rearrangements involving *IGH*. Subsequent investigation confirmed that the break points were not part of the approximately 10% of the *IGH* region in the OncoPanel bait set. For 2 samples with *IGH-BCL2* rearrangements, sequencing identified the structural variant because of hybrid capture of sequences of *BCL2*. An *IGH-MYC* rearrangement was also detected in another specimen because of capture of sequences associated with *MYC*. Sequencing did not detect any of the 3 samples with *IGH-CCND1* rearrangements or a single sample with an *IGH-FGFR3* rearrangement. BreakMer also did not detect an *ALK* rearrangement for a patient who had an *ALK* rearrangement detected by fluorescence in situ hybridization in a different specimen. Assay-specific limitations, such as a rearrangement involving an untargeted intron of *ALK*, or specimen-specific factors, such as tumor purity, could explain the discrepant results. Overall, the sensitivity for structural variant detection was 74% (20 of 27; 95% CI, 53%–88%) in this study.





**Figure 6.** Copy number variant assessment by next-generation sequencing (NGS) compared with array comparative genomic hybridization. For each gene-level alteration, the presence of the variant is indicated by shading according to the accompanying color legend as either a low- or high-copy gain, or a 1- or 2-copy loss. The absence of color indicates the alteration was not detected. Light gray color indicates no data were available for that location. For each chromosome-level alteration (see lower panel), the blue shading indicates the presence of the alteration.

## DISCUSSION

We describe the analytical validation of OncoPanel, a targeted assay to detect genomic alterations in 282 genes implicated in cancer biology. Our assay achieves a sensitivity of 98% for single-nucleotide variants, 84% for indels, 86% for copy number variants, and 74% for structural variants, compared with those determined using orthogonal methods. The reproducibility of this assay for single-nucleotide variants and indels is high, with concordance of 98.4% across triplicate samples for variants with coverage of greater than 50× and greater than 10% allele fraction.

The detection of somatic alterations by NGS poses specific challenges due to admixture of neoplastic and nonneoplastic cells within each specimen. Tumor purity is a major consideration in the interpretation of low-allele fraction variants as well as copy number analysis. Because the validation establishes the reproducible limit of detection at 10% allele fraction at 50× coverage, our laboratory has set a minimum tumor content of 20% neoplastic cell nuclei based on histologic evaluation as a preanalytic criterion for sequencing. Heterozygous somatic variants in a diploid tumor population would be expected to be identified in specimens meeting this criterion.

As a part of the validation, we identify recurrent limitations of our targeted NGS panel and bioinformatics methods to detect alterations in select genomic regions. Our assay is limited specifically in the detection of indels in *FLT3* and *NPM1*. To improve assay sensitivity for *FLT3* and *NPM1* indels, we have implemented manual review of *FLT3* and *NPM1* in IGV as a part of the interpretation and reporting process. In addition, we recognize limited sensitivity of this assay to detect genomic rearrangements involving *IGH* based on limitations of hybrid capture design. These limitations highlight current challenges in NGS cancer panel testing and show that testing of some genomic regions by complementary single-gene assays may be necessary.

Overall, the performance characteristics of this assay are comparable to those previously described.<sup>8–10</sup> Our validation

methods are unique in that the assay is validated across a variety of fresh, frozen, and formalin-fixed specimens, and that the assay is compared against multiple validated molecular testing methods. In addition, we describe comprehensive validation of copy number analysis compared with aCGH and demonstrate the ability of NGS to detect copy number alterations, including low copy gains and single-copy losses, while acknowledging limitations in copy number detection in tumors with a poorly defined copy-neutral baseline. Finally, we demonstrate the ability of capture-based targeted NGS to detect structural variants across multiple specimen types. Although our assay is able to detect copy number variants and structural variants with reasonable accuracy, the ability to detect these alterations is dependent on library preparation methods, assay design, and the availability of bioinformatics tools. Correlation of NGS results with orthogonal methods, such as fluorescence in situ hybridization, is useful for some alterations, and additional standardization in the implementation and reporting of copy number variants and structural variants is needed.

In addition to the validation data described here, our laboratory has implemented quality control steps within the laboratory workflow. In the preanalytic phase, a hematoxylin-eosin slide of every solid tumor sample is reviewed by a pathologist for adequacy. The hematoxylin-eosin slide is subsequently scanned, and the digital image is made available for review at the time of molecular interpretation. Every sample is screened for DNA contamination from an unrelated source. After extraction, the isolated DNA is tested in parallel by mass spectrometry genotyping at 48 sites of known single-nucleotide polymorphisms, and these results are compared to the genotype assessed by NGS. The results verify that molecular bar codes are associated with the correct sample, and they assess for cross-contamination during library preparation. After sequencing, a laboratory scientist manually reviews every variant identified by the pipeline to evaluate for the possibility of a technical artifact and makes initial copy number interpretations. Finally, an attending

pathologist confirms key elements of the technical review and provides clinical variant interpretation for the final report.

