Validation of a Next-Generation–Sequencing Cancer Panel for Use in the Clinical Laboratory

Birgitte B. Simen, PhD; Lina Yin, PhD; Chirayu P. Goswami, MS; Kathleen O. Davis, MB(ASCP); Renu Bajaj, PhD; Jerald Z. Gong, MD; Stephen C. Peiper, MD; Erica S. Johnson, PhD; Zi-Xuan Wang, PhD

• ***Context.***—**Next-generation sequencing allows for high-throughput processing and sensitive variant detection in multiple genes from small samples. For many diseases, including cancer, a comprehensive mutational profile of a targeted list of genes can be used to simultaneously inform patient care, establish eligibility for ongoing clinical trials, and further research.**

Objective.—**To validate a pan-cancer, next-generation–sequencing assay for use in the clinical laboratory.**

Design.—**DNA was extracted from 68 clinical specimens (formalin-fixed, paraffin-embedded; fine-needle aspirates; peripheral blood; or bone marrow) and 5 normal controls. Sixty-four DNA samples (94%; 64 of 68) were successfully processed with the TruSeq Amplicon Cancer Panel (Illumina Inc, San Diego, California) and sequenced in 4 sequencing runs. The data were analyzed at 4 different filter settings for sequencing coverage and variant frequency cutoff.**

Results.—**Libraries created from 40 specimens could be successfully sequenced in a single run and still yield sufficient coverage for robust data analysis of individual samples. Sensitivity for mutation detection down to 5% was demonstrated using dilutions of clinical specimens and control samples. The test was highly repeatable and reproducible and showed 100% concordance with clinically validated Sanger sequencing results. Comparison to an alternate next-generation sequencing technology was performed by also processing 9 of the specimens with the AmpliSeq Cancer Hotspot Panel (version 2; Life Technologies, Grand Island, New York). Thirty of the 31 (97%) TruSeq-detected variants covered by the designs of both panels were confirmed.**

Conclusions.—**A sensitive, high-throughput, pan-cancer mutation panel for sequencing of cancer hot-spot mutations in 42 genes was validated for routine use in clinical testing.**

(*Arch Pathol Lab Med.* 2015;139:508–517; doi: 10.5858/arpa.2013-0710-OA)

Current molecular approaches for individualized and precision diagnosis of cancer typically interrogate one or a few genes as companion diagnostics. The dramatic increase in insight into mutational profiles of malignancies and the rapid development of antagonists for activated oncogenes have spawned more-aggressive genomic tactics to identify individual tumor genotypes for predicting responsiveness to molecularly targeted agents. While current methods employ real-time polymerase chain reaction (PCR) and Sanger sequencing, which are typically performed separately for each exon of each gene to be analyzed within a given sample, next-generation sequencing, with its increased sensitivity and remarkable throughput, is rapidly entering the clinical testing space. Validation for use in a Clinical Laboratories Improvement Amendments of 1988 (CLIA)–approved testing laboratory requires rigorous validation and ongoing performance monitoring. Guidelines are now available from the College of American Pathologists, American College of Medical Genetics,[1] and a comprehensive work group convened by the Centers for Disease Control.[2] These will likely continue to evolve as individual laboratories share ideas and results from tailoring of commercially available tests and custom panels.[3–8]

Somatic mutations in cancer can be challenging to detect by Sanger sequencing. Often, these mutations are found at low prevalence because they can be differentially present in distinct cell populations or because the tumor material is mixed with normal tissue. Robust DNA library preparation and high-sensitivity detection are important for successful detection of such molecular changes. We decided to base our first, routine, next-generation sequencing test offering on a commercially available kit that has been designed to cover a fairly comprehensive list of cancer-associated mutational hot spots.

The resulting pan-cancer panel is well suited for clinical applications to guide advanced treatment and to predict the clinical course of different types of cancers. The immediate concern for clinicians is to obtain data that will guide

Accepted for publication June 3, 2014.

Published as an Early Online Release October 30, 2014.

Supplemental digital content is available for this article at www.archivesofpathology.org in the April 2015 table of contents.

From the Genomic Pathology Laboratory, Thomas Jefferson University Hospital (Dr Simen, Mr Goswami, and Ms Davis), Philadelphia, Pennsylvania; and the Departments of Pathology, Anatomy, & Cell Biology (Drs Yin, Bajaj, Gong, Peiper, Johnson, and Wang); and Surgery (Dr Wang), Thomas Jefferson University, Philadelphia. Drs Simen and Yin, Mr Goswami, and Ms Davis contributed equally to this article.

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

Reprints: Zi-Xuan Wang, PhD, Department of Surgery, Thomas Jefferson University, 117 South 11th Street, Pavilion Building, Suite 207, Philadelphia, PA 19107 (e-mail: zi-xuan.wang@jefferson.edu).

treatment with US Food and Drug Administration–approved drugs currently used as the standard of care according to National Comprehensive Cancer Network guidelines. These include somatic mutations in *KRAS* (codons 12, 13) and *BRAF* (V600E) in colorectal cancer (CRC) that predict poor prognosis and nonresponse to anti-EGFR antibodies, for example, cetuximab or panitumumab.[9,10] In melanoma, *BRAF* V600E is predictive of a positive response to the *BRAF* V600-specific inhibitor vemurafenib.[11] Somatic mutations in *EGFR* in exons 18, 19, and 21 are predictive of a clinical response to the EGFR tyrosine kinase inhibitor drugs gefitinib and erlotinib, and the T790M point mutation in exon 20 has been associated with resistance to these drugs.[12] In the absence of a companion diagnostic result that permits therapy with an approved, targeted agent, the emerging paradigm calls for profiling tumors for an extended panel of mutations to identify alternative driver mutations, including those predictive of sensitivity/clinical responsiveness to novel molecularly targeted drugs. Clinical trial information related to many of the 42 validated genes can be found on http://clinicaltrials.gov (accessed November 1, 2013) and http://www.mycancergenome.org (accessed November 1, 2013). For instance, somatic mutations in *NRAS* codons 12, 13, and 61 are used as biomarkers for advanced solid tumors and are associated with eligibility for several ongoing clinical trials, for example, a phase II trial of pimasertib versus dacarbazine in the management of *NRAS*-mutated, locally advanced or metastatic, malignant, cutaneous melanoma (NCT01693068), and a phase II trial of MEK1/2 inhibitor in patients with advanced melanoma and mutations in *NRAS* or *BRAF* V600 (NCT01320085).[13] Another important use of the data is to predict the expected clinical course. For instance, mutations in *TP53* leading to altered function or expression of p53 are typically predictive of a poor clinical course in many cancer types.[14]

After removing 6 genes from the data analysis because they each included amplicons with inferior performance in library preparation or sequencing, we established performance criteria for mutational analysis of 42 genes with full or partial coverage for 157 exons. Here, we describe the process and the results from the validation, showing robust coverage statistics, high precision, sensitivity down to 5% variant frequency, and excellent concordance with other clinical and experimental data produced in our CLIA-approved and College of American Pathologists–accredited testing laboratory.

