# NKI: mutation - Illumina HiSeq 2000, WES

## Molecular Methods Description:

Genomic DNA (1µg) was fragmented with a Covaris S220 sonicator and DNA fragment libraries were prepared using the TruSeq DNA Sample Preparation Kit (Illumina, Eindhoven, the Netherlands). Library pools (8 libraries/ pool) were hybridized to the V4 Exome + UTR kit (Agilent) and sequenced on an Illumina HiSeq (50bp PE).

## Analysis Description:

The reads were trimmed using Cutadapt (Martin, 2011) to remove any remaining adapter sequences, filtering reads shorter than 60 bp after trimming to ensure good mappability. The trimmed reads were aligned to the human (GRCh38) and mouse (GRCm38) reference genome using BWA. The human alignment was processed for duplicate marking, indel realignment, and base recalibration using Picard Tools and GATK, as recommended by GATK best practices, and filtered to remove contaminating mouse reads using AstraZeneca's tool disambiguate (Ahdesmäki et al., 2016). QC statistics from Fastqc (Andrews S., 2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and the above-mentioned tools were collected and summarized using Multiqc (Ewels et al., 2016). Freebayes was used for SNP calling. SNPs that had a coverage of less than 10, a allele frequency of less than .01, or called as synonymous variants were excluded from the included data set.