The Journal of Molecular Diagnostics, Vol. 15, No. 5, September 2013

the Journal of Molecular Diagnostics
jmd.amjpathol.org

Clinical Validation of a Next-Generation Sequencing Screen for Mutational Hotspots in 46 Cancer-Related Genes

Rajesh R. Singh,* Keyur P. Patel,* Mark J. Routbort,* Neelima G. Reddy,* Bedia A. Barkoh,* Brian Handal,* Rashmi Kanagal-Shamanna,* Wesley O. Greaves,* L. Jeffrey Medeiros,* Kenneth D. Aldape,† and Rajyalakshmi Luthra*

From the Departments of Hematopathology and Pathology,† The University of Texas MD Anderson Cancer Center, Houston, Texas*

Accepted for publication May 3, 2013.

Address correspondence to Rajyalakshmi Luthra, Ph.D., Department of Hematopathology, The University of Texas MD Anderson Cancer Center, 8515 Fannin St., NAO1.046c, Houston, TX 77054. E-mail: rluthra@mdanderson.org.

Transfer of next-generation sequencing technology to a Clinical Laboratory Improvement Amendments–certified laboratory requires vigorous validation. Herein, we validated a next-generation sequencing screen interrogating 740 mutational hotspots in 46 cancer-related genes using the Ion Torrent AmpliSeq cancer panel and Ion Torrent Personal Genome Machine (IT-PGM). Ten nanograms of FFPE DNA was used as template to amplify mutation hotspot regions of 46 genes in 70 solid tumor samples, including 22 archival specimens with known mutations and 48 specimens sequenced in parallel with alternate sequencing platforms. In the archival specimens, the IT-PGM detected expected nucleotide substitutions ($n = 29$) and four of six insertions/deletions; in parallel, 66 variants were detected. These variants, except a single nucleotide substitution, were confirmed by alternate platforms. Repeated sequencing of progressively diluted DNA from two cancer cell lines with known mutations demonstrated reliable sensitivity at 10% variant frequency for single nucleotide variants with high intrarun and inter-run reproducibility. Manual library preparation yielded relatively superior sequencing performance compared with the automated Ion Torrent OneTouch system. Overall, the IT-PGM platform with the ability to multiplex and simultaneously sequence multiple patient samples using low amounts of FFPE DNA was specific and sensitive for single nucleotide variant mutation analysis and can be incorporated easily into the clinical laboratory for routine testing. *(J Mol Diagn 2013, 15: 607–622; http://dx.doi.org/10.1016/j.jmoldx.2013.05.003)*

Oncogenesis is characterized by the accumulation of stable DNA sequence abnormalities in tumors, resulting in deregulation of multiple cell signaling pathways. Detailed profiling of these aberrations in tumors will improve our understanding of the genetic basis of disease and aid in prognostication. Comprehensive information regarding these genetic changes is also valuable in choosing suitable treatment options, a hallmark of personalized cancer therapy, which aims to maximize therapeutic benefits and minimize therapy-associated risks.

In recent years, tremendous improvements in sequencing technology and computational methods have led to the emergence of next-generation sequencing (NGS) platforms that have drastically decreased the time and cost associated with comprehensive genome analysis.[1–3] NGS represents a major departure from so-called first-generation sequencing by allowing whole-genome and whole-exome sequencing along with the flexibility of multiplexing and screening specific panels of genes for mutations. These advances are of high relevance in a clinical diagnostic setting, where mutational screening of genes by first-generation sequencing assays has been largely restricted to singleplex analysis of mutational hotspot regions or entire coding regions of an individual gene. The drawbacks of a single-gene or single-exon approach include high cost, more labor, and slower turnaround time. These limitations have been circumvented by the high multiplexing capacity of NGS platforms, making comprehensive mutational screening of tumors achievable. Furthermore, NGS technologies facilitate screening of multiple genes with limited starting material, a significant advantage over conventional sequencing platforms that require relatively larger DNA quantities.

http://dx.doi.org/10.1016/j.jmoldx.2013.05.003

Currently, several NGS platforms are commercially available for sequencing either targeted genomic regions or whole genomes/exomes to analyze a variety of disease-associated changes, such as point mutations, insertions, deletions, and copy number variations.[3] These NGS applications have facilitated the discovery of novel genetic aberrations in several diseases.[4–7] In addition, NGS technology also has been applied for comprehensive transcriptome (RNA-Seq) analysis,[8,9] for recruitment of chromatin-modifying complexes and transcription factors (ChIP-Seq),[10] and for assessment of epigenetic genome methylation signatures (Methyl-Seq).[11] This broad applicability has established NGS as a valuable tool for mutation screening for routine diagnosis and discovery.

In our Clinical Laboratory Improvement Amendments–certified Molecular Diagnostic Laboratory (The University of Texas MD Anderson Cancer Center, Houston, TX), we tested the clinical applicability of the Ion Torrent Personal Genome Machine (IT-PGM; Life Technologies, Carlsbad, CA), an NGS platform, to routinely and simultaneously screen tumors for mutations in cancer-related genes using the IT AmpliSeq cancer panel genomic library preparation protocol (Life Technologies). This platform performs DNA sequencing-by-synthesis and is based on a nonoptical detection of a hydrogen ion released during the formation of phosphodiester bonds when a nucleotide is incorporated into the course of complementary strand synthesis. The instrument uses an ion-sensitive field-effect transistor for hydrogen ion detection in specially designed semiconductor device interfaced with the sequencing reaction. The change in hydrogen ion concentration is detected and recorded as change in voltage. This technique has been used recently to successfully sequence three bacterial genomes and one human genome.[12] Additional disparate applications include mutation screening in hereditary diseases, such as cystic fibrosis,[13] and characterization of microbial genomes.[14,15] Comparisons of the IT-PGM platform with similar benchtop sequencers also have been reported recently.[16,17]

The IT AmpliSeq cancer panel is a multiplex PCR-based library preparation method by which 190 regions (70-150 bp) that encompass 740 mutational hotspots in the coding sequence of 46 cancer-related genes are selectively amplified. In our high-volume clinical laboratory, we validated the AmpliSeq cancer panel using a variety of tumor types. Mutations detected by IT-PGM platform were confirmed by traditional mutation detection assays, such as Sequenom MassARRAY (Sequenom Inc, San Diego, CA) or Sanger sequencing. We show that the IT-PGM platform is sensitive and specific and can be used for routine mutational screening of patient tumors in a molecular diagnostics laboratory.

Materials and Methods

Tumor Samples

Solid tumor samples from 70 patients were analyzed, including melanoma ($n = 36$); colorectal ($n = 16$), lung ($n = 5$), gastrointestinal tract ($n = 5$), papillary thyroid ($n = 4$), and endometrial serous ($n = 3$) adenocarcinomas; and squamous cell carcinoma ($n = 1$). Twenty-two samples with known mutations were selected from the archives of our laboratory. These 22 samples had been sequenced using traditional methods before sequencing using IT-PGM. An additional 48 cases were analyzed by IT-PGM in parallel with Sequenom MassARRAY-based custom-designed 11-gene mutation detection assay. Mutations detected by IT-PGM that were not covered by the Sequenom MassARRAY assay were confirmed by Sanger sequencing.

DNA Extraction

DNA was extracted from formalin-fixed, paraffin-embedded (FFPE) tumor samples. Unstained 4-μm-thick tissue sections were deparaffinized, and tumor tissue was manually microdissected using as a guide an H&E-stained slide from the same block. H&E-stained tissue sections were reviewed by pathologists on clinical service in the Department of Pathology, who circled the tumors and indicated the tumor percentage. Only specimens with >20% tumor in the circled area were included in the study. DNA was extracted using a PicoPure DNA extraction kit (Arcturus, Mountain View, CA) and was purified using an Agencourt AMPure XP kit (Agencourt Biosciences, Beverly, MA). A Qubit DNA high-sensitivity assay kit (Life Technologies) was used to quantify purified DNA.

Library Preparation

Library preparation for each sample was performed using the IT AmpliSeq 2.0 Beta kit and IT AmpliSeq cancer panel primers (Life Technologies) following the manufacturer's instructions. Briefly, 10 ng of DNA was used as a template to generate the amplicon library for sequencing hotspot mutations in 46 genes. The genes included in this panel are *AKT1, BRAF, FGFR1, GNAS, IDH1, FGFR2, KRAS, NRAS, PIK3CA, MET, RET, EGFR, JAK2, MPL, PDGFRA, PTEN, TP53, FGFR3, FLT3, KIT, ERBB2, ABL1, HNF1A, HRAS, ATM, RB1, CDH1, SMAD4, STK11, ALK, SRC, SMARCB1, VHL, MLH1, CTNNB1, KDR, FBXW7, APC, CSF1R, NPM1, SMO, ERBB4, CDKN2A, NOTCH1, JAK3,* and *PTPN11*. To interrogate a mutational hotspot at codon 17 of *AKT1*, we added a primer pair that was designed by Life Technologies at our request to the vendor-provided 190 primer pair pool. Genomic regions to be sequenced were PCR amplified using the 191 primer pairs. Sequencing adaptors with short stretches of index sequences (barcodes) that enabled sample multiplexing were ligated to the amplicons using the IT Xpress barcode adaptor kit (Life Technologies). The library prepared was quantified using the Bioanalyzer high-sensitivity DNA chip (Agilent Technologies Inc, Santa Clara, CA).

Emulsion PCR

The library prepared for each patient sample with distinct barcoding was diluted in nuclease-free water to obtain

a library stock of 160 pmol/L. From this stock, either four patient samples (on a Ion 316 chip) or eight samples (on a Ion 318 chip) were pooled and diluted further to generate a working library concentration of 16 pmol/L. To clonally amplify the library DNA onto the IonSpheres (ISPs; Life Technologies), the library pool was subjected to emulsion PCR (E-PCR) by following either a manual method using the Ion Xpress template kit (Life Technologies) or an automated method using an IT OneTouch template kit on an IT OneTouch system (Life Technologies) following the manufacturer's protocol. After manual E-PCR, the ISPs were isolated by manual breaking of the emulsion as per the manufacturer's instructions. Breaking of the emulsion and isolation of ISPs was automated using an Ion OneTouch system. As a measure of the efficiency of the E-PCR, estimation of the percentage of ISPs with DNA in the background of blank ISPs was performed using the Qubit Ion Sphere quality control kit (Life Technologies). Enrichment of ISPs, which involves selective isolation of ISPs with clonally amplified DNA, was achieved using the IT OneTouch ES (Life Technologies) and the IT OneTouch kit following the manufacturer's protocol.

