LIH: mutation - Illumina MiSeq, Targeted capture

Molecular Methods Description:

500ng of extracted gDNA were diluted in 130µl low TE buffer (Qiagen) and sheared via sonication on a Bioruptor® UCD-200 (Diagenode) to an average fragment size of 150-300 bp. DNA fragment size was determined using the DNA 1000 Kit on the Bioanalyzer 2100 (Agilent Technologies). A custom-made Agilent SureSelect XT Target Enrichment Library (Cat No. G9612B) was used for Illumina Paired-End Multiplexed Sequencing on a MiSeq® instrument (Illumina). The panel design called panel 1 for the Target Enrichment Library was fully adapted from 2 . After design changes that were made using SureDesign - Agilent eArray (Agilent Technologies) we refer to it as panel 2 containing additional regions. Samples have been sequenced with version 1 of the panel or with version 2 of the panel (additional 53 genes). Library preparation was performed according to manufacturers' instruction. The Illumina MiSeq® Reagent Kit v3 (Cat No. MS-102-3003) was selected applying the Illumina reagent selection algorithm (Coverage Calculator?).

1. Stieber D, Golebiewska A, Evers L, et al. Glioblastomas are composed of genetically divergent clones with distinct tumourigenic potential and variable stem cell-associated phenotypes. Acta neuropathologica 2014; 127(2): 203-19.
2. Sahm F, Schrimpf D, Jones DT, et al. Next-generation sequencing in routine brain tumor diagnostics enables an integrated diagnosis and identifies actionable targets. Acta neuropathologica 2016; 131(6): 903-10.

Analysis Description:

The analysis of sequencing reads was performed step-wise: Raw sequencing reads (fastq) were quality trimmed using the tool fastp (v. 0.20.0) 3 . Trimmed reads were aligned to an in silico fused reference genome (ICRG) containing the human genome GRCh37.75 (ENSEMBL) and the mouse genome mm10 using BWA mem (v. 0.7.17) 4 . Reads that mapped to human chromosomes were extracted from the bam file using SAMtools (v.1.9) and realigned to the human reference genome only 5 . Duplicates were annotated and removed using MarkDuplicates under GATK (v.4.0.5.1). Bam statistics were assessed using SAMtools and compared between the initial mapping to the ICRG, the realignment to the human genome and after removing duplicates. Single nucleotide variants (SNVs) and smaller insertions and/or deletions (indels) were called in the CLC Genomics Workbench (v.12.0.3) using deduplicated mappings. Variants were only called in regions with a minimum coverage of 10 reads and a minimal allele frequency of 5 %. All variants that were likely to be polymorphisms and occurred in more than 1 % of the gnomAD (v.2.0.2) data base were filtered out. SNVs were annotated with COSMIC (v.89), ClinVar and dbSNP (v.150) 6 . The primary focus in SNV calling was to determine coding changes (missense and inframe mutations), truncating (stop and frameshift mutations) and splice site mutations, as well as mutations in the promotor region of TERT. For SNV calling, all filtered variants were manually checked to exclude artefacts and variants were further classified according to the American College of Medical Genetics and Genomics (ACMG). Only pathogenic, likely pathogenic or variants of uncertain significance (VUS) were included.

1. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018; 34(17): i884-i90.
2. Callari M, Batra AS, Batra RN, et al. Computational approach to discriminate human and mouse sequences in patient-derived tumour xenografts. BMC genomics 2018; 19(1): 19.
3. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009; 25(16): 2078-9.
4. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic acids research 2018; 46(D1): D1062-D7.

5. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 2016; 32(8): 1220-2.
6. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS computational biology 2016; 12(4): e1004873.
7. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 2012; 6(2): 80-92.
8. Koster J, Rahmann S. Snakemake-a scalable bioinformatics workflow engine. Bioinformatics 2018; 34(20): 3600.