Targeted High Depth Next Generation Sequencing of Tumor Specimens

OVERVIEW

We have developed a targeted next generation sequencing (NGS) methodology for sensitive DNA variant detection that is highly optimized for clinical specimens and enables the accurate detection of clinically actionable mutations from low input DNA quantities. This strategy provides reliable, uniform, and high depth (>1000x) sequencing across gene regions representing >500 known cancer-associated mutations, and is compatible with the types of "real world" patient samples that are commonly used for cancer diagnostics, including formalin-fixed paraffin-embedded (FFPE) and fine needle aspiration (FNA) specimens. The performance of several pre-designed gene panels was validated using orthogonal methods, including Sanger sequencing and a highly sensitive probe-based mutation assay. Last, we have demonstrated the value of distinct sequencing technologies, such as those from Illumina and Ion Torrent, to provide accurate confirmation of mutation positives from primary NGS data.

CONCLUSIONS

The combination of targeted PCR gene enrichment and ultra deep NGS provides high sensitivity to detect mutations in cancer-relevant genes using low inputs of residual clinical sample DNA. Our data demonstrate:

- Uniform coverage of targeted regions across cell line, FFPE and FNA DNA
- High concordance of variant detection by targeted NGS compared to traditional Sanger sequencing and Luminex analysis
- Excellent sensitivity and reliable quantification of variant detection coupled with an optimized variant caller algorithm for FFPE and FNA samples that minimizes the false positive rate and detects mutations that may be missed by Sanger sequencing
- Importance of a validated and automated bioinformatics analysis pipeline to maximize sensitivity and positive predictive value

In conclusion, our data shows that NGS can be used for screening large numbers of cancer-associated genes to uncover benign and clinically relevant variants as well as variants of unknown significance. We further demonstrate that variants can be rapidly confirmed with an orthogonal platform, as available in Asuragen's Genomic Services laboratory.

INTRODUCTION

Advancements in next generation sequencing (NGS) technologies have proved invaluable in deepening our understanding of human disease at the molecular level. The use of genomic profiling to identify novel disease-related mutations has enormous potential to improve and individualize the molecular diagnosis of cancer and to guide clinical decision-making. Translation of NGS technologies into routine clinical use, however, requires that the technology be compatible with routine clinical specimens. FFPE tumors contain fragmented and cross-linked DNA, while FNA biopsies often result in very low DNA yields. Both sample types may present cellular heterogeneity (i.e., mixtures of tumor and stroma cells) and molecular heterogeneity (i.e., subclones of varying genotype).[1,2]

We developed a targeted high depth sequencing methodology for mutation detection that is highly optimized for interrogating these challenging clinical specimens. We demonstrate that the PCR-based targeted enrichment of specific gene regions (up to 540 mutational hotspots, Table 1) coupled with either of two NGS platforms—the Illumina GAIIx or the Ion Torrent Personal Genome Machine (PGM)—enables accurate detection of low-level variants in FFPE and FNA samples.

This workflow requires as little as 10 ng of sample DNA, and generates uniformly high coverage (>1000 reads per target) to ensure the detection of lower frequency variants in both sample types. Furthermore, we show that the sequencing data is highly concordant with mutations identified using Sanger sequencing and Asuragen's validated Signature® KRAS assay.[3] The experimental workflow and bioinformatics analysis pipeline is offered by Asuragen Genomic Services to detect variants within a panel of cancer-related genes and pathways.

FOCUSED GENE PANEL

ABL1	FGFR1	**HRAS**	MET
AKT1	FGFR3	JAK2	**NRAS**
BRAF	FLT3	KIT	PDGFRA
EGFR	RET	**KRAS**	**PIK3CA**

Table 1. Asuragen's Focused Gene Panel targets mutational hotspots in 16 oncogene targets. The five genes highlighted in bold, provide equivalent analytical performance with greater throughput per sequencing run. These panels are available commercially as SuraSeq™ 500 and SuraSeq™ 200 Cancer Panels, respectively.