Sequencing Using IT-PGM

Enriched ISPs were subjected to sequencing on a Ion 316 Chip or a Ion 318 Chip to sequence pooled libraries with four or eight samples, respectively. Because IT-PGM provides the flexibility of using these two chips of different capacities for scaling, we tested the performance of both chips during validation. Sequencing was performed using the sequencing kit (Life Technologies) as per the manufacturer's instructions. A cutoff point of 300,000 reads with a quality score of AQ20 (one misaligned base per 100 bases) was used as a measure of successful sequencing of a sample. For a sequence variant to be considered authentic, sequencing coverage of 250× and a variant frequency of at least 10% in the background of wild type were used as minimum requirements in this study. The cutoff point of 10% was chosen to facilitate the confirmation of mutations discovered by IT-PGM by alternate sequencing platforms available in our laboratory. The mutational status of amplicons that failed to attain the minimum coverage of 250× was recorded as indeterminate.

Sensitivity Analysis

The sensitivity of the IT-PGM platform for mutation detection was determined by sequencing serially diluted DNA from two human cancer cell lines: H2122 (CRL-5985; ATCC, Manasass, VA) and DLD1 (CCL-221; ATCC) fixed in formalin and embedded in paraffin. For H2122, DNA was diluted into DNA from FFPE HL60 (CCL-20; ATCC) in ratios of 1:4, 1:9, and 1:19 (H2122:HL60), resulting in 20%, 10%, and 5% dilutions, respectively. Similarly, DLD1 DNA was serially diluted into DNA from paraffin-embedded H460 (HTB-177; ATCC) in ratios of 1:3, 1:9, 1:19, and 1:39, resulting in 25%, 10%, 5%, and 2.5% dilutions of DLD1 DNA, respectively. Confidence Intervals were estimated using the efficient score method (corrected for continuity).[18]

Inter-run and Intrarun Assay Reproducibility

Inter-run assay reproducibility was assessed by sequencing a patient sample indexed with different barcodes randomly across 25 independent multiplexed sequencing runs. To test the intrarun assay reproducibility, libraries prepared from a patient DNA sample indexed with 10 different barcodes were multiplexed and sequenced on the same IonChip.

Confirmation of Mutations by Sanger Sequencing and the Sequenom MassARRAY System

Sanger sequencing assays were designed to cover all the genomic regions interrogated by the AmpliSeq cancer panel to serve as a confirmatory assay for mutations detected by IT-PGM. Regions of interest for Sanger sequencing were PCR amplified using forward and reverse primers tagged

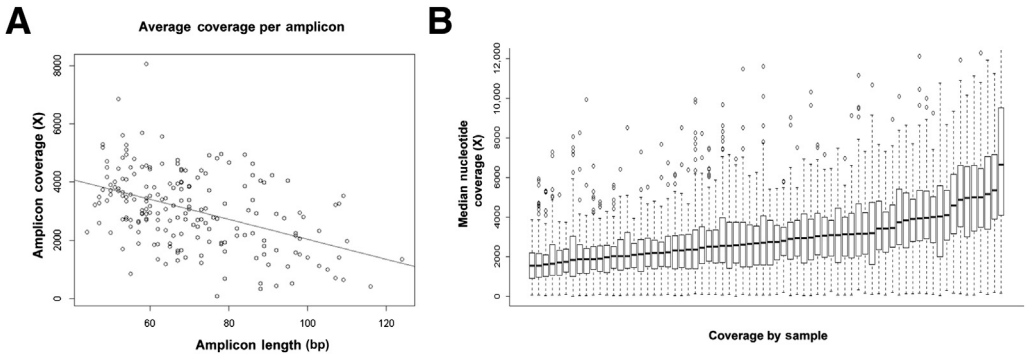


Figure 1 Sequencing efficiency on the IT-PGM platform as assessed by sequencing coverage of amplicons. **A:** Average coverage (including forward and reverse strands) per amplicon during sequencing of a 70-sample set used in the study showed high variability with amplicon size, but a trend toward inverse correlation of overall sequencing coverage to the length of the amplicon was observed. **B:** Coverage of each nucleotide (including both directions) in all the amplicons per sample showed median coverage of >1000×, indicating efficient sequencing in each sample.

Table 1 List of Amplicons that Consistently Failed in ≥20% of Samples (Count of ≥14) to Reach 250× Coverage Along with the Hotspots as per COSMIC Database Interrogated

Amplicon	Gene	No. of samples failed	Codons covered	Codons with COSMIC hotspots (recurrent mutation count ≥5)
AMPL508777	*FGFR3*	70	248–268	248, 249
AMPL483802	*CDKN2A*	56	52–75	57, 58, 61, 69
AMPL341576	*NOTCH1*	53	1567–1604	1575, 1586, 1593, 1594, 1599, 1601
AMPL559669	*RET*	53	610–626	None
AMPL226369	*TP53*	42	150–180	Every codon covered
AMPL495041	*MPL*	41	500–522	505, 515
AMPL178642	*KIT*	35	551–585	557–561, 565,576
AMPL145156	*APC*	27	1431–1466	1447, 1450, 1464
AMPL255119	*RB1*	25	351–370	358
AMPL64101	*SMO*	15	400–417	None
AMPL490549	*STK11*	14	326–359	354

COSMIC, Catalogue of Somatic Mutations in Cancer.

with M13 universal sequences: 5′-TGTAAAACGACGGCCAGT-3′ and 5′-CAGGAAACAGCTATGACC-3′. Primer pairs for amplicons were designed to work under the same PCR conditions to facilitate simultaneous PCR amplification of different genes. PCR was performed using a 10-ng template of DNA, and the amplified products were purified using AMPure magnetic beads (Agencourt Biosciences) according to the manufacturer's protocol. Purified amplicons were diluted (1:5) in water, and 5 μL of diluted DNA was used for Sanger sequencing on a 3730 DNA analyzer (Applied Biosystems, Carlsbad, CA). The resulting data were analyzed by SeqScape version 2.5 or version 2.7 software (Applied Biosystems). For the 9-well, 11-gene mutation hotspot screen developed by our laboratory using the MassARRAY platform, DNA were PCR amplified and subjected to single-base primer extension using the iPLEX Gold kit and were analyzed using MALDI-TOF (matrix-assisted laser desorption/ionization–time of flight) mass spectrometry (Sequenom). Our 9-well multiplex primer extension assay was designed to assess the mutational status of a total of 82 hotspot regions in *AKT1, BRAF, GNAS, GNAQ, IDH1, IDH2, KRAS, MET, NRAS, PIK3CA,* and *RET*. Briefly, the regions of interest were PCR amplified using 10 ng of DNA per well. Subsequently, the PCR products were treated with shrimp alkaline phosphatase to dephosphorylate unincorporated nucleotides. The PCR products were used as templates for locus-specific single-base extension with mass-modified dideoxy nucleotides. Primers for PCR amplification and single-base extension were designed using MassARRAY Assay Design software Typer 4.0 (Sequenom) and were obtained from Integrated DNA Technologies (Coralville, IA). The mass of the products of single-base extension was analyzed by MALDI-TOF for single nucleotide polymorphism detection.

Data Analysis

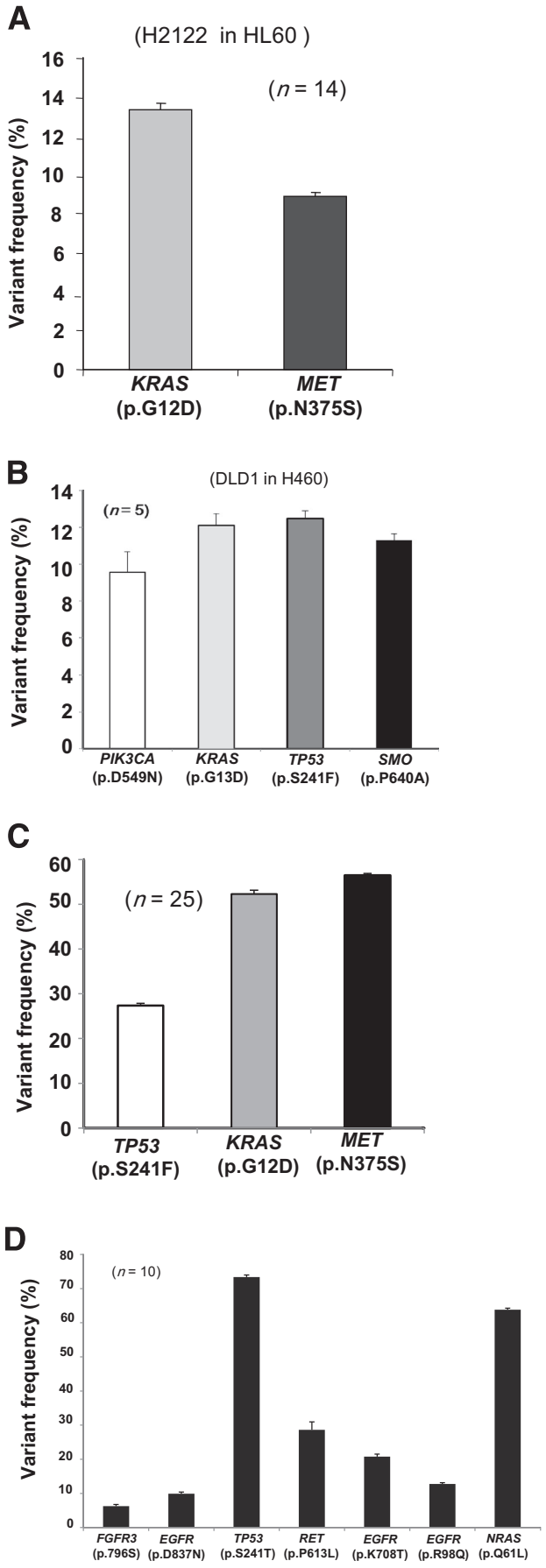
Alignment of the sequences to human genome build 19 reference genome and base calling was performed using Torrent Suite software version 2.0.1 (Life Technologies). Identification of variants was facilitated by IT Variant Caller plugin software version 1.0 (Life Technologies) and Integrative Genomics Viewer (IGV).[19] The IGV was used to visualize the read alignment and the presence of variants against the reference genome and to confirm the veracity of the variant calls by checking for possible strand biases and sequencing errors. A software custom developed by one of the authors (M.J.R.), designated as OncoSeek,[20] was used to interface the data generated by IT Variant Caller with IGV to visualize the alignment and mutation detected in IGV, filter repeat errors due to nucleotide homopolymer regions, compare sequencing replicates, and annotate the sequencing results.