MATERIALS AND METHODS

Sample Information

Sixty-eight clinical tumor specimens and 5 normal controls were analyzed in this study. Of the 68 clinical specimens, 4 (6%) did not pass sample quality control (3 failed real-time PCR quality control; 1 failed quality control after PCR cleanup; see "DNA Extraction and Sample Quality Control" and "Library Preparation and Quality Control"). The 64 successful tumor specimens included melanoma (n = 1; 2%), brain (n = 1; 2%), colon (n = 19; 30%), lung (n = 10; 16%), thyroid (n = 23; 36%), and leukemia (n = 10; 16%). By sample type, the 64 specimens consisted of formalin-fixed, paraffin-embedded tissue (FFPE; n = 32; 50%), fine-needle aspirates (FNA; n = 22; 34%), and peripheral blood or bone marrow (PM/BM; n = 10; 16%) (Supplemental Table S1; see supplemental material file at www.archivesofpathology.org in the April 2015 table of contents). The 5 normal controls were peripheral blood samples from 5 healthy individuals. Institutional review board approval was not required for test validation under CLIA using leftover clinical specimens. Results for these specimens were only used for correlation analysis.

Study Design

The 64 successful clinical specimens and 5 normal controls were processed with the TruSeq Amplicon Cancer Panel (Illumina, San Diego, California) and sequenced in 4 MiSeq (Illumina) instrument runs to test the assay precision (reproducibility and repeatability), sensitivity, and concordance with 2 alternate sequencing platforms, that is, with Ion Torrent Personal Genome Machine (PGM; Life Technologies, Carlsbad, California) and Sanger sequencing. Assay sensitivity was determined by sequencing serially diluted DNA from a clinical specimen and, separately, a mixture of normal control samples. The clinical specimen was diluted to contain expected percentages of a *TP53* variant present at nearly 100% in the original sample. Four dilutions, targeted at 5%, 3.7%, 2.5%, and 1.2% variant content, were each processed in triplicate. For the normal control sample mixture, DNA from 2 healthy volunteers of white and Asian origin were mixed to contain expected percentages of heterozygous single nucleotide polymorphisms (SNPs) present in the white sample. Three dilutions, targeted to 10%, 5%, and 2.5%, were each processed in duplicate. Assay reproducibility (interrun precision) was assessed by sequencing 6 specimens (2 FFPE [33%], 2 FNA [33%], and 2 PB/BM [33%]) in duplicates across separate runs. Assay repeatability (intrarun precision) was assessed by sequencing one clinical specimen (FFPE) with a range of variant frequencies. This specimen was indexed with different barcodes and sequenced in quadruplicate within a single run. Nine clinical specimens (FFPE) were sequenced using the Ion Torrent Hotspot V2 panel (Life Technologies) to provide a comparison with an alternate sequencing platform. Fifty-three of the 64 samples had known relevant clinical variant results done by in-house Sanger sequencing, which were confirmed by the assay across separate runs (Supplemental Table S1). The relevant clinical variants included *BRAF* V600E (colorectal and thyroid), *NRAS* Q61 (thyroid), *EGFR* exons 18 to 21 (lung), and *KRAS* G12 and G13 (colorectal). In addition, Sanger-sequencing assays were designed to cover *TP53* exons 4 to 9 identified by the pan-cancer mutation panel (our 42 gene version of the TruSeq Amplicon Cancer Panel) to serve as a confirmatory assay.

Counting all specimens, dilutions, and replicates described above, 100 samples were sequenced in the 4 runs. Additionally, 11 samples were included in run 2 and 20 in run 3 but were not analyzed in this validation study because they were concurrent tests of decreased test volumes that will not be further pursued at this time. Contamination controls (blanks, water controls) were included in the first 3 runs. A positive (SNP) control was included in all runs. This control employed a separate probe set and was intended for the TruSeq Amplicon Cancer Panel and MiSeq instrument vendor, Illumina, to perform troubleshooting in case of library preparation failure.

DNA Extraction and Sample Quality Control

Blood and bone marrow samples were extracted according to the manufacturer's protocol with the QIAamp DNA Mini Kit (Qiagen, Valencia, California) via automated QIAcube (Qiagen) extraction or manual kit extraction. Two hundred microliters of PB/BM samples was used, and DNA was eluted in a total volume of 100 μL. The FFPE tissue sections (10 μm thick) were microdissected to enrich for tumor cells. Paraffin was removed with xylene, and the tissue was washed with 100% ethanol. The FFPE tissue DNA was extracted according to the manufacturer's instructions using the QIAamp DNA FFPE kit (Qiagen) and was eluted in a 25 μL volume. Cell pellets from FNA samples were extracted according to the manufacturer's instructions using the QIAamp DNA Micro Kit (Qiagen) and were eluted in a 25 μL volume. The extracted DNA was quantified using the Qubit dsDNA BR assay (Life Technologies), and 250 ng of DNA was used as a template for sequencing-library preparation for all tissue types. For FFPE and FNA samples, the DNA quality and quantity were further assessed using the Illumina FFPE QC Kit according to the manufacturer's instructions,

with samples with changes in cycle threshold (ΔCt) value of 2 or less (versus a control) considered acceptable for sequencing.

Library Preparation and Quality Control

Sequencing amplicon libraries were prepared using the TruSeq Amplicon Cancer Panel, according to the manufacturer's instructions. The panel contains probes to generate 212 amplicons from 48 cancer-related genes: *ABL1*, *AKT1*, *ALK*, *APC*, *ATM*, *BRAF*, *CDH1*, *CDKN2A*, *CSF1R*, *CTNNB1*, *EGFR*, *ERBB2*, *ERBB4*, *FBXW7*, *FGFR1*, *FGFR2*, *FGFR3*, *FLT3*, *GNA11*, *GNAQ*, *GNAS*, *HNF1A*, *HRAS*, *IDH1*, *JAK2*, *JAK3*, *KDR*, *KIT*, *KRAS*, *MET*, *MLH1*, *MPL*, *NOTCH1*, *NPM1*, *NRAS*, *PDGFRA*, *PIK3CA*, *PTEN*, *PTPN11*, *RB1*, *RET*, *SMAD4*, *SMARCB1*, *SMO*, *SRC*, *STK11*, *TP53*, and *VHL*. Briefly, a probe set containing pairs of oligonucleotides specific to the targeted regions was hybridized to each genomic DNA sample. Amplicons were generated by connection of bound oligonucleotides by extension and ligation using a DNA polymerase and ligase, followed by PCR amplification. The PCR primers were flanked by index sequences for sample multiplexing as well as common adapters for sequencing cluster generation. After PCR cleanup, library quality was assessed on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California). Each sample library was normalized according to the manufacturer's instructions, and equal volumes were pooled to generate the final sequencing library.

