Video Article

Detecting Somatic Genetic Alterations in Tumor Specimens by Exon Capture and Massively Parallel Sequencing

Helen H Won[1], Sasinya N Scott[1], A. Rose Brannon[1], Ronak H Shah[1], Michael F Berger[1,2]

[1]Department of Pathology, Memorial Sloan-Kettering Cancer Center

[2]Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center

Correspondence to: Michael F Berger at bergerm1@mskcc.org

URL: http://www.jove.com/video/50710
DOI: doi:10.3791/50710

Keywords: Molecular Biology, Issue 80, Molecular Diagnostic Techniques, High-Throughput Nucleotide Sequencing, Genetics, Neoplasms, Diagnosis, Massively parallel sequencing, targeted exon sequencing, hybridization capture, cancer, FFPE, DNA mutations

Date Published: 10/18/2013

Citation: Won, H.H., Scott, S.N., Brannon, A.R., Shah, R.H., Berger, M.F. Detecting Somatic Genetic Alterations in Tumor Specimens by Exon Capture and Massively Parallel Sequencing. *J. Vis. Exp.* (80), e50710, doi:10.3791/50710 (2013).

Abstract

Efforts to detect and investigate key oncogenic mutations have proven valuable to facilitate the appropriate treatment for cancer patients. The establishment of high-throughput, massively parallel "next-generation" sequencing has aided the discovery of many such mutations. To enhance the clinical and translational utility of this technology, platforms must be high-throughput, cost-effective, and compatible with formalin-fixed paraffin embedded (FFPE) tissue samples that may yield small amounts of degraded or damaged DNA. Here, we describe the preparation of barcoded and multiplexed DNA libraries followed by hybridization-based capture of targeted exons for the detection of cancer-associated mutations in fresh frozen and FFPE tumors by massively parallel sequencing. This method enables the identification of sequence mutations, copy number alterations, and select structural rearrangements involving all targeted genes. Targeted exon sequencing offers the benefits of high throughput, low cost, and deep sequence coverage, thus conferring high sensitivity for detecting low frequency mutations.

Video Link

The video component of this article can be found at http://www.jove.com/video/50710/

Introduction

The identification of "driver" tumor genetic events in key oncogenes and tumor suppressor genes plays an essential role in the diagnosis and treatment of many cancers[1]. Large-scale research efforts utilizing massively parallel "next-generation" sequencing have enabled the identification of many such cancer-associated genes in recent years[2]. However, these sequencing platforms typically require large quantities of DNA isolated from fresh frozen tissues, thus posing a major limitation in characterizing and analyzing DNA mutations from preserved tissues, such as formalin-fixed paraffin embedded (FFPE) tumor samples. Improved efforts to efficiently and reliably characterize "actionable" genomic information from FFPE tumor samples will enable the retrospective analysis of previously-banked specimens and further encourage individualized approaches to cancer management.

Traditionally, molecular diagnostic laboratories have relied on time-consuming, low-throughput methodologies such as Sanger sequencing and real-time PCR for DNA mutation profiling. More recently, higher-throughput methods utilizing multiplexed PCR or mass spectrometric genotyping have been developed to investigate recurrent somatic mutations in key cancer genes[3-5]. These approaches, however, are limited in that only predesignated "hotspot" mutations are assayed, making them unsuitable for detecting inactivating mutations in tumor suppressor genes. Massively parallel sequencing offers several advantages over these strategies including the ability to interrogate entire exons for both common and rare mutations, the ability to reveal additional classes of genomic alterations such as copy number gains and losses, and greater detection sensitivity in heterogeneous samples[6, 7]. Whole genome sequencing represents the most comprehensive approach for mutation discovery, though it is relatively expensive and incurs large computational demands for data analysis and storage.

For clinical applications, where only a small fraction of the genome may be of clinical interest, two particular innovations in sequencing technology have been transformative. First, through hybridization-based exon capture, one can isolate DNA corresponding to key cancer-associated genes for targeted mutation profiling[8,] . Second, through ligation of molecular barcodes (*i.e.* DNA sequences 6-8 nucleotides in length), one can pool hundreds of samples per sequencing run and fully take advantage of the ever-increasing capacity of massively parallel sequencing instruments[10]. When combined, these innovations enable tumors to be profiled for lower cost and at higher throughput, with smaller computational requirements[11]. Further, by redistributing sequence coverage to only those genes most critical to the particular application, one can achieve greater sequencing depth for higher detection sensitivity for low allele frequency events.

Here we describe our IMPACT assay (Integrated Mutation Profiling of Actionable Cancer Targets), which utilizes exon capture on barcoded sequence library pools by hybridization using custom oligonucleotides to capture all protein-coding exons and select introns of 279 key cancer-associated genes (**Table 1**). This strategy enables the identification of mutations, indels, copy number alterations, and select structural

Copyright © 2013 Journal of Visualized Experiments

IDT Exhibit 1063.0001

rearrangements involving these 279 genes. Our method is compatible with DNA isolated from both fresh frozen and FFPE tissue as well as fine needle aspirates and other cytology specimens.