Table 2 Sensitivity Studies with H2122 Cell Line DNA

	KRAS, p.G12D (homozygous)				*MET*, p.N375S (heterozygous)			
	Variant frequency (%)		Coverage (×)		Variant frequency (%)		Coverage (×)	
H2122 in HL60	Tech 1	Tech 2	Tech 1	Tech 2	Tech 1	Tech 2	Tech 1	Tech 2
Undiluted (100%)	99.2	99.2	499	390	52.6	49.5	1697	1302
1:4 (20%)	25.8	25.6	659	543	12.0	13.3	1901	1286
1:9 (10%)	13.6	15.4	684	608	7.3	7.5	1739	1464
1:19 (5%)	11.0	7.0	636	564	4.0	4.3	2131	1401

The variant frequencies and sequencing coverage of mutations detected in sequentially diluted FFPE H2122 cell line DNA samples in two runs by different techs are summarized.

Tech, technician.



Results

Sequencing on IT-PGM and Detection of Variants

All 70 samples tested were sequenced with a minimum of 300,000 reads and an average of 520,961 AQ20 reads, indicating efficient sequencing. Amplicon sequencing coverage for the 70 samples assessed via OncoSeek showed an overall inverse correlation between amplicon coverage and amplicon length (Figure 1A). Analysis of the median coverage for each nucleotide in 191 amplicons across the 70 samples showed median coverage depth of $>1000\times$ for all the samples, indicating very efficient sequencing coverage on the whole. A plot summarizing the median sequencing coverage for the samples tested is provided in Figure 1B. The list of amplicons that consistently failed to sequence with minimum coverage of $250\times$ in $\geq 20\%$ of the samples, the gene, the codons covered, the number of counts when this occurred during the sequencing of the 70 samples, and the mutation hotspots interrogated by them as listed in the Catalogue of Somatic Mutations in Cancer database[21,22] are summarized in Table 1. The overall workflow and the quality control metrics are summarized in Supplemental Figure S1.

Sensitivity Studies

In the serial dilution studies using DNA from the H2122 cell line with two mutations (homozygous *KRAS,* p.G12D and heterozygous *MET*, p.N375S), sequencing performed by two different technicians (N.G.R. and B.A.B.) showed mean $\pm$ SE variant frequencies of 99.2%, 25.7% $\pm$ 0.1%, 14.31% $\pm$ 1.15%, and 9.05% $\pm$ 3.92% for the *KRAS,* p.G12D mutation for undiluted (100%) and 1:4 (20%), 1:9 (10%), and 1:19 (5%) dilutions of H2122 into HL60, respectively. In the case of the *MET,* p.N375S mutation, mean $\pm$ SE variant frequencies of 51.08% $\pm$ 1.54%, 12.6% $\pm$ 0.62%, 7.4% $\pm$ 0.07%, and 4.2% $\pm$ 0.13% were detected for undiluted and 1:4, 1:9, and 1:19 dilutions, respectively (Table 2). Furthermore, a 10% dilution of H2122 sequenced as one of the multiplexed samples in 14 different independent runs showed mean $\pm$ SE variant frequencies of 13.25% $\pm$ 0.4% for *KRAS* p. G12D and

Figure 2 Inter-run and intrarun sequencing reproducibility. **A:** A 1:9 dilution of H2122 in HL60 DNA (10% H2122) was sequenced multiple times ($n = 14$) as a part of multiplexed runs, and the results showed very little variability in the variant frequencies detected for the *KRAS* and *MET* mutations, indicating very reproducible mutation detection sensitivity. **B:** Four heterozygous mutations detected in a 25% (1:3) dilution of DLD1 DNA in H460 DNA showed very little variation in their variant frequencies across five different multiplexed runs. **C:** A patient sample harboring three mutations was sequenced in 25 different multiplexed runs. Consistent detection of the mutations with minimum variation of the variant frequencies showed a high degree of inter-run sequencing. **D:** Intrarun reproducibility was tested by using a single patient DNA as a template and preparing 10 different libraries, each with a distinct barcode. The 10 libraries were multiplexed and sequenced on the same IonChip. The results showed consistent detection of mutations with $\geq 10\%$ variant frequency and coverage $\geq 250\times$. Data are given as means $\pm$ SE.

Table 3 Sensitivity Studies with DLD1 Cell Line DNA

		DLD1 (100%) undiluted	1:3 (25%) (DLD1:H460)		1:9 (10%) (DLD1:H460)		1:19 (5%) (DLD1:H460)		1:39 (2.5%) (DLD1:H460)	
Run No.	Variants	Frequency found (%)	Expected (%)	Found (%)	Expected (%)	Found (%)	Expected (%)	Found (%)	Expected (%)	Found (%)
1	*PIK3CA,* p.D549N	38.1	9.5	8.5	3.8	NC	1.9	NC	0.95	NC
	KRAS, p.G13D	50.5	12.6	11.8	5.0	5.1	2.5	NC	1.26	NC
	TP53, p.S241F	51.3	12.8	8.0	5.1	4.0	2.5	NC	1.28	NC
	SMO, p.T640A	43.9	10.9	10.0	4.3	3.7	2.1	NC	1.09	NC
2	*PIK3CA,* p.D549N	37.6	9.3	9.3	3.7	NC	1.8	NC	0.93	NC
	KRAS, p.G13D	46.8	11.7	15.0	4.6	NC	2.3	NC	1.17	NC
	TP53, p.S241F	48.7	12.1	8.8	4.8	NC	2.4	NC	1.2	NC
	SMO, p.T640A	46.1	11.5	NC	4.6	NC	2.3	NC	1.1	NC
3	*PIK3CA,* p.D549N	32.4	8.1	5.4	3.2	NC	1.6	NC	0.81	NC
	KRAS, p.G13D	50.1	12.5	14.9	5.0	6.1	2.5	NC	1.25	NC
	TP53, p.S241F	51.5	12.8	10.4	5.1	NC	2.5	NC	1.28	NC
	SMO, p.T640A	44.3	11.0	8.5	4.4	NC	2.2	NC	1.10	NC
4	*PIK3CA,* p.D549N	42.3	10.5	11.3	4.2	NC	2.1	NC	1.05	NC
	KRAS, p.G13D	49.6	12.4	13.8	4.9	5.4	2.4	NC	1.24	NC
	TP53, p.S241F	49.5	12.3	9.5	4.9	3.2	2.4	NC	1.23	NC
	SMO, p.T640A	45.2	11.3	9.4	4.5	NC	2.2	NC	1.13	NC
5	*PIK3CA,* p.D549N	41.2	10.3	9.9	4.1	NC	2.0	NC	1.03	NC
	KRAS, p.G13D	45.2	11.3	14.1	4.5	4.7	2.2	NC	1.13	NC
	TP53, p.S241F	48.3	12.0	9.3	4.8	4.2	2.4	NC	1.20	NC
	SMO, p.T640A	46.3	11.5	9.8	4.6	NC	2.3	NC	1.15	NC
Sensitivity (%)		1		0.95		0.40		NA		NA
95% confidence interval (%)		0.79—1.0		0.73—0.99		0.19—0.63		NA		NA

DNA from DLD1 was sequentially diluted into H460 DNA to obtain different levels of four heterozygous mutations. Details regarding the various dilutions used and the expected and detected variant frequencies of the four mutations are summarized. Sensitivity and 95% confidence intervals were calculated for various dilutions and are provided for each dilution over the five runs.

NA, not applicable; NC, not called by variant caller.

8.35% ± 0.2% for *MET,* p.N375S, indicating that the assay sensitivity was reproducible (Figure 2A).

In a second sensitivity assessment, sequentially diluted DNA from an additional cell line, DLD1, with four heterozygous mutations was used. Five separate runs of undiluted DLD1 DNA showed reproducible mean ± SE variant frequencies for the mutations at *PIK3CA,* p.D549N (38.2% ± 1.9%); *KRAS,* p.G13D (48.4% ± 1.1%); *TP53,* p.S241F (49.8% ± 1.4%); and *SMO*, p.T640A (38.2% ± 0.5%) (Table 3 and Figure 2B). At 1:3 (25% DLD1) dilution, *PIK3CA*, *KRAS*, *TP53*, and *SMO* mutations were observed at mean ± SE variant frequencies of 9.57% ± 1.1%, 12.11% ± 0.63%, 12.47% ± 0.43%, and 11.29% ± 0.34% for the expected frequencies of 10.3%, 11.3%, 12.0%, and 11.5%, respectively. The mutation calls were less consistent at 1:9 (10% DLD1) dilution, and no mutations were called in subsequent dilutions tested (1:19 or 5% and 1:39 or 2.5%). This showed that the assay had consistent detection sensitivity at 10% variant frequency (Table 3).

Inter-run Sequencing Performance

The inter-run reproducibility was assessed by including a patient sample with mutations in *TP53*, *KRAS*, and *MET* in 25 different sequencing runs as 1 of the 10 multiplexed samples using Ion 318 chips. The detection and calling of the three mutations in all of the 25 sequencing runs was highly reproducible, with very little variation in mean ± SE detected variant frequencies [*TP53*, p.S241F (27.3% ± 0.5%); *KRAS*, p.G12D (52.2% ± 0.8%); and *MET*, p.N375S (56.4% ± 0.4%) (Figure 2C).