MiSeq Sequencing and Data Analysis

Each pooled library was sequenced on a MiSeq instrument using a 2 × 150 paired-end sequencing design. Image processing and fastq file generation from raw read data were accomplished with CASAVA version 1.8.2 and RTA version 1.17.28 (Illumina). Alignment of paired-end raw reads to the human hg19 genome assembly was performed with a banded Smith-Waterman alignment algorithm. The Somatic Variant Caller algorithm version 3.1.6.4 (Illumina) was used for identification of variants from aligned reads, and coverage analysis was performed in parallel with Bamtools[15] (open-source; https://github.com/pezmaster31/bamtools; accessed June 1, 2013). Variants were annotated by Illumina Variant Studio version 1.0 (Illumina) and were subsequently filtered and reported using an in-house reporting Web-based application, ClinMut Reporter (Thomas Jefferson University Hospitals, Philadelphia, Pennsylvania). Four filter settings for variant frequency cutoffs were tested in this validation: 10%, 7.5%, 5%, or 2.5%, with a minimum variant coverage of 10 reads (ie, 10 reads that contained the actual variant sequence).

Variants with a global minor allele frequency greater than 1.0% were removed because they were considered common SNPs. Only coding variants and variants in the 2 nucleotides immediately preceding and following exons were considered in this validation study. The Integrative Genomics Viewer[16] (Broad Institute, Cambridge, Massachusetts) was used to visualize variants against the reference genome.

PGM Sequencing and Data Analysis

Libraries were prepared using the AmpliSeq Cancer Hotspot Panel v2 according to the manufacturer's instructions. Briefly, targets were amplified using 207 primer pairs, then partially digested to facilitate bar-coded adapter ligation. Following purification, libraries were quantified using a Qubit dsDNA HS Assay Kit and a Qubit 2.0 fluorometer, diluted to a concentration of 3 ng/mL, and pooled in equal volumes. The library pool was clonally amplified in an emulsion PCR reaction using Ion Sphere Particles on the OneTouch 2 instrument (Life Technologies) according to the manufacturer's protocol. Template-positive ion sphere particles were enriched on the Ion OneTouch ES (Life Technologies) according to the manufacturer's protocol. Following enrichment, sequencing primers and polymerase were added. The libraries were loaded onto an Ion 318 chip (Life Technologies) and sequenced on an Ion Torrent PGM instrument according to the manufacturer's protocol. Variants were identified using the Torrent Somatic Variant Caller (versions 3.4.5 and 3.6.6; Life Technologies) and annotated using Ion Reporter software (Life Technologies).

Sanger Sequencing Verification

Twenty-six samples showed *TP53* variants in exons 4 to 9 at a variant frequency above 5%. Twenty-one samples (81%) with sufficient remaining sample material were examined by Sanger sequencing using the Big Dye Terminator v3-1 cycle sequencing kit (Life Technologies) and a 3130xL Genetic Analyzer (Life Technologies) according to the manufacturer's instructions. The primers used were (5′–3′): TP53exon4F TGCCGTCTTCCAGTTGCTTT, TP53exon4R AGCAATCAGTGAGGAATCAGAGG, TP53exon5F TGGGGCTGGAGAGACGACA, TP53exon5R GGAGGTCAAATAAGCAGCAGGAG, TP53exon6F TTGCCACAGGTCTCCCCAAG, TP53exon6R TGGGTAGTAGTATGGAAGAAATCGG, TP53exon7F GGGAGTAGATGGAGCCTGGTT, TP53exon7R GGAAAGAGGCAAGGAAAGGTGA, TP53exon8F CAGGACAAGAAGCGGTGGAG, TP53exon8R ACAGTCAAGAAGAAAACGGCA, TP53exon9F TTTGTACCGTCATAAAGTCAAACAA, and TP53exon9R CCTTTGACCATGAAGGCAGGA. *TP53* primers for exons 4, 6, and 7 were combined for multiplex PCR, and exons 5, 8, and 9 were individually amplified.

For detection of deletion variants in *APC* and *CSF1R*, the following primers were used: APCexon15-1F GGATGTAATCAGACGACACAGGAT, APCexon15-1R GAACATAGTGTTCAGGTGGACT, APCexon15-2F CAGGAGACCCCACTCATGTT, APCexon15-2R GCAGCTTGCTTAGGTCCACT, CSF1Rexon21F GTAAAGCACGTTGGGCTGGG, and CSF1Rexon21R CCCCATCCATGGAGGAGTTGA. Thermocycling was performed as follows: 94°C (10 minutes); 35 cycles at 94°C (30 seconds), 58°C (30 seconds), 72°C (30 seconds); 72°C (7 minutes); and hold at 4°C. The primer concentration was 2 μM for all primers, and $MgCl_2$ concentration was 25mM.

RESULTS

Regions of Interest Covered by Targeted Sequencing Panel

The TruSeq Amplicon Cancer Panel consists of 212 amplicons, representing portions of 48 genes, which are used to generate paired-end reads approximately 125 base pairs (bp) in length. Sequenced regions of amplicons range from 109 bp to 141 bp. Each sequence read begins precisely at one of the 2 ends of the region. Therefore, complete bidirectional sequence is obtained when this region is shorter than 125 bp, whereas longer amplicons generate as much as 16 bp of monodirectional sequence on both ends. Two broad classes of genes are targeted by the panel: tumor suppressor genes and genes in which activating or dominant-negative mutations are concentrated in short hot spots. Supplemental Table S2 shows, for each gene, the percentage of somatic mutations in the Catalogue of Somatic Mutations in Cancer (COSMIC) database[17] that are covered by the panel. The panel covers all of the major cancer-associated hot spots in the targeted genes (Supplemental Table S2), although, in some cases, the hot spot is present in a region of monodirectional coverage. For tumor-suppressor genes, in which inactivating mutations are often distributed fairly evenly throughout much of the gene, the panel covers varying fractions of previously observed mutations (Supplemental Table S2). For example, 95% of somatic mutations in *TP53* are covered, but, in several other genes, less than 50% of relevant mutations are covered (Supplemental Table S2). Thus, although many mutations in tumor suppressor genes can be detected using this panel, absence of detected variants in one of these genes does not exclude the presence of a deleterious mutation in an uncovered portion of the gene.

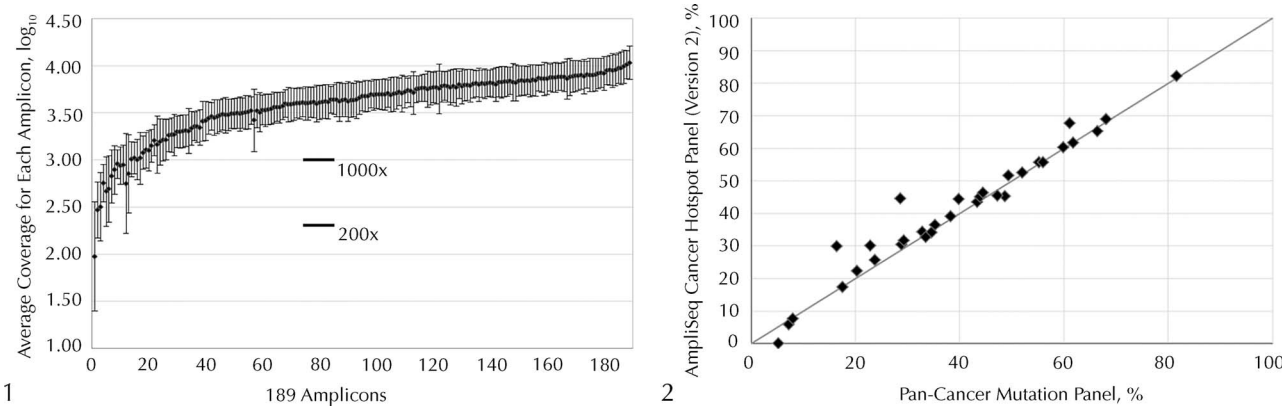

Figure 1. *Average coverage for each of the 189 amplicons across all samples. Data are shown as mean ± standard deviation.*

