IRCCS-CRC: mutation - Illumina CASAVA v1.8, WES

Analysis Description:

Primary processing of next-generation sequencing data and identification of putative somatic mutations. Somatic mutations were identified using VariantDx34 custom software for identifying mutations in matched tumour and normal samples. Before mutation calling, primary processing of sequence data both for tumour and for normal samples was performed using Illumina CASAVA software (version 1.8), including masking of adaptor sequences. Sequence reads were aligned against the human reference genome (version hg18) using ELAND with additional realignment of select regions using the Needleman-Wunsch method [ref]. Candidate somatic mutations, consisting of point mutations, insertions, and deletions, were then identified using VariantDx across the either the whole exome or regions of interest. VariantDx examines sequence alignments of tumour samples against a matched normal while applying filters to exclude alignment and sequencing artefacts. In brief, an alignment filter was applied to exclude quality failed reads, unpaired reads, and poorly mapped reads in the tumour. A base quality filter was applied to limit inclusion of bases with reported Phred quality score. 30 for the tumour and. 20 for the normal. A mutation in the tumour was identified as a candidate somatic mutation only when 1. distinct paired reads contained the mutation in the tumour, 2. the number of distinct paired reads containing a particular mutation in the tumour was at least 2% of the total distinct read pairs for targeted analyses and 10% of read pairs for exome, 3. the mis- matched base was not present in. 1% of the reads in the matched normal sample as well as not present in a custom database of common germline variants derived from dbSNP, and 4. the position was covered in both the tumour and normal. Mutations arising from misplaced genome alignments, including paralogous sequences, were identified and excluded by searching the reference genome. Potential alterations were compared with mouse sequences from experimentally obtained mouse whole-exome and targeted sequence data as well as the reference mouse genome (mm9) to remove mouse-specific variants. Candidate somatic mutations were further filtered on the basis of gene annotation to identify those occurring in protein coding regions. Functional consequences were predicted using snpEff and a custom database of CCDS, RefSeq and Ensembl annotations using the latest transcript versions available on hg18 from UCSC . Predictions were ordered to prefer transcripts with canonical start and stop codons and CCDS or Refseq transcripts over Ensembl when available.

Finally, mutations were filtered to exclude intronic and silent changes, while retaining mutations resulting in missense mutations, nonsense mutations, frame- shifts, or splice-site alterations. A manual visual inspection step was used to further remove artefactual changes. Amplification analyses were performed using the digital karyotyping approach [ref]. by comparing the number of reads mapping to a particular gene with the average number of reads mapping to each gene in the panel, along with a minor allele fraction analysis of heterozygous single nucleotide polymorphisms contained within each gene. For comparison of somatic altera- tions in tumour graft and pre-implantation material, we considered all alterations where the mutation was present in at least 20% of the read pairs in the tumour graft samples. To evaluate whether mutant genes observed in individual cases could be clinically actionable using existing or investigational therapies, we examined altered genes that were associated with (1) US Food and Drug Administration- approved therapies for oncological indications, (2) therapies in published pro- spective or retrospective clinical studies, and (3) ongoing clinical trials for patients with CRC or other tumour types.