Intrarun Sequencing Performance

To test the intrarun sequencing performance, 10 libraries indexed with different barcodes were prepared using DNA from a single patient with seven mutations in five genes (*FGFR3, EGFR, RET, TP53*, and *NRAS*) at different variant frequencies. The libraries were multiplexed and sequenced on a Ion 318 Chip. Five of the seven mutations were detected reproducibly with minimal variation in mean ± SE variant frequencies [*EGFR,* p.D837N (10.2% ± 1.3%); *TP53,* p.S241T (72.9% ± 2.0%); *EGFR,* p.K708T (20.9% ± 1.8%); *EGFR,* p.R98Q (12.6% ± 1.2%); and *NRAS,* p.Q61L (63.5% ± 1.4%) (Figure 2D). Detection was inconsistent in relation to *FGFR3,* p.796S and *RET,* p.P613L. *FGFR3,* p.796S was present at approximately 6% (<10%) variant frequency and was not called in 3 of 10 samples. Mutation in *RET,* p.P613L was in an amplicon that consistently had sequencing coverage under 250× and was

Table 4 Intrarun Reproducibility

Mutation	Barcoded samples 1	2	3	4	5	6	7	8	9	10	Mean	SE	Sensitivity (%)	95% CI (%)
Variant frequency (%)														
FGFR3, p.P796S	6.95	3.55	NC	6.22	6.55	6.92	6.9	8.15	NC	NC	6.46	1.31	0.70	0.35–0.91
EGFR, p.D837N	10.3	8.8	8.9	8.3	10.7	9.3	12.7	10.4	11.0	12.2	10.27	1.39	1	0.65–1.0
TP53, p.S241T	72.8	73.5	71.7	74.5	70.2	75.9	75.1	73.1	69.1	73.1	72.91	2.00	1	0.65–1.0
RET, p.P613L	25.7	23.6	38	30.6	30.2	NC	36.1	28.9	NC	21.8	29.36	5.31	0.80	0.44–0.96
EGFR, p.K708T	17.4	19	18.5	22.2	23.4	23.1	21.4	21.3	21.5	21.7	20.95	1.89	1	0.65–1.0
EGFR, p.R98Q	13.3	11.8	11.1	13.7	14.6	12.1	12.0	13.5	13.9	10.7	12.67	1.23	1	0.65–1.0
NRAS, p.Q61L	65.6	63.8	62.3	65.1	64.6	61.5	63.1	64.4	63.8	60.9	63.52	1.46	1	0.65–1.0
Coverage (×)														
FGFR3, p.P796S	734	789	620	836	687	1055	632	940	505	649	786.63	142.46		
EGFR, p.D837N	1430	1496	1124	886	1453	1925	1312	1964	975	1426	1399.10	337.05		
TP53, p.S241T	1544	1753	1376	1701	1959	2175	1420	1904	1062	1466	1636.00	310.75		
RET, p.P613L	136	110	121	137	139	149	130	183	87	110	150.63	41.77		
EGFR, p.K708T	1378	1432	1189	1422	1450	1966	1129	2027	891	1128	1401.20	341.90		
EGFR, p.R98Q	1206	1194	981	1486	1137	1487	874	1498	697	908	1146.80	268.55		
NRAS, p.Q61L	3170	3509	3494	4322	3551	3780	2775	3164	2086	2991	3284.20	574.20		

Ten library preparations were made from the DNA of a single patient sample where each was indexed with a distinct barcode. They were multiplexed and sequenced on a single Ion 318 Chip. The variants called, their frequencies, and the respective sequence coverage are summarized. The sensitivity and 95% confidence intervals for the variant detection were also calculated.

CI, confidence interval; NC, not called by Variant Caller software.

not detected and called in 2 of the 10 samples. The variant frequencies of the mutations, respective sequencing coverage, sensitivity, and 95% confidence intervals are summarized in Table 4.

Sequencing of Archival Samples with Known Mutations

A set of 22 archival tumor samples was sequenced using IT-PGM. The criteria for the selection of these samples were i) availability of adequate DNA and ii) presence of mutations detected by other sequencing platforms. The mutations in this set included 29 single nucleotide missense mutations and six insertions/deletions. Sequencing of this sample set using IT-PGM had an average of 447,741 AQ20 reads, indicating adequate sequencing. Each of the expected mutation was clearly detected by the IT-PGM platform with a median coverage depth of 2771×. Details of the tumor type tested, mutations known to be present in these tumors, and their detection using the IT-PGM platform, along with the variant frequency and coverage, are summarized in Table 5.

Representative examples of two mutations detected by IT-PGM and their confirmatory detection on Sequenom MassARRAY analysis and Sanger sequencing are shown in Figure 3. A *BRAF* mutation was identified in one of the melanoma samples tested, which was estimated to be composed of 90% tumor cells. The IT-PGM platform clearly identified a single base pair substitution in codon 600 of *BRAF* [c.1799 A>T, CAC>CTC (− strand, in blue), or GTG>GAG (+ strand, in red)], resulting in substitution of a valine with a glutamic acid residue (*BRAF*, pV600E). The reliability of the mutation call was evident by the high total sequencing coverage of 2326× with comparable mutational frequency on both the strands. The Sequenom MassARRAY platform also detected the presence of this mutation (Figure 3A). Similarly, in another example, a *PIK3CA* mutation (c.1624 G>A), GAA>AAA in codon 542 (p.E542K), was identified by IT-PGM, resulting in the substitution of glutamic acid by lysine at total coverage of 5552× (Figure 3B). This mutation was previously detected by Sanger sequencing in this sample (Figure 3B).

Detection of Insertions and Deletions

Six tumor samples with known insertions or deletions detected previously using other sequencing methods were also analyzed using the IT-PGM platform. Four tumor samples had *KIT* abnormalities, three with deletions in exon 9 and one with an insertion in exon 11, and two tumors had EGFR deletions involving exon 19. The IT-PGM platform successfully identified 3- and 6-bp deletions in exon 11 of *KIT* (Figure 4A). However, a 12-bp deletion in exon 11 and a 6-bp tandem duplication in exon 9 of *KIT* were not recognized by the IT Variant Caller plugin software. Nevertheless, these deletions were present in the sequence information generated by IT-PGM as they were evident in the alignment when viewed through IGV, albeit only in the forward direction (Figure 4, A and B). In the two tumor samples with *EGFR* abnormalities, a 15-bp deletion in exon 19 was successfully detected and called by the IT-PGM platform in both samples (Figure 4C). A summary of expected insertions and deletions and the results of sequencing using the IT-PGM platform with the variant frequencies and sequencing coverage are provided in Table 5.

Table 5 Sequencing of Samples in Retrospect by IT-PGM

Sample No.	Tumor type (%)	Mutations expected	Detected by IT-PGM	Coverage (×)	Variant frequency (%)
1	Colorectal adenocarcinoma (30)	*KRAS*, c.436G>A, p.A146T	Yes	2008	10.76
		MET, c.1124A>G, p.N375S	Yes	3818	49.45
2	Colorectal adenocarcinoma (80)	*KRAS*, c.38G>A, p.G13D	Yes	1610	14.61
3	Colorectal adenocarcinoma (70)	*KRAS*, c.35G>C, p.G12A	Yes	1653	30.19
		KIT, c.1621A>C, p.M541L	Yes	1454	47.46
4	Colorectal adenocarcinoma (90)	*KRAS*, c.35G>T, p.G12V	Yes	912	39.47
		PIK3CA, c.1624G>A, p.E542K	Yes	4106	30.39
5	Colorectal adenocarcinoma (20)	*KRAS*, c.38G>A, p.G13D	Yes	1847	23.5
		MET, c.1124A>G, p.N375S	Yes	5604	57.74
6	Colorectal adenocarcinoma (30)	*KRAS*, c.34G>A, p.G12S	Yes	3302	54.63
7	Colorectal adenocarcinoma (40)	*KRAS*, c.35G>T, p.G12V	Yes	2771	22.16
		MET, c.3029C>T, p.T1010I	Yes	3912	36.5
8	Colorectal adenocarcinoma (50)	*KRAS*, c.35G>A, p.G12D	Yes	1093	35.86
9	Melanoma (90)	*NRAS*, c.34G>A, p.G12S	Yes	9958	69.09
		MET, c. 1124A>G, p.N375S	Yes	4994	46.1
10	Melanoma (80)	*NRAS*, c.35G>C, p.G12A	Yes	3298	78.14
		MET, c. 1124A>G, p.N375S	Yes	3992	49.2
11	Melanoma (95)	*BRAF*, c.1799T>A, p.V600E	Yes	1005	27.16
		FBXW7, c.1394G>C, p.R465P	Yes	1157	29.47
12	Melanoma (90)	*PIK3CA*, c.3129G>A, p.M1043I	Yes	1746	27.49
13	Endometrial adenocarcinoma (90)	*KRAS*, c.34G>A, p.G12S	Yes	1196	66.47
		MET, c.1124A>G, p.N375S	Yes	2261	51.22
14	Rectal serous carcinoma (35)	*TP53*, c.722C>T, p.S241F	Yes	4485	63.46
		KRAS, c.35G>A, p.G12D	Yes	2962	30.79
		MET, c.1124A>G, p.N375S	Yes	8427	57.32
15	Squamous cell carcinoma (90)	*PIK3CA*, c.1624G>A, p.E542K	Yes	5552	18.7
16	Lung adenocarcinoma (80)	*TP53*, c.742C>T, p.R248W	Yes	2741	56.18
		KRAS, c.35G>A, p.G12D	Yes	2765	37.83
		PIK3CA, c.1633G>A, p.E545K	Yes	6504	30.21
17	Lung adenocarcinoma (60)	*EGFR*, 15-bp del, exon 19	Yes	1901	44.1
18	Lung adenocarcinoma (60)	*EGFR*, 15-bp del, exon 19	Yes	1869	24.4
19	GI stromal tumor (90)	*KIT*, 3-bp del, exon 11	Yes	220	40.5
20	GI Stromal tumor (95)	*KIT*, 6-bp del, exon 11	Yes	211	73.5
21	GI stromal tumor (90)	*KIT*, 12-bp deletion, exon 11	No	NA	NA
22	GI stromal tumor (95)	*KIT*, exon 9, 6-bp insertion	No	NA	NA

The details of mutations detected in 22 samples sequenced in retrospective by IT-PGM, their coverage, and variant frequencies are summarized.
GI, gastrointestinal; NA, not applicable.

