STO Society for Translational Oncology®
The Oncologist® Is The Society's Official Journal

2017 Society for Translational Oncology Fellows' Forum

The Cancer Genome Atlas (TCGA)

Faculty Presenter
William Kim, MD, Associate Professor of Medicine and Genetics Division of Hematology/Oncology, University of North Carolina, Chapel Hill, NC, USA

Scholars' Summaries

*Authored by **Saurin Chokshi, MD**, Yale University School of Medicine, New Haven, CT, USA*

Dr. William Kim gave an enlightening and pragmatic overview of the science behind and applications for The Cancer Genome Atlas (TCGA). The TCGA is a comprehensive molecular and genomic analysis of over 30 different cancers derived from tumor tissue sources from over 150 different global sites. This has led to the creation of an expansive atlas of genomic alterations. The raw data from this atlas provides a rich and deep look into many aspects of the genomic make up of cancers. Specifically it offers the ability to review multiple molecular platforms for each tumor; these include mRNA, micro RNA, DNA copy number, DNA methylation, mutations, and protein. Accordingly, this offers the opportunity for investigators to pursue "pan-cancer" analyses. Such analyses have been conducted and published across major biomedical journals. Notably, these efforts have demonstrated the heterogeneity of mutational burden across cancers. They have also created an avenue for conducting integrative analyses across cancer types – "cluster of cluster" analyses – that is providing unique insights into the common thread of cancers.

The TCGA, however, has a few limitations. Firstly, the tumor samples collected for the atlas were from untreated patients. Hence the samples provide baseline tumor information, but no insights into genomic changes as a result of therapy. This point alone highlights a major hurdle in the clinical applicability of the atlas data. Secondly, the atlas did not evaluate any immuno-oncologic aspects such as PDL1 status or HLA or T/B cell sequencing. This omission was rather unfortunate, as advances in immunotherapy and the coupling of immune check point inhibitors with conventional chemotherapy have come to the forefront of cancer therapy.

In the endeavor to create and successfully establish the TCGA, cancer researchers now have a powerful platform that has raised the standard for tumor data processing and acquisition, created a pipeline for genomic data and analysis and ultimately therapeutic exploration, and fostered collaboration among basic and translational scientists and clinicians.

---

*Authored by **Xin Gao, MD**, Beth Israel Deaconess Medical Center, Boston, MA, USA*

The Cancer Genome Atlas (TCGA) is a large multi-institutional collaborative initiative sponsored by the National Cancer Institute (NCI) to comprehensively investigate the molecular and genomic basis of 33 types of human cancers. Tumor specimens were collected from over 150 different sites worldwide over the past several decades, with the vast majority of tissue samples derived from untreated patients and

selected for high tumor content and purity. TCGA specimens have been subjected to rigorous pathology review and characterize the same tumor specimen via multiple "omic" platforms at the DNA, RNA, and protein levels. However, there are minimal clinical data associated with these samples, and those samples that had associated clinical data generally had a short length of follow up. Therefore, TCGA provides an excellent database of genomic alterations in a variety of common cancers, but the ability to make clinical associations with those alterations is limited.

The sequencing efforts of TCGA concluded in 2017. TCGA data is publicly available on the Genomic Data Commons Data Portal, and additional analyses of the primary TCGA data may be expected for many years to come. TCGA has added tremendous value to our understanding of the molecular and genomic characteristics of human cancers, but many questions remain unanswered. Newer collaborative efforts in precision medicine are already underway via the NCI Center for Cancer Genomics, including the Clinical Trials Sequencing Project, the Cancer Driver Discovery Program, and an Exceptional Responders Initiative. Precision oncology will continue to improve as understanding of the genomic features of various cancers expands via these large-scale endeavors.

At the University of North Carolina Lineberger Comprehensive Cancer Center, UNCseq is one example of how precision oncology is currently being used to improve patient care. Through UNCseq, tumor specimens and normal samples from the same patient are analyzed via a next-generation sequencing platform to yield a tumor "omic" profile. The genomic data and clinical case is presented at a weekly Molecular Tumor Board, during which a recommendation for standard-of-care therapy, clinical trial enrollment, or use of an off-label novel agent may be offered to the treating oncologist. The UNCseq team has also started working with IBM Watson to help improve identification of actionable genomic alterations. Thus far, over 1000 patients have enrolled in the UNCseq program, and 70% of enrolled patients have been found to have an actionable genomic alteration, including 32% identified by IBM Watson. The recommendations of the Molecular Tumor Board have led to a change in therapy in 14% of the patients enrolled. While further advancements are needed to take advantage of the resources available in large-scale "omic" data, the future is bright for the ability of precision oncology to improve outcomes for cancer patients.