Figure 2. *Correlation between variant frequencies detected by the Pan-Cancer Mutation Panel (our panel based on the TruSeq Amplicon Cancer Panel [Illumina, San Diego, California]) and the AmpliSeq Cancer Hotspot Panel (Version 2) (Life Technologies, Carlsbad, California). Frequencies are shown for each variant that was detected by both panels.*

Coverage

To confidently detect low-frequency variants represented by at least 10 reads out of the total reads from the corresponding amplicon, we assessed coverage statistics for each of the 212 amplicons in the TruSeq Amplicon Cancer Panel (Supplemental Table S3 and Supplemental Figure S1). This analysis showed that 6 genes (*CDKN2A*, *FGFR3*, *HRAS*, *RB1*, *STK11*, and *VHL*) contained amplicons having significantly lower-than-average coverage, with many samples covered at less than 200× (Supplemental Table S3). This contrasted with average coverage of greater than 1000× for 88% of amplicons. Low coverage increases the levels of background noise in the assay (see "Comment") as well as increasing the probability of allele dropout. These issues make data from low-coverage amplicons less reliable. Five of these poorly performing genes (*CDKN2A*, *FGFR3*, *RB1*, *STK11*, and *VHL*) were not high-priority targets for our clinical practice and were excluded from the validation. These genes were designated for research use only. *HNF1A* was also excluded and designated for research use only, even though coverage was acceptable, because it contained a C_8 sequence that sequenced poorly. These exclusions reduced the number of amplicons from 212 to 189 and brought the number of reported genes to 42. Although coverage was also low for *HRAS*, this gene was retained in our assay because of its potential clinical importance.

For the remaining 42 genes, hereafter referred to as the *pan-cancer mutation panel*, coverage statistics relevant to the chosen frequency filters of 10%, 7.5%, 5%, and 2.5% were compiled to examine the proportion of amplicons yielding a minimum number of reads. The total coverage per sample ranged from 428 551 to 2 103 971 reads. Average coverage per amplicon across the 100 samples ranged from 197 to 11 724 reads (Figure 1). (Note that average coverage in run 1, which included 16 samples, was approximately 2-fold higher than in runs 2–4, which each included approximately 40 samples [Supplemental Table S3]). As seen in Table 1, the vast majority (96.0%–99.9%) of the amplicons met each minimum coverage criterion (100×, 134×, 200×, and 400×, respectively). This was true for all 3 specimen types, FFPE, FNA, and PB/BM. The generally higher numbers of amplicons with low read count for PB/BM specimens were likely more related to the fact that most of these samples were grouped on the same run rather than a true property of the specimen type. For unknown reasons, run 4 had overall poorer coverage for a small subset of amplicons than the preceding 3 runs.

Sensitivity

To examine the ability of the assay to detect low-frequency variants, dilutions of a sample containing the *TP53* variant c.844C>T (p.R282W) at a variant frequency close to 100% were performed, targeting final variant frequencies of 1.2%, 2.5%, 3.7%, and 5.0%, respectively. As seen in Table 2, the variant was detected in all replicates of each dilution. In a second experiment, normal control samples were used to assess sensitivity. Two normal control samples of white and Asian origins, respectively, were mixed to obtain expected percentages of a heterozygous SNP (EGFR rs1050171) present in the white sample but not in the Asian sample. Three dilutions, targeted to 10%, 5%,

Table 1. Coverage Statistics

Minimum Coverage[a]	All Samples (n = 18 900), No. (%)	PB/BM (n = 7938), No. (%)	FFPE (n = 6615), No. (%)	FNA (n = 4347), No. (%)
100×	18 820 (99.6)	7871 (99.2)	6610 (99.9)	4339 (99.8)
134×	18 786 (99.4)	7842 (98.8)	6607 (99.9)	4337 (99.8)
200×	18 693 (98.9)	7768 (97.9)	6592 (99.7)	4333 (99.7)
400×	18 442 (97.6)	7622 (96.0)	6523 (98.6)	4297 (98.8)

Abbreviations: FFPE, formalin-fixed, paraffin-embedded; FNA, fine-needle aspirate; PB/BM, peripheral blood/bone marrow.

[a] The number of amplicons with coverage above 100×, 134×, 200×, and 400× are shown for all samples and separately for the each specimen type as a fraction of the total number of amplicons. The corresponding percentage of amplicons that met each coverage criterion is shown in parentheses.

Table 2. Limit-of-Detection Analysis			
Target 1,[a] %	Replicate 1, %	Replicate 2, %	Replicate 3, %
5.0	5.8	3.9	5.1
3.7	2.4	3.9	4.5
2.5	3.0	2.8	2.6
1.2	0.9	1.1	1.1
Target 2,[b] %			
10	10.4	9.3	
5	6.0	5.9	
2.5	2.7	3.2	

[a] Dilutions of a clinical sample with the variant *TP53* c844C>T (p.R282W) were processed in triplicate. Target frequencies and the observed variant frequency for each replicate are listed.

[b] Mixes of 2 normal control samples were processed in duplicate, and the observed frequency of an originally heterozygous single-nucleotide polymorphism (dbSNP rs1050171 in *EGFR*) was tracked.

and 2.5%, respectively, were each processed in duplicate. As seen in Table 2, the SNP was detected in all replicates of each dilution, providing additional evidence that the test can detect predefined variants down to at least 2.5%.

At very low variant frequencies, some noise, (ie, false-positive calls because of PCR and sequencing errors, as well as artifacts from FFPE processing) might be expected. To examine background noise in all sample types (FFPE, FNA, PB/BM), we plotted variant frequency versus coverage for all software-called variants in all samples, including those in excluded genes (Supplemental Figure S2). This analysis showed that, in amplicons with at least 200× coverage, the vast majority of the background noise variants were present at less than 5% in all 3 sample types (see below).

Precision

One clinical specimen (FFPE) with a range of variant frequencies was processed in quadruplicate within a single run to assess repeatability. As seen in Table 3, 3 variants at different frequencies were detected in all 4 replicates within a tight range. No other variants were found. There was no change in the data between the 4 sensitivity settings of 10%, 7.5%, 5%, and 2.5%, so these data were applicable to all 4 conditions. An FFPE sample was chosen because it represents a specimen type that typically yields DNA with lower-amplification efficiency than PB/BM does.

Two clinical specimens of each type (PB/BM, FNA, and FFPE, respectively) were predefined for assessment of reproducibility. The first replicate of the samples were present on one of the first 3 runs (with all 3 runs represented) and was compared with a second replicate on run 4.

As seen in Table 4, each variant was seen at similar frequencies in 2 separate runs using a 5% sensitivity setting. At 2.5%, a number of variants were observed that could not be repeated; most of these variants were derived from elevated noise in run 2 (data not shown).