Sequencing of Melanoma Tumor Samples Using the IT-PGM Platform in Parallel with the Sequenom MassARRAY Platform

Thirty-two melanoma tumor samples were sequenced on the IT-PGM platform in parallel with the Sequenom MassARRAY cancer panel for confirmation. Adequate sequencing of these samples was evident by AQ20 reads >300,000, with an average of 541,792 reads. Mutations detected by the IT-PGM platform in genes and hotspots not assessed by the Sequenom MassARRAY panel were confirmed by Sanger sequencing. In the 32 tumor samples sequenced, 47 missense mutations and one insertion were identified, with median coverage of 1872×. Sequencing using the IT-PGM platform identified *BRAF* mutations in 17 of 32 tumor samples (53.1%), including 13 samples with single nucleotide substitutions [p.V600E ($n = 12$) and p.G466E ($n = 1$)], 3 samples with double nucleotide substitutions [p.V600K ($n = 2$) and p.V600R ($n = 1$)], and 1 sample with a 6-bp insertion. Each of the missense mutations was also confirmed by the Sequenom MassARRAY 11-gene mutational screen, and the 6-bp insertion was confirmed by Sanger sequencing (Table 6). Furthermore, using the IT-PGM platform, mutations in other genes were identified in these 32 tumor samples, including mutations in *TP53* ($n = 9$), *NRAS* ($n = 7$), *EGFR* ($n = 3$), *KIT* ($n = 2$), *RB1* ($n = 2$), *ATM* ($n = 2$), *FGFR3* ($n = 1$), *RET* ($n = 1$), *KDR* ($n = 1$), *ERBB4* ($n = 1$), *JAK3* ($n = 1$), and *FBXW7* ($n = 1$). Note that all three *EGFR* mutations were identified in a single tumor sample that also had four mutations in additional genes (Table 2). In this melanoma set, all mutations detected by IT-PGM were confirmed by either Sequenom or Sanger analysis, except for a single base substitution mutation in exon 21 of *RB1* (c.2112G>T, p.M704I) detected by IT-PGM, which was found to be

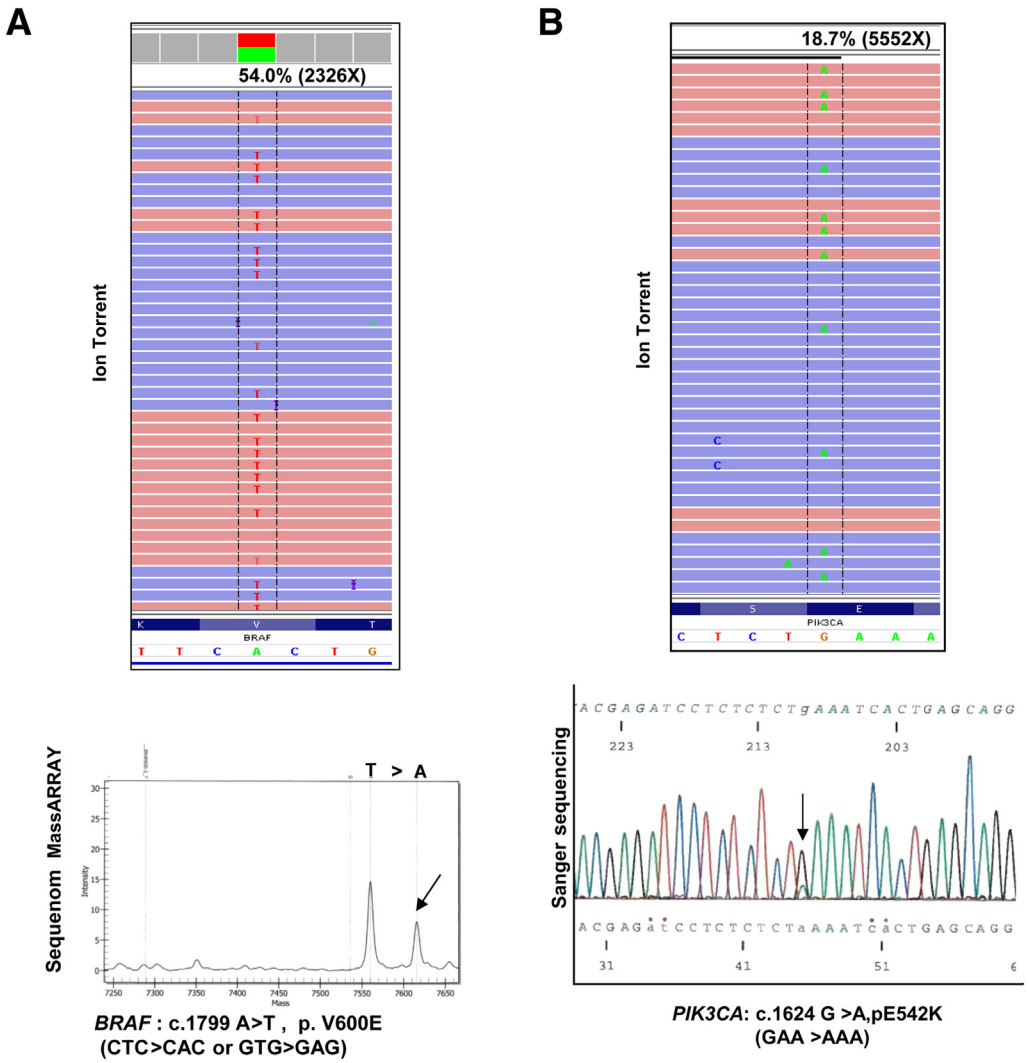


Figure 3 Examples of mutations detected using the IT-PGM platform and confirmed by alternate sequencing platforms. **A:** A *BRAF* substitution mutation (c.1799A>T, p.V600E) was identified in a melanoma, which resulted from a GTG>GAG (valine to glutamic acid) change at codon 600. The mutation was evident in the forward (+ strand, red in IGV) and reverse (− strand, blue in IGV) strands and was confirmed by Sequenom MassARRAY by the appearance of a peak corresponding to nucleotide **A** and indicated by the **arrow**. **B:** A *PIK3CA* mutation (c.1624G>A, p.E542K) was identified by IT-PGM in a squamous cell carcinoma sample resulting in GAA>AAA (glutamic acid to lysine) at codon 542, which was confirmed by Sanger sequencing (as indicated by the **arrow**).

a 16-bp deletion in the intron-exon boundary by Sanger sequencing (sample 28) (Table 6). Although not called by IT-PGM, this deletion could be seen clearly in the forward but not in the reverse strands in the IGV alignment (Supplemental Figure S2).

Parallel Sequencing of Additional Tumor Types

In addition to the melanoma cases, 16 adenocarcinoma samples were sequenced using the IT-PGM system in parallel with the Sequenom MassARRAY platform. Sanger sequencing was used to confirm additional mutations identified by IT-PGM that were not covered by the 11-gene Sequenom panel. These samples included seven colorectal, four papillary thyroid, two serous carcinoma, two lung, and one gastric adenocarcinoma. All the samples had >300,000 AQ20 reads, with an average of 509,180 AQ20 reads, indicating efficient sequencing. Eighteen single nucleotide substitution mutations were identified in this set, with median coverage of 1687×, including mutations in *TP53* ($n = 5$), *BRAF* ($n = 4$), *KRAS* ($n = 4$), *NRAS* ($n = 4$), and *FBXW7* ($n = 1$). Each of these 18 mutations was confirmed by either Sequenom MassARRAY or Sanger sequencing (Table 7).

Presence of Homopolymer Stretches Leads to Erroneous Mutation Calls

At a homopolymer stretch, nonlinear correlation between signal generated and the number of nucleotides incorporated leads to erroneous calls when using the IT-PGM platform. For example, the presence of seven cytidines in exon 9 of *FGFR3* resulted in two mutations, p.S400T and p.S400N, whereas these mutations were not detected using Sanger

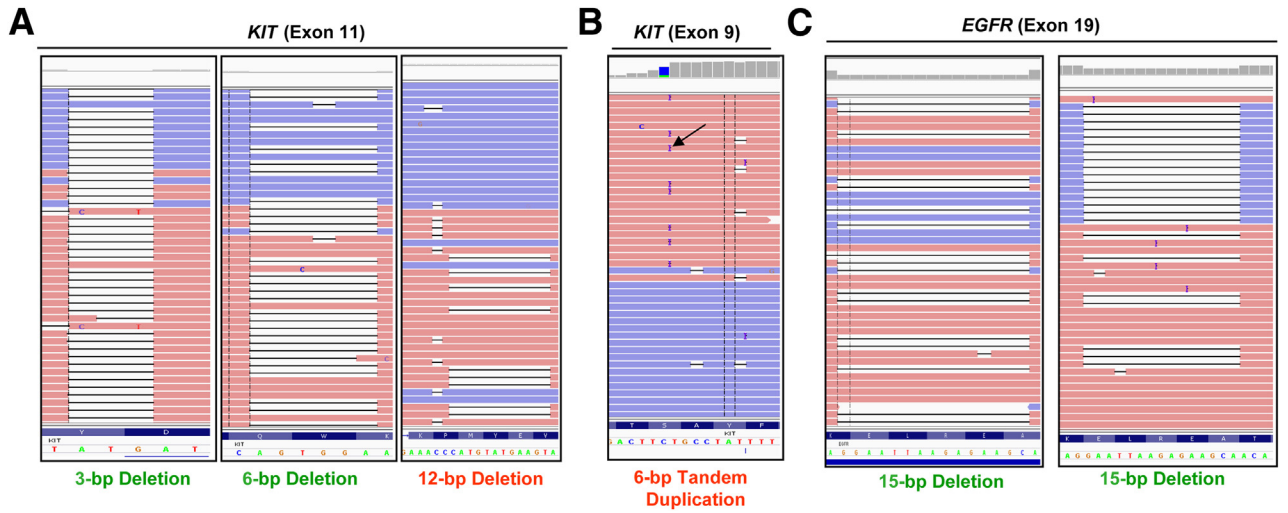


Figure 4 Detection of insertions and deletions by the IT-PGM platform. **A:** The 3- and 6-bp deletions in exon 11 of *KIT* were identified in the forward and reverse strands, which resulted in their successful calling by IT-PGM. In another sample, a 12-bp deletion in exon 11 of *KIT* was not identified because it was seen to be present in only the forward strand and, therefore, was not called by IT-PGM (**A**). **B:** Tandem duplication of a 6-bp sequence (TGCCTA) in exon 9 of *KIT* was seen as present in only the forward strand (**arrow**) and not in the reverse strand and, hence, was not called by IT-PGM. **C:** A 15-bp deletion in exon 19 of *EGFR* was successfully identified in the forward and reverse strands and was called in two tumor samples by IT-PGM.

sequencing (Figure 5A). Similarly, the presence of homopolymer stretches resulted in erroneous calling of a deletion in exon 4 of *TP53* and a missense mutation in exon 21 of *ERBB2*, which were not confirmed by Sanger sequencing (Figure 5, B and C).

Summary of Mutations Detected by Using IT-PGM Platform and Concordance with Alternate Sequencing Platforms

In 70 solid tumor samples sequenced using the IT-PGM platform and the AmpliSeq cancer panel, 100 sequence variants were detected. These variants included 95 missense mutations, 1 insertion (in *BRAF*), and 4 deletions (2 in *EGFR* and 2 in *KIT*). Most of the mutations were detected in genes commonly mutated in a variety of cancers, such as *BRAF, NRAS, KRAS, TP53, PIK3CA*, and *KIT*. In addition, low-frequency mutations were also found in other genes, including *RB1, MET, ERBB4, ATM, JAK3,* and *FBXW7*. The number of samples sequenced in each tumor type and the distribution of these gene mutations are summarized in Figure 6.

