ABSTRACT

Somatic single nucleotide variants (SNVs) are mutations resulting from the substitution of a single nucleotide in the genome of cancer cells. Somatic SNVs are numerous in the genomes of most types of cancers. SNVs can contribute to the malignant phenotype of cancer cells, though many SNVs likely have negligible selective value. Because many SNVs are selectively neutral, their presence in a measurable proportion of cells is likely due to drift or genetic hitchhiking. This makes SNVs an appealing class of genomic aberrations to use as markers of clonal populations and ultimately tumour evolution. Advances in sequencing technology, in particular the development of high throughput sequencing (HTS) technologies, have made it possible to systematically profile SNVs in tumour genomes. We introduce three probabilistic models to solve analytical problems raised by experimental designs that leverage HTS to study cancer biology.

The first experimental design we address is paired sequencing of normal and tumour tissue samples to identify somatic SNVs. We develop a probabilistic model to jointly analyse data from both samples, and reduce the number of false positive somatic SNV predictions.

The second experimental design we address is the deep sequencing of SNVs to quantify the cellular prevalence of clones harbouring the SNVs. The key challenge we resolve is that allele abundance measured by HTS is not equivalent to cellular prevalence due to the confounding issues of mutational genotype, normal cell contamination and technical noise. We develop a probabilistic model which solves these problems while simultaneously inferring the number of clonal populations in the tissue.

The final experimental design we consider is single cell sequencing. Single cell sequencing provides a direct means to measure the genotypes of clonal populations. However, sequence data from a single cell is inherently noisy which confounds accurate measurement of genotypes. To overcome this problem we develop a model to aggregate cells by clonal population in order to pool statistical strength and reduce error. The model jointly infers the assignment of cells to clonal populations, the genotype of the clonal populations, and the number of populations present.

BIOGRAPHICAL NOTES

Place of Birth: Toronto, Canada

Academic Studies: B. Sc. University of British Columbia, 2009

GRADUATE STUDIES

Field of Study: Cancer genomics

Courses

BIOF 501A	Special Topics in Bioinformatics	Dr. R. Brinkman
MBB 505	Problem Based Learning in Bioinformatics	Dr. S. Jones
CPSC 540	Machine Learning	Dr. K. Murphy
CMPT 711	Bioinformatics Algorithms	Dr. C. Sahinalp
MBB 741	Bioinformatics	Dr. J. Chen
STAT 801	Statistics	Dr. J. Cao

AWARDS

2013 CIHR CGS-D
2009 NSERC CGS-M
2009 CIHR Bioinformatics Training Grant
2005 NSERC USRA

PUBLICATIONS

Roth et al. PyClone: Statistical Inference Of Clonal Population Structure In Cancer From Deep Digital Sequencing. Nature Methods, 2014.
Roth et al. JointSNVMix: A Probabilistic Model For Accurate Detection Of Somatic Mutations In Normal/Tumour Paired Next-generation Sequencing Data. Bioinformatics, 2012.

PRESENTATIONS

Probabilistic models for identification and interpretation of somatic single nucleotide variants in cancer genomes, Wellcome Trust Centre for Human Genetics, Oxford, UK, 2015.
Inference Of Clonal Genotypes From Single Cell Sequencing Data, HiTSeq, Dublin, Ireland, 2015.
Inferring The Clonal Population Structure of Cancers Using Deeply Sequenced Mutations, Bertinoro Computational Biology, Bertinoro, Italy, 2013.
Inferring The Cellular Frequency Of Somatic Mutations From Next Generation Sequencing Data, HiTSeq, Long Beach, USA, 2012.
JointSNVMix : A Probabilistic Model For Accurate Detection Of Somatic Mutations In Normal/Tumour Paired Sample Sequence Data, HiTSeq, Vienna, Austria, 2011.

SUPERVISORY COMMITTEE
Dr. Sohrab Shah
Dr. Samuel Aparicio
Dr. Alexandre Bouchard-Côté
Dr. Steven Jones

a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA
Graduate and Postdoctoral Studies

PROGRAMME

The Final Oral Examination
For the Degree of

DOCTOR OF PHILOSOPHY
(Bioinformatics)

ANDREW ROTH

B.Sc., University of British Columbia, 2009

Thursday, November 26, 2015, 9:00 am
Room 200, Graduate Student Centre
Latecomers will not be admitted

"Probabilistic Models for the Identification and Interpretation of Somatic Single Nucleotide Variants in Cancer Genomes"

EXAMINING COMMITTEE

Chair:
Dr. John Petkau (Statistics)

Supervisory Committee:
Dr. Sohrab Shah, Research Supervisor (Pathology and Laboratory Medicine)
Dr. Samuel Aparicio (Pathology and Laboratory Medicine)
Dr. Alexandre Bouchard-Côté (Statistics)

University Examiners:
Dr. Daniel Coombs (Mathematics)
Dr. Paul Pavlidis (Bioinformatics)

External Examiner:
Dr. Niko Beerenwinkel
Department of Biosystems Science and Engineering
ETH Zurich
Zurich, Switzerland