A complex rearrangement of *PDGFRB* elucidated by long-read sequencing

Sean D. McGrath, Huachun Zhong, Anthony R. Miller, Amy Wetzel, James R. Fitch, Kathleen M. Schieffer, Natalie Bir, Selene Koo, Richard K. Wilson, Elaine R. Mardis, Catherine E. Cottrell, Vincent Magrini

Steve and Cindy Rasmussen Institute for Genomic Medicine
The Abigail Wexner Research Institute at Nationwide Children's Hospital

Background

We have implemented a translational protocol evaluating the genomic landscape of rare and refractory cancer and hematological disorders within our pediatric patient population. Our approach uses an enhanced whole exome sequencing (eWES) of disease-associated and normal tissues to identify germline and somatic alterations. Complementing our eWES approach, we perform whole transcriptome analysis for gene expression profiling, pathway analysis, and detection of fusion transcripts as a result of genomic rearrangements.

We highlight a case of a female born with a right thigh mass enrolled on our protocol. Illumina methods identified an intra-genic, in-frame, fusion in the juxtamembrane and kinase domains of *PDGFRB* (platelet-derived growth factor receptor-beta) and is a predicted activating alteration.

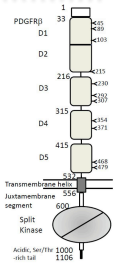
To further characterize the genomic rearrangement contributing to the observed *PDGFRB* intragenic fusion transcript, we report on the use of Pacific Biosciences (PacBio) long read RNA-Seq (Iso-Seq) and High Fidelity (HiFi) Whole Genome Sequencing (WGS) to elucidate the structure of this genomic rearrangements and resulting expressed full-length transcripts.

PDGFRB (platelet derived growth factor receptor beta)

- Complex alteration spanning exon 12 (around amino acid 578) through the start of exon 15 (amino acid 675)
- Mutation affects the juxtamembrane domain and part of the kinase domain

Juxtamembrane domain – inhibition of catalytic activity; mutations in this domain result in ligand-independent activation

Kinase domain – catalytic function; mutations in this domain that result in an active conformation lead to constitutive activation



Chen *et al.* 2013 *Biochim. Biophys. Acta.* 1834:2176-86

PacBio Iso-Seq

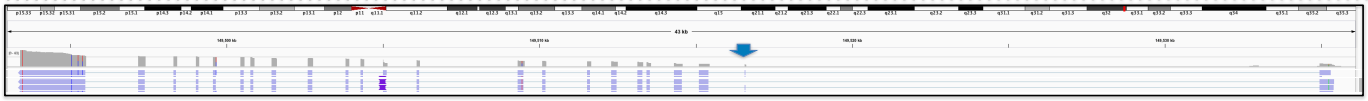


Figure 1A: IGV alignments of a representative *PDGFRB* Iso-Seq reads.

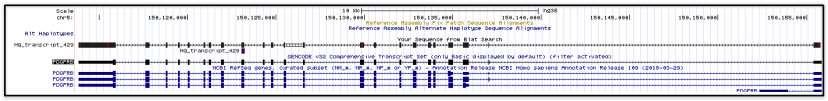


Figure 1B: UCSC Blat alignment of a representative *PDGFRB* Iso-Seq read.

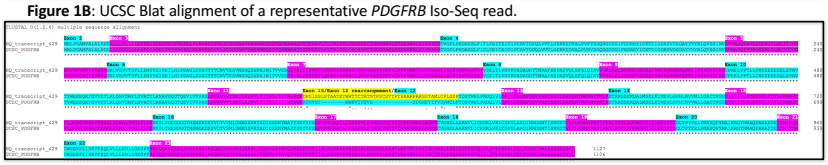


Figure 1C: Clustal Ω comparison of *PDGFRB* wild-type and rearranged amino acid sequence.

Illumina exome and RNA-Seq revealed the replacement of *PDGFRB* Exon 12 with combination of *PDGFRB* Exon 15 (5') and Exon 12 (3'). We prepared Takara SMARTer cDNA from 1 µg tumor total RNA and sequenced the resulting SMRTbell Template Prep 1.0 library on four Sequel 1M SMRT Cells. The circular consensus (CCS) reads were analyzed with PacBio Iso-Seq application in SMRT Link.

Figure 1A: IGV alignments of Iso-Seq full-length wild-type and differentially spliced rearrangement transcripts.
Figure 1B: UCSC Blat alignment of representative Iso-Seq transcript showing the loss of 5' Exon 12 and duplication of the 5' portion of Exon 15..
Figure 1C: Wild-type and rearrangement *PDGFRB* amino acid sequences were aligned with Clustal Ω. The rearrangement transcript results in an open reading frame.

