Delineation of Complex Genomic Alterations via Iso-Seq in a Comprehensive Genomic Profiling Protocol for Pediatric Cancer

Vincent Magrini , Sean D. McGrath , Anthony Miller, Kathleen Schieffer, Kyle Voytovich, James Fitch, Natalie Bir, Amy Wetzel, Elizabeth Varga, Selene Koo, Peter White, Richard K. Wilson , Elaine R. Mardis, Catherine E. Cottrell

Institute for Genomic Medicine, Nationwide Children's Hospital

Introduction

The Institute for Genomic Medicine developed a translational protocol to evaluate the genomic landscape of cancer and hematologic disease amid our patient population. This comprehensive approach allows for detection of germline and somatic single nucleotide variants, small indels, copy number alterations, structural variation and expression. Additional methodologies to further delineate complex genomic alterations are employed including long-read technologies. To date, we have enrolled 63 subjects and have sequenced 55 cases using Illumina short-read chemistries.

Following protocol enrollment, genomic profiling methodologies are employed using an **N-of-1** approach tailored to the patient and disease status.

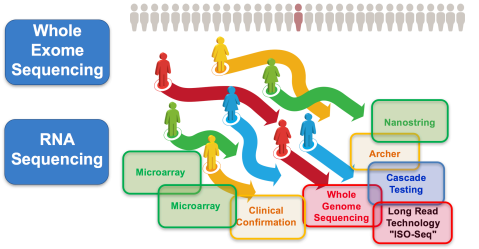


Figure 1. Our N-of-1 sequencing approach uses Whole Exome and RNA Sequencing for an initial genomic evaluation. Additional genomic profiling is performed on case-by-case basis including long read sequencing technologies.

Enhanced Whole Exome Sequencing :
Disease and comparator normal samples target 250X average coverage depth. The "enhanced exome" includes the IDT xGen exome plus two additional IDT probe panels: the IDT xGen CNV backbone panel (9385 probes) to help identify copy number variants, and >2400 probes among clinically relevant cancer loci implicated in copy number alteration and mutation.

RNA-Sequencing:
Whole transcriptome profiling by short read RNA-Seq. Total RNA is processed using the NEBNext Ultra II Directional library prep kit for Illumina and targeting 80-100 million paired-end reads per library. The RNA analysis pipeline uses an ensemble approach for fusion detection. Current tools in the fusion pipeline include JAFFA, FusionMap, MapSplice, SOAPfusion, FusionCatcher, StarFusion, and TopHat-Fusion.

Somatic events of diagnostic, prognostic, or therapeutic relevance included inter or intra-genic structural variation in 16 patient samples. A subset of these were evaluated by Pacific Biosciences Iso-Seq to characterize novel isoforms and resolve complex rearrangements (Table 1).

Table 1. Somatic fusions detected by RNA-Seq

Diagnosis	Fusion	Junction
Alveolar rhabdosarcoma	*PAX3-FOXO1*	GTGTAGGGACAGATTATGACGAATTGAATT@ctgag gtgagaggccattgccaatggtggg
Ependymoma	*C11orf95-RELA*	CGCTGTCTGAGCTCACCCAGTCCCCACCAG@aactgtt cccctcatcttcccggcaggta
Large cell lymphoma	*NPM1-ALK*	CTTGCAGCTCCTGGTGCTTCCGGCGGTACA@ctactaa gtgctgtccactaatatgcactg
Spindle cell lesion	*COL3A1-PLAG1*	TTTGCTGCTGGTTGGGTATAACTTGATAAC@ctgttcca gcacatccaacttttaactgag
Infantile fibrosarcoma	*RBPMS-MET*	AACCGTTCTGAGATGAATTAGGAAACTGAT@atggctc tctggcaatgaactgaggtacag
Meningeal sarcoma	*PDGFRB*	△GATGGAAGGTGATTG
Spindle cell neoplasm	*PDGFRB*	Complex

Pacific Biosciences Iso-Seq

Iso-Seq is an RNA-Seq protocol aiming to represent full-length RNA transcripts using the Takara SMARTer cDNA library prep (Figure 2). We are incorporating Iso-Seq into our translational protocol to identify expressed RNA isoforms and further characterize structural variation inferred by short read sequencing technologies.

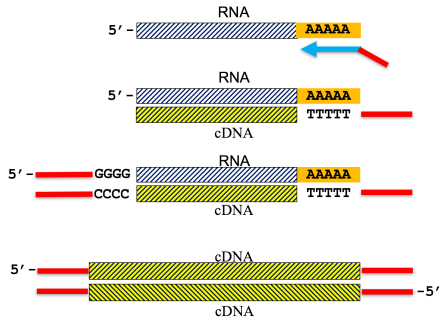
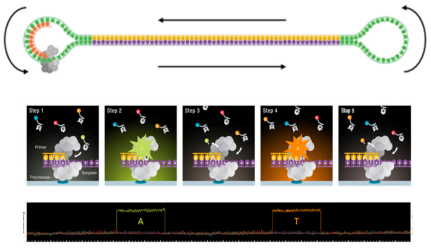


Figure 2. Total RNA is reverse transcribed with a tailed polyA primer (Blue/Red arrow). The RT reaction generates first strand cDNA (yellow hatch //) and an extended RNA template (red bar GGGG). Template switching synthesizes the extended template into first strand cDNA (red bar CCCC). PCR of tailed cDNA templates generates double-stranded cDNA.

Iso-Seq Single Molecule Real Time (SMRT) bell libraries

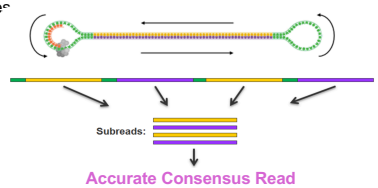
One microgram of double-stranded cDNA is size selected between 500bp and 2kbp and greater than 2kbp. Size-fractioned cDNAs are ligated with PacBio hair-pin adapters generating a SMRTbell configuration. Libraries are primed and complexed with DNA polymerase and loaded into zero mode wave guides (ZMWs).



Iso-Seq SMRT Sequencing

Within the ZMW, DNA replication occurs producing nucleotide-specific fluorescent signals. Circular consensus sequencing (ccs) allows the polymerase to replicate the circularized SMRTbell template for up to 20 hours producing long reads (N50 = 15kb).

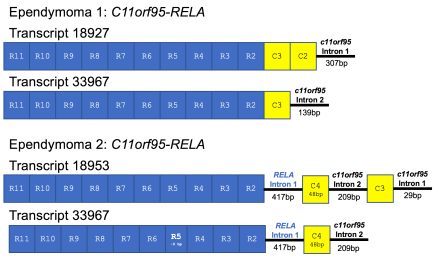
Post sequencing, the SMRTbell adapters are trimmed and the single-molecule fragments are aligned generating an accurate consensus read from a single molecule with multiple sequencing passes.



Accurate Consensus Read

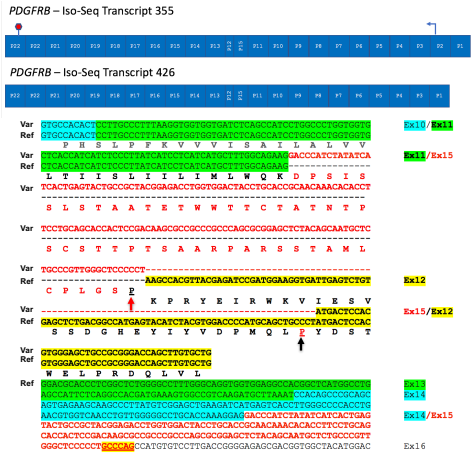
Iso-Seq Results

Identification of *C11orf95-RELA* fusion isoforms in two patients diagnosed with ependymoma allowed for refinement in diagnosis enabling classification as a genetically defined ependymoma subtype: Ependymoma, *RELA* fusion-positive (2016 WHO CNS guidelines).



A complex *PDGFRB* genomic alteration

An exon 11-15 intragenic rearrangement was identified by RNA-Seq in an infant with a spindle cell lesion. Iso-Seq identified two expressed isoforms with a complex rearrangement. We observed one isoform harboring an exon 2 deletion, while both isoforms carried a 154bp repeat of exon 15 sequence inserted between exon 11 and an altered exon 12 sequence (with a 50bp deletion). This alteration results in an in-frame insertion/deletion altering the amino acid sequence of the juxtamembrane domain. Identification of this *PDGFRB* alteration, predicted to be activating in nature, allows for refinement in diagnosis (now best classified as infantile myofibromatosis) and enables potential for use of targeted therapeutics.



Summary

Use of our N-of-1 pediatric cancer-heme protocol has allowed for identification and characterization of known and novel genomic alterations in our patient population using short-read Illumina sequencing methodologies. We are able to take advantage of long-read sequencing technologies to characterize full-length cDNA and resolve the structure of expressed alterations in disease-associated tissue. Complex genomic alterations will continue to be examined with PacBio Iso-Seq at IGM, with future directions to include advanced sequencing chemistries to further enhance resolution of complex genomic alteration allowing for refinement in patient diagnosis and management.