EUR. J. ONCOL.; Vol. 20, Suppl n. 1, pp. 5-6, 2015

Next-generation sequencing: instructions for the medical oncologist

Federica Di Nicolantonio[1, 2]

[1] Department of Oncology, University of Torino, Candiolo (TO), Italy; [2] Candiolo Cancer Institute-FPO, IRCCS, Candiolo (TO), Italy

Cancer is essentially a disease of genes. Accumulation of molecular alterations in the genome of somatic cells is at the basis of tumorigenesis and underlies tumor response to therapy. Common cancers such as those arising in the lung, colon, breast, skin and prostate harbor a cocktail of mutated (or otherwise altered) oncogenes and tumor suppressors that work in concert to determine the molecular pathways that lead to their genesis, maintenance, and progression (1). Oncology research has benefited immensely from the worldwide efforts to characterize the genomes of all major cancer types. Technologic and analytic advances have enabled a comprehensive catalogue of cancer genes that can be exploited as diagnostic, prognostic and predictive biomarkers.

The implementation of genomics-driven oncology - also known as precision oncology - involves several stages: initially, genomes and transcriptomes of patient tumors must be analyzed using state-of-the-art technologies; second, the molecular profiles of each patient must be integrated using sophisticated bioinformatics tools with knowledge of existing and emerging anticancer drugs; finally, an annotated list of genomic alterations must be provided to the medical oncologist so that it can be incorporated into clinical decision making (2).

This seemingly straightforward process is ridden by several challenges. To begin with, high-quality genomic information must be obtained consistently in the diagnostic setting - often from sparse amounts of archival tumor tissue. Secondly, different layers of data (genomic, transcriptional and epigenetic profiles) must be annotated and interpreted. Finally, scientists often trained in different fields and clinicians should work in multi-disciplinary teams so that genomic information can be used for evidence-based therapies and innovative clinical trials.

I will briefly review here the steps undertaken as well as the hurdles that we have encountered to build an infrastructure which could support genomics-driven cancer medicine for clinical practice and research purposes at our institution.

In order to analyze and interpret next-generation sequencing (NGS) data, we formed a multi-skilled team composed of scientists with a background in biology, informatics, pharmacology, mathematics, medicine or engineering. The team met regularly once every 1-2 weeks to set tasks and worked together to develop NGS data analytical pipelines. No golden standard or validated commercial software exist to analyze complex tumor NGS data. Therefore, we strived to build internal pipelines to identify point mutations, copy number changes as well as more complex structural variations. All pipelines include a first mapping step, during which reads are aligned against the reference human genome and attributed the best genomic coordinates. Our first home-built pipeline allows the identification and annotation of single nucleotide variants by comparing the base present at each position in the tumor DNA with the germ-line DNA of the same patient. The user has the option to sieve variants at a-given threshold. When sequencing at a depth of 100x (this means that each nucleotide is sequenced on average 100 times), we usually filter out the variants

that are present in less than 5% sequence reads, as this takes into account the error rate accumulated by PCR amplification as well as sequencing itself. A second bioinformatics pipeline was then devised to extract information on gene copy number variations. As a third step, we are currently working on a tool to identify and annotate indels. This has revealed the most challenging task since sequences with insertions or deletions tend to be misaligned with the reference genome and discarded as low-quality sequence reads when mapping is initially performed.

In parallel to bioinformatics, we proceeded with the implementation of standardized wet laboratory protocols to capture the DNA regions of interest, prepare NGS libraries and perform sequencing. Commercially available kits include exomes and gene panels, which include targeted sequencing of mutation hotspots in genes relevant for cancer prognosis or treatment. Custom panel including dozens to hundreds genes can also be designed. Under optimized conditions and starting from 100-500 ng of DNA, turnaround time for performing NGS is 2-3 weeks, including the analytical procedures for DNA extraction and quality control. Depending upon the complexity of data and number of samples, bioinformatics analysis can take as long as the wet procedures, if not longer, totaling up to 6 weeks since sample submission.

Finally, data should be reported to the medical oncologist in an intelligible format, with reference to background literature so that it can inform on the functional and potentially clinical implications of the identified molecular variants.

References

1. Vogelstein B, Papadopoulos N, Velculescu VE, *et al.* Cancer genome landscapes. Science 2013; 339 (6127): 1546-58.
2. Meric-Bernstam F, Farhangfar C, Mendelsohn J, *et al.* Building a personalized medicine infrastructure at a major cancer center. J Clin Oncol 2013; 31 (15): 1849-57.