

## The future of cancer genomics

**With the completion of The Cancer Genome Atlas, it is time to evaluate its impact and mine its data to gain a better understanding of cancer biology and therapy.**

**T**he Cancer Genome Atlas (TCGA) will wind down in 2015, bringing to a close one of the most ambitious large-scale initiatives spearheaded by the US National Institutes of Health. Starting as a pilot project in 2006, its mandate was to generate a comprehensive landscape of alterations in all tumor types, with the aim of gaining novel insights into cancer biology that could be applied to develop better therapies. The high-yield approach was a departure from the traditionally funded hypothesis-driven projects, and its lofty goal of capturing the whole spectrum of cancer alterations was initially met with a mix of excitement and skepticism by the scientific community. It is now time to take stock of TCGA and determine how its insights can be used to benefit the cancer community.

In terms of data generation, the project has been an undisputed success. In the almost ten years since its inception, and after a total \$375 million investment, TCGA has incorporated scientific contributions from more than 150 researchers from 16 countries, characterizing 10,000 tumors from more than 25 different cancer types. Its 20 petabytes of data include 10 million mutations, and they have been reported in 17 (so far) publications from the TCGA Research Network and mined in hundreds of articles. These staggering numbers reflect the exponential growth of the project, which has been enabled by the rapid evolution of technologies for sample collection, sequencing and analysis.

A wealth of information has been steadily pouring in from the TCGA pipeline. TCGA data has been used to discover new mutations, define intrinsic tumor types, identify pan-cancer similarities and differences, uncover mechanisms of therapy resistance and gather evidence of tumor evolution. Undoubtedly, we can now look at cancer with unprecedented detail, but we are still far from interpreting the full picture of this disease and unraveling its mechanisms.

Some TCGA investigators think that additional insights can be gained by continuing to search for novel cancer alterations. But recent estimates underscore the daunting task of achieving saturation in cancer sequencing: depending on background mutation rate, some tumor types would require the characterization of more than 10,000 samples in order to detect alterations with a 1% frequency. Thus, Louis Staudt, the director of the Office for Cancer Genomics at the US National Cancer Institute (NCI), has announced that the TCGA Research Network will now focus its efforts on applying whole-genome sequencing, a technique that was not available at the inception of the initiative, to expand the

characterization of three selected tumor types: lung adenocarcinoma, colon and ovarian cancer. The aim is to uncover alterations present only in 2% of tumors, as well as to discover types of alterations that may have been previously missed, such as translocations.

This pilot project will also seek to overcome past financial and logistic hurdles. Sample acquisition, one of the biggest monetary burdens of TCGA, will now be coordinated with ongoing clinical trials of targeted cancer therapies, allowing for a more integrated characterization of genotype and phenotype in different cancer stages. Importantly, the NCI will devote resources to ensure the accessibility and proper analysis of the sequencing data. The newly minted NCI Genomics Data Commons will provide a portal offering interactive support and best practices to genomic data users. The outcome of this pilot study will determine whether a similar approach is applied to a broader spectrum of tumors.

While sequencing continues, albeit on a smaller scale, it is important to tackle the next step, the systematic functional validation of genetic “hits.” This will require renewed effort, creativity, and grit from the cancer community, as well as the strong commitment of funding agencies.

Several challenges and solutions are being charted out for the translation of TCGA data. First, better computational models are being developed to identify relevant alterations (drivers) from background genetic noise (passengers). This is likely to reduce the complexity of the data, but functional studies must also be scaled up to the dimension of genetic studies. For example, recent advances in genome editing tools such as CRISPR-Cas9 provide unprecedented power to study genetic variations in a rapid, scalable and more cost-effective way. But to obtain meaningful insights, we need to study genetic alterations within the complex and heterogeneous context of the physiological tumor environment. This will require the incorporation of cell lines, organoids and patient-derived models into a pipeline that enables high-throughput functional testing of genetic alterations. Additionally, a better integration between cancer genomics and clinical practice will allow for direct phenotype-to-genotype characterization. Initiatives like the NCI Exceptional Responder program, which prioritizes the study of patients with outlier responses to treatment, are expected to funnel clinically relevant insights into the genomic and functional pipelines.

TCGA represents a formidable effort for the cancer community. The translation of the cancer genome into mechanistic insights and future therapies will take its achievements to the next level and usher in a new era in cancer research.