Editorial: using cancer ‘omics’ to understand cancer

Health & Medicine

Editorial on the Research Topic

Using Cancer ‘ Omics' to Understand Cancer

The notion of using the “ big data” approach to study human disease is not new. Scientists have been tapping data from studies of genomics, proteomics, transcriptomics, metabolomics, and microbiomics since the initial mapping of the human genome (1). What has changed, however, is a fundamental shift in how we think about these technologies. The “ omics” field is expanding in scope, blending biology, technology (radiomics), and clinical observations (electronic health records), as well as size. This amplification of content and quantity has required parallel development and application of novel informatic tools. The need to accommodate the ever-larger datasets critical to our understanding of cancer omics has instigated a movement toward development of high-performance computing, including both hardware and software to analyze the massive, generated big data. The manuscripts contained in this volume reflect this constantly evolving panel of bioinformatic programs and resources with capacity to carry out large-scale data analysis.

Most of the papers in this issue report findings that share the common feature that all distill a select number of biomarkers from a large spectrum of potential markers from an analysis of large datasets. This volume of Frontiers broadens its approach to include papers dealing directly with the attributes, management, and clinical application of big data. Focusing on some of the key databases, projects and methodologies developed to implement such analyses, emphasizing the ever-expanding scale of big data, exascale computing is discussed. At the initiation of marker discovery, the

patients and other individuals who serve as the source of big data are highlighted, while encouraging big data researchers to keep in mind the humanity inherent in these data (Helzlsouer et al.).

To date much of our focus has been on comparing the omics information of cancer patients with that of “ normal” controls, i. e., healthy individuals, and looking for genotypic or phenotypic differences that set the patients apart. This rudimentary approach has led to practical applications, including offering targets for early detection, prognosis, and treatment. Along these lines, in this special issue of Frontiers, several authors address genomic (and epigenomic) abnormalities that characterize specific cancers and may thus have practical applications at the clinical level.

The manuscript by Yang et al. offers an example of the application of omics research to biomarker discovery. This paper describes potential diagnostic markers and therapeutic targets for leiomyosarcoma (LMS). This cancer is particularly aggressive, with invasive clinical characteristics and often a poor prognosis. Finding new biomarkers to assess malignancy and prognosis of LMS is critical. Yang et al. used Weighted Gene Co-expression Network Analysis (WGCNA), a systematic molecular clustering approach, to look for gene expression patterns that are associated with LMS and thereby should help to improve our understanding of the molecular mechanisms of this cancer. Their results showed that the expression of CDK4, CCT2, and MGAT1 in LMS tissues was significantly higher than that in adjacent tissues, suggesting that these genes may be part of the cancer signaling pathway. Such findings could pave the way for new strategies for diagnosing and treating LMS.

Another cancer, glioblastoma multiforme (GBM), is the focus of two articles in this special volume, by Cheng et al. and by Stajkovska et al. Cheng et al. employed a data mining approach by tapping into The Cancer Genome Atlas (TCGA). TCGA is managed by the Genomic Data Commons (GDC) (2) funded by the National Cancer Institute (NCI) which provides the cancer research community with a unified data repository that enables data sharing across cancer genomic studies in support of precision medicine. They then applied various bioinformatic tools aimed at discovery of relevant genes and pathways. They examined the gene expression patterns of transcription factors associated with GBM and identified four potential candidates based on their differential expression between tumor and adjacent tissue: *LHX2, MEOX2, SNAI2* , and *ZNF22* . By clustering transcription factors that are differentially expressed in GBM and screening these clusters using appropriate bioinformatic programs, they identified cancer pathways primarily associated with cell migration, cell adhesion, epithelial-mesenchymal transition (EMT), cell cycle, as well as other signaling pathways. Combining these results with patient characteristics, such as risk score, age, gender, type of treatment, and treatment response, these authors showed that their model was able to precisely predict the outcome of patients with GBM. GBM was further explored in the study by Stajkovska et al. , in their description of a case report of a pediatric patient. Using targeted gene panel testing in blood and tumor tissue, these researchers identified a heterozygous frameshift mutation (c. 333_334delTC; p. His112CysfsTer9) in the *MLH1* gene in addition to a known heterozygous missense variant of unknown significance/VUS (c. 847C > T; p. Arg283Cys) in the *TP53* gene. Screening of the patient's parents revealed the presence of

the *MLH1* abnormality in the father and the *TP53* variant in the mother. They report for the first time the co-occurrence of a genetic mutation in the *MLH1* gene of the mismatch repair pathway, often associated with Lynch syndrome, accompanied by a rare variant in the *TP53* gene. The authors stress that co-occurrence of multiple gene abnormalities should be considered as a possible contributory cause of a cancer. However, caution must be exercised in interpreting a VUS as contributing to the cancer phenotype, as these variants are of unproven pathogenicity, a subject addressed in Helzlsouer et al. in this volume.

Biomarkers also are the focus of the study by Wang Y. et al. , who looked at new ways of predicting the progression and prognosis of bladder cancer (BC) using a big data approach. Through a series of screenings and WGCNA they identified “ hub” genes (i. e., a hub gene serves as the focal point of interaction with other genes; in general, the genes connected to the hub are critical to gene regulation and other biological processes). Gene-set enrichment analysis (GSEA) revealed that the sets of highly expressed hub genes were mainly enriched in “ bladder cancer,” “ cell cycle,” and “ ubiquitin-mediated proteolysis” related pathways. They further honed their results to two genes (*ANLN, HMMR*), which had prognostic value for different stages and grades of BC. These genes not only could accurately predict the overall survival of patients with BC, but also the progression-free survival, a common outcome measure in clinical trials.

In another biomarker study included in this volume, Wang X. et al. showed how a set of small nucleolar RNAs (snoRNAs), which guide the modification of other RNAs and which have been implicated in alternative splicing, can

predict overall survival of gastric cancer patients. An eight-snoRNA risk signature serves as a prognostic factor in gastric cancer. The authors validated the expression patterns of these eight snoRNAs, both in cell lines and patients' tissues. The authors point out that seven of these snoRNAs correlate with survival, suggesting relevance of these markers to the clinical behavior of the bladder cancer. One snoRNA, U66, was linked to cell proliferation. These findings provide potential prognostic and therapeutic clues into gastric cancer.

2. Genomic Data Commons (GDC). Available online at: https://gdc. cancer. gov .

3. Cancer Research Data Commons/CRDC. Available online at: https://datascience. cancer. gov/data-commons .

4. Meerzaman D, Dunn BK. Value of collaboration among multi-domain experts in analysis of high-throughput genomics data. *Cancer Res.* (2019) 79: 5140–5. doi: 10. 1158/0008-5472