STUDY DESIGN

Sample Preparation for NGS

Samples included 22 cancer cell line DNAs, 38 colorectal cancer (CRC) FFPE tissue specimens and 10 thyroid FNA specimens. DNA was purified from FFPE samples and total nucleic acid was isolated from FNA samples. The workflow for sample preparation (Figure 1) included a two-step PCR procedure to amplify 35 target regions across 16 oncogenes, followed by a "tagging" PCR to append adapter sequences specific to either the GAIIx or PGM platforms. The amplified regions totaled 1283 nucleotides, excluding primers, and included 168 dbSNP IDs (v132), and 540 COSMIC registry variants (v54).

NGS

The Illumina GAIIx and Ion Torrent PGM are orthogonal methods which utilize different chemistries for massively parallel sequencing. For GAIIx analysis, the enriched samples were normalized and pooled into respective multiplexed mixtures comprising 24–68 per lane. For PGM analysis, ~500-750 million copies of PCR templates were input for manual emulsion PCR, and ~100-150 million copies of templates were used for the Ion OneTouch (Life Technologies) automated emulsion PCR. All results presented in this study were obtained using 314 or 316 chips, and 5-20 multiplexed samples per run. Current capacities represent scaling due to technology improvements.

Bioinformatics

The sequence read data generated from the GAIIx were demultiplexed, trimmed of adapter sequences and gene-specific primer sequences, and filtered for high quality reads. Alignments to the human genome (hg19) sequence were performed using the BWA aligner (v0.5.9-r16).[4] PGM data were similarly pre-processed using Torrent Suite 1.5.1 from the Ion Torrent Browser, and aligned to hg19 using TMAP. Aligned reads were further processed using GATK.[5,6] Variant calling was performed using proprietary software following application of a coverage filter and background variant threshold.

Confirmatory Analysis

Sequence confirmation was obtained using BigDye Direct (Applied Biosystems) for the variants KRAS Codons 12/13, BRAF V600E, and PIK3CA Exon 10 and Exon 20. Prior to NGS analysis, the CRC samples were also assessed using Asuragen's Signature® KRAS assay, a validated liquid bead array method (Luminex platform) for detection of variants in KRAS codon 12 and 13.[3]

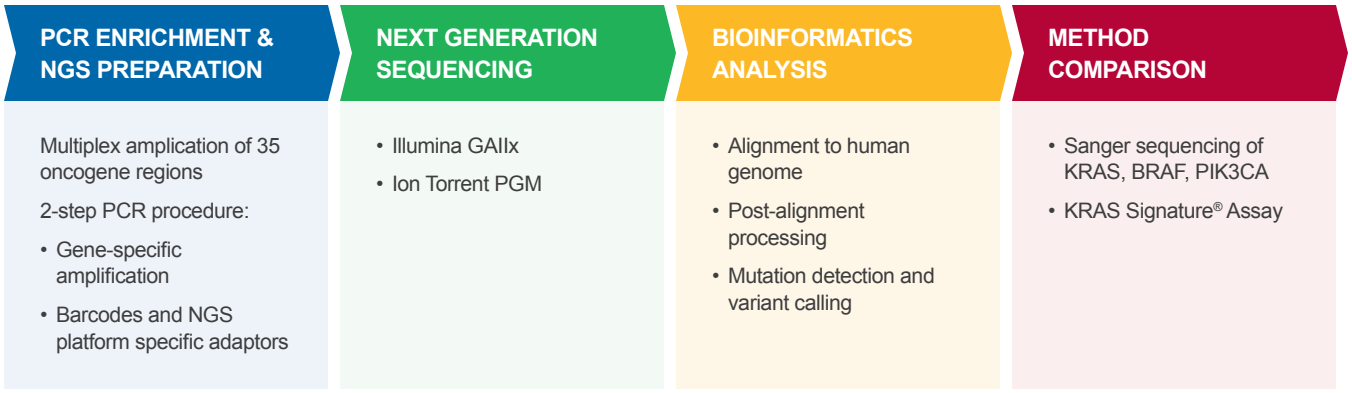

Figure 1. Overview of the targeted NGS experimental workflow. Content is enriched using a two-step PCR enrichment procedure that first amplifies 35 "hotspot" regions within 16 oncogenes and then appends platform-specific adapters for sequencing on either Illumina or Ion Torrent NGS instruments.

