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

Annai Systems, Inc., Burligname, CA, USA
Email: Francisco.dlv@gmail.com

ADAM MARGOLIN

Oregon Health & Science University, Portland, OR, USA
Email: margolin@ohsu.edu

BENJAMIN J. RAPHAEL

Brown University, Providence, RI, USA
Email: braphael@brown.edu

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

Targeted cancer treatment is becoming the goal of newly developed oncology medicines and has already shown promise in some spectacular cases such as the case of BRAF kinase inhibitors in BRAF-mutant (e.g. V600E) melanoma.[1] These developments are driven by the advent of high-throughput sequencing, which continues to drop in cost, and that has enabled the sequencing of the genome, transcriptome, and epigenome of the tumors of a large number of cancer patients in order to discover the molecular aberrations that drive the oncogenesis of several types of cancer.[2] Applying these technologies in the clinic promises to transform cancer treatment by identifying therapeutic vulnerabilities of each patient's tumor.[3] These approaches will need to address the panomics of cancer – the integration of the complex combination of patient-specific characteristics that drive the development of each person's tumor and response to therapy. This in turn necessitates new computational methods to integrate large-scale "omics" data for each patient with their electronic medical records, and in the context of the results from large-scale pan-cancer research studies, to select the best therapy and/or clinical trial for the patient at hand.

This session of the Pacific Symposium on BioComputing 2015 features papers that address a range of issues from the analytical methods to discover molecular aberration from high-throughput sequencing

data, the prediction of pharmacogenetic effects in oncology drugs, the implications of clinical trials in personalized oncology, and the resources and IT infrastructure required to support the analysis of patient data.

The first set of contributions target novel analysis methodologies. One of the challenges of analyzing sequencing data from tumor samples is the fact that tumors are heterogeneous, and often composed of many sub-clones with different somatic mutations. The ability to deconvolute these mixtures and understand what somatic mutations co-occur in a given sub-clone, is important to understand the drivers of drug resistance of a particular tumor lineage. The work of Deshwar *et al.* aims to compare novel Bayesian methods for sub-clonal reconstruction of tumors that provides faster execution time and better resolution than other commonly used methods. On the subject of the analysis of transcriptome datasets, Lehmann *et al.* tackle the analysis and interpretation of the variation of alternative splicing in tumor samples, and are able to correlate this variability with QTLs that map to variants previously implicated in susceptibility to cancer and other traits, information which could be helpful when integrating a patient's germline with their tumor genome and corresponding transcriptome. Jang *et al.* focus on the development of pharmacogenetic predictive models utilizing domain-specific priors, and demonstrate that stepwise group sparse regression performs more accurately and provides better interpretability than purely data driven methods. On the other hand, Wu *et al.*, deal with the upcoming challenge of combination therapy - how to design and interpret trials where two drugs are provided in combination and reported in ClinicalTrials.gov.

On the practical side, performing panomic analysis of patient samples in the clinic would require IT infrastructures and resources that can deliver results in a fast time frame and can support the analyst to interpret and validate new molecular biomarkers. Nasser *et al.* present an integrated framework to analyze and present genome/transcriptome data of patients focused on clinical interpretation. They describe the challenges in developing a platform that can deliver results in a 24-hour timeframe. Ching *et al.* focus on the development of an online resource that integrates expression, copy number variation, mutation, compound activity, and meta data from cancer cells coming from publicly available projects, the Cell Index Database.

Precision oncology will require the analysis of data collected from a single patient. It will need to draw timely hypotheses about treatment in the context of prior knowledge of the specific form of cancer and based on information about the individual patient. This "n=1" approach is significantly different from that of pattern discovery on large patient cohorts. Continued development of new methodology to meet the demand is clearly still an active area to pursue by the computational biology community.

References

1. Ribas, A. & Flaherty, K. T. BRAF targeted therapy changes the treatment paradigm in melanoma. *Nature Publishing Group* **8,** 426–433 (2011).
2. Omberg, L. *et al.* Enabling transparent and collaborative computational analysis of 12 tumor types within The Cancer Genome Atlas. *Nat. Genet.* **45,** 1121–1126 (2013).
3. Craig, D. W. *et al.* Genome and transcriptome sequencing in prospective refractory metastatic triple negative breast cancer uncovers therapeutic vulnerabilities. *Mol. Cancer Ther.* **12,** 104–116 (2012).