Table 4. Reproducibility			
Specimen	Variant[a]	Replicate 1, %	Replicate 2, %
PB/BM-1	*TP53* p.E198*	86.2	87.7
PB/BM-2	*TP53* p.G245S	43.2	44.5
FNA-1	*NRAS* p.Q61R	9.7	9.6
FNA-2	*KDR* p.G873E	52.2	49.8
FNA-2	*NRAS* p.Q61R	24.1	21.0
FFPE-1	*KRAS* p.G13C	16.8	17.1
FFPE-2	*KRAS* p.G12C	37.0	38.6
FFPE-2	*CSF1R* p.S928_G934del	54.0	60.7

Abbreviations: FFPE, formalin-fixed, paraffin-embedded; FNA, fine-needle aspirate; PB/BM, peripheral blood/bone marrow.

[a] Two samples of each specimen type (PB/BM, FNA, and FFPE) were processed in separate runs. The variant frequency for each replicate is shown.

Concordance With Clinical Sanger Sequencing

Concordance with existing Sanger data was investigated by examining the validation data for mutations in *KRAS*, *BRAF*, *EGFR*, and *NRAS*. FFPE-preserved colon cancer samples are routinely tested for clinically actionable mutations at codons 12 and 13 in *KRAS* and codon 600 in *BRAF*. Thyroid tumor aspirates (FNA) are tested for *BRAF* and *NRAS*. Various *EGFR* mutations, deletions, and insertions (exons 18–21) are tested in lung cancer or metastases. The concordance data were calculated for each specimen type. As seen in Table 5, FFPE specimens with preexisting clinical Sanger data for *KRAS*, *BRAF*, and *EGFR* (n = 31) showed 100% concordance with these data. Similarly, FNA specimens (n = 22) showed 100% concordance with preexisting clinical Sanger data for *NRAS* and *BRAF* (Table 6). At the 10% sensitivity cutoff, 2 of the variants detected by Sanger sequencing were not reported by next-generation sequencing (ie, they were observed at <10% frequency). Sanger sensitivity is highly variable dependent on background noise and sequence context, and these 2 samples had small positive peaks, demonstrating that a test with 10% sensitivity may miss certain Sanger-detectable variants.

Further Verification of Detected Variants

In addition to the concordance analysis with clinical data, we performed Sanger sequencing for *TP53* on all samples that had available material and where a *TP53* mutation was detected by the pan-cancer mutation panel. Because of the inherently lower sensitivity of population-based Sanger sequencing, there was no expectation of being able to confirm low-frequency variants. When only variants detected at greater than 5% frequency were considered, the agreement with Sanger sequencing results was 100% (Table 7).

Table 3. Repeatability					
Variant[a]	Replicate 1, %	Replicate 2, %	Replicate 3, %	Replicate 4, %	Mean (SD), %
APC c.4505G>A (pA1474T)	20.3	19.1	19.9	19.5	19.7 (0.5)
PTPN11 c.179G>T (p.G60V)	34.0	33.8	32.9	30.4	32.8 (1.6)
TP53 c.722C>G (p.G60V)	69.2	68.7	66.9	67.8	68.1 (1.0)

[a] One formalin-fixed, paraffin-embedded specimen was processed in quadruplicate and sequenced within a single run. Three variants were detected across all 4 samples. No other variants greater than 2.5% were detected. The variant frequency for each replicate and calculated means and standard deviations are shown.

Table 5. Concordance in Formalin-Fixed, Paraffin-Embedded Samples[a]

	Pan-Cancer Mutation Panel							
	KRAS				*BRAF*		*EGFR* Exons 18–21	
Sanger Sequencing	G12V	G12D	G13D	WT	V600E	Wild Type	Positive	Wild Type
KRAS								
G12V	4							
G12D		2						
G13D			3					
Wild type				10				
BRAF								
V600E					2			
Wild Type						7		
EGFR exons 18–21								
Positive							0	
Wild type								10

[a] A 2 × 2 comparison with clinical Sanger sequencing results is shown for the 31 formalin-fixed, paraffin-embedded specimens.

A subset of specimens was additionally sequenced on a PGM using the Ion AmpliSeq Cancer Hotspot v2 panel to provide a comparison with an alternate technology. Although both technologies fall in the category of next-generation sequencing, the error profiles of the 2 platforms, MiSeq and PGM, are very different because of the chemistries employed.[18] The AmpliSeq panel amplifies segments of the same genes as the pan-cancer samples as well as 2 additional genes, but the exact positions covered in some cases differ between the 2 panels. Nine FFPE clinical specimens were chosen for sequencing with the AmpliSeq panel for concordance analysis. Note that the concordance was not expected to be 100% because of differences in primer design for the 2 panels.

At the 10%, 7.5%, and 5% sensitivity levels, 38 variants were identified in the 9 specimens by the pan-cancer panel test. Thirty-one of these positions (82%) were also covered by AmpliSeq panel amplicons. Thirty of the thirty-one variants were recapitulated in the AmpliSeq panel data (97% confirmatory outcome). At the 2.5% sensitivity level, 40 mutations were identified by the pan-cancer panel test. The 2 additional mutations were not covered by the AmpliSeq panel, and the outcome remained at 97%. The observed frequencies of the 30 variants that were detected by both methodologies are shown in Figure 2. Pan-cancer panel sequencing of the single discordant specimen at the 5% sensitivity level (variant *PDGFRA* c2360 T>A at 5.25%) was repeated in the final validation run, but the variant call was not reproduced. Analysis of the common SNPs in both samples ruled out sample swap with high confidence. It is likely that the original variant call resulted from a mutation generated during probe extension or PCR (see "Comment"). Separate analysis of the PGM data gave very similar results to those seen with MiSeq. Four mutations affecting protein sequence were found in the PGM data that were not identified using MiSeq (not shown). In all cases, the relevant codon was in a region not covered by the TruSeq Amplicon Cancer Panel. (No mutations in 3 of these codons are present in the COSMIC database[17]; the remaining codon is represented by 16 entries in COSMIC.)

Insertion and deletion variants were identified in 10 samples (Supplemental Table S1). Sufficient material remained for specimens 3, 12, 41, and 53, and primers were designed to amplify the regions of interest. The deletions in these 4 samples, ranging from 1 to 21 bases, were all confirmed by Sanger sequencing of the generated amplicons (data not shown). Additionally, because the validation set did not contain any specimens with clinically relevant *EGFR* deletions, we included 2 such specimens in a subsequent run and successfully detected the p.E476_A750delELREA mutation (Supplemental Figure S3).

Table 6. Concordance in Fine-Needle Aspirate Samples[a]

	Pan-Cancer Mutation Panel				
	NRAS			*BRAF*	
Sanger	Q61K	Q61R	Wild Type	V600E	Wild Type
NRAS					
Q61K	2				
Q61R		4			
Wild type			11		
BRAF					
V600E				8	
Wild type					14

[a] A 2 × 2 comparison with clinical Sanger sequencing results is shown for the 22 fine-needle aspirate specimens. Note that some samples contain more than one mutation.

