ABSTRACT

Genome aberrations and chromosomal instability are hallmarks of malignant human cancers. These mutational abnormalities, which encompass copy number alterations (CNA) and loss of heterozygosity (LOH), can have a measurable effect on the structure and dosage of chromosomal regions. Tumour suppressors and oncogenes that are altered by CNAs can contribute to a tumourigenic phenotype. Branched evolution throughout tumour progression results in genomic heterogeneity, such that divergent clones with distinct aberrations are often present at diagnosis. Measuring and modeling subclonal CNA/LOH events can elucidate the abundance of clonal populations. This enables the study of clonal evolution dynamics, which will have far-reaching implications for understanding modes of selection, and the genetic basis of metastatic potential and therapeutic resistance. Single nucleotide polymorphism (SNP) genotyping arrays and whole genome sequencing (WGS) are high-resolution genome-wide assays to profile the genomes of cancer cell populations. However, accurate and statistically robust computational methods for inferring CNA/LOH in these data remain under-developed.

I present three novel probabilistic approaches that apply hidden Markov models (HMM) to analyze CNA/LOH in tumour genomes. The first method is HMM-Dosage, which distinguishes somatic and germline events. This tool was used to profile the largest breast cancer dataset to date, reporting the copy number landscapes in 2000 patient. The second method is APOLLOH, which was one of the earliest methods developed to profile LOH in tumour WGS data. Its application to WGS of 23 triple negative breast cancers (TNBC) represents the first time that LOH and its effects on allelic expression were jointly analyzed from sequencing data for a cohort of this magnitude. The third method is TITAN, which is the first method to simultaneously infer CNA/LOH and estimate their cellular prevalence in tumour WGS. This framework provides an analytical route to studying the degree of clonal evolution, driven from the CNA/LOH perspective. I applied TITAN to a novel set of primary breast tumours and corresponding mouse xenografts, presenting the results of distinct modes of temporal clonal selection patterns. These approaches are built on fundamental principles immediately relevant to current technologies, and address the demand for robust algorithms in cancer genomics research.

BIOGRAPHICAL NOTES

Place of Birth: Vancouver, Canada

Academic Studