mRNA Sequencing

Clear Cell Sarcoma of the Kidney (CCSK)

*Protocols were performed at NCI Center for Cancer Research through the laboratory of Dr. Javed Khan.

RNA-seq Library construction and sequencing by Illumina HiSeq2000:

RNA-seq libraries were prepared using Illumina TruSeq Stranded Total RNA Sample Preparation kits according to the manufacturer's protocol. Briefly, ribosomal RNA was removed using Ribo-Zero Gold beads. After purification, total RNA was fragmented to 200nt pieces and then reverse-transcribed using reverse transcriptase and random primers. Second strand cDNA was synthesized using DNA polymerase I and RNase H. These cDNA fragments were added as a single base and ligated with adaptors. The products were purified and enriched with PCR to create the final RNA-seq libraries. RNA libraries were sequenced on Illumina HiSeq2000 using 100bp paired-end sequencing according to the manufacturer's protocol.

Reads Alignment:

Align reads to reference genome (GRCh37) using Tophat version 2.0.8b with default options, expect for options specifying number of processor threads and fusion search. An example code for alignment with fastq files is shown below.
-o tophat.out –p 6 --fusion-search –fusion-min-dist 100000 GRCh37 read_1.fq read_2.fq

Gene and isoform expression:

Gene and isoform expression from RNA-seq data was generated using Cufflinks version 2.1.1. with default options and supplied reference annotation (Homo_sapiens.GRCh37.71.gtf) for estimation of expression. Cufflinks will not assemble novel transcripts, and it will ignore alignments not structurally compatible with any reference transcript.

Exon expression:

RPKM for a given ExonX is determined by: ((raw base counts / median read length) * 10^9) / (total reads * exon length). The raw base counts for a given ExonX is the total number of bases aligned to that genomic segment. Raw base counts are used instead of raw read counts because in many cases only a portion of a read will align to a given exon.

Gene fusion:

Gene fusion file was generated using defuse version 0.6.1 with default parameters and with reference annotation Homo_sapiens.GRCh37.69.