ABSTRACT

The rapidly falling cost and increasing quality of DNA and RNA sequencing associated with the "Next-Generation" sequencing platforms has required the development of new techniques and algorithms for processing and analyzing the ever increasing volume of data being generated, expanding the opportunity for bioinformatics research in the area of high-throughput sequencing.

The development of these new sequencing platforms has also enabled the development of new protocols, including ChIP-Seq, which uses Chromatin Immunoprecipitation and massively parallel sequencing to identify genomic binding sites of nuclear-associated proteins with high specificity. The availability of this protocol has been one of the driving forces behind the growing field of epigenetics, in which gene expression changes and heritable traits are produced without modifications to the DNA of a cell. One of first software applications, and one of the most widely used for the protocol is the FindPeaks software described in this thesis.

A second challenge associated with the dramatic increase in data made available by the adoption of next-generation sequencing is the storage and retrieval of variations across a wide variety of samples collected. We also describe the Variation Database, a scalable open source database schema and API that has been deployed to provide rapid access to single nucleotide variations across a wide variety of sequencing data from next-generation platforms. The Variation Database deployed at the Genome Sciences Centre has been utilized to gain insight into a variety of diseases and to identify recurrent variations in several forms of cancer.

Finally, we describe work done to gain insight into the mechanisms of breast cancer through the sequencing and analysis of eight mammary ductal carcinoma cell lines and a set of four B-cell derived matched normal cell lines, illustrating similarities and differences between them. With both RNA and DNA data available, a rich background of information, including RNA-expression and single nucleotide substitutions, allows for a diverse set of metrics to be explored. This work also includes the proposal of a pathway exhibiting differential expression between the cancer-derived and normal-derived cell lines and shows a wide variety non-synonymous disruptions.

BIOGRAPHICAL NOTES

Born:	March 7th, 1977, Winnipeg, Manitoba
Academic Studies:	B.Sc., University of Waterloo, Honours Co-op Biochemistry, 2000 B.I.S., University of Waterloo, 2001 M.Sc., University of British Columbia, Microbiology & Immunology, 2004
Current Position:	Ph.D. Candidate, University of British Columbia Bioinformatician, CLCbio A/S, Århus, Denmark

GRADUATE STUDIES

Field of Study:	Bioinformatic applications for high-throughput DNA and RNA sequencing

SELECTED AWARDS

- Michael Smith Foundation for Health Research, Research Trainee Award
- Genome BC Scientific Travel Award
- Award of Excellence (Silver category), Canadian Institutes of Health Research National Research Poster Competition.
- BIOTECanada TVG Student Bursary 2010

SELECTED PUBLICATIONS

Fejes AP, Hadj Khodabakhshi A, Birol I, Jones SJM, Human Variation Database: An open source database template for genomic discovery. Bioinformatics, 2011 April 15, 27(8):1155-1156

Fejes AP, Robertson G, Bilenky M, Varhol R, Bainbridge M, and Jones SJM. 2008. FindPeaks 3.1: A Tool for Identifying Areas of Enrichment from Massively Parallel Short-Read Sequencing Technology, Bioinformatics 2008, 24(15):1729-30

Heravi-Moussavi, A, Anglesio, MS, Cheng, SW, Senz, J, Yang, W, Prentice, L, Fejes, AP, Chow, C, Tone, A, Kalloger, SE, Hamel, N, Roth, A, Ha, G, Wan, AN, Maines-Bandiera, S, Salamanca, C, Pasini, B, Clarke, BA, Lee, AF, Lee, CH, Zhao, C, Young, RH, Aparicio, SA, Sorensen, PH, Woo, MM, Boyd, N, Jones, SJ, Hirst, M, Marra, MA, Gilks, B, Shah, SP, Foulkes, WD, Morin, GB, Huntsman, DG,. Recurrent somatic DICER1 mutations in nonepithelial ovarian cancers. 2012 New England Journal of Medicine, 366, 3:234-42.

SUPERVISORY COMMITTEE

Dr. Steven J.M. Jones
Dr. Marco A. Marra
Dr. Angela Brooks-Wilson
Dr. Paul Pavlidis

Faculty of Graduate Studies

THE UNIVERSITY OF BRITISH COLUMBIA

PROGRAMME

The Final Oral Examination
For the Degree of

DOCTOR OF PHILOSOPHY
(Bioinformatics)

ANTHONY PETER FEJES

B.Sc., University of Waterloo, 2000
B.I.S., University of Waterloo, 2001
M.Sc., University of British Columbia, 2004

Friday, March 30, 2012, 12:30 pm
Room 200, Graduate Student Centre
Latecomers will not be admitted

"Algorithms and Applications of Next-Generation DNA Sequencing"

EXAMINING COMMITTEE

Chair:
Dr. Patricia Schulte (Zoology)

Supervisory Committee:
Dr. Steven J.M. Jones, Research Supervisor (Medical Genetics)
Dr. Angela Brooks-Wilson (Medical Genetics)
Dr. Paul Pavlidis (Psychiatry)

University Examiners:
Dr. Phil Hieter (Medical Genetics)
Dr. I. Robert Nabi (Cell and Developmental Biology)

External Examiner:
Dr. Gary Bader
Donnelly Centre for Cellular and Biomolecular Research
University of Toronto
Toronto, Ontario