NKI: copy number alteration - Illlumina HiSeq 2500, WGS

Molecular Methods Description:

The total amount of DNA was quantified on the Nanodrop 2000 (Thermofisher). The amount of double stranded DNA in the genomic DNA samples was quantified by using the Qubit dsDNA HS Assay Kit (Invitrogen, cat no Q32851). A max of 2000 ng of double stranded genomic DNA were fragmented by Covaris shearing to obtain fragment sizes of 160-200 bp. Samples were purified using 2X Agencourt AMPure XP PCR Purification beads according to manufacturer■s instructions (Beckman Coulter, cat no A63881). The sheared DNA samples were quantified and qualified on a BioAnalyzer system using the DNA7500 assay kit (Agilent Technologies cat no. 5067- 1506). With an input of maximum 1 mg sheared DNA, library preparation for Illumina sequencing was performed using the KAPA Hyper Prep Kit (KAPA Biosystems, KK8504). During library amplification 6-8 PCR cycles were used to obtain enough yield for the exome capture. After library preparation, the libraries were cleaned up using 1X AMPure XP beads. All DNA libraries are analyzed on a BioAnalyzer system using the DNA7500 chips for determining the molarity. Up to 13 uniquely indexed samples are mixed together by equimolar pooling. The pools are analyzed on the Agilent Technologies 2100 Bioanalyzer. Pools are diluted to 10 nM, and measured on the qPCR. The pool is subjected to sequencing on an Illlumina Hi- Seq2500 machine, each pool in one lane of a single read 65 bp run, according to manufacturer■s instructions.

Analysis Description:

The resulting reads were trimmed using Cutadapt61 to remove any remaining adapter sequences. The trimmed reads were aligned to the GRCh38 version 97 and GRCm38 version 89 reference genome using BWA aln.62 Mouse reads were filtered out by AstraZeneca■s tool disam- biguate.63 The resulting alignments were sorted and marked for duplicates using Picard tools. QC statistics from Fastqc,64 Samtools65 and the above-mentioned tools were collected and summarized using Multiqc.66 The copy-number data was segmented using QDNAseq (version 1.22.0)67 from Bioconductor. The entire analysis was implemented by Julian de Ruiter using Snakemake (snakemake version 7.2.1; wrapper version 0.60.0)68 and is freely available on GitHub (https://github.com/jrderuiter/snakemake- cnvseq). Unsupervised clustering was performed on the segmented copy-number data. Copy-number instability was scored by calculating the fraction of bins with copy-number values above or below a threshold of respectively 2.5 and 1.5 in the segmented copy-number data. KCsmart R-package (version 2.48.0)79 and GISTIC2.069 were used to determine focal copy number groupwise aberrations.