Protocol

1. DNA and Reagent Preparation

Note: This protocol describes the simultaneous processing and analysis of 24 samples (*e.g.* 12 tumor/normal pairs) but can be adapted for smaller and larger batches. DNA samples may derive from FFPE or fresh frozen tissue, cytological specimens, or blood. Typically, both tumor and normal tissue from the same patient will be profiled together in order to distinguish somatic mutations from inherited polymorphisms. The protocol begins immediately following DNA extraction.

1. Aliquot 50-250 ng (250 ng recommended) of extracted DNA per sample, diluted in 1x Tris-EDTA (pH 8.0) buffer to a final volume of 50 µl, into separate Covaris round bottom tubes.
2. Shear the DNA on the Covaris E220 instrument at the following shear settings: 360 sec at 10 duty factor, 175 PIP, and 200 cycles per burst.
3. Thaw AMPure XP Beads (**Table 2**) to room temperature. Thaw all buffers and enzymes on ice prior to the appropriate step.

2. Library Preparation - End Repair Module

1. Prepare the end repair master mix in a sterile microcentrifuge tube using the following volumes per sample: 10 µl of 10x NEBNext End Repair Reaction Buffer (**Table 2**), 5 µl NEBNext End Repair Enzyme Mix (**Table 2**), 35 µl sterile nuclease free H_2O.
2. Aliquot the 50 µl of each sheared DNA into separate wells of a 96-well plate. Add 50 µl of the prepared end repair master mix to each reaction. Incubate in a thermocycler for 30 min at 20 °C.

3. Post End Repair Clean-up

1. Add 2x volume (200 µl) of AMPure XP Beads (**Table 2**) to each sample and gently pipette the entire volume up and down 10x using a multichannel pipettor. Incubate at RT for 15 min.
2. Place on the magnetic stand (**Table 2**) at RT for 15 min until the sample appears clear.
3. Gently remove and discard clear supernatant taking care not to disturb beads. Some liquid may remain in wells.
4. With tubes on the stand, gently add 200 µl of freshly prepared 80% ethanol to each sample and incubate plate at room temperature for 30 sec. Carefully, remove ethanol by pipette.
5. Repeat step 5, for a total of 2 ethanol washes. Ensure all ethanol has been removed. Remove from the magnetic stand and let dry at RT for 5 min.
6. Resuspend dried beads with 44.5 µl sterile H_2O. Gently, pipette entire volume up and down 10x mixing thoroughly. Ensure beads are no longer attached to the side of the well.
7. Incubate resuspended beads at RT for 2 min. Place on magnetic stand for 5 min until the sample appears clear.
8. Gently transfer 42 µl of clear supernatant containing the sample to a new well.

   The procedure may be safely stopped at this step and samples stored at -20 °C for up to 7 days. To restart, thaw frozen samples on ice before proceeding.

4. Library Preparation - dA-Tailing Module

1. Prepare the dA-Tailing master mix in a sterile microcentrifuge tube using the following volumes per sample: 5 µl of 10x NEBNext dA-Tailing Reaction Buffer (**Table 2**), 3 µl (3'-5' exo-) Klenow Fragment (**Table 2**).
2. Add 8 µl of the prepared dA-Tailing master mix to each well containing 42 µl of end-repaired DNA. Incubate in a thermocycler for 30 min at 37 °C.
3. Perform a post dA-Tailing clean-up: Repeat steps 3.1-3.8 using 2x volume (100 µl) of beads, but resuspend in sterile H_2O to a final volume of 33.75 µl.

   The procedure may be safely stopped at this step and samples stored at -20 °C for up to 7 days. To restart, thaw frozen samples on ice before proceeding.

5. Library Preparation - Adapter Ligation Module

1. Prepare the Ligation master mix in a sterile microcentrifuge tube using the following volumes per sample: 10 µl of 5x NEBNext Quick Ligation Reaction Buffer (New England Biolabs, **Table 2**), 5 µl Quick T4 DNA Ligase (**Table 2**).
2. Add 15 µl of the Ligation master mix to each well containing the dA-Tailed DNA.
3. Add 1.25 µl of the appropriate 25 µM NEXTflex barcoded adapter (**Table 2**) to each well. Each sample should receive a separate barcoded adapter. Incubate in a thermocycler for 15 min at 20 °C.
4. Perform a post adapter ligation clean-up: Repeat steps 3.1-3.8 using 1x volume (50 µl) of beads, but resuspend in sterile H_2O to a final volume of 50 µl.
5. Repeat steps 3.1-3.8 for a second 1x volume (50 µl) clean-up, but resuspend in sterile H_2O to a final volume of 23 µl.

Copyright © 2013 Journal of Visualized Experiments

IDT Exhibit 1063.0002

The procedure may be safely stopped at this step and samples stored at -20 °C for up to 7 days. To restart, thaw frozen samples on ice before proceeding.

6. Library Amplification

1. Prepare the KAPA HiFi master mix in a sterile microcentrifuge tube using the following volumes per sample: 25 µl 2x KAPA HiFi HS RM (**Table 2**), 1 µl each of 100 µM Primer 1 and Primer 2 (**Table 3**).
2. Add 27 µl of the master mix to each well containing the ligation product. Set pipette to 50 µl and gently pipette the entire volume up and down 10x.
3. Place in the thermocycler for the following PCR cycles: initial denaturation of 45 sec at 98 °C, 10 cycles of 15 sec at 98 °C, 30 sec at 60 °C, 30 sec at 72 °C, and final extension of 1 min at 72 °C.
4. Perform a post amplification clean-up: Repeat steps 3.1-3.8 using 1x volume (50 µl) of AMPure XP Beads, and resuspend in sterile water to a final volume of 30 µl.
5. Quantify the concentration of each library using Qubit Broad Range Assay (Life Technologies, **Table 2**) according to the manufacturer's instructions.

