THE FUTURE OF GENOME-BASED MEDICINE

QUAID MORRIS
University of Toronto, Donnelly Centre,160 College Street, Toronto, ON M5S 3E1, Canada
Email: QUAID.MORRIS@UTORONTO.CA

STEVEN E. BRENNER
Department of Plant & Microbial Biology, 111 Koshland Hall, University of California, Berkeley 94720
Email: brenner@compbio.berkeley.edu

JENNIFER LISTGARTEN
MICROSOFT RESEARCH, 110 GLENDON AVENUE, SUITE PH1, LOS ANGELES, CA
EMAIL: JENNL@MICROSOFT.COM

OLIVER STEGLE
MAX PLANCK INSTITUTES TÜBINGEN, 72076 TÜBINGEN, GERMANY
EMAIL: *OLIVER.STEGLE@TUEBINGEN.MPG.DE*

There has been unprecedented public investment in sequencing human and cancer genomes in the hopes of understanding disease [1, 2]. At the same time, large genome-wide association studies have helped elucidating the genetic underpinning of common diseases, identifying thousands of putative disease relevant loci [7, 8]. Complementary molecular profiling studies have revealed that several of these loci are co-associated with individual mRNA levels, suggesting candidate pathways that are putative mediators of genetic signals [7]. Coupled with the public investment, there is considerable personal investment in genetic profiling, being offered by companies such as 23andMe, deCODEme and others. Although this work has led to amazing discoveries, such as the surprising genetic, subclonal diversity within tumor populations (e.g., [3-5]), it's not clear how much how these insights will improve personalization of medicine.

In this workshop, we hope to address questions about how much genome sequencing has helped our understanding of the causal factors in disease and how much will these data change the way we treat disease in the clinic. Are genome clinics even realistic? If not, what other data will we need on individual patients before genome-based personalized medicine is possible?

Establishing causal mutations

Genetic variants cause disease through their impact on cell function. Despite promising recent efforts to predict phenotype from genotype in single cells [6], we are still far from being able to connect each mutation with its molecular and, ultimately, gross phenotypic consequences in humans. Establishing the causal mutation(s) is often a prerequisite to subsequent therapeutic and preventive actions especially for disease caused by a rare or somatic variant. So, a major challenge in genome-based medicine is distinguishing disease-causing polymorphisms among a large number of candidates, many of which are spuriously correlated.

In cancer, the causal mutations are called driver mutations because they contribute to the progression of cancer; these variants must be distinguished from passenger mutations that have little effect on cell function but appear as a result of greatly increased somatic mutation rates in cancer cells. In our workshop, we have

two invited speakers who will present different strategies to solving the problem of identifying driver mutations. One approach is to identify those regions that are aberrant more often than expected across different tumors of the same type. This approach assumes that driver mutations appear in regions that are under positive selection during cancer biogenesis. But limits on the resolution of this analysis make it difficult to identify driver regions that are small enough to contain a single gene or single functional element. **Dr Dana Pe'er** will introduce Helios, a new method that incorporates multiple layers of cancer profiling data to increase the effective resolution of the analysis, allowing to identifying driver mutations. Another approach to find driver mutations is to infer the evolutionary history of subclonal populations of cancer cells by genetically profiling multiple sites within primary and metastatic tumors and deconvolving the subclonal evolutionary structure from these data. **Dr Sohrab Shah** will discuss analysis approaches of ovarian tumors that embody this strategy.

Connecting genetic variants to molecular and disease phenotypes
Even if the causal mutation(s) can be identified, their phenotypic impact often remains unclear, especially if they do not directly affect the encoded protein sequence. Fortunately, genomes are being sequenced along with their products on the level of RNA or other molecular layers, allowing connections to be made between genotype and molecular phenotypes. In some cases, the RNA profile alone can be used to select therapy. **Dr Lars Steinmetz** will discuss efforts to use yeast as a model system for identifying molecular signals that suggest targets for therapeutic intervention. Ultimately, as we learn more and more about genome function in coding and non-coding regions [9], we should be able to design algorithms that predict the functional consequences of individual mutations. **Dr Steven Brenner** will discuss the Critical Assessment of Genome Interpreation (CAGI) project, a community experiment to objectively assess computational methods for predicting the phenotypic impact of genome variation.