Comparability of Sequencing on Ion 316 and 318 Chips

Sequencing efficiency of barcode-indexed samples with various levels of multiplexing was gleaned by comparing the sequencing of four samples on both Ion 316 and 318 Chips as a part of four and eight multiplexed sample pools, respectively. A summary of the sequencing runs, including loading density, percentage of polyclonal ISPs, and final library reads, is shown in Supplemental Figure S3, A and B. Sample 1 was sequenced as part of comparison set 1, and samples 2, 3, and 4 were sequenced as part of comparison set 2. Comparable sequencing efficiency was observed for these samples on either Chip as evident by minimal variability of the total AQ20 reads for each sample on Ion 316 and 318 Chips (Supplemental Figure S3C). This was also evident in the congruence of mutations detected and their variant frequencies in the samples tested on Ion 316 and 318 Chips (Supplemental Figure S3D).

Comparison of Manual and Automated E-PCR and ISP Recovery

To compare the manual and automated methods for clonal amplification of library DNA on ISPs by E-PCR and their subsequent recovery, two sets of two samples each were sequenced after sample preparation by the manual method and the IT OneTouch system. The efficiency of each method was evaluated by the extent of polyclonal ISPs (with more than one clonally amplified amplicon on an ISP, which does not provide useful sequence information) and the overall efficiency of sequencing. This comparison showed a trend toward better sample preparation by manual E-PCR and breaking compared with the IT OneTouch system. This was evident by the consistently decreased levels of polyclonal ISPs and low-quality reads obtained for the samples prepared by manual methods, which resulted in consistently higher amounts of the total sequence information obtained (AQ17 bases) and their quality (AQ20 and perfect reads) compared with samples prepared by the IT OneTouch system. In the first sample set, sample preparation by IT OneTouch had ISPs with 37% and 9% polyclonal and low-quality signal, respectively, generating total sequence information with 87.75-Mb (AQ17), 68.5 Mb (AQ20), and 67.8 Mb (perfect) base calling. In comparison, manual sample preparation had ISPs with 24% polyclonal and 5% low-quality signal, generating 222.5 Mb (AQ17),

Table 6 Sequencing of Melanoma Samples by IT-PGM

Sample No.	Tumor percentage	Mutation detected by IT-PGM	Coverage (×)	Variant frequency (%)	Confirmed by Sequenom	Confirmed by Sanger
1	90	*TP53*, c.742C>T, p.R248W	4015	20.5		Yes
		KIT, c.1621A>C, p.M541L	4351	66.56		Yes
2	80	*BRAF*, c.1799T>A, p.V600E	770	40.39	Yes	
3	70	*BRAF*, c.1799T>A, p.V600E	3696	40.99	Yes	
					Yes	
4	95	*BRAF*, c.1799T>A, p.V600E	1545	64.6	Yes	
5	90	*TP53*, exon 5, TCC>TTC, p.S127F	1937	44.71		Yes
		KIT, exon 10, ATG>CTG, p.M541L	2294	54.1		Yes
6	90	*BRAF*, c1798_1799, TG>AA, p.V600K	1228	41.12	Yes	
7	30	*BRAF*, c.1799T>A, p.V600E	4948	79.14	Yes	
8	80	*NRAS*, c.182A>T, p.Q61L	4092	45.97	Yes	
		BRAF, c.1398G>A, p.G466E	6843	18.37	Yes	
9	90	*BRAF*, c.1799T>A, p.V600E	2326	54.08	Yes	
10	80	*BRAF*, c1798_1799 TG>AA, p.V600K	1414	60.75	Yes	
11	90	*BRAF*, c1798_1799 TG>GA, p.V600R	1528	43.06	Yes	
12	75	*ATM*, c.1810C>T, p.P604S	2428	43.7		Yes
		BRAF, c1799T>A, p.V600E	1757	13.55	Yes	
13	70	*ATM*, c.2572T>C, p.F858L	1213	53.09		Yes
		BRAF, c1799T>A, p.V600E	1523	51.67	Yes	
14	70	*BRAF*, c1799 T>A, p.V600E	728	43.27	Yes	
15	80	*BRAF*, c1799 T>A, p.V600E	1853	26.66	Yes	
16	40	*NRAS*, c.182 A>G, p.Q61R	2141	32.41	Yes	
17	95	*BRAF*, c1799 T>A, pV600E	2281	36.21	Yes	
18	90	*JAK3*, c.2164G>A, p.V722I	2132	51.83	Yes	Yes
		BRAF, c.1799T>A, p.V600E	1725	32.7		
19	90	*NRAS*, c.181C>A, p.Q61K	2781	44.16	Yes	
20	90	*NRAS*, c.182A>G, p.Q61R	5427	58.43	Yes	
21	90	*FGFR3*, c1150T>C, p.F384L	700	46.71		Yes
		NRAS, c181C>A, p.Q61K	2637	29.28	Yes	
22	90	*BRAF*, c1799T>A, p.V600E	1119	54.07	Yes	
		TP53, c.470T>C, p.V157A	422	42.18		Yes
23	90	*KDR*, c.1444T>C, p.C482R	4372	56.79		Yes
24	80	*ERBB4*, c.722C>T, p.P241L	1144	16.87		Yes
25	90	*TP53*, c.742T>C, p.R248W	2458	77.3		Yes
26	90	*TP53*, c.734G>A, p.R248Q	1066	67.54		Yes
		RB1, c.411A>T, p.137D	401	16.21		Yes
27	80	*NRAS*, c.182A>G, p.Q61R	3637	44.93	Yes	
28	75	*RB1*, c.2112G>T, p.M704I	585	59.15		No
		TP53, c.596C>T, p.190L	181	88.4		Yes
29	90	*TP53*, c.715A>G, p.N239D	2361	74.93		Yes
30		*TP53*, c.637C>T, p.R213*	1209	12.49		Yes
31	90	*FGFR3*, c.2386C>T, p.P796S	744	6.72		Yes
		EGFR, c.2509G>A, p.D837N	1775	14.42		Yes
		TP53, c.721T>A, p.S241T	1891	71.71		Yes
		RET, c.1838C>T, p.P613L	192	26.04		Yes
		EGFR, c.2123A>C, p.K708T	1171	23.65		Yes
		EGFR, c.293G>A, p.R98Q	2027	16.58		Yes
		NRAS, c.182A>T, p.Q61L	4359	62.35		Yes
32	90	*BRAF*, exon 15, 6-bp insertion	1952	33.40		Yes

The details of mutations detected in 32 melanoma samples sequenced on IT-PGM in parallel with Sequenom MassARRAY are shown. The confirmation status by either Sequenom MassARRAY or Sanger sequencing is also listed.

175.7 Mb (AQ20), and 170.5 Mb (perfect) base calls (Supplemental Figure S4A). In the second sample set, sample preparation by IT OneTouch had 24% polyclonal and 11% low-quality ISPs, generating 122.9 Mb (AQ17), 101.5 Mb (AQ20), and 99.7 Mb (perfect) base calling. In comparison, the manual sample preparation had ISPs with 23% polyclonal and 6% low-quality ISPs, respectively, generating 158.1 Mb (AQ17), 126.6 Mb (AQ20), and 125.0 Mb (perfect) base

Table 7 Sequencing of Additional Tumor Types by IT-PGM in Parallel with Sequenom MassARRAY

Sample No.	Tumor type (%)	Mutation detected by IT-PGM	Coverage (×)	Variant frequency (%)	Confirmed by Sequenom	Confirmed by Sanger
1	Colorectal adenocarcinoma (30)	*TP53*, c.817C>T, p.R273C	1161	29.89		Yes
		KRAS, c.35G>T, p.G12V	1477	20.65	Yes	
2	Colorectal adenocarcinoma (30)	*KRAS*, c.38G>A, p.G13D	1864	15.4	Yes	
		FBXW7, c.1393C>T, p.R465C	2912	15.35		Yes
3	Colorectal adenocarcinoma (40)	*KRAS*, c.35G>A, p.G12D	1024	29.2		Yes
4	Colorectal adenocarcinoma (60)	*KRAS*, c.436G>A p.A146T	1136	63.2	Yes	
5	Colorectal adenocarcinoma (50)	*NRAS*, c.35G>A, p.G12D	3050	15.21	Yes	
6	Colorectal adenocarcinoma (40)	*TP53*, c.498C>T, p.R273H	2302	9.6		Yes
7	Colorectal adenocarcinoma (80)	*NRAS*, c.181C>A, p.Q61K	1407	25.59	Yes	
8	Gastric adenocarcinoma (20)	*TP53*, c.742C>T, p.R248W	3798	49.84		Yes
9	Lung adenocarcinoma (70)	*TP53*, c. 637C>T, p.R213*	313	31.63		Yes
10	Lung adenocarcinoma (50)	*TP53*, c.734G>T, p.G245V	3176	21.13		Yes
11	Serous carcinoma (75)	*NRAS*, c.182A>G, p.Q61R	1639	20.44	Yes	
12	Serous carcinoma (75)	*NRAS*, c.182A>G, p.Q61R	2754	56.68	Yes	
13	Papillary thyroid carcinoma (70)	*BRAF*, c1799T>A, p.V600E	1462	16.01	Yes	
14	Papillary thyroid carcinoma (60)	*BRAF*, c1799T>A, p.V600E	1873	21.78	Yes	
15	Papillary thyroid carcinoma (85)	*BRAF*, c1799T>A, p.V600E	1735	32.16	Yes	
16	Papillary thyroid carcinoma (50)	*BRAF*, c1799T>A, p.V600E	1110	26.04	Yes	

The details of mutations detected in 16 samples of different tumor type sequenced on IT-PGM in parallel with Sequenom MassARRAY are shown. The confirmation status by either Sequenom or Sanger sequencing is also listed.

calls (Supplemental Figure 4B). Better sample preparation by the manual methods also resulted in the higher levels of total AQ20 reads obtained per sample by the manual methods compared with the IT OneTouch method (Supplemental Figure S4).