Analysis of 64 Validation Samples

Using the 5% sensitivity cutoff established above, we detected 123 variants in the 64 validation samples (Supplemental Table S1). Of these, 108 (88%) were convincing and could be reported out clinically, whereas 13 (11%) appeared to be artifacts, and 2 (2%) would have to be reproduced to be called with confidence. All potential artifact variants were present at a frequency of 6.8% or less. The Somatic Variant Caller software assigns quality scores (maximum, 100) for all variants, and 8 (of 13; 62%) of the likely artifact variants were the only ones in our analysis with a score less than 100 (Supplemental Table S1). Furthermore, 8 (62%) of the artifactual variants were found in just 2 samples (samples 43 and 54). Both of these samples only marginally met our quality control criteria, and it is likely that poor sample quality was responsible for the presence of artifacts (see "Comment"). Two other questionable variants were detected in *HRAS*, which performs poorly in library preparation (see above). Two of the other artifactual variants were identical (MPL p.W515*) and involved an apparent sequencing artifact that we have seen in many samples in

Table 7. Confirmation of *TP53* Variants by Sanger Sequencing[a]

Specimen	Gene	Nucleotide Variant	Amino Acid Variant	Detected by Sanger	Pan-Cancer Panel Variant Frequency, %
PB/BM-1	*TP53*	c.758C>T	p.T253I	No	3.7
PB/BM-2	*TP53*	c.758C>T	p.T253I	No	4.6
FFPE-1	*TP53*	c.713C>A	p.C238F	Yes[b]	8.5
FFPE-2	*TP53*	c.847G>A	p.R283C	Yes	20.4
FFPE-2	*TP53*	c.839C>G	p.R280T	Yes	33.7
FFPE-3	*TP53*	c.437C>T	p.W146*	Yes	34.9
FFPE-4	*TP53*	c.524C>T	p.R175H	Yes	35.3
PB/BM-2	*TP53*	c.733C>T	p.G245S	Yes	43.2
FFPE-2	*TP53*	c.569G>A	p.P190L	Yes	55.2
FFPE-5	*TP53*	c.523G>A	p.R175C	Yes	59.9
FFPE-6	*TP53*	c.817G>A	p.R273C	Yes	61.1
FFPE-5	*TP53*	c.818C>T	p.R273H	Yes	61.8
FFPE-7	*TP53*	c.428A>C	p.V143G	Yes	62.7
PB/BM-1	*TP53*	c.722G>C	p.S241C	Yes	65.7
FFPE-8	*TP53*	c.524C>T	p.R175H	Yes	74.9
FFPE-9	*TP53*	c.839C>A	p.R280I	Yes	76.1
FFPE-10	*TP53*	c.475C>G	p.A159P	Yes	84.2
PB/BM-3	*TP53*	c.592C>A	p.E198*	Yes	86.2
FFPE-11	*TP53*	c.481C>T	p.A161T	Yes	88.4
FFPE-12	*TP53*	c.523G>C	p.R175G	Yes	94.1
PB/BM-4	*TP53*	c.844G>A	p.R282W	Yes	96.3

Abbreviations: FFPE, formalin-fixed, paraffin-embedded; PB/BM, peripheral blood/bone marrow.

[a] Four PB/BM and 12 FFPE specimens with *TP53* variants detected by the Pan-Cancer Mutation Panel were sequenced with a Sanger TP53 assay.
[b] Peak only slightly above background.

different sequencing runs. These results suggested that a few artifactual variants may still be identified using the 5% sensitivity cutoff. Therefore, it is important to exercise caution when analyzing variants with low coverage, low quality scores, or with frequencies just above 5%, particularly those in poorly performing amplicons or in lower-quality samples (see "Comment").

COMMENT

Next-generation sequencing enables parallel processing of samples for high-sensitivity detection of mutations in multiple genes. Our clinical laboratory performs weekly processing of variable numbers of specimens by Sanger sequencing for several genes. Additional, low-volume, mutational analyses in less-commonly requested genes are sent to commercial reference laboratories. We found it attractive to streamline processing of all these specimens with a single assay. Such batch processing allows for efficient wet-laboratory time management and a more-convenient reporting system because all samples are processed simultaneously and identically.

We validated a pan-cancer mutation panel based on Illumina's TruSeq Amplicon Cancer Panel by sequencing 68 clinical samples and normal controls. DNA extracted from different sources of input material (FFPE, FNAs, PB/BM) was processed. As depth of sequencing and sensitivity is intimately related, we analyzed the data at 4 different combinations of these variables (Table 8), all based on a requirement for a minimum of 10 reads that contain the variant sequence. Both factors also contribute to noise in next-generation sequencing, but the sensitivity increases with sequencing coverage until a lower barrier caused by cross-linking artifacts in FFPE specimens and PCR error from the library preparation is reached.[19,20] At the 2.5% sensitivity level, predefined variants could still be comfortably detected, but background noise in one of the 4 runs decreased reproducibility of other variants when specimens were compared between runs. Thus, the test was validated to 5% sensitivity with a requirement for at least 10 reads that contain the variant allele. Therefore, at least 200 total reads must be present to call a variant at 5%. Using these cutoffs, we found 100% repeatability and reproducibility when tested directly in a small number of samples (Tables 3 and 4). Concordance with existing clinical Sanger sequencing was 100%. Typical Sanger sensitivity is about 15%, but, in some cases, variants at lower frequencies can be detected depending on run quality and sequence context.[21] When 9 samples were tested by sequencing with an alternate next-

Table 8. Summary of Validation Data[a]

	400× Coverage/ 2.5% Sensitivity	200× Coverage/ 5% Sensitivity	134× Coverage/ 7.5% Sensitivity	100× Coverage/ 10% Sensitivity
Coverage	97.6% > 400×	98.9% > 200×	99.4% > 134×	×99.6% > 100×
Reproducibility and repeatability	**Poor**	100	100	100
LOD (Sensitivity)	Pass	Pass	Pass	Pass
Concordance/Sanger detectable variants	100	100	100	**<100**[b]
Confirmation by alternate NGS technology	96.8	96.8	96.8	96.8

Abbreviations: LOD, limit of detection; NGS, next-generation sequencing.

[a] Results from each test are shown for each filter setting. Results with unacceptable false-positive or false-negative levels are shown in bold type.
[b] When the sensitivity filter was limited to 10%, 2 of the variants detected by clinical Sanger sequencing would have been missed, reflecting the varied sensitivity of Sanger sequencing by position depending on run quality and sequence context.

generation sequencing technology,[18] one of 31 variants (3%) was not detected. This variant also was not reproduced when the sample was resequenced using MiSeq. It is likely that the original variant call was artifactual and resulted from poor sample quality.

