

CANCER PANOMICS: COMPUTATIONAL METHODS AND INFRASTRUCTURE FOR INTEGRATIVE ANALYSIS OF CANCER HIGH-THROUGHPUT “OMICS” DATA

SØREN BRUNAK

*Center for Biological Sequence Analysis Department of Systems Biology,
Technical University of Denmark, Copenhagen, Denmark
Email: brunak@cbs.dtu.dk*

FRANCISCO M. DE LA VEGA

*Real Time Genomics, Inc., San Bruno, CA, USA
Email: francisco@realtimegenomics.com*

GUNNAR RÄTSCH

*Computational Biology Center,
Memorial Sloan-Kettering Cancer Center, New York City, NY, USA
Email: raetsch@cbio.mskcc.org*

JOSHUA M. STUART

*Center for Biomolecular Science and Engineering,
University of California Santa Cruz, CA, USA
Email: jstuart@soe.ucsc.edu*

Precision medicine promises to transform cancer treatment in the next decade through the use of high-throughput sequencing and other technologies to identify telltale molecular aberrations that reveal therapeutic vulnerabilities of each patient's tumor [1]. This session will address the "panomics" of cancer – the complex combination of patient-specific characteristics that drive the development of each person's tumor and response to therapy [2]. The realization of this vision will require novel infrastructure and computational methods to integrate large-scale data effectively and query it in real-time for therapy and/or clinical trial selection for each patient.

The session will explore the computational needs to enable precision oncology from both the academic, industrial, and healthcare viewpoints. New methods and infrastructure to integrate multiple "omics" datasets (e.g., proteome, genome, exome, transcriptome), as well as existing clinical data types to enable precision medicine (e.g., medical literature, electronic medical records, clinical trial data, histopathology) will be discussed. The session is particularly interested in discussing pathway disruption analysis by combining data from different "omics" sources in single patients; joint analysis of "omics" data, literature, clinical trial data, and medical records; data structures & systems to enable big-data integrative analysis in patients. A summary of the accepted papers in this volume is below.

One of the most successful bioinformatics applications to cancer diagnosis and prognosis has been the identification and development of biomarkers that can distinguish disease subtypes, predict mutation status, or predict outcomes or treatment responses. However, the field is still in need of strategies that develop robust signatures as current methodologies often fail to translate

across studies and platforms (e.g., microarray- to RNA-Sequencing-based signatures). Two methods for novel biomarker discovery will be presented including an integrative approach by Min *et al.* as well as a method by Morgan *et al.* that combines multiple expression studies to identify more reliable robust gene expression-based signatures. In addition to the biomarker studies, machine-learning models for predicting the sensitivity of a cell to a drug based on its omics profile will be discussed including a comparison of methods in a comprehensive cell line panel by Jang *et al.*, the description of a new ensemble-based methods called Stream described by Chaibub Neto *et al.*, and an integrative method introduced by Mayba *et al.*

Interpreting the role of specific mutations in somatic cells is a fundamental problem in the individualized treatment of cancer. Identifying driver from passenger events and the assessment of the gain- or loss-of-function of specific proteins may offer important clues for drug targeting. Two papers investigate omics-derived statistical patterns to assess the functional role of somatic variants including connecting such events to the germline by Hu *et al.* and one that leverages protein-protein interactions to identify possibly important driving events by Badea *et al.* A new method for assembling haplotypes that is key for the interpretation of the combined influence of multiple variant alleles on the cancer phenotype is described in Aquilar *et al.*

Finally, to maximize the benefit of the cancer panomics endeavor, findings in the n=1 setting must be distributed in a way to empower the next n=1 analysis. Approaches that can interlink the findings of patients, doctors, trials, and researchers in one system would enable a new era of integrative approaches. Gitter *et al.* in this session describe one such strategy for approximating the influence of genetic pathways in disease.

Cancer panomics as applied to the individual patient is an emerging area driven by the lowering in cost of sequencing a patient's tumor and germline tissues. There is every expectation that the costs will continue their downward spiral once the competitive landscape of the industry and the maturity of 3rd or 4th generation sequencing technologies improve. In the very near future it will be feasible to sequence the complete cancer tumor genome and transcriptome as a routine procedure rather than just a targeted set of genes. The result of all this sequencing will mean that the bottleneck for the treatment of patients will transition from data production to the computational analysis of these massive information troves. Thus, it is critical to continue the discussion of novel bioinformatics ideas and strategies to empower the development of new cancer panomics approaches in the near future.

References

1. Craig, D.W. *et al.* *Mol. Cancer Ther.*, **12**, 104–116 (2012).
2. ASCO Board of Directors. *American Society of Clinical Oncology*, 1–16 (2012).