Clinical OMICs
Vol. 1, No. 4, May 2014

pp. 23–24
DOI: 10.1089/clinomi.01.04.08

Detection and Interpretation of Rare Sub-Clonal Cancer Driver Mutations

Anika Jöcker, Rupert Yip, and Salim Essakali

Within the last decade, our understanding of the molecular biology of cancer and its underlying complexity has experienced unparalleled and exponential growth (Hanahan et al. 2000, Hanahan et al. and 2011) [1, 2]. Both patients and clinicians have started benefiting from this new level of understanding, which can also be seen in the FDA approval within the last two years of approximately 20 new drugs that target a specific mutation for cancer.

With next-generation sequencing technologies, pathologists and clinicians are now empowered to test for several actionable mutations, as well as identify potential drug targets in parallel. This can significantly improve the diagnostic rate, as well as the drug treatment. Furthermore, data from large patient cohorts can be used to develop even better biomarkers, which can be used for prognosis, diagnosis, and drug treatment.

However, the bioinformatics analysis tools required to identify and interpret the plethora of variants—which in some cancers may amount to 50 variants per megabase (Vogelstein et al. 2013) [3]—still remain highly fragmented and require multiple experts.

ANIKA JÖCKER, *is Global Product Manager, Qiagen Aarhus. (ajoecker@clcbio.com)*

RUPERT YIP, *is Director, Product Management at QIAGEN Silicon Valley. (Rupert.Yip@qiagen.com)*

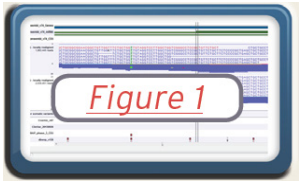
SALIM ESSAKALI, *is Senior Global Product Manager at QIAGEN. (Salim.Essakali@qiagen.com)*

Two new bioinformatics platforms have been specifically designed to target and streamline the identification (CLC Cancer Research Workbench) and the interpretation (Ingenuity® Variant Analysis) of somatic variants and driver mutations in cancer. This case study was aimed at testing both platforms in combination to identify cancer driver mutations that are present in a low percentage of tumor cells. These mutations are of special interest in a clinical setting as they can drive tumor spread and recurrence.

Materials, Methods, and Discussion

For this study, we used a publicly available tumor/normal dataset from a patient with massive acinic cell carcinoma, which was published last year by Nichols et al. in Case Reports in Oncological Medicine[4]. The data is targeted whole exome sequencing prepared using Agilent's Sure Select for the enrichment of exonic regions in the genome then sequenced using Illumina's HiSeq® 2000.

Raw reads from the HiSeq 2000 were imported into CLC Cancer Research Workbench for streamlined data analysis and visualization of results. The sequencing reads from the patient's tumor and normal tissues were mapped to the human reference genome and locally realigned to improve sensitivity and specificity of variant calling. Furthermore, variants were called down to an allele frequency of 5%, while potential false positives were filtered out. In addition, a detailed and summary coverage report was generated to check for coverage and uniformity of the targeted regions. All these steps were facilitated by running the "Identify variants" workflow in CLC Cancer Research Workbench. Next, detected somatic variants were separated from germline variants using the "Identify somatic variants from tumor/normal pair" workflow. Instead of using variants identified in the normal samples and subtracting them from the list of variants detected in the tumor, the workflow uses the mapped sequencing reads directly. This approach increases the sensitivity for removing germline variants. As no circulating tumor cells were expected and the coverage was quite low with 40x on average for an exon dataset, the cutoff for reads supporting the tumor variant in the normal (= potential germline variant) was adjusted to 1. This approach reduced the number of potential somatic variants from over 61,000 to 136. Manual filtering using the Genome Browser view in CLC Cancer Research Workbench (Figure 1) additionally removed 25 larger insertions and deletions, which were due to a different alignment of sequencing reads to the reference genome not automatically detected as germline. The Genome Browser view enables comparative visualization of detected somatic variants together with mapped reads from tumor and normal, known human genes, human reference sequences, as well as variants from clinical databases.



An additional filtering step using variants frequently detected in a population further reduced the number of potential somatic variants to 67. Next, Ingenuity Variant Analysis was used to detect potential cancer driver mutations. QIAGEN's Ingenuity has a manually curated knowledge base with detailed information on variants, genes, pathways, cancer-relevant phenotypes, and treatments (including insights on population allele frequencies and broader evolutionary conservation; known/likely effects on gene product structure, function, expression, and interaction; and relations among gene products, phenotypes, and drugs).

Using Ingenuity Variant Analysis, 25 variants were selected as potential driver mutations and 7 of them were directly linked to the disease. Six of the mutations linked to the disease were not described in the published paper and should be validated in further experiments.

Conclusions

This study demonstrates the power and ease of using both platforms to quickly identify driver mutations in cancer. CLC Cancer Research Workbench leverages state-of-the art pipelines for detecting somatic variants in pre-assembled and ready to use workflows, which can detect rare sub-clonal somatic variants that are usually found only in a small proportion of primary tumor cells, and which drives tumor recurrence if ignored when choosing treatment. At the same time, CLC Cancer Research Workbench remains an open yet flexible platform to customize pre-assembled workflows. Visualization using the genome browser and other features optimize specificity and sensitivity of somatic variant calling. Ingenuity Variant Analysis enables fast and powerful interpretation by contextualizing variants against a knowledge base that has been manually curated from peer-reviewed literature sources, thus always keeping the user up-to-date with the newest clinical and research advancements.

Reference

1 Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell 100, 57–70.

2 Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell 144, 646-674.

3 Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A.Jr., and Kinzler, K.W. (2013). Cancer genome landscapes. Science 339, 1546-1558.

4 Nicholset al. 2013 PMID 23653877