Application Note

Low-frequency somatic variants detection and interpretation in a paired tumor/normal sample
Identifying and characterizing somatic cancer driver variants in deep genome sequence data from tumor samples remains challenging and time consuming. The detection and annotation of low-frequency somatic variants, which are just in a small proportion of cells of the primary tumor, is of special interest in cancer research. Such variants can drive tumor spread and recurrence. This application note describes how to identify low-frequency somatic cancer driver variants from a tumor/normal matched sample pair using CLC Cancer Research Workbench and interpret them using Ingenuity® Variant Analysis™ in one solution.

Data
We use publicly available Illumina® HiSeq® 2000 exome sequencing data from Nichols et al. available from the European Nucleotide Archive (ENA) under the accession number SRP018669. The data represents a tumor/normal matched sample pair from a massive acinic cell carcinoma of the parotid gland. Enrichment of target DNA regions was done using Agilent's 50 MB Whole Exome panel. Software: CLC Cancer Research Workbench 2.0 and Ingenuity Variant Analysis™ plugin.

Preparation
Target regions from the enrichment kit (Agilent's 50 MB whole exome panel) used by Nichols et al., were imported into CLC Cancer Research Workbench using "Import -> Tracks -> bed import". Sequencing reads were downloaded from the European Nucleotide Archive (http://www.ebi.ac.uk/ena/data/view/PRJNA184172) and imported into CLC.

Figure 1. New identified potential causal mutation in LTN1
Cancer Research Workbench using “Import->Illumina...”. “Prepare raw data” ready to use workflow with the Illumina adapter list has been used to trim Illumina adaptors from sequencing reads and generate a QC Sequencing report, which can be used to check the sequencing run quality.

Analysis

CLC Cancer Research Workbench provides ready-to-use workflows for NGS data analysis. To identify somatic variants in the tumor sample, the “Identify somatic variants from tumor normal pair” ready-to-use workflow for Whole Exome Sequencing is used for analysis. This workflow uses the “Low Frequency Variant Caller” to detect variants down to 1%. As this data set does not have high overall coverage, the parameters are adjusted accordingly (frequency cutoff is set to 5% and minimum coverage is set to 10, minimum count is set to 5, broken paired reads were ignored) and as no contaminations in the normal sample with tumor cells are expected, the “Keep variants with control read count below” parameter is set to 1.

Identification of cancer driver mutations

Ingenuity Variant Analysis (IVA) allows rapid identification of causal variants using QIAGENs comprehensive Knowledge Base. Variant Analysis is accessible directly from CLC Cancer Research Workbench after installing the plugin. When a subscription for Variant Analysis is available, variants are sent from CLC Cancer Research Workbench to Variant Analysis, filtered, annotated, and reimported into CLC Cancer Research Workbench. By right clicking on the variant track with the filtered and annotated variants, the analysis in Variant Analysis can be opened and modified. In this case in the “Confidence filter” the variability filters were switched on, the Common Variants filter settings were decreased to 1%, the Genetic Analysis Filter was removed and instead a Biological Context filter with association to Acinic Cell Carcinoma added at the end of the filter cascade. The updated list of variants were retrieved from Variant Analysis by a right click on the former Variant Analysis result in CLC Cancer Research Workbench and choosing “Variant Analysis Update”.

Results

Identified somatic variants

Using the “Identify somatic variants from tumor normal pair” ready-to-use workflow, we identified 94 variants found in the tumor sample, which are not present in the read mapping of the normal sample. Of these, 59 were SNVs, 20 were deletions, 12 were insertions and 3 MNVs. Ingenuity Variant Analysis identified 87 as high confidence variants and 57 were found to be relatively uncommon within a healthy population. 39 variants were identified as plausible drivers and 20 variants were directly linked to the disease.
Executive Summary
The integrated CLC-Ingenuity solution detected more novel variants than the original report by Nichols et al. Of these, 15 were found to be directly linked to the disease described by Nichols et al.

References

Comparison of results
Nichols et al. (2013) found 14 somatic variants using GATK Unified Genotyper tool. Using CLC Cancer Research Workbench, the “Identify Somatic Variants from Tumor Normal Pair” ready-to-use workflow with manual filtering identified 94 somatic variants. Of these, 11 overlapped the results obtained by Nichols et al. The other three variants reported by Nichols et al. were also detected by CLC Cancer Research Workbench but were not classified as somatic variants as these variants were found in both the tumour and normal samples. Only two of the 17 variants identified using Ingenuity Variant Analysis as being directly linked to the disease were described by Nichols et al.

After manual filtering using the mapped sequencing reads of the tumor and the normal sample, 3 additional variants could be excluded.