The limiting factor on the sensitivity of our assay was noise in the sequence data (ie, artifactual sequence variants that appear in sequencing runs but that are not derived from the sample genome) (Supplemental Figure S2). Noise variants that appeared at significant levels (>2.5%) could be grouped into 2 main types. One type appeared to be caused by consistent sequencing errors at specific DNA sequences. These variants were seen repeatedly in multiple samples and in different runs, although their frequencies varied from run to run. Because sequencing errors are often strand-specific, this type of noise was less of a problem in regions of bidirectional coverage. We are currently assembling a database of common sequencing errors so that they can be filtered out before analysis. The second type of noise appeared to correspond to mutations generated during the probe extension and PCR steps of library preparation. These variants were most frequent in amplicons with low coverage and were different between samples and even among different runs of the same sample. It is possible for such mutations to be present at high frequency (>2.5%) only if very few of the almost 10^5 sample-derived copies of the gene present in the library preparation step are actually used as templates in the amplification. This would allow a mutation from a single nucleotide misincorporation at an early step to become overrepresented in the final library. This type of noise was an issue for the genes that were excluded from our validation and was also seen when DNA samples that did not meet our quality control criteria were used in the assay (not shown). However, because these noise variants differed between runs, it was possible to distinguish between noise and genuine variants simply by running the sample in duplicate: noise variants changed between duplicates, but genuine mutations were present in both (not shown). In clinical practice, we will sequence marginal quality specimens in duplicate to distinguish genuine mutations from noise. Some amplicons consistently showed lower levels of both of these types of noise, so it is possible that, in the future, different sensitivity cutoffs could be used for different amplicons.

The sensitivity of the assay did not depend on the sample type. In theory, sample types could differ either in coverage or in noise levels. Coverage for FFPE samples was comparable to that for FNA and PB/BM samples (Table 1). Furthermore, a known mutation in an FFPE sample was detected at the expected frequency when it was diluted into a normal PB sample, indicating that the FFPE DNA was not underrepresented in the library, even when amplified in competition with PB DNA. The FFPE samples also did not exhibit higher levels of noise above the 5% limit than did FNA or PB/BM samples (Supplemental Figure S2).

Different cancer types have been found to be more frequently associated with mutations in particular genes, which underlies the fundamental idea of personalized medicine. To assess whether the detected mutations conformed to expectations based on prior research, we grouped specimens by diagnosis and compared the results to well-known, disease-dependent mutation profiles. A summary of the mutation profile and the 6 cancer types covered by the validation specimens can be found in Supplemental Table S1.

Recurrently mutated genes identified at the 5% sensitivity level are among the most frequently mutated genes for the respective cancer types according to the COSMIC database.[17] For instance, in CRC specimens (n = 19), recurrently mutated genes identified included *APC* (n = 14; 74%), *TP53* (n = 11; 58%), *KRAS* (n = 11; 58%), *PIK3CA* (n = 4; 21%), *BRAF* (n = 2; 11%), and *SMAD4* (n = 2; 11%). Somatic mutations causing loss of function of *APC* have been reported in most sporadic CRCs, indicating a role in initial clonal expansion.[22] The prognostic significance of *APC* in CRC is currently unclear. A study among Taiwanese patients with CRC indicated that *APC* mutation carriers had lower rates of clinical survival after 5-fluorouracil adjuvant chemotherapy[23]; *TP53* mutations have been found in 43% of CRC cases,[17] with the prognostic value being complex and dependent on the primary tumor site.[24] Hot-spot mutations in *KRAS* (codons 12 and 13) and *BRAF* (V600E) are routine clinical testing targets before prescribing anti-EGFR antibody for CRC. *PIK3CA* mutations have been reported to be associated with poor prognosis among patients with curatively resected colon cancer.[25] In a recent study of 964 patients with CRC, it was found that aspirin therapy was associated with longer survival among patients with mutated *PIK3CA* by down-regulating PIK3CA signaling activity.[26] *SMAD4* mutations have been found in 14% of CRC cases and indicate poor prognosis.[27]

A major remaining challenge for clinical next-generation sequencing is to accurately classify each variant according to its function in disease. Gene panels with a limited disease association scope, such as the one presented here, contain only genes with known connections to the condition(s) of interest. This makes it less burdensome, but by no means easy, to classify variants. In some cases, information can be gleaned from references collected in databases such as COSMIC, and in other cases, the annotation can be inferred from basic knowledge of the protein functional domains and pathways. However, in many instances, the functional effect of a given amino acid change will remain inconclusive until more information becomes available.[1] To facilitate the physical process of variant assessment and reporting, we built an in-house, PHP-MySQL–based, software application designated ClinMut Reporter that was validated as part of the pan-cancer mutation panel described here. This application extracts information on variant identity, position, frequency, and coverage from the instrument generated alignment files and removes synonymous variants as well as common SNPs with a minor allele frequency greater than 1% in the general population. The remaining mutations are displayed along with annotations, such as COSMIC ID and the functional change, that is, missense, frameshift, among others. Laboratory directors can further classify each variant according to predicted pathogenicity based on available literature. After review and approval by the responsible pathologist, a report can be instantly generated for each patient's clinician that further details information about the mutated genes, protein pathways, disease links, and ongoing clinical trials related to the detected variations.

Several multicenter studies have shown sufficient robustness of even highly complex next-generation sequencing assays for clinical use.[28,29] Coupled with multiplex library preparation and/or automation, the complexity is drastically reduced, and especially the smaller platforms are making

rapid headway into clinical laboratories. The clonal nature of next-generation sequencing allows for increased sensitivity as well as the ability to phase *cis/trans* variants. Increased sensitivity is critical for applications such as detecting somatic variants in mixed tissues,[19] tracking minimal residual disease for lymphocyte clonalities,[30] and finding drug-resistant retroviral quasispecies.[31] Phasing is crucial to experiments involving 16S and human leukocyte antigen, but resolution at the molecular level can also give important indicators of pathogenicity when 2 deleterious variants can be shown to occur on separate alleles. In the present data set, we could distinguish 2 truncating *TP53* mutations in 2 distinct read populations from the same specimen (data not shown). Although it is possible that the 2 mutations occurred in separate subpopulations, it is likely that both copies of this important tumor suppressor gene were affected.

The parallelization inherent in next-generation sequencing means that, for the technology to be cost effective, a threshold number of specimens must be collected. To decrease turnaround times for the processing of clinical specimens, we intend to also validate specific combinations of assays on a single instrument run, which will use the large capacity and minimize the cost per sample. This can be easily accomplished by assigning specific groups of molecular barcode indices to each assay ensuring that the bioinformatics pipeline can correctly sort and analyze the data.

Multiplex generation of amplicons, such as the extension/ligation process underlying both TruSeq technology and molecular inversion probes,[32] greatly reduces labor time and the potential for human error, both typical pain points for clinical laboratory operations. However, because such multiplex libraries are created in a single reaction, all amplicons must be generated under identical PCR conditions in the amplification step. This leads to a bias against difficult regions, for example, those with high GC content. Most of the low coverage amplicons that we detected had more than 64% GC content (Supplemental Table S3; Supplemental Figure S4), suggesting that high GC content was a factor in poor amplicon performance. In the case of TruSeq probes, assay development is further hampered by a need to resynthesize the entire pool of oligonucleotides to replace a single probe. The cost of a large, custom probe library makes it prohibitive to perform empirical optimization. As described in this report, even using a commercially available probe set in our hands yielded suboptimal results for several amplicons. For our future assays, it is likely that an intermediate strategy involving several different library amplifications per sample may be needed, or, alternatively, a combination of probe library and standard PCR amplicons to be pooled before the next-generation sequencing step.

