Whole Exome Sequencing

Acute Lymphoblastic Leukemia Phase 2 (ALL P2)
Acute Myeloid Leukemia (AML)
Neuroblastoma Models (NBL MDLS)
Pediatric Preclinical Testing Program (PPTP)
Wilms Tumor (WT)

*Protocols were performed at Baylor College of Medicine.

Library construction

Specimen processing, DNA extraction, standard QC and Illumina paired-end pre-capture libraries were prepared according to the manufacturer's protocol (Illumina Inc, San Diego, CA) with the following modifications: 0.5 - 1ug genomic DNA in 100ul volume was sheared into fragments of approximately 300 base pairs in a Covaris E210 system (Covaris, Inc. Woburn, MA). The setting was 10% duty cycle, intensity of 4,200 cycles per burst for 120 seconds. Fragment size was checked using a 2.2% Flash Gel DNA Cassette (Lonza, Walkersville, MD, Cat. No.57023). End-repair of fragmented DNA was performed in 90ul total reaction volume containing sheared DNA, 9 ul 10X buffer, 5 ul END Repair Enzyme Mix and H2O (NEBNext End-Repair Module, New England BioLabs, Ipswich, MA, Cat. No. E6050L), incubated at 20°C for 30 minutes. A-tailing was performed in a total reaction volume of 60ul containing end-repaired DNA, 6ul 10X buffer, 3ul Klenow fragment (NEBNext dA-Tailing Module; Cat. No. E6053L) and H2O followed by incubation at 37°C for 30 minutes. Illumina multiplex adapter ligation (NEBNext Quick Ligation Module Cat. No. E6056L) was performed in a total reaction volume of 90ul containing 18ul 5X buffer, 5ul ligase, 0.5ul 100uM adaptor and H2O at room temperature for 30 minutes. After ligation, PCR with Illumina PE 1.0 and modified barcode primers (manuscript in preparation) was performed in 170ul reactions containing 85ul of 2x Phusion High-Fidelity PCR master mix, adaptor ligated DNA, 1.75ul of 50uM primers and H2O. PCR was performed using a 5-minute initial denaturation at 95°C, 6-10 cycles of 15 seconds at 95°C, 15 seconds at 60°C and 30 seconds at 72°C followed by a final extension for 5 minute at 72°C. Agencourt XP Beads (Beckman Coulter Genomics, Inc., Danvers, MA, Cat. No. A63882) were used to purify DNA after each enzymatic reaction. After purification, PCR product quantification and size distribution was determined using the Caliper GX 1K/12K/High Sensitivity Assay Labchip (Hopkinton, MA, Cat. No. 760517).

Exome capture

Illumina pre-capture libraries (1ug DNA input) were hybridized in solution to SeqCap EZ Human Exome 2.0 (Nimblegen, Madison, WI) probes targeting approximately 44Mbs of sequence from approximately 30K genes according to the manufacturer's protocol with the following modifications: hybridization enhancing oligos IHE1, IHE2 and IHE3 replaced oligos HE1.1 and HE2.1 and post-capture LM-PCR was performed using 14 cycles. Capture libraries were quantified using Caliper GX 1K/12K/High Sensitivity Assay Labchip (Hopkinton, MA, Cat. No. 760517). The efficiency of the capture was evaluated by performing a qPCR-based quality check

on the built-in controls (qPCR SYBR Green assays, Applied Biosystems, Grand Island, NY). Four standardized oligo sets, RUNX2, PRKG1, SMG1, and NLK, were employed as internal quality controls. The enrichment of the capture libraries was estimated to range from 7- to 9-fold over background.

Library templates were prepared for sequencing using Illumina's cBot cluster generation system with TruSeq PE Cluster Generation Kits (Part no. PE-401-3001). Briefly, these libraries were denatured with sodium hydroxide and diluted to 6-9 pM in hybridization buffer in order to achieve a load density of ~800K clusters/mm2. Each library pool was loaded in a single lane of a HiSeq flow cell, and each lane was spiked with 2% phiX control library for run quality control. The sample libraries then underwent bridge amplification to form clonal clusters, followed by hybridization with the sequencing primer. Sequencing runs were performed in paired-end mode using the Illumina HiSeq 2000 platform. Using the TruSeq SBS Kits (Part no. FC-401-3001), sequencing-by-synthesis reactions were extended for 101 cycles from each end, with an additional 7 cycles for the index read. Sequencing runs generated approximately 300-400 million successful reads on each lane of a flow cell, with approximately 9-10 Gb produced per sample. With these sequencing yields, samples achieved an average of 95% of the targeted exome bases covered to a depth of 20X or greater.

Real Time Analysis (RTA) software was used to process the image analysis and nucleotide base calling. On average, about 80-100 million successful reads, consisting of 2X 100 bp, were generated on each lane of a flow cell.

Mapping Reads

Illumina HiSeq bcl files were processed using BCLConvertor v1.7.1. All reads from the prepared libraries that passed the Illumina Chastity filter were formatted into fastq files. The fastq files were aligned to human reference genome build37 (NCBI) using BWA (bwa-0.5.9-R16) with default parameters with the following exceptions: seed sequence: 40 bpseed mismatch: 2, total mismatches allowed: 3. BAM files generated from alignment were preprocessed using GATK (v1.3-8-gb0e6afe) [1] to recalibrate and locally realign reads.

Mutation Detection

Sequence variants were called from tumor and matched normal BAM files using Atlas [2] an integrative variant analysis suite of tools specializing in the separation of true SNPs and insertions and deletions (indels) from sequencing and mapping errors in whole exome capture sequencing (WXS) data. The suite implements logistic regression models trained on validated WXS data to identify the true variants. ATLAS-SNP-2 (v1.3) [3] and ATLAS-Indel-2 (v0.3.1) along with Pindel (v0.2.4q) [4] were run on the BAM files producing variant data that were further filtered to remove all those observed fewer than 5 times or were present in less than 0.08 of the reads (e.g., variant allele fraction must be greater than 0.08 to undergo validation). At least one variant read of Q30 or better was required, and the variant had to lie in the central portion of the read (15% from the 5' end of the read and 20% from the 3' end). In addition, reads

harboring the variant must have been observed in both forward and reverse orientations. Finally, the variant base was not observed in the normal tissue. Indels were discovered by similar processing except indels must have been observed in at least 10 of the reads.