The promise of image-guided cancer genomics

Posted by Biome on 16th March 2015

**Cancer tissue banks and large cohort studies provide a vast resource of human tumor samples for genomic and transcriptomic analysis. However even with the greatest surgical precision, these tumor samples contain neighboring non-tumor tissue and other cells of the tumor microenvironment. Tumor purity significantly impacts the relative number of tumor cells analysed in a given study and therefore the strength of the subsequent findings. In a**[**Research Highlight in Genome Biology**](http://dx.doi.org/10.1186/s13059-014-0563-3)**,Kosuke Yoshihara from the Niigata University Graduate School of Medical and Dental Science, Japan, and Roel Verhaak from the University of Texas MD Anderson Cancer Center, USA, discuss how image-guided genomics may provide a solution. Here Verhaak explains the importance of tackling the issue of tumor purity, how image analysis can help, and the challenges raised by a combined imaging and genomics led approach.**

**Why is tumor purity so important when analyzing tumor samples to identify driver mutations in cancer?**

It is important to note that tumor-associated normal cells should not be equated to contamination. The tumor microenvironment is a hallmark of cancer and plays important roles in regulating tumor cell proliferation, possibly through secretion of growth factors.

Having said that, driver mutations, including single nucleotide variants, DNA copy number changes, chromosomal rearrangements and transcript fusions, are commonly detected through DNA and/or RNA sequencing. The molecules being sequenced are extracted from all cells in the tumor sample, including tumor-associated normal cells such as fibroblast and endothelial cells. Non-cancer cells, which are present in any solid tumor, decrease the number of sequence reads carrying tumor driving mutations and thereby lead to a lesser sensitivity for identifying driver mutations. Intratumoral heterogeneity, as a result of which driver mutations may be present in only a subset of tumor cells, further affects detection sensitivity.

It is important to consider factors such as the expected fraction of tumor cells and the expected level of subclonality before designing a sequencing experiment and determining a sequencing coverage threshold. While the relation between sequencing coverage and the number of somatic alterations is not linear and varies per tumor type, a commonly recommended (but not necessarily adhered to) coverage threshold is 60x for whole genome sequencing to account for tumor purity levels and subclonality.

Array-based approaches (such as the Affymetrix SNP arrays and the Agilent CHG arrays) are unable to account for these factors and may be less effective in detecting DNA copy number changes from tumor samples with low purity. An example of this limitation was shown in a DNA copy number study by Barbara Weir and Matthew Meyerson from the Dana-Farber Cancer Institute, USA, and colleagues (Nature. 2007, Dec, 450, 7171, 893-8), where they separated their lung adenocarcinoma cohort into three groups based on tumor purity and the dilution effect was clearly visible. This landmark paper motivated Scott Carter and Gad Getz to develop the ABSOLUTE computational method to infer tumor purity from genomic profiles, which is now commonly applied in analysis of for example The Cancer Genome Atlas data (Nat Biotechnol. 2012, May, 30, 5, 413-21). Since genomic profiles are not always available, my lab developed a gene expression method with the same purpose, called ESTIMATE (Nat Commun. 2013, 4, 2612). These methods can aid in determining tumor purity and setting expectations for the design of new experiments.

**How can advances in image analysis help address issues of tumor purity?**

Histology (immunohistochemistry stain) images are the basis for pathology review, which is standard practice in the clinical management of cancer patients. While a pathologist may be limited by the area of the tumor that can be reviewed within a reasonable timeframe, machine based approaches for image evaluation do not know this limitation. Pathologists are experts in quickly recognizing cell morphology and patterns related to tumor characteristics, which are challenging to capture in computer algorithms. Amongst others, the lab of Florian Markowetz has pioneered machine based image review approaches and their recent studies have shown that non-tumor cells can be recognized in an automated fashion and with high precision (published in this [Genome Biology study](http://dx.doi.org/10.1186/s13059-014-0442-y)). Applying their methods to sets of histology images from The Cancer Genome Atlas, which are available through the [Cancer Imaging Archive](http://www.cancerimagingarchive.net/), they were able to identify cases of ovarian carcinoma with low tumor purity. Restricting their further analysis to high purity tumor samples led to uncovering the role of PTEN as a tumor suppressor in ovarian carcinoma (published in [this Genome Biology study](http://dx.doi.org/10.1186/s13059-014-0526-8)). Their work is a clear example of how image processing can provide an additional layer of information on top of what can be inferred from genomic and transcriptomic profiles.

**What confounding factors may affect a combined imaging and genomics led approach to identifying cancer drivers in tumor samples?**

The most important possible confounder is the spatial heterogeneity of solid tumors. The tumor sector from which the histology slide is obtained may be geographically distinct from the tumor sample that is genomically analysed. Furthermore, while improvements have been made to optimize computer-based feature recognition, machine based slide review does not yet lead to the same quality of diagnosis as would be obtained from an expert pathologist.

Additional factors to consider are differences in the quality of staining, batch effects and image or slide artifacts. The studies that have been published so far were performed on high quality, curated image collections and may not be representative of the slide sets that pass through the daily pathology practice. A lot more and more comprehensive studies are needed to determine the possible impact of these confounding factors.

**In the context of image-guided genomics, what computational challenges arise when dealing with the vast quantities of genomics and imaging data?**

In most parts of oncology, the maximum turnaround time for supporting clinical decisions is one to two weeks. Approaches like whole genome sequencing take more than that, even without considering analysis. However, like the cost of sequencing has dropped dramatically, the time (and associated cost) to analyse a genome are similarly decreasing rapidly. The processing time and data volume of a single immunohistochemistry/histology image is minimal compared to whole genome sequencing and are not as rate limiting. However image processing does contribute to the required disk and compute capacity.

**Does the future herald a fully automated approach to screening tumor samples using image-guided genomics?**

The use of ‘Big Data’ to learn patterns and use that information to predict behavior or future actions is noticeable everywhere, including in oncology. The MD Anderson Cancer Center has partnered with IBM and together they have created the Oncology Expert Advisor (OAE), a cognitive clinical decision support tool powered by IBM’s Watson. The OAE was trained using MD Anderson’s data bases that go back all the way to first patients from 1941. The features provided to the system are comprehensive and include patient demographics and general disease details, but also imaging, genomics and so forth. It makes personalized treatment recommendations that support the oncology team in making clinical decisions. While patients that seek treatment at academic institutions such as MD Anderson are guaranteed access to get the most advanced therapies, approaches like the OAE may be able to extend the latest and greatest knowledge from the top cancer centres to institutions that are more remote, thus democratizing healthcare. OAE is not intended to replace the oncologist, but to guide and assist treatment decision making. I think it is unlikely that we will see clinical treatments entirely based on algorithms, just because so much complex information goes into that process that it will take years and lots of data to appropriately train OAE and similar approaches.