CRL: mutation - Illumina HiSeq 2000, WES

Molecular Methods Description:

Exonic regions from Charles River DNA samples were targeted using Agilent SureSelect Human All Exon kits V1 38MB, V4 51MB, V5 50MB or V6 60MB. Regarding mouse models, exonic regions were targeted using the Agilent SureSelect Mouse All Exon kit V1 50MB. Enriched genomic DNA was sequenced with Illumina HiSeq-2000/2500/4000 in 100 bp, 126 bp or 150 bp paired-end (PE) reads and an expected coverage on targets of ~100X.

Analysis Description:

Isolate Human Reads

Xenome, with an xenome-index built from GRCh38 and GRCm38, was utilized to classify reads from xenograft samples into human, mouse, both, ambiguous and neither. Only human reads were used for further downstream analyses. This step was skipped for non-xenograft samples.

Alignment to Reference Genome

Following the GATK best practices, PE reads were mapped to the respective reference genome (GRCh38 or GRCm38) using the BWA (Burrows-Wheeler Aligner). Next, duplicated reads were marked and removed using PicardTools and the mapping was recalibrated within the targeted exonic regions (manufacturer enrichment kits) with GATK, resulting in a recalibrated .bam file.

Coverage

Per exon, coding region, transcript and gene coverage were calculated from the realigned bam file using BEDTOOLS.

Variant Calling

Within the targeted exonic regions (manufacturer enrichment kits) variants were detected using GATK haplotyper and filters were indicated using GATK VariantFiltration based on variant quality (QUAL < 30), genotype alternate coverage (AD < 3), genotype quality (GQ < 20) and genotype alternate allele frequency (< 0.05). Next, all variants were annotated with ENSEMBL VEP. A filtered list of variants was compiled based on the variant filter (PASS), the protein impact (HIGH and MODERATE), and a population frequency filter (gnomAD: MAX_AF < 0.01). Furthermore, the filtered variants were used to calculate the tumor mutational burden [number of somatic mutations per Mb sequenced].

Quality Assessment

Raw reads were subjected to fastQC to calculate read quality metrics. After the alignment to the Human reference genome and Mouse reads removal, the quality of BAM files was assessed by Qualimap to obtain the mapped reads on target, target GC content, mean coverage and standard deviation of coverage. Variant detection analysis was QC-evaluated with bcftools stats by computing and validating the transition/transversion ratio from SNVs.