

The Cancer Genome Atlas (TCGA) and Big Data

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 **Siddhartha Devarakonda, MD**, Barnes-Jewish/Washington University in St. Louis School of Medicine, St. Louis, MO, USA*

The Cancer Genome Atlas (TCGA) is an NCI-sponsored collaborative initiative, conceived with the objective of understanding the genomic landscape of cancer. Several hundred cancer samples have been characterized on multi-omic platforms as a part of this initiative to date. These studies have offered valuable insights into the complexity and heterogeneity of the molecular mechanisms that drive the malignant phenotype. These data are public and accessible (although access to Level 1 and 2 data is controlled and requires special permission) for analysis. Apart from helping us gain an understanding of the types of mutations and their burden, pathways deregulated, and various structural and epigenetic alterations, in individual tumor types - data generated from TCGA samples has facilitated "pan-cancer" analyses. These pan-cancer analyses have provided the ability to identify novel driver alterations that are otherwise encountered at a low frequency in different cancers by boosting sample size and statistical power and to identify common molecular subtypes across histologically different malignancies and distinct molecular subtypes within histologically identical cancers.

*Authored by **Keigo Komine, MD, PhD**, Tohoku University, Seiryomachi, Aoba-ku, Sendai, Japan*

The Cancer Genome atlas (TCGA) is a National Cancer Institute-sponsored large collaborative initiative to comprehensively study the molecular and genomic bases of over 30 types of cancer. Tumor tissues have been collected from over 150 different source sites from around the world over the past several decades. DNA and RNA are being sequenced and evaluated for mutations, copy number alterations, LOH, and epigenetic changes. TCGA established multiple "omic" platforms and has provided the ability to perform "pan-cancer" analyses. These analyses data are publically available, so that many studies have been conducted based on these data. However, TCGA has some limitations. Because the collected samples are untreated tumors, the data are annotated with very limited clinical information.

Recently precision medicine has become a topic in cancer therapy. UNC provides treatment based on the omic profile as UNCseq™. In this process it is important to feed back the clinical outcome. Accumulation of the clinical information contributes to further development of the cancer treatment and enables precision medicine. Although it is hard to detect a target for treatment from vast analysis data, computers may assist in finding rational therapy. Watson, the first mononymous computer, identified the actionable gene in 32% cases in which physicians were not able to detect an actionable gene. UNC has collaborated with IBM to develop this computer.

Big data becomes more and more important. It is necessary to be strategic about how we collate and use it.

Authored by **Tomohiro Nishijima, MD**, University of North Carolina Hospitals, Chapel Hill, NC, USA

The Cancer Genome Atlas (TCGA) is an NCI-sponsored collaborative initiative that has comprehensively studied the molecular and genomic basis of more than 30 types of cancer. The strengths of TCGA include that rigorous pathology review was done when selecting samples and that multiple "omic" platforms were performed on the same tumor, which led to a number of "pan-cancer" analyses regarding mutations, copy number alterations and immuno-oncology etc. On the other hand, TCGA dataset has limitations; the samples are primarily untreated, data on response to treatment is not available, and follow up for survival data was short and incomplete.

Dr. William Kim provided a summary of his work on intrinsic molecular subtypes of high-grade bladder cancer. His group performed consensus clustering on gene expression data from a meta-dataset of bladder cancer and identified "luminal" and "basal-like" subtypes. These two intrinsic subtypes have characteristics of different stages of urothelial differentiation and have clinically meaningful differences in outcome. More recently, Dr. Kim's group discovered a claudin-low molecular subtype of high-grade bladder cancer in addition to the "luminal" and "basal-like" subtypes. Claudin-low tumors showed high expression of immune gene signatures with active immunosuppression. This was found to be associated with upregulation of cytokine and chemokine levels from low PPARG activity, allowing unopposed NFkB activity.

Dr. Kim also talked about the challenges of running molecular tumor boards (MTB) based on his experience at UNC. The MTB discusses actionability of genomic alterations found on the genomic sequencing. Currently, only a subset of genomic alterations is clinically actionable, and the majority of somatic events remains classified as variants of unknown significance and will require functional characterization. One of the major challenges is how to update the "actionability atlas" based on scientific and clinical knowledge that is updated on a daily basis. As there is a limitation on how much the MTB members read papers and update the "actionability atlas" regularly, UNC is collaborating with IBM Watson to use the power of cognitive computing to improve translation of cancer-sequencing results into potential treatment options based on the best available data.