Clinical Reporting of Mutations

The steps involved in clinical reporting are summarized in Supplemental Figure S5. Briefly, the sequencing metrics of a run, sequencing coverage at hotspots for each sample, and the variants called for sample generated by the IT Variant Caller software are parsed into the in-house designed software OncoSeek for further annotation and reporting. OncoSeek facilitates compilation of the patient information (patient identifiers, tumor type, tumor percentage, etc) with the variant calls from IT-PGM, viewing of the aligned reads and sequence variants using IGV, linking of the variants to the Catalogue of Somatic Mutations in Cancer and dbSNP databases, annotation of the position of the variant, and amino acid change according to the Human Genome Variation Society–compliant nomenclature and reference gene NM numbers. During reporting, each sequence variant identified for the sample is visualized in IGV and is decided by several criteria, including variant frequency of $\geq$10%, adequate sequencing coverage (250×), presence of the variant in both forward and reverse reads, and comparable variant frequency in reads of both directions. Any mutation that seems real in IGV but has <10% variant frequency is subjected to confirmation on alternative sequencing platforms and is reported only when confirmed. Silent and nonexonic mutations are not reported. The mutational statuses of the genes for which the amplicons failed to reach the minimum coverage of 250× are marked as indeterminate. It is important to note that OncoSeek does not filter out any variants called by the IT Variant Caller software. However, OncoSeek's database has a self-updating population analysis feature that keeps track of the patient samples analyzed and the frequency of variants in the samples analyzed. Using this, spurious mutation calls, such as sequencing artifacts at homopolymer regions or variants due to the reference genome polymorphisms that result in high population frequency, can be easily identified and filtered, by choice, along with mutations that are silent or intronic and in untranslated regions. In the clinical report, the mutation status of the genes ordered for testing was listed first, followed by mutations observed in the remaining genes. The steps involved in clinical reporting are summarized in Supplemental Figure S5.

Discussion

The massive parallel genome sequencing capability of NGS technologies has made simultaneous screening of multiple genes for mutations in multiple samples possible. Application of NGS technologies for routine screening of patients, however, requires rigorous validation in a Clinical Laboratory Improvement Amendments diagnostic laboratory approved to establish consistent and reliable performance. To this end, in this study, we sequenced genomic DNA from 70 solid tumor samples using the IT AmpliSeq cancer panel to screen mutational hotspots in 46 cancer-related genes. With this approach, we could efficiently amplify areas of interest, obtain sequence information, and identify mutations with consistent sensitivity and specificity using as little as 10 ng of FFPE DNA as template. The quality of sequencing is evident by the overall amplification and sequencing coverage

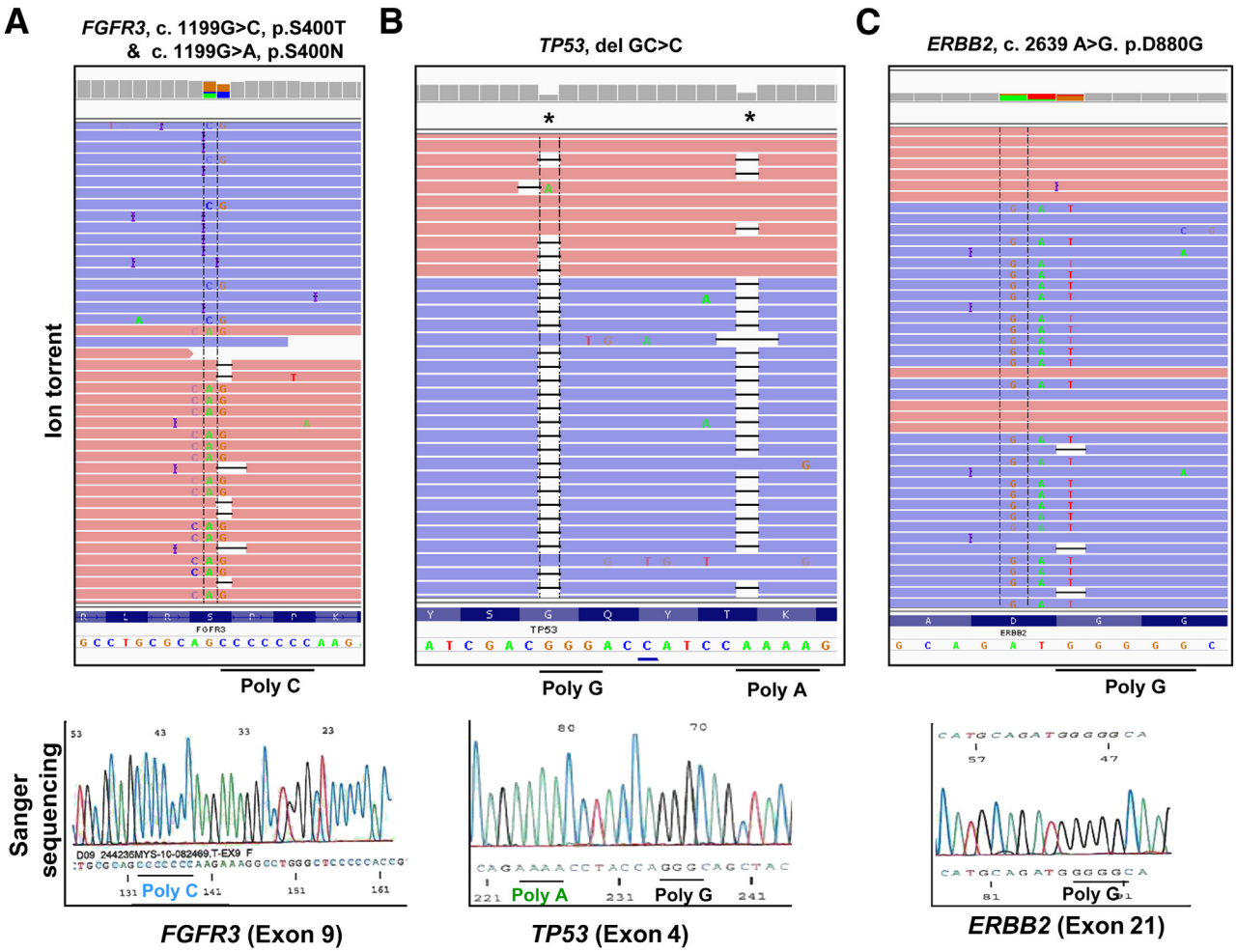


Figure 5 Homopolymer nucleotide stretches were a source of spurious mutation calls. The presence of homopolymer stretches in the DNA sequence often results in disproportionate and lower change in the signal (pH change), resulting in anomalous detection and calling of either deletions or nucleotide substitutions. **A:** A poly C homopolymer stretch in *FGFR3*, exon 9 resulted in calling of two missense mutations by IT-PGM, which were unconfirmed by Sanger sequencing. **B:** A poly G stretch in exon 4 of *TP53* resulted in calling of an erroneous single nucleotide deletion by IT-PGM, which was not found by Sanger sequencing. **Asterisks** indicate the positions in the sequence where spurious single nucleotide deletions were detected and called by IT-PGM in the areas of homopolymer stretches. **C:** A poly G track in exon 21 of *ERBB2* resulted in misalignment and calling of a missense mutation at codon 880 (**top panel**), which was not evident by Sanger sequencing.

summary shown in Figure 1. Efficient sequencing coverage with few exceptions was seen for all amplicons. Dilution studies using cell lines established a consistent detection sensitivity at 10% variant frequency. A high degree of inter-run reproducibility was also evident by comparing the variants detected in 1:9 diluted H2122 cell line DNA compared across 14 runs and a patient DNA sample compared across 25 runs. A high degree of intrarun reproducibility was also observed at a variant frequency of $\geq$10% in an amplicon with coverage $\geq$250×. This helped us establish minimum coverage and variant frequency cutoff values of 250× and 10%, respectively, as reliable for clinical reporting of a mutation. A recently published study testing the sequencing of FFPE tumor samples on IT-PGM with library preparation using the AmpliSeq cancer panel also reported 100% assay sensitivity at 8% variant frequency,[23] which supports the sensitivity limit we established in this study.

The IT AmpliSeq cancer panel library preparation we used generates 191 amplicons interrogating mutational hotspots across 46 genes. For validation purposes, it was a daunting task to have a control (cell line or patient sample) positive for mutation in each amplicon. The option of having a synthetic oligo or a plasmid control was not desirable owing to the high potential for contamination and the inability to mimic the quality of FFPE samples. For these reasons, the approach we chose was to validate the overall performance of this sequencing platform using samples of several tumor types positive for mutations (indels and SNVs) across several genes. To this end, we initially sequenced 22 archival tumor samples with mutations detected previously in our laboratory by alternate sequencing methods. Using this approach, the mutations present in these 22 samples had a minimum mutation level of 10%, which is the known sensitivity of Sequenom MassARRAY analysis or Sanger sequencing used

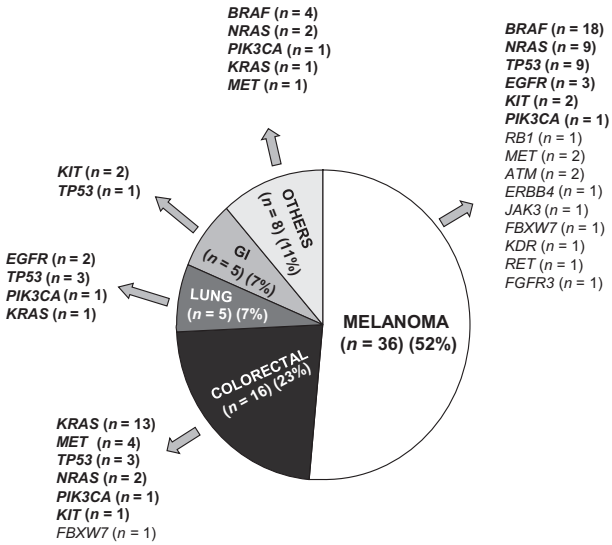


Figure 6 Summary of the tumor types tested in the validation set and the number of mutations detected mostly in routinely tested genes (in bold) and in genes not routinely tested (in normal text). Others includes thyroid, squamous cell, and endometrial tumors.

earlier to detect these mutations. This approach allowed us to initially establish the ability of the IT-PGM platform to detect mutations simultaneously in several multiplexed samples. In these samples, the IT-PGM platform efficiently detected each of the known single nucleotide mutations.