In conclusion, we have successfully validated a next-generation sequencing–based pan-cancer panel that provided sensitive detection of variants across 42 genes. This powerful approach to identifying alternative therapies to molecular targets can be effectively implemented in specialized clinical laboratories.

This project was funded, in part, by Thomas Jefferson University Hospital, the Division of Genomic Pathology, the Department of Pathology, Anatomy, & Cell Biology, Thomas Jefferson University, and a grant with the Pennsylvania Department of Health. The Pennsylvania Department of Health specifically disclaims responsibility for any analyses, interpretations, or conclusions.

References

1. Rehm HL, Bale SJ, Bayrak-Toydemir P; Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med*. 2013;15(9):733–747.

2. Gargis A, Kalman L, Berry M, et al. Assuring the Quality of next-generation sequencing in clinical laboratory practice. *Nat Biotechnol*. 2012;30(11):1033–1036.

3. Grossmann V, Roller A, Klein HU, et al. Robustness of amplicon deep sequencing underlines its utility in clinical applications. *J Mol Diagn* 2013;15(4):473–484.

4. McCourt CM, McArt DG, Mills K, et al. Validation of next generation sequencing technologies in comparison to current diagnostic gold standards for BRAF, EGFR, and KRAS mutational analysis. *PLoS One* 2013;8(7):e69604. doi:10.1371/journal.pone.0069604.

5. Singh RR, Patel KP, Routbort MJ, et al. Clinical validation of a next-generation sequencing screen for mutational hotspots in 46 cancer-related genes. *J Mol Diagn*. 2013;15(5):607–622.

6. Luthra R, Patel KP, Reddy NG, et al. Next-generation sequencing-based multigene mutational screening for acute myeloid leukemia using MiSeq: applicability for diagnostics and disease monitoring. *Haematologica*. 2014;99(3):465–473.

7. 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*. 2011;2(1):82–93.

8. Beadling C, Neff TL, Heinrich MC, et al. Combining highly multiplexed PCR with semiconductor-based sequencing for rapid cancer genotyping. *J Mol Diagn*. 2013;15(2):171–176.

9. Cheng YD, Yang H, Chen GQ, Zhang ZC. Molecularly targeted drugs for metastatic colorectal cancer. *Drug Des Devel Ther*. 2013;7:1315–1322.

10. Hoyle M, Peters J, Crathorne L, et al. Cost-effectiveness of cetuximab, cetuximab plus irinotecan, and panitumumab for third and further lines of treatment for KRAS wild-type patients with metastatic colorectal cancer. *Value Health*. 2013;16(2):288–296.

11. Ravnan MC, Matalka MS. Vemurafenib in patients with BRAF V600E mutation-positive advanced melanoma. *Clin Ther*. 2012;34(7):1474–1486.

12. Sakai K, Horiike A, Irwin DL, et al. Detection of epidermal growth factor receptor T790M mutation in plasma DNA from patients refractory to epidermal growth factor receptor tyrosine kinase inhibitor. *Cancer Sci*. 2013;104(9):1198–1204.

13. Ascierto PA, Schadendorf D, Berking C, et al. MEK162 for patients with advanced melanoma harbouring NRAS or Val600 BRAF mutations: a non-randomised, open-label phase 2 study. *Lancet Oncol*. 2013;14(3):249–256.

14. Robles AI, Harris CC. Clinical outcomes and correlates of TP53 mutations and cancer. *Cold Spring Harb Perspect Biol*. 2010;2(3):a001016. doi:10.1101/cshperspect.a001016.

15. Barnett DW, Garrison EK, Quinlan AR, Strömberg MO, Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*. 2011;27(12):1691–1692.

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. Forbes SA, Tang G, Bindal N, et al. COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res*. 2010;38(database issue):D652–D657. doi:10.1093/nar/gkp995.

18. Mardis ER. Next-generation sequencing platforms. *Annu Rev Anal Chem (Palo Alto Calif)*. 2013;6:287–303. doi:10.1146/annurev-anchem-062012-092628.

19. Thomas RK, Nickerson E, Simons JF, et al. Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nat Med* 2006;12(7):852–858.

20. Milbury CA, Correll M, Quackenbush J, Rubio R, Makrigiorgos GM. COLD-PCR enrichment of rare cancer mutations prior to targeted amplicon resequencing. *Clin Chem*. 2012;58(3):580–589.

21. Thiatis AC, Norris-Kirby A, Rich RG, et al. Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of *KRAS* mutations. Diagnostic and clinical implications. *J Mol Diagn*. 2010;12(4):425–432.

22. Fodde R. The APC gene in colorectal cancer. *Eur J Cancer*. 2002;38(7):867–871.

23. Chen SP, Wu CC, Lin SZ, et al. Prognostic significance of interaction between somatic APC mutations and 5-fluorouracil adjuvant chemotherapy in Taiwanese colorectal cancer subjects. *Am J Clin Oncol*. 32(2):122–126.

24. Russo A, Bazan V, Iacopetta B, et al; TP53-CRC Collaborative Study Group. The *TP53* colorectal cancer international collaborative study on the prognostic and predictive significance of p53 mutation: influence of tumor site, type of mutation, and adjuvant treatment. *J Clin Oncol*. 2005;23(30):7518–7528.

25. Ogino S, Nosho K, Kirkner GJ, et al. *PIK3CA* mutation is associated with poor prognosis among patients with curatively resected colon cancer. *J Clin Oncol*. 2009;27(9):1477–1484.

26. Liao X, Lochhead P, Nishihara R, et al. Aspirin use, tumor *PIK3CA* mutation, and colorectal-cancer survival. *N Engl J Med*. 2012;367(17):1596–1606.

27. Xie W, Rimm DL, Lin Y, Shih WJ, Reiss M. Loss of *SMAD* signaling in human colorectal cancer is associated with advanced disease and poor prognosis. *Cancer J*. 2003;9(4):302–312.

28. Holcomb CL, Höglund B, Anderson MW, et al. A multi-site study employing high-resolution HLA genotyping by next generation sequencing. *Tissue Antigens*. 2011;77(3):206–217.

29. Kohlmann A, Klein HU, Weissmann, S et al. The Interlaboratory Robustness of Next-generation sequencing (IRON) study: a deep sequencing investigation of *TET2*, *CBL* and *KRAS* mutations by an international consortium involving 10 laboratories. *Leukemia*. 2011;25(12):1840–1848.

30. Boyd SD, Marshall EL, Merker JD, et al. Measurement and clinical monitoring of human lymphocyte clonality by massively parallel V-D-J pyrosequencing. *Sci Transl Med*. 2009;1(12):12ra23. doi:10.1126/scitranslmed.3000540.

31. Simen BB, Simons JF, Hullsiek KH, et al; Community Programs for Clinical Research on AIDS. Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. *J Infect Dis*. 2009;199(5):693–701.

32. Hedges DJ, Guettouche T, Yang S, et al. Comparison of 3 targeted enrichment strategies on the SOLID sequencing platform. *PLoS One*. 2011;6(4):e18595. doi:10.1371/journal.pone.0018595.