Appendix: Content Summaries of Selected Best Papers for the IMIA Yearbook 2018, Section Cancer Informatics

Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, Rudolph JE, Yaeger R, Soumerai T, Nissan MH, Chang MT, Chandarlapaty S, Traina TA, Paik PK, Ho AL, Hantash FM, Grupe A, Baxi SS, Callahan MK, Snyder A, Chi P, Danila D, Gounder M, Harding JJ, Hellmann MD, Iyer G, Janjigian Y, Kaley T, Levine DA, Lowery M, Omuro A, Postow MA, Rathkopf D, Shoushtari AN, Shukla N, Voss M, Paraiso E, Zehir A, Berger MF, Taylor BS, Saltz LB, Riely GJ, Ladanyi M, Hyman DM, Baselga J, Sabbatini P, Solit DB, Schultz N

OncoKB: a precision oncology knowledge base

JCO Precis Oncol 2017 Jul;2017

The practice of oncology is increasingly informed by biologic factors beyond the traditional biomarkers, in particular those obtained through clinical genomic sequencing. Variants have prognostic and predictive implications and comprise a large and growing knowledge space. Chakravarty et al., have built a large public web resource, OncoKB (http://oncokb.org/) with a goal of providing evidence-based information for clinicians and researchers. The content is curated and stored in an internal data model which is exposed via a public API on demand. As of this writing, OncoKB contains information on 477 genes, 3,855 variants, 60 tumor types, and 86 drugs. Twenty-five genes are linked to FDA-approved or standard of care treatment evidence, and 39 are linked to more limited clinical or biological (non-human) evidence. OncoKB is a cornerstone in the emergent ecosystem of cancer genomics content management, and is developed and maintained by the Knowledge Systems group in the Marie Josée and Henry R. Kravis Center for Molecular Oncology at the Memorial Sloan Kettering Cancer Center (MSK), in partnership with Quest Diagnostics.

Newton Y, Novak AM, Swatloski T, McColl DC, Chopra S, Graim K, Weinstein AS, Baertsch R, Salama SR, Ellrott K, Chopra M, Goldstein TC, Haussler D, Morozova O, Stuart JM

TumorMap: exploring the molecular similarities of cancer samples in an interactive portal

Cancer Res 2017 Nov 1;77(21):e111-e114

Outside of the clinical setting, high-dimensional -omics data are being routinely generated on large numbers of cancer samples. This includes genomic, transcriptomic, proteomic, and epigenomic profiles that bring data to terabyte and petabyte scales. Exploring these data in an intuitive and meaningful way is an unmet need; despite the rapidly falling costs of technology, analysis remains a major bottleneck. In this Focus on Computer Resources article, the authors describe TumorMap (https://tumormap.ucsc.edu/), an interactive portal for the exploration of molecular similarities across cancer samples. TumorMap uses Google's Map technology to arrange samples in a hexagonal grid based on their similarity, after applying user-selected dimensionality reduction techniques. The resulting maps can be colored by various attributes such as clinical, molecular, phenotype, and outcome data and metadata. This novel and practical methodology allows not only the assessment of similarity and dissimilarity but also supports experimental research purposes.

Seyednasrollah F, Koestler DC, Wang T, Piccolo SR, Vega R, Greiner R, Fuchs C, Gofer E, Kumar L, Wolfinger RD, Winner KK, Bare C, Neto EC, Yu T, Shen L, Abdallah K, Norman T, Stolovitzky G, Soule HR, Sweeney CJ, Ryan CJ, Scher HI, Sartor O, Elo LL, Zhou FL, Guinney J, Costello JC, and Prostate Cancer DREAM Challenge Community

A DREAM challenge to build prediction models for short-term discontinuation of docetaxel in metastatic castration-resistant prostate cancer

JCO Clin Cancer Inform 2017 Aug 4;(1):1-15

Discontinuation of chemotherapy treatment for the reason of toxicity is as common as discontinuation for lack of efficacy. This may have major consequences when treatments have a very narrow therapeutic index. Predicting which patients are likely to experience early treatment failure is a critical issue in the oncology domain. One approach to this intractable problem is to engage the larger community to solve towards a common problem. The DREAM challenges are crowd-sourced competitions meant to accelerate progress in the resolution of various biomedical informatics problems; importantly they are built on the FAIR data principles: Findable, Accessible, Interoperable and Reusable. The article by Seyednasrollah, et al., describes the process of the Prostate Cancer DREAM Challenge, the first crowd-sourced competition in metastatic prostate cancer and possibly in any cancer, which was launched in 2015; primary results are described elsewhere. The challenge attracted 34 independent teams from around the world, and has led to post-challenge community collaborations. While the best results are only slightly better than reference, integrated time-dependent area under the curve (iAUC) of 0.791 versus 0.743, the successful operation of the challenge, which included data from four independent phase III randomized controlled trials, suggests a paradigm for crowd-sourced tasks in the oncology domain.