

## **VHIO-CRC: mutation - Illumina HiSeq 2500, WES**

### **Molecular Methods Description:**

WES libraries were prepared from 500ng of DNA per sample using the SureSelect XT Target Enrichment for Illumina Paired-End Multiplexed Sequencing Library (Agilent Technologies) using the Agilent SureSelect XT Library Prep Kit. Target enrichment was carried out using the Agilent SureSelect XT Human All Exon v5 capture set, and sequencing with 100-base paired-end reads of targeted enrichment libraries was carried out on the HiSeq 2500 sequencer (chemistry v3, high-output mode).

### **Analysis Description:**

A quality check of the raw data was carried out by the FastQC tool. Reads were mapped to the Sanger human reference (hg19) by bwa (v. 0.6.2) with default settings. The resulting BAM files were processed using SAMtools (v. 0.1.19) and the Genome Analysis ToolKit (GATK) release 3.2.0. In brief, BAM files were binary compressed, sorted and indexed by SAMtools (samtools view, sort and index tools), duplicated reads were then removed by the SAMtools function rmdup, and base quality score recalibration and local realignment around indels followed the recommended workflow of the GATK toolkit (RealignerTargetCreator, IndelRealigner, BaseRecalibrator and PrintReads). Using pileup files from a matched tumour-normal pair, somatic variants were called by VarScan (v2.3.7) function "somatic" with the following parameters: minimum variant allele frequency (VAF) of 5%, a minimum coverage of 10 reads, at least five reads that confirm the mutation and a P-value  $<0.05$ . Annotation of the vcf files was carried out with the software ANNOVAR. Variants were filtered: variant position must be annotated as exonic by RefSeq (Release 45), and synonymous/non-synonymous calls were made and the synonymous excluded from further analysis. All filtering was carried out using in-house parsers.