RESULTS

Workflow

The primary goal of this study was to assess the feasibility of applying NGS technologies for variant detection in residual clinical FFPE and FNA cancer specimens. Two NGS platforms were compared: the established Illumina GAIIx platform and the more recently developed Ion Torrent PGM. An overview of the study design, including sample specimens, PCR-based enrichment, NGS workflows, data analysis, and comparison with traditional sequencing methods is shown in Figure 1. The enriched oncogene regions (Table 1) represent greater than 95% of mutations within the 16 genes as reported in the COSMIC database.[7] A bioinformatics pipeline was developed and employed for genome alignment and variant identification. Finally, variants identified via NGS were confirmed using the established methods of Sanger sequencing and liquid bead array analysis.

Uniformity and Read Coverage

Recent studies have shown that low incidence therapyresistant subclones within a tumor may have important clinical implications.[10,11] High read coverage and uniformity are necessary to ensure detection of lower abundance tumor cell variants within FFPE and FNA specimens. To establish the quantitative performance of our approach, we evaluated coverage depth and uniformity following PCR-based enrichment of FFPE and FNA DNA. Greater than 90% of all amplicons were within 5X of the median read coverage for the FFPE and FNA samples, comparable to results using cancer cell line DNA (Table 2). FNA DNA had lower read depths which is attributed to the fact that samples were analyzed on a separate, more highly multiplexed GAIIx run.

Sensitivity of Variant Detection in FFPE Specimens

Tumor biopsies typically vary in the number of tumor cells present and in the molecular composition of different subclones.[8] We assessed the ability of this deep sequencing methodology to detect known cancer variants over a wide range of variant frequency on both NGS platforms.

A 2-fold dilution series of FFPE tumor specimens containing known variants was prepared in a background of wild-type FFPE DNA. One specimen (FFPE1) harbored mutations in KRAS (A146T) and PIK3CA (H1047R), while a second specimen (FFPE2) contained a BRAF V600E mutation.

	GAIIX			PGM		
	Cell Line	FFPE	FNA	Cell Line	FFPE	FNA
Average # of reads/ amplicons	23,570	26,628	12,200	NA	2,242	2,829
Amplicon Coverage*	97.5%	95.2%	97.3%	NA	100%	100%

* within 5X of the median

Table 2. The average total number of reads per sample and percent coverage reflects the uniform coverage and high depth of the sequencing results. The target amplicons were well represented in all sample types and with both platforms. Uniform coverage is important to provide the appropriate statistical power for variant calling across all regions of interest.

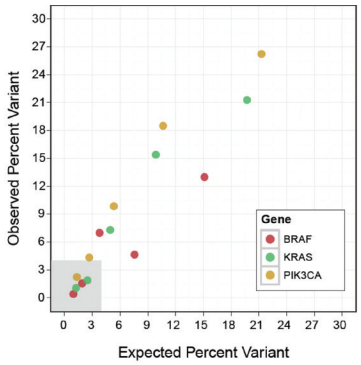

Figure 2. Observed versus expected percent variant from a tumor FFPE DNA dilution series. Tumor samples presented 50% to 3.1% relative tumor burden in a wild-type FFPE DNA background. The background threshold was set at 4%. Five variants were observed under the threshold (shaded area), but above the background variant noise.

