

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

Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, Colaprico A, Wendl MC, Kim J, Reardon B, Ng PK, Jeong KJ, Cao S, Wang Z, Gao J, Gao Q, Wang F, Liu EM, Mularoni L, Rubio-Perez C, Nagarajan N, Cortés-Ciriano I, Zhou DC, Liang WW, Hess JM, Yellapantula VD, Tamborero D, Gonzalez-Perez A, Suphavitai C, Ko JY, Khurana E, Park PJ, Van Allen EM, Liang H; MC3 Working Group; Cancer Genome Atlas Research Network, Lawrence MS, Godzik A, Lopez-Bigas N, Stuart J, Wheeler D, Getz G, Chen K, Lazar AJ, Mills GB, Karchin R, Ding L

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.

Hosny A, Parmar C, Coroller TP, Grossmann P, Zeleznik R, Kumar A, Bussink J, Gillies RJ, Mak RH, Aerts HJWL

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

PLoS Med 2018 Nov 30;15(11):e1002711

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 convolution 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.

Low CA, Dey AK, Ferreira D, Kamarck T, Sun W, Bae S, Doryab A

Estimation of symptom severity during chemotherapy from passively sensed data: Exploratory study

J Med Internet Res 2017 Dec 19;19(12):e420

With smart phones and other wearable devices now nearly ubiquitous, it is natural to wonder if they can be utilized as health care collection tools. In particular, can passive data collection yield insights similar to those collected directly from patients? Low, *et al.*, undertook

a prospective pilot study of passively collected data from patients undergoing chemotherapy treatment. The aim of this study was to explore whether passively sensed mobile phone and Fitbit® data could be used to estimate daily symptom burden during chemotherapy. Passively collected data matched patient-reported symptom burden with an accuracy of 88%. This was a small study (n=14), but proof of concept that mobile phone accelerometer and usage and Fitbit®-assessed activity and sleep were related to daily symptom burden during chemotherapy. These findings highlight opportunities for long-term monitoring of cancer patients during chemotherapy, including the possibility of obviating invasive and tedious surveys.

Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, Tusquets I, Albanell J, Rodon J, Tabernero J, de Torres C, Dienstmann R, Gonzalez-Perez A, Lopez-Bigas N

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

Genome Med 2018 Mar 28;10(1):25

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.