PacBio *PDGFRB* amplicon sequencing

To further examine the *PDGFRB* exon 12 rearrangement, PCR primers were designed to the end of *PDGFRB* exon 11 and the middle of *PDGFRB* exon 12. Amplicons were converted into SMRTbell Template Prep 1.0 libraries and sequenced on Sequel 1M. Single CCS reads were aligned to the human reference in the UCSC Genome Browser.

Figure 2: UCSC Blat alignment of an annotated PacBio amplicon CCS read. The Blat alignments are numbered in the linear order of the CCS read components. The *PDGFRB* wild-type exon 11-intron 12 locus (716 bp) is replaced by a complex rearrangement of *PDGFRB* segments (1374 bp). The rearrangement contains a large, intron 11 inversion (627 bp), deletion of a portion of exon 12 (31 bp) and duplications of intron 14, exon 15 and intron 15.

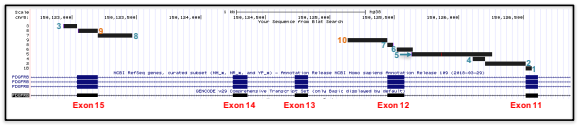


Figure 2: UCSC Blat alignment of a representative *PDGFRB* amplicon CCS read.

PacBio WGS HiFi 15 kb circular consensus sequencing

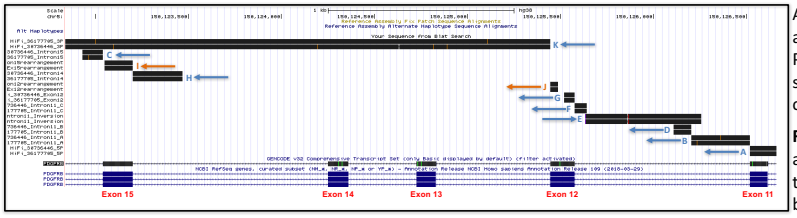


Figure 3: UCSC Blat alignment of two representative *PDGFRB* HiFi CCS reads.

All previous examinations of this tumor, either Illumina or PacBio, required enzymatic amplification. As a final exploration of this complex rearrangement, we prepared a PacBio HiFi library from 3.8 µg genomic DNA that was sheared to 15 kb and sequenced without size-fractionation using the Sequel II 8M and v2 Early Access chemistry to achieve 39X mapped sequence coverage.

Figure 3: UCSC Blat alignment of two representative *PDGFRB* HiFi CCS reads (19.0 and 20.1 kb). These reads encompass the *PDGFRB* exon 12 rearrangement as well as the start codon (in Exon 2) and the stop codon (in Exon 23). The aligned sequence, between exon 11 and intron 15, is labeled in the order represented in the CCS reads.

Discussion

- Genomic profiling identified a complex alteration within the *PDGFRB* juxtamembrane and kinase domains predicting a gain of function mutation that is therapeutically targetable with tyrosine kinase inhibitors.
- Short-read sequencing was unable to accurately resolve this complex genomic rearrangement. To elucidate the structure of this alteration, we performed PacBio long-read sequencing of normal and tumor derived DNA and the tumor derived RNA.
- PacBio Iso-Seq of the tumor RNA revealed full-length wild-type and aberrant *PDGRFB* transcripts. We observed RNA molecules with allele-specific isoform expression that maintain frame predicted to encode an aberrant protein products.
- Long-range amplicon and long read (>15kb) whole genome sequencing (HiFi) support this somatic alteration maintaining both genomic order and orientation (Figure 4).
- The combination of rapid, cost-effective short-read sequencing and comprehensive long-read DNA and RNA sequencing provide the opportunity to diagnose and explain novel, disease-causing genetic alterations.

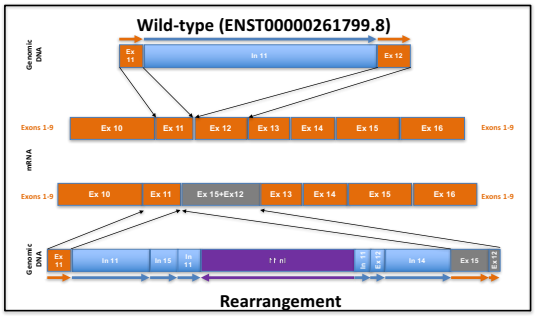


Figure 4: Schematic of *PDGFRB* wild-type and rearrangement genomic DNA and mRNA.