Figure 2 shows the observed versus expected percent variant for each variant as a function of the dilution series for each FFPE sample. A dose-dependent recovery of both variants was observed from 50% to 3.1% relative tumor burden in the wild-type FFPE background. This illustrates the quantitative capabilities of this NGS methodology and the potential for identifying variants that are present at a low frequency in clinical specimens. Importantly, all known variants input at >4% relative abundance were correctly detected using the bioinformatic calling algorithms.

Precision of Mutation Quantification in FFPE Tumor DNA

To assess the reproducibility and precision of the NGS procedures to quantify known mutations, DNA from two colon FFPE blocks was processed for targeted NGS across multiple operators and PGM instruments. These two FFPE DNA samples were previously characterized by Sanger sequencing and known to present either BRAF V600E or PIK3CA H1047R mutations.

Across 6 independent runs, the BRAF V600E mutation was detected at 14.5±2.8%, whereas the PIK3CA H1047R mutation was detected at 31.8±2.5%. Thus, mutation quantification was reliable, and was determined within 20% CV from operator-to-operator, instrument-to-instrument, and day-to-day, even for lower abundance variants.

NGS Analysis of Thyroid Specimens

The mutation status of 10 thyroid specimens was assessed with the 16 gene panel on the GAIIx and with the modular five gene panel (Table 1) on the PGM. Specimens were classified as malignant, suspicious or benign after histopathological examination of FNA. Mutations were identified in 7 of the 10 samples, including BRAF V600E, NRAS Q6IR, and HRAS G13R. All variants in the malignant and suspicious specimens were confirmed with Sanger sequencing. In a specimen that was classified as benign, a complex KRAS variant was identified at a low level (8%) on the GAIIx which was not detected on the PGM.

FNA samples were found to be well suited for NGS analysis, particularly when using targeted, multiplexed panels with low DNA input requirements (≤50 ng) that can accommodate the poor nucleic acid yields that may be obtained from such biopsies. NGS offers the potential for detecting rare or unusual cancer-related variants at a high sensitivity which may warrant additional testing, more aggressive treatment or closer monitoring of the patient.

Concordance with Sanger Sequencing and Signature® KRAS Mutation Assays

Variants in the regions most commonly associated with cancer (mutations in KRAS codons 12 and 13, BRAF 600, and PIK3CA exons 9 and 20) identified using the NGS workflow were confirmed via two methods, Sanger sequencing and, for a subset of these mutations, the Signature® KRAS assay. A total of 342 loci (i.e. 9 loci * 38 specimens) were evaluated.

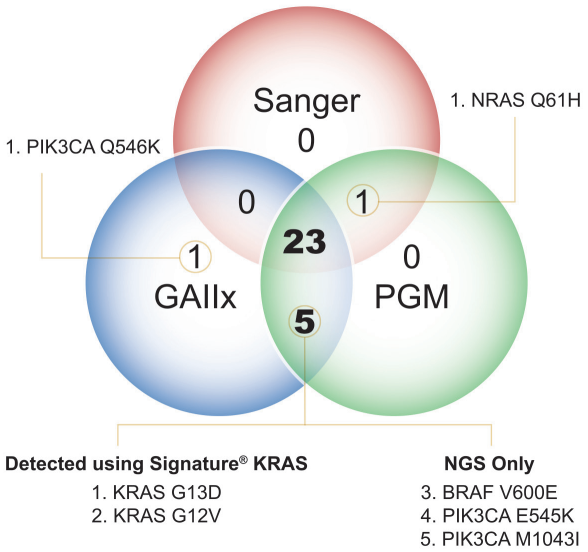

Figure 3. The majority of CRC FFPE variants (23 out of 29) that were identified by the GAIIx were concordant with Sanger sequencing. One variant was only detected by the GAIIx. Five calls that were not detected by Sanger sequencing were detected by both NGS platforms. Of the five, two could be detected with the Signature® KRAS assay, indicating that the greater sensitivity of NGS resulted in the detection of true mutations that were false negatives by Sanger sequencing.