Since completing the validation of OncoPanel in 2013, this assay has been implemented in a molecular epidemiology project to characterize genomic alterations across multiple cancer types and has been performed on more than 12 000 samples. In addition, OncoPanel has been used clinically to identify targetable variants in lung cancer with planned clinical implementation in other tumor types.

This report and others have demonstrated that NGS technology has matured and now provides a reliable method for the detection of a variety of somatic alterations in cancer, including single-nucleotide variants, insertions and deletions, copy number changes, and structural variants. As the number of actionable gene targets increases because of an enhanced understanding of cancer biology, we anticipate that targeted NGS will continue to be an accurate and efficient tool for the analysis of cancer genomes.

The authors acknowledge Lauren Crosby, BS, Mark Byrne, MS, Ruchi Joshi, MS, and Jodie Conneely, BA, at the Center for Advanced Molecular Diagnostics for technical assistance; Ryan Abo, PhD, Trevor Pugh, PhD, Phani Davineni, MS, Larry Chung, MS, Chesley Leslin, PhD, Matthew Temple, MS, and Jay Fink, AS, for bioinformatics analysis and infrastructure development; Dimity Hall, BS, and Eric Reed, BS, for assistance in identification and procurement of validation samples; and Paul Van Hummelen, PhD, and the Center for Cancer Genome Discovery for shared expertise. The authors acknowledge Patricia Vasalos, BS, for providing support and coordination for all of the NGS validation manuscripts in this series; she is an employee of the College of American Pathologists (Northfield, Illinois). This project is supported by the Dana Farber Cancer Institute and Brigham and Women's Hospital.

# References

1. Mody RJ, Wu YM, Lonigro RJ, et al. Integrative clinical sequencing in the management of refractory or relapsed cancer in youth. *JAMA*. 2015;314(9):913–925.
2. Paez JG, Jänne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*. 2004;304(5676):1497–1500.

3. Sosman JA, Kim KB, Schuchter L, et al. Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. *N Engl J Med*. 2012;366(8):707–714.
4. Wagle N, Berger MF, Davis MJ, et al. High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. *Cancer Discov*. 2012;2(1):82–93.
5. Gagan J, Van Allen EM. Next-generation sequencing to guide cancer therapy. *Genome Med*. 2015;7(1):80.
6. MacConaill LE, Garcia E, Shivdasani P, et al. Prospective enterprise-level molecular genotyping of a cohort of cancer patients. *J Mol Diagn*. 2014;16(6):660–672.
7. Su Z, Dias-Santagata D, Duke M, et al. A platform for rapid detection of multiple oncogenic mutations with relevance to targeted therapy in non-small-cell lung cancer. *J Mol Diagn*. 2011;13(1):74–84.
8. Frampton GM, Fichtenholtz A, Otto GA, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol*. 2013;31(11):1023–1031.
9. Cottrell CE, Al-Kateb H, Bredemeyer AJ, et al. Validation of a next-generation sequencing assay for clinical molecular oncology. *J Mol Diagn*. 2014;16(1):89–105.
10. Cheng DT, Mitchell TN, Zehir A, et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J Mol Diagn*. 2015;17(3):251–264.
11. Hagemann IS, Cottrell CE, Lockwood CM. Design of targeted, capture-based, next generation sequencing tests for precision cancer therapy. *Cancer Genet*. 2013;206(12):420–431.
12. Aziz N, Zhao Q, Bry L, et al. College of American Pathologists' laboratory standards for next-generation sequencing clinical tests. *Arch Pathol Lab Med*. 2015;139(4):481–493.
13. Craig JM, Vena N, Ramkissoon S, et al. DNA fragmentation simulation method (FSM) and fragment size matching improve aCGH performance of FFPE tissues. *PLoS One*. 2012;7(6):e38881.
14. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31(3):213–219.
15. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24–26.
16. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14(2):178–192.
17. National Heart, Lung, and Blood Institute GO Exome Sequencing Project. Exome Variant Server. <http://evs.gs.washington.edu/EVS>. Accessed May 2013.
18. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–1303.
19. Abo RP, Ducar M, Garcia EP, et al. Breakmer: detection of structural variation in targeted massively parallel sequencing data using kmers. *Nucleic Acids Res*. 2014;43(3):e19.