Appendix: Summary of Best Papers Selected for the 2019 Edition of the IMIA Yearbook, Section Cancer Informatics


Comprehensive characterization of cancer driver genes and mutations
Cell 2018 Apr 5;173(2):371-385.e18

Understanding which genes and which gene mutations are cancer drivers is an essential first step towards contemplating ways to disable the cancer machinery through pharmacologic intervention. Generally speaking, somatic mutations observed in cancers are felt to either be drivers of the cancer or passive passengers; most drug development is focused on disabling drivers. Despite much work in this area, automated algorithms often do not agree on candidate driver genes and mutations, requiring expert manual curation. This broad application of 26 bioinformatic software tools to 10,000 TCGA tumor samples (representing 33 cancer types) is the most comprehensive discovery of cancer driver genes and mutations to date. The data generated lay the groundwork for years of basic, translational, and clinical efforts. All data generated are publicly available.


Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study

Despite several years of anticipation, artificial intelligence methods such as deep learning have yet to enter the clinical cancer setting. In general, findings based on a single institution retrospective study must be replicated across institutions before prospective trials can be considered. Hosny et al., have met the second mark through their multi-site retrospective study of lung cancer prognostication using radiomics. Despite being the deadliest cancer, there are scant prognostic tools to determine lung cancer prognosis outside of the traditional anatomic staging systems. This study was an integrative analysis on seven independent radiographic datasets across five institutions, using a 3D convolutional neural network. The authors successfully predicted survival for lung cancer patients with an AUC of 0.70. While the performance is only modest, radiography is standard of care for lung cancer and this study has immediate clinical applicability for a very common and highly lethal disease. Over the coming years, we expect that such approaches will become more comprehensive and accurate and will be tested in the prospective setting.


Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations

Modern cancer DNA sequencing tests generate vast amounts of data, with most commercial panels easily generating over 1 000 000 base pairs of data. These results must be filtered, interpreted, and presented to clinicians who will undertake medical decisions, frequently in the setting of multiple possible courses of action. Cancer Genome Interpreter is a software tool that streamlines and automates the process of identifying and annotating variants. The tool accepts several data formats and provides a user-friendly output. Also described is a new knowledge base of 5,314 validated mutations (the Catalog of Validated Oncogenic Mutations). As a proof of concept, 72% of AACR Project GENIE tumors (~17k) have at least one biomarker of drug response in the system. This proportion is much higher than what has been reported in older studies of genomically-informed treatment decisions, suggesting that the match between mutation and drug continues to improve.