Acting on mutations
Finally, once the causal mutations and their phenotypic consequence are identified, their remains the problem of how to design targeted treatment. This remains a significant, unsolved problem but some progress has been made in identifying actionable mutations based on known (or assumed) targets of current drugs [10].

Workshop contributions
The workshop includes invited talks from researchers active in genome-based clinical research.

Dana Pe'er is an Associate Professor in Biology and Systems Biology at Columbia University, New York. She is the director of the laboratory on Computational Systems Biology and is an active researcher in both the systems biology and machine learning communities. Dana pioneered the application of Bayesian networks and Bayesian modeling techniques to molecular profiling data. She will discuss new computational methodology to integrate multiple cancer genome profiling layers to identify copy number-based driver mutations.

Sohrab Shah is Assistant Professor in the Dept. of Pathology at the University of British Columbia and a research scientist at the BC Cancer Agency. He heads a computational biology laboratory that combines deep expertise in genome sequencing and analysis with machine learning methodology development to make discoveries in breast and ovarian cancer, including two recent Nature papers. In this talk, he will discuss recent approaches to study the variation between spatially and temporally distinct tumor specimens in ovarian cancer. Understanding the variation between tumor specimens within individual patients can be used to detect driver mutations and the evolution of mutational accumulation.

Lars Steinmetz is co-chair of the genome biology unit at the European Molecular Biology Laboratory in Heidelberg, which consists of over 100 scientists and 9 research teams. In parallel, Dr. Steinmetz leads a focused research team at the Stanford Genome Technology Centre in the USA. Lars is a leading scientists at the forefront of genetic and genomics research. His lab pioneered the development and the application of high-throughput approaches to functionally profile genetic and molecular systems at a genome-wide scale. His lecture will address how observational molecular and genetic profiling information can be used to identify causal molecular mediators that confer genetic signals to phenotype. The hypotheses generated by computational modeling are systematically validated in a yeast model.

Steven E. Brenner is Professor in the Department of Plant and Microbial Biology at the University of California, Berkeley with adjunct appointments in Bioengineering and Therapeutic Sciences. He has won numerous awards for his research including the prestigious ICSB Overton Prize reflecting his broad, seminal contributions to computational biology in diverse areas including alternative splicing, protein evolution, critical assessment of bioinformatics methodology, and, most recently, genome-based medicine. His presentation will give an overview of recent community efforts to assess genome interpretation. The Critical Assessment of Genome Interpreation (CAGI) is a community experiment to objectively assess computational methods for predicting the phenotypic impact of genome variation. CAGI has revealed the relative strengths of different prediction approaches, showing some that worked consistently well, while other classes worked only on special types of problems. Even with the simplest dataset, involving nonsynonymous mutations in a human metabolic enzyme, yielded great variability of the result. Overall, CAGI revealed very significant biomedical insights into the implications of genetic variation are embodied in current algorithms, but that the ability of generic methods to make clinically important decisions is presently limited.

References

1. International Cancer Genome, C., et al., *International network of cancer genome projects.* Nature, 2010. **464**(7291): p. 993-8.
2. Siva, N., *1000 Genomes project.* Nat Biotechnol, 2008. **26**(3): p. 256.
3. Gerlinger, M., et al., *Intratumor heterogeneity and branched evolution revealed by multiregion sequencing.* N Engl J Med, 2012. **366**(10): p. 883-92.
4. Shah, S.P., et al., *The clonal and mutational evolution spectrum of primary triple-negative breast cancers.* Nature, 2012. **486**(7403): p. 395-9.
5. Schuh, A., et al., *Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns.* Blood, 2012.
6. Karr, J.R., et al., *A whole-cell computational model predicts phenotype from genotype.* Cell, 2012. **150**(2): p. 389-401.
7. Visscher, P.M., et al., *Five years of GWAS discovery.* Am J Hum Genet, 2012. **90**(1): p. 7-24.
8. Hu, X. and M. Daly, *What have we learned from six years of GWAS in autoimmune diseases, and what is next?* Curr Opin Immunol, 2012.
9. Consortium, E.P., et al., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
10. Dancey, J.E., et al., *The genetic basis for cancer treatment decisions.* Cell, 2012. **148**(3): p. 409-20.