However, detection of insertions and deletions was not consistent. A 15-bp deletion in exon 19 of *EGFR*, smaller deletions of 3 and 6 bp in exon 11 of *KIT,* and a 6-bp insertion in exon 15 of *BRAF* (Tables 5 and 6) were detected successfully. However, IT-PGM did not call a 12-bp deletion in *KIT* exon 11 or a 6-bp insertion in exon 9. Nonetheless, the presence of these insertions was evident when the read alignments were visualized in IGV, albeit only in the forward direction (Figure 4). These data indicated that the insertion and deletion were present in the sequence information but were not called by the IT Variant Caller software. A similar trend of indel detection was also reported in a recent study in which most indels tested were not called by IT Variant Caller but their presence was evident in the sequencing reads.[23] This is not surprising as the mutation-calling algorithm of the IT-PGM platform is designed for single nucleotide mutations at gene hotspots and not for large insertions or deletions. Hence, this could be a reliable mutation screening assay for SNVs and not for indels. However, sequencing comparison of several additional samples not included in the validation set using the IT-PGM platform updated with the recent release of Torrent Suite software version 3.0 and with Torrent Suite software version 2.0.1 demonstrated that version 3.0 enables the IT-PGM platform to identify and call several insertions/deletions that were not called by version 2.0.1 (Supplemental Figure S6).

The foremost advantage of mutation screening using the IT-PGM platform and the IT AmpliSeq cancer panel was its capability to efficiently sequence multiple genes with very low quantities (10 ng) of DNA derived from FFPE tumor samples obtained by various methods of tumor tissue isolation (Supplemental Figure S7). Additional advantages include the flexibility of choosing the number of samples to be analyzed per sequencing run on chips of varying capacities with multiplexing. It supports sample preparation with 32 barcodes, allowing sequencing of 32 samples simultaneously in a single run. Both Ion 316 and 318 Chips support multiplexed samples, and our comparison of the same tumor samples sequenced in parallel on both chips showed comparable sequencing with similar AQ20 reads and no variation in either the mutations detected or their variant frequencies (Supplemental Figure S3).

In the process of validation, we also compared the manual method of E-PCR and sample breaking with the automated IT OneTouch system. The manual method is a multistep, labor-intensive procedure, an obvious drawback, but in our laboratory, it consistently yielded a better quality of ISPs and sequencing compared with the automated IT OneTouch system (Supplemental Figure S4). One of the major problems we found to be associated with sequencing using the IT-PGM platform was the high level of sequencing errors in the areas of homopolymer nucleotide repeats, where usually spurious deletions or substitutions were called. The issue is an integral part of the sequencing chemistry, which involves flowing of one nucleotide at a time and has been measured and reported by others for the IT platform and another benchtop sequencer, the 454 GS Junior (Roche Applied Science, Indianapolis, IN), which applies similar sequencing chemistry.[16,22] However, we found that the newer version of the Torrent Suite software (version 3.0) considerably decreased calling of these homopolymer-induced spurious mutations (Supplemental Figure S6).

In the set of 70 different tumor samples tested, we detected 99 sequence variations that were known before sequencing on the IT-PGM platform (a retrospective study set) or were confirmed by parallel testing by Sequenom MassARRAY analysis or subsequently by Sanger sequencing. Note that some of the mutations detected could be germline polymorphisms, such as *MET*, p.N375S and *KIT,* p.M541L. To address this issue, we would need to analyze paired normal samples from these patients, which were not available. Nonetheless, these variants still contributed toward validating the performance of the IT-PGM platform to correctly identify variants. Note that the IT-PGM platform is more sensitive compared with more traditional sequencing methods. Thus, validation of this platform in the clinical laboratory setting was challenging because traditional, time-tested Sanger sequencing or other methods are less sensitive to confirm low allelic frequency variants. For this reason, clinical laboratories may need to develop two different NGS methods simultaneously that can then be used to validate each other.

Most of the mutations detected using the IT-PGM platform were found in genes known to be commonly mutated in cancers; however, mutations were also identified, albeit at lower frequency, in genes that would not generally be

tested, such as, *RB1*, *MET*, *ERBB4*, *ATM*, *EGFR*, *JAK3*, and *FBXW7* (Figure 6). Also, the IT-PGM platform detected multiple mutations in some tumor samples. For example, in a melanoma (sample 31 in Table 6), seven mutations were detected and confirmed. These findings also highlight the advantages of simultaneous screening of multiple genes in tumor samples, facilitating the discovery of mutations in key oncogenes that might not be suspected to be involved in the pathobiology of the tumor type being tested. Furthermore, in this melanoma sample, mutations in different genes were found at different frequencies, perhaps an indication of clonal evolution of neoplastic cells.

To facilitate clinical reporting of the IT-PGM sequencing results, we developed a software (OncoSeek) to interface with the sequencing output from IT-PGM and to link the results with additional information, such as patient identifiers, requesting physician, tumor type, and tumor percentage, and with several public databases of mutations, such as the Catalogue of Somatic Mutations in Cancer and dbSNP.[24] The software also helps in annotating the sequence variants according to Human Genome Variation Society guidelines. Most importantly, OncoSeek facilitates visualization of the sequencing reads via IGV, which is one of the most important steps in determining the authenticity of the variants detected for reporting. Furthermore, OncoSeek uses the entire report of variants from IT without any filtering, which ensures that all the variants are available for scrutiny before reporting. OncoSeek also facilitates tracking the frequency of each sequence variant seen in this data set. This is recorded as "population frequency," which is the percentage of the samples in which the mutation has been detected in the database of samples sequenced. This important parameter helps to filter out frequently occurring germline polymorphisms and homopolymer-related artifacts that occur with abnormally high frequency. With our experience, we have observed that any variant that has a population frequency of $\geq$20 (detected in $\geq$20% of the samples sequenced) is most likely spurious, and this criteria could be used to identify and filter them before reporting. Oncoseek is used to summarize the variant calls along with the patient information and to generate a well-annotated and detailed clinical report, which has the mutations detected in ordered gene sets for each tumor type listed first, followed by mutations found in additional genes tested in the panel.

On the whole, this validation study has shown that screening of solid tumor samples for mutation hotspots using the IT AmpliSeq cancer panel IT-PGM is a valid screening assay that is sensitive, specific, convenient, and can be used routinely in a clinical molecular diagnostics laboratory. As IT-PGM platform uses only 10 ng of genomic DNA from FFPE tissue samples, this assay can be used for archival tumor samples retrospectively as well as prospectively for diagnostic specimens, including highly limited specimens such as fine-needle aspirates. We have implemented this as a routine mutation screening test for solid tumors in our laboratory.

Supplemental Data

Supplemental material for this article can be found at *http://dx.doi.org/10.1016/j.jmoldx.2013.05.003*.

References

1. Mardis ER: Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 2008, 9:387–402
2. Ross JS, Cronin M: Whole cancer genome sequencing by next-generation methods. Am J Clin Pathol 2011, 136:527–539
3. Metzker ML: Sequencing technologies: the next generation. Nat Rev Genet 2010, 11:31–46
4. Choi BO, Koo SK, Park MH, Rhee H, Yang SJ, Choi KG, Jung SC, Kim HS, Hyun YS, Nakhro K, Lee HJ, Woo HM, Chung KW: Exome sequencing is an efficient tool for genetic screening of Charcot-Marie-Tooth Disease. Hum Mutat 2012, 33:1610–1615
5. Coonrod EM, Durtschi JD, Margraf RL, Voelkerding KV: Developing genome and exome sequencing for candidate gene identification in inherited disorders: an integrated technical and bioinformatics approach. Arch Pathol Lab Med 2013, 137:415–433
6. Doherty D, Bamshad MJ: Exome sequencing to find rare variants causing neurologic diseases. Neurology 2012, 79:396–397
7. Johnson JO, Gibbs JR, Megarbane A, Urtizberea JA, Hernandez DG, Foley AR, Arepalli S, Pandraud A, Simon-Sanchez J, Clayton P, Reilly MM, Muntoni F, Abramzon Y, Houlden H, Singleton AB: Exome sequencing reveals riboflavin transporter mutations as a cause of motor neuron disease. Brain 2012, 135:2875–2882
8. Wang Z, Gerstein M, Snyder M: RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009, 10:57–63
9. Costa V, Aprile M, Esposito R, Ciccodicola A: RNA-Seq and human complex diseases: recent accomplishments and future perspectives. Eur J Hum Genet 2012, 21:134–142
10. Wold B, Myers RM: Sequence census methods for functional genomics. Nat Methods 2008, 5:19–21
11. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR: Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009, 462:315–322
12. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al: An integrated semiconductor device enabling non-optical genome sequencing. Nature 2011, 475:348–352
13. Elliott AM, Radecki J, Moghis B, Li X, Kammesheidt A: Rapid detection of the ACMG/ACOG-recommended 23 CFTR disease-causing mutations using Ion Torrent semiconductor sequencing. J Biomol Tech 2012, 23:24–30
14. Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska M, Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM, Karch H: Prospective genomic characterization of the German enterohemorrhagic Escherichia coli O104:H4 outbreak by rapid next generation sequencing technology. PLoS One 2011, 6:e22751
15. Vogel U, Szczepanowski R, Claus H, Junemann S, Prior K, Harmsen D: Ion Torrent Personal Genome Machine sequencing for genomic typing of Neisseria meningitidis for rapid determination of multiple layers of typing information. J Clin Microbiol 2012, 50:1889–1894
16. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ: Performance comparison of benchtop high-throughput sequencing platforms. Nature Biotechnol 2012, 30:434–439
17. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y: A tale of three next generation

sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 2012, 13: 341
18. Newcombe RG: Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med 1998, 17:857−872
19. Thorvaldsdottir H, Robinson JT, Mesirov JP: Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 2013, 14:178−192
20. Routbort M, Handal B, Patel KP, Singh R, Aldape K, Reddy N, Barkoh B, Riben M, Medeiros J, Luthra R: OncoSeek: a versatile annotation and reporting system for nest generation sequencing-based clinical mutation analysis of cancer specimens. AMP 2012 Meeting Abstracts. J Mol Diagn 2012, 14:637−748, Abstract TT79
21. Forbes SA, Tang G, Bindal N, Bamford S, Dawson E, Cole C, Kok CY, Jia M, Ewing R, Menzies A, Teague JW, Stratton MR, Futreal PA: COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. Nucleic Acids Res 2010, 38:D652−D657
22. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, Teague JW, Campbell PJ, Stratton MR, Futreal PA: COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 2011, 39:D945−D950
23. Beadling C, Neff TL, Heinrich MC, Rhodes K, Thornton M, Leamon J, Andersen M, Corless CL: Combining highly multiplexed PCR with semiconductor-based sequencing for rapid cancer genotyping. J Mol Diagn 2013, 15:171−176
24. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K: dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001, 29:308−311