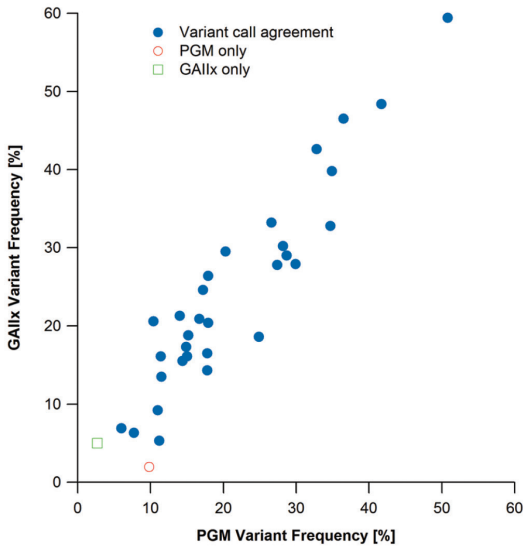

Figure 4. Variant frequency comparison of GAIIx calls (y-axis) versus PGM calls (x-axis) for positive calls above a 4% threshold shows high NGS platform concordance. The two discordant calls were below the 4% threshold.

Of the 29 positively called variants, 23 variants identified by Sanger sequencing were concordant with the NGS variant calls using both platforms. Overall, a 98.5% [95%CI: 96.6-99.5%] agreement (23 positive and 317 wild-type calls) was observed (Figure 3). Importantly, 5/6 mutations not detected by Sanger sequencing were detected by both NGS platforms (Figure 4).

We used the more sensitive Signature® KRAS assay to evaluate 7 variants in KRAS codons 12 and 13. Of the 266 distinct tests (7 loci * 38 specimens) measured using the Signature® KRAS assay, there were 16 positive calls across 16 samples. One sample was not detected by NGS, therefore 15/16 samples were in agreement with targeted NGS analysis. A total of 250/250 negative calls were in agreement. The overall percent agreement of the NGS assay with respect to Signature® KRAS assay was 99.6% [95%CI: 96.1-100%, while the negative percent agreement was 100% [95%CI: 98.5-100%]. This result indicates that variants not detected by the Signature® assay were highly unlikely to be called positive by the NGS assay. As a result, the comparative analysis with the Signature® assay confirms the greater sensitivity of target NGS compared to the Sanger sequencing, and the risk of false negative calls using this “gold reference” methodology.

REFERENCES

1. Nowell, P.C., Cancer Res, 1986. **46**: p. 2203-7.
2. Szerlip, N.J., et al., Proc Natl Acad Sci U S A, 2012. **109**: p. 3041-6.
3. Laosinchai-Wolf, W., et al., J Clin Pathol, 2011. **64**: p. 30-6.
4. Li, H. and R. Durbin, Bioinformatics, 2009. **25**: p. 1754-60.
5. DePristo, M.A., et al., Nat Genet, 2011. **43**: p. 491-8.
6. McKenna, A., et al., Genome Res, 2010. **20**: p. 1297-303.
7. Forbes, S.A., et al., Nucleic Acids Res, 2010. **38**: p. D652-7.
8. Gerlinger, M., et al., N Engl J Med, 2012. **366**: p. 883-92.
9. van Beers, E.H., et al., Br J Cancer, 2006. **94**: p. 333-7.
10. Yancovitz M, et al., PLoS One, 2012. **7**:e29336.
11. Taniguchi K, et al., Cancer Sci, 2008. **99**:929-35.

About Asuragen

Asuragen is a molecular diagnostics company using genomics to drive better patient management through best-in-class clinical solutions. The company uses a breadth of technologies and talent to discover, develop, and commercialize diagnostic products and clinical testing services with efficiency and flexibility both internally and for our companion diagnostic partners. Asuragen’s products, services, and technologies drive countless patient management decisions across oncology, genetic disease and other molecular testing modalities.

ASURAGEN, INC. 2150 Woodward Street, Suite 100 • Austin TX 78744 • tel 512 681 5200 • fax 512 681 5201 • asuragen.com