---

*Authored by **Amar Patel, MD**, UCSF Clinical Fellow, Department of Medicine, Division of Hematology/Oncology, San Francisco, CA, USA*

The Cancer Genome Atlas (TCGA) is a large, collaborative effort sponsored by the National Cancer Institute to study the molecular and genomic underpinnings of 33 different cancer subtypes. Tumor tissues collected over several decades from over 150 different sites around the world make this a tremendous resource for referencing genomic alterations in cancer. Samples that comprise the TCGA dataset are largely derived from untreated patients who were selected for high tumor content and purity. Rigorous pathology review of samples, the ability to integrate data across multiple "-omics" platforms, unique opportunities for investigator collaboration, and the ability to perform "PanCancer" analyses across tumor types are clear strengths of the TCGA. It is important to note, however, that generally short duration of follow-up, limited clinical annotation data, and absence of critical immuno-oncology parameters such as direct HLA typing and TCR/BCR sequencing are limitations to the current utilization of this data set.

The TCGA pipeline integrates DNA, RNA, and protein tumor sample data and organizes it at 3 levels for investigator use:

- Level 1 consists of FASTQ sequencing data files
- Level 2 includes a compressed binary format that stores sequencing alignment data called BAM files. Data at these levels is controlled.
- Level 3 open access data can be obtained through the NCI portal.

Large scale profiling through TCGA has allowed for PanCancer analyses, looking at data such as copy number alterations and mutational burden across a range of tumor subtypes. Integrative analysis on genomic signatures has also allowed unique taxonomy which extends beyond traditional tissue-of-origin grouping and organizes tumors according to molecular signatures.

Though the TCGA will conclude in 2017, it has set the stage for expanded efforts in precision medicine which integrate genomic information into clinical decision making. Next generation sequencing platforms are now being implemented to sequence tumor and normal tissue from patients in an effort to define the unique molecular drivers of tumor growth. This data can now be analyzed and reviewed by multidisciplinary tumor boards to evaluate for possible actionable targets and provide novel avenues for tumor treatment. A major challenge remains in learning what targets are truly actionable, particularly when there is such heterogeneity across various institutional sequencing platforms. One method for building a better "Actionability Atlas" will be iterative processing of "-omics" data and clinical outcomes by cognitive computing platforms such as IBM Watson. Attempts at integrating precision medicine data are underway as evidenced by the execution of the SHIVA trial, an open-label, randomized controlled phase 2 which assigned patients to either molecular therapy based on specific tumor alterations in pathways that may be targeted with existing drugs or physicians' choice of therapy. Though this use of targeted agents outside their indications did not improve PFS in this trial, future studies are now building on this targeted therapy paradigm in the ultimate hope of improved selection of targeted agents and combination therapies for cancer treatment.

---

*Authored by **Jia Wei, MD**, Nanjing University, Jiangsu Sheng, China*

The Cancer Genome Atlas is a NCI-sponsored large collaborative initiative and comprehensive study on the molecular and genomic basis of 30+ types of cancer. Tumor tissues in this project were collected from over 150 different source sites from around the world over the past several decades.

Six "omic" platforms (mRNA, miRNA, protein, DNA copy number, mutation, and methylation) were used on one tumor, making a great dataset for an atlas of genomic alterations and the ability to perform "Pan-Cancer" analyses. There are three different levels of data descriptions in TCGA, and data are available to the public through NCI's Genomic Data Commons Data Portal. However, there are some limitations of TCGA: 1) the clinical data is spotty because almost all of the samples are primarily untreated, without any response data and short follow up. 2) There is no immune-oncology data; for example, TCGA didn't include direct HLA or TCR/BCR sequencing data. 3) Samples in the TCGA project are all fresh frozen samples, which are not commonly used in clinical settings.

Dr. Kim also introduced the process and challenges of molecular tumor boards (MTB) at UNC. In the process of accumulating updated scientific and clinical information, the informatics support is crucial. To

this end, they have collaborated with IBM's Watson to identify all actionable genes to ensure all therapeutic options have been explored.