7. Roche Nimblegen SeqCAP EZ Library Hybridization, Wash, and Amplification

1. Thaw universal and index oligonucleotide blockers (**Table 3**), SeqCap EZ Library, and Nimblegen capture components (COT DNA, 2x hybridization buffer, and hybridization component A, **Table 2**) on ice. All oligonucleotide blockers are stored in working stocks of 1 mM. Our SeqCap EZ Library is a custom-designed pool of capture probes encompassing all protein-coding exons of 279 cancer-associated genes, though other custom and catalog designs may be used.
2. Prepare a 1 mM pool of index oligonucleotide blockers corresponding to the specific barcoded adaptor sequences used in library preparation. Combine 1 µl of each blocker, and vortex.
3. In a new sterile 1.5 ml microcentrifuge tube, add 5 µl of 1mg/ml COT DNA, 2 µl of 1 mM universal blocker and 2 µl of the 1 mM pool of index oligonucleotide blockers.
4. Pool 24 barcoded libraries into a single reaction. Add a total of 1-3 µg of pooled, barcoded sequence library (prepared above) to the capture mixture. For 24 samples, 100 ng per sample is recommended. When pooling tumors and matched normal samples, we recommended adding more tumor library (*e.g.* 100 ng per tumor and 50 ng per normal) so that they are sequenced to greater depth of coverage.
5. Close the tube's cap and make 15-20 holes in the top of the cap with an 18-20 G or smaller needle. Speed Vac the amplified sample library/ COT DNA/oligos in a DNA vacuum concentrator on high heat (60 °C).
6. Once completely dried down, rehydrate with 7.5 µl of 2x hybridization buffer and 3 µl of hybridization component A. The tube should now contain the following: 5 µg COT DNA, variable µg amplified sample library, 1000 pmol of Universal and Index Oligos, 7.5 µl 2x hybridization buffer, 3 µl hybridization component A in a total volume of 10.5 µl.
7. Cover the holes in the cap with a small piece of laboratory tape. Vortex the sample for 10 sec and centrifuge at maximum speed for 10 sec.
8. Incubate the sample at 95 °C for 10 min to denature the DNA, followed by centrifugation at maximum speed for 10 sec at RT.
9. In a new 0.2 ml PCR strip tube, prepare 2.25 µl aliquot of SeqCap EZ Library capture probes and 2.25 µl of sterile nuclease free H_2O and add the contents of step 7.7. Gently pipette the entire volume up and down 10x.
10. Incubate in a thermal cycler at 47 °C for 48-96 hr. Set and maintain the thermal cycler's lid temperature to 57 °C.
11. Follow steps in Chapter 6 of the NimbleGen SeqCap EX Library SR User's Guide v3.0 (**Table 2**) for washing and recovery of the captured DNA.
12. Follows steps in Chapter 7 of the Roche NimbleGen protocol for amplification of the captured DNA with the following modifications:
    - Use 2x KAPA HiFi Polymerase (**Table 2**) instead of Phusion High-Fidelity PCR Master Mix
    - Use 100 µM Primer 1 and Primer 2 listed in **Table 3**.
    - Run the following PCR conditions: initial denaturation at 98 °C for 30 sec, 11 cycles of 98 °C for 10 sec, 60 °C for 30 sec, 72 °C for 30 sec, and final extension at 72 °C for 5 min. Hold at 4 °C.
13. Quantify the concentration of the amplified captured DNA using Qubit High Sensitivity Assay (**Table 2**) according to the manufacturer's instructions.
14. Analyze the captured DNA quality using Agilent Bioanalyzer DNA HS chip (**Table 2**) according to the manufacturer's instructions.
    - The PCR yield should be at least 100 ng (range 100-1,000 ng).
    - The average fragment length should be between 150-400 base pairs.

8. Illumina Hi-Seq Sequencing

1. Dilute the amplified captured DNA to 2 pM, and load the sample on a single lane of an Illumina HiSeq 2000 flow cell. Sequence paired-end 75 base pair reads according to the manufacturer's instructions. In our experience, one lane is sufficient to produce 500-700x unique sequence coverage per sample across 279 genes for a pool of 24 samples.
2. Illumina software (Real Time Analysis) will convert high-resolution images to cluster intensities to base calls and quality scores. Use CASAVA to demultiplex data according to barcode identity and generate individual FASTQ files for each sample.

9. Data Analysis

1. Trim adapter sequences from the 3' end of sequence reads in FASTQ files using Cutadapt (http://code.google.com/p/cutadapt/). The minimum overlap length (-O) is 3 with an error rate (-e) of 1%. The minimum base length for retention of paired reads after trimming is 25.
2. Align sequence reads to the reference human genome hg19 using the Burrows-Wheeler Aligner tool (BWA)[12]. Perform post-processing steps including duplicate marking, local multiple sequence realignment, and base quality score recalibration using the Genome Analysis

Copyright © 2013 Journal of Visualized Experiments

IDT Exhibit 1063.0003

Toolkit (GATK) according to their standard best practices (http://www.broadinstitute.org/gatk/guide/topic?name=best-practices)[13]. When profiling multiple samples from the same patient (*e.g.* matched tumor and normal), local multiple sequence realignment should be performed jointly. The output of these post-processing steps is a separate BAM file for each sample, which contains all sequence, quality, and alignment information and can be loaded directly into the Integrative Genomics Viewer (IGV) for visualization of reads and sequence variants[14].

3. Calculate sequence performance metrics using Picard tools (http://picard.sourceforge.net/). Informative metrics include alignment rate, fragment size distribution, GC-content associated coverage bias, on-target capture specificity, PCR duplicate rate, library complexity, and mean target coverage. Representative values are displayed in **Figure 1**. Allele frequencies at common polymorphism sites calculated using DepthOfCoverage from GATK are used to monitor contamination from unrelated DNA and to ensure that matched samples came from the same patients.

The following computational analyses depend on the presence of matched tumor and normal samples for the detection of somatic genetic alterations. Unmatched tumors may also be analyzed using a separate normal control sample, but most somatic mutations cannot easily be distinguished from inherited sequence variants.

4. Call somatic single nucleotide variants (SNVs) for each tumor-normal pair using a somatic mutation caller, such as muTect[15], SomaticSniper[16], or Strelka[17]. We use muTect with modified filtering criteria, allowing for variants with low (non-zero) allele counts in the normal sample as long as the allele frequency is at least 5x greater in the tumor. Variants recurrently called in a panel of normal DNAs are removed as likely systematic artifacts of sequencing or alignment. Each SNV passing filters is then carefully reviewed using IGV.
5. Call somatic indels for each tumor-normal pair using algorithms such as SomaticIndelDetector[13] , Dindel[18], or SOAPindel[19]. We use SomaticIndelDetector with the following criteria: at least 10% allele frequency with at least 3 supporting reads, or at least 2% allele frequency with at least 10 supporting reads. We allow for indels with low (non-zero) allele counts in the normal sample as long as the allele frequency is at least 5x greater in the tumor. Indels called in a panel of normal DNAs are removed, as above. Each indel passing filters is then carefully reviewed using IGV.
6. Extrapolate copy number status from the sequence coverage at target exons. For each individual sample, perform a Loess normalization of mean sequence coverage for all target exons, regression over GC percentage. Adjust the unnormalized coverage values by subtracting the Loess fit and adding the sample-wide median coverage. Calculate the ratio in normalized sequence coverage across all target exons for tumor-normal pairs. In the absence of a matched normal sample and/or to decrease noise, other gender-matched diploid normal controls lacking germline copy number variants may be used. Determine somatic copy number gains and losses based on increases or decreases in the coverage ratio.
7. Call somatic rearrangements involving cancer genes where at least one genomic breakpoint falls within or near a targeted interval of the genome. Suggested algorithms include GASV[20], Breakdancer[21], CREST[22] , and dRanger[23]. Intrachromosomal inversions, deletions, and tandem duplications may be distinguished based on the relative position and orientation of supporting reads in discordant pairs. All candidate rearrangements should be manually reviewed using IGV. We recommend experimental validation by PCR and Sanger sequencing across the putative rearrangement breakpoints.

Representative Results

One pool of 24 barcoded sequence libraries (12 tumor-normal pairs) was captured using probes corresponding to all protein-coding exons of 279 cancer genes and sequenced as 2 x 75 bp reads on a single lane of a HiSeq 2000 flow cell. Tumor and normal libraries were pooled in a 2:1 ratio. Sample performance metrics for a pool of frozen tumor DNA samples are shown in **Figure 1**, including alignment rate, fragment size distribution, on-target capture specificity, and mean target coverage. Example somatic mutations, insertions, deletions, and copy number alterations are shown in **Figures 2-4**.

Copyright © 2013 Journal of Visualized Experiments

IDT Exhibit 1063.0004

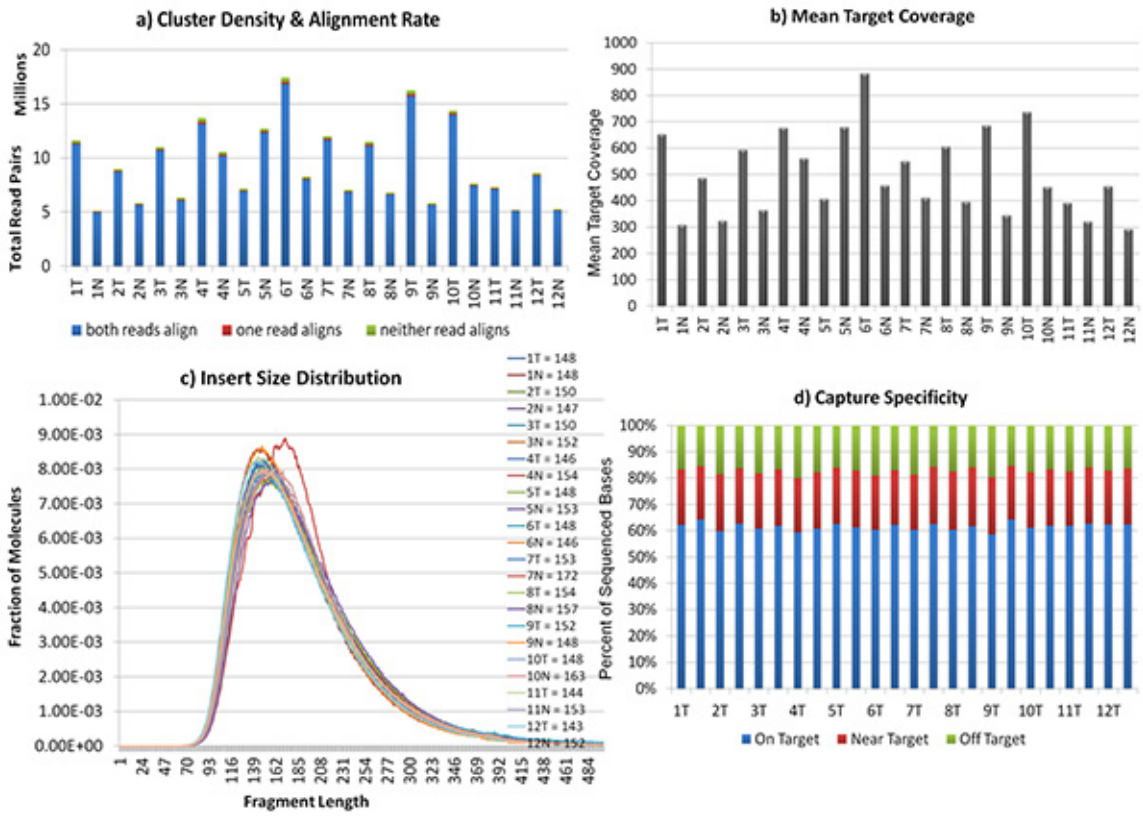

Figure 1. Representative figures of the sequence performance metrics. a) cluster density and alignment rate, **b**) mean target coverage, **c**) insert size distribution, and **d**) capture specificity. Click here to view larger image.

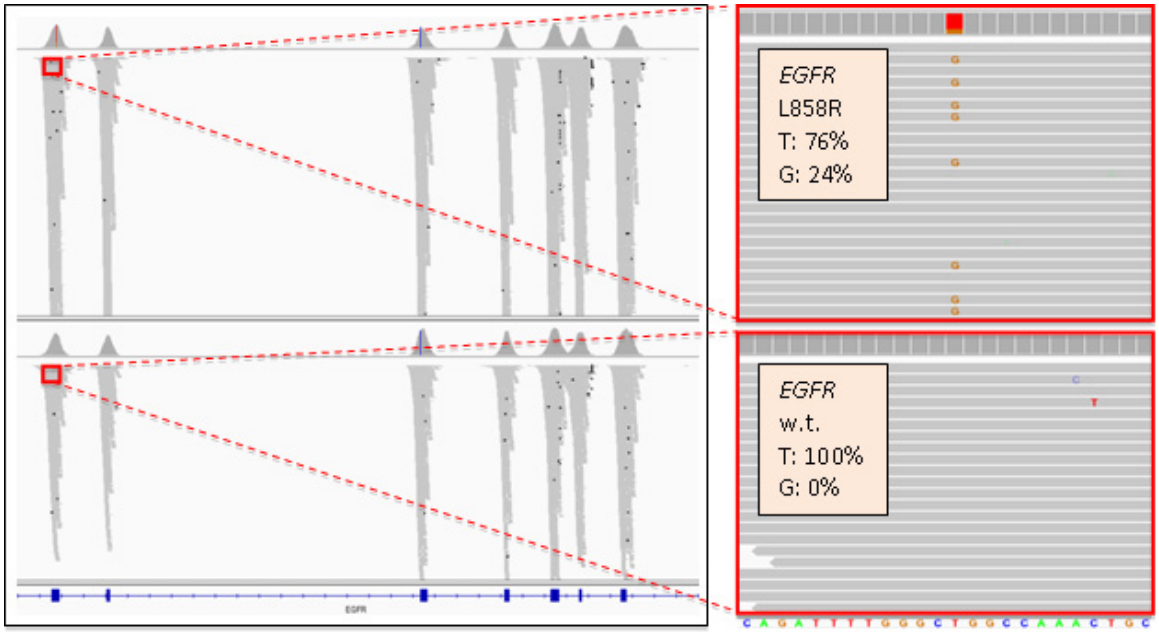

Figure 2. Integrative Genomics Viewer (IGV) image showing specificity of sequence coverage at exons of EGFR in a lung cancer (top) and matched normal tissue (bottom). Gray bars represent unique sequence reads. At right, a heterozygous T>G mutation is present in 24% of reads in the tumor and 0% of reads in the normal, indicating that it is a somatic L858R amino acid substitution. Click here to view larger image.

IDT Exhibit 1063.0005

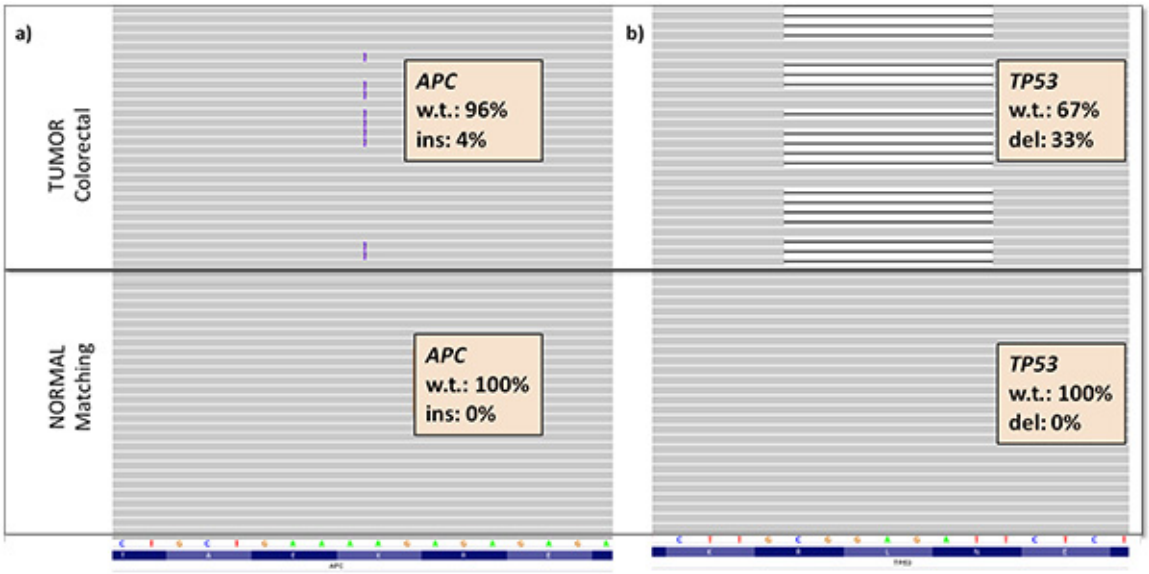

Figure 3. Examples of indels in a colorectal cancer tumor-normal pair: a) Somatic frameshift insertion in APC. **b**) Somatic frameshift deletion of 7 base pairs in TP53. Click here to view larger image.

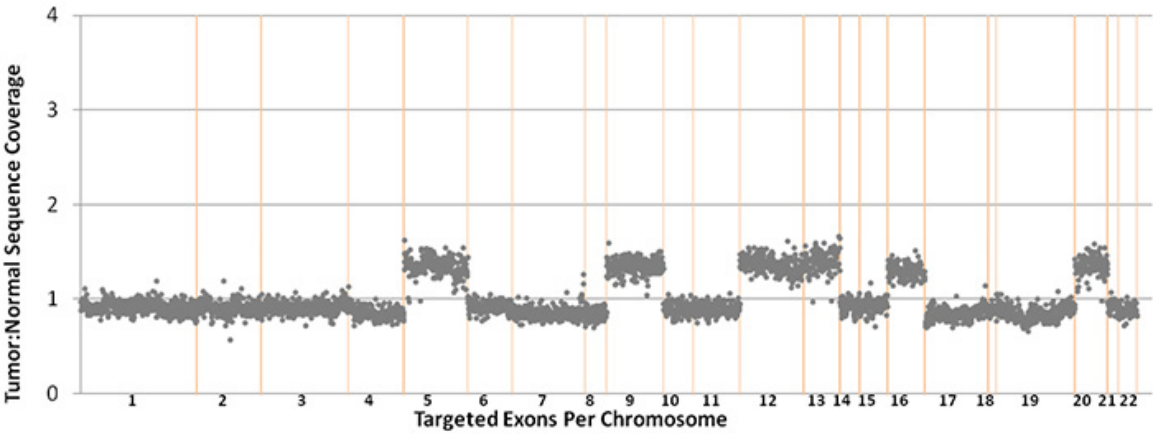

Figure 4. Copy number alterations in a tumor-normal pair. Each data point represents a single exon from 279 target genes. Copy number gains and losses are inferred from increases and decreases in tumor sequence coverage. Click here to view larger image.

Table 1: Probe Design Features for Target Capture.

Total exons	4,535
Total introns	14
SNPs*	1042
Total target territory	879,966 basepairs
Total probe territory	1,400,415 basepairs

Table 2. Reagents and Kits.

IDT Exhibit 1063.0006

Name of Reagent/Material	Company	Catalog Number
NEBNext End Repair Module	New England Biolabs	E6050L
NEBNext dA-Tailing Module	New England Biolabs	E6053L
NEBNext Quick Ligation Module	New England Biolabs	E6056L
Agencourt AMPure XP	Beckman Coulter Genomics	
NEXTflex PCR-Free Barcodes – 24	Bioo Scientific	514103
HiFi Library Amplification Kit	KAPA Biosystems	KK2612
COT Human DNA, Fluorometric Grade	Roche Diagnostics	05 480 647 001
NimbleGen SeqCap EZ Hybridization and Wash kit	Roche NimbleGen	05 634 261 001
SeqCap EZ Library Baits	Roche NimbleGen	
QIAquick PCR Purification Kit	Qiagen	28104
Qubit dsDNA Broad Range (BR) Assay Kit	Life Technologies	Q32850
Qubit dsDNA High Sensitivity (HS) Assay Kit	Life Technologies	Q32851
Agilent DNA HS Kit	Agilent Technologies	5067-4626, 4627
Agilent 2100 Bioanalyzer	Agilent Technologies	
Covaris E220	Covaris	
Magnetic Stand-96	Ambion	AM10027
Illumina Hi-Seq 2000	Illumina	

Table 3. Index Oligo Blockers & Primers Sequences.

Copyright © 2013 Journal of Visualized Experiments

IDT Exhibit 1063.0007

TS-HE Universal Blocker	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG
TS-INV-HE Index 1 Blocker	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 2 Blocker	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 3 Blocker	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 4 Blocker	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 5 Blocker	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 6 Blocker	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 7 Blocker	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 8 Blocker	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 9 Blocker	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 10 Blocker	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 11 Blocker	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 12 Blocker	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 13 Blocker	CAAGCAGAAGACGGCATACGAGATTTGACTGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 14 Blocker	CAAGCAGAAGACGGCATACGAGATGGAACTGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 15 Blocker	CAAGCAGAAGACGGCATACGAGATTGACATGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 16 Blocker	CAAGCAGAAGACGGCATACGAGATGGACGGGTGACTGGAGTTCAGACGTGTG
TS-INV-HE Index 18 Blocker	CAAGCAGAAGACGGCATACGAGATGCGGACGTGACTGGAGTTCAGACGTGTG
TS-INV-HE Index 19 Blocker	CAAGCAGAAGACGGCATACGAGATTTTCACGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 20 Blocker	CAAGCAGAAGACGGCATACGAGATGGCCACGTGACTGGAGTTCAGACGTGTG
TS-INV-HE Index 21 Blocker	CAAGCAGAAGACGGCATACGAGATCGAAACGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 22 Blocker	CAAGCAGAAGACGGCATACGAGATCGTACGGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 23 Blocker	CAAGCAGAAGACGGCATACGAGATCCACTCGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 25 Blocker	CAAGCAGAAGACGGCATACGAGATATCAGTGTGACTGGAGTTCAGACGTGTGC
TS-INV-HE Index 27 Blocker	CAAGCAGAAGACGGCATACGAGATAGGAATGTGACTGGAGTTCAGACGTGTGC
Primer 1	**5'-**AATGATACGGCGACCACCGAGATCT-3'
Primer 2	**5'-**CAAGCAGAAGACGGCATACGAGAT-3'

Discussion

Our IMPACT assay produces a high alignment rate, high on-target rate, high target coverage, and high sensitivity for detecting mutations, indels, and copy number alterations. We have demonstrated the capability of our IMPACT assay to sequence DNA from both fresh frozen and archived FFPE samples of low DNA input. By performing targeted exon sequencing of key cancer-associated genes, one can achieve very deep sequence coverage for the exons of these most critical genes thereby maximizing the ability to detect low frequency mutations. The use of barcoding and multiplexing enables higher throughput, lower cost per sample, and lower input amounts of DNA.

A benefit of the targeted approach is that it is feasible to optimize capture probes such that all exons are covered uniformly and sufficiently for mutation detection. As shown in **Table 1**, our design encompasses 4,535 exons in 279 genes. Following iterative improvements to our design, a typical sequencing experiment produces no more than 2% of exons at less than 20% of the median coverage for the entire sample. Thus, for a tumor sequenced to 500x sequence coverage (a conservative estimate as displayed in **Figure 1**), >98% of exons will be covered by >100x depth, guaranteeing high sensitivity for identifying low frequency mutations. This uniformity of coverage can be largely attributed to the flexibility of the Nimblegen SeqCap system, where probes may be synthesized to different lengths in regions of different nucleotide composition. We have also successfully used individually synthesized biotinylated oligonucleotides from Integrated DNA Technologies (IDT) to add new content to our panel and to boost the coverage of regions that are difficult to capture and/or sequence. By also capturing select introns of recurrently rearranged genes, we are able to identify structural rearrangements producing fusion genes even when only one fusion partner is captured.

This targeted sequencing approach does possess two main limitations in the ability to detect "driver" genetic alterations in tumors. First, only a fraction of the genome is interrogated. Targeted sequencing is, by nature, limited. As new cancer genes emerge from systematic genomics efforts, they may be rapidly added to a capture panel, but functionally important mutations may be missed that would be detected by whole genome or whole exome sequencing. Second, it is limited to DNA-based alterations. The detection of aberrantly expressed transcripts, novel splice isoforms, and gene fusions requires RNA-Seq or similar techniques [24]. Epigenetic alterations such as chromatin modifications and DNA methylation may also play key roles in tumorigenesis and tumor progression. As sequencing technologies continue to improve, one can envision integrated strategies employing whole genome sequencing, RNA-Seq and complementary approaches to comprehensively characterize the genetic and epigenetic make-up of individual tumors on a routine basis [25]. As complete integrated strategies are currently cost- and

IDT Exhibit 1063.0008

computationally-prohibitive, targeted sequencing presents a practical alternative to capture the most salient genomic features for clinical and translational applications.

Our protocol assumes that tumors are sequenced alongside matched normal tissue taken from the same patient. The purpose of this normal control is to determine whether sequence variants detected in the tumor are inherited or somatically acquired. In the absence of a matched normal, one can infer that variants occurring at mutational hotspots (*i.e.* sites of recurrent somatic mutations in other cancers) are likely to be somatic. Further, analysis of copy number gains and losses can be performed with a non-genetically matched diploid normal sample. For all other novel genetic alterations, the simultaneous analysis of matched normal tissue greatly simplifies the interpretation of the resulting mutation data and is highly recommended.

The IMPACT protocol described here demonstrates consistency and high quality performance for frozen samples. We have observed occasional variability when working with FFPE samples, though. Depending on the age and quality of the FFPE sample, the DNA may be severely degraded or chemically-modified, and as a result may render sequencing and analysis difficult [26]. That being said, we have found these experimental conditions to be reliable for the large majority of FFPE samples. We have also successfully applied this protocol to characterize other clinically relevant specimens such as biopsies, fine needle aspirates, and cytologic specimens. It is for this reason-the flexibility to accommodate different types of specimens of variable quality, quantity, and heterogeneity-that this method stands to impact the diagnosis and treatment of cancer patients in the clinical arena.

Disclosures

Dr. Berger has received consulting fees and research funding from Foundation Medicine, a cancer diagnostics company.

Acknowledgements

We thank Dr. Agnes Viale and the MSKCC Genomics Core Laboratory for technical assistance. This protocol was developed with support from the Geoffrey Beene Cancer Research Center and the Farmer Family Foundation.

References

1. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature.* **458**(7239), 719-724 (2009).
2. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat. Rev. Genet.* **11**(10), 685-696 (2010).
3. Thomas RK, Baker AC, Debiasi RM *et al.* High-throughput oncogene mutation profiling in human cancer. *Nat. Genet.* **39**(3), 347-351 (2007).
4. Macconaill LE, Campbell CD, Kehoe SM *et al.* Profiling critical cancer gene mutations in clinical tumor samples. *PLoS One.* **4**(11), e7887 (2009).
5. Dias-Santagata D, Akhavanfard S, David SS *et al.* Rapid targeted mutational analysis of human tumours: a clinical platform to guide personalized cancer medicine. *EMBO Mol. Med.* **2**(5), 146-158 (2010).
6. Macconaill LE, Van Hummelen P, Meyerson M, Hahn WC. Clinical implementation of comprehensive strategies to characterize cancer genomes: opportunities and challenges. *Cancer Discov.* **1**(4), 297-311 (2011).
7. Taylor BS, Ladanyi M. Clinical cancer genomics: how soon is now. *J. Pathol.* **223**(2), 318-326 (2011).
8. Gnirke A, Melnikov A, Maguire J *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **27**(2), 182-189 (2009).
9. Mamanova L, Coffey AJ, Scott CE *et al.* Target-enrichment strategies for next-generation sequencing. *Nat. Methods.* **7**(2), 111-118 (2010).
10. Craig DW, Pearson JV, Szelinger S *et al.* Identification of genetic variants using bar-coded multiplexed sequencing. *Nat. Methods.* **5**(10), 887-893 (2008).
11. 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.* **2**(1), 82-93 (2012).
12. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* **25**(14), 1754-1760 (2009).
13. Depristo MA, Banks E, Poplin R *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**(5), 491-498 (2011).
14. Robinson JT, Thorvaldsdottir H, Winckler W *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**(1), 24-26 (2011).
15. Cibulskis K, Lawrence MS, Carter SL *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**(3), 213-219 (2013).
16. Larson DE, Harris CC, Chen K *et al.* SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics.* **28**(3), 311-317 (2012).
17. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: Accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics.* **28**(14), 1811-1817 (2012).
18. Albers CA, Lunter G, Macarthur DG, Mcvean G, Ouwehand WH, Durbin R: Dindel. Accurate indel calls from short-read data. *Genome Res.* **21**(6), 961-973 (2011).
19. Li S, Li R, Li H *et al.*: SOAPindel. Efficient identification of indels from short paired reads. *Genome Res.* **23**(1), 195-200 (2013).
20. Sindi S, Helman E, Bashir A, Raphael BJ. A geometric approach for classification and comparison of structural variants. *Bioinformatics.* **25**(12), i222-230 (2009).
21. Chen K, Wallis JW, Mclellan MD *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat. Methods.* **6**(9), 677-681 (2009).
22. Wang J, Mullighan CG, Easton J *et al.* CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat. Methods.* **8**(8), 652-654 (2011).

Copyright © 2013 Journal of Visualized Experiments

IDT Exhibit 1063.0009

23. Drier Y, Lawrence MS, Carter SL *et al.* Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability. *Genome Res.* **23**(2), 228-235 (2012).
24. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**(1), 57-63 (2009).
25. Roychowdhury S, Iyer MK, Robinson DR *et al.* Personalized oncology through integrative high-throughput sequencing: a pilot study. *Sci. Transl. Med.* **3**(111), 111ra121 (2011).
26. Kerick M, Isau M, Timmermann B *et al.* Targeted high throughput sequencing in clinical cancer settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity. *BMC Med. Genomics.* **4**, 68 (2011).

Copyright © 2013 Journal of Visualized Experiments

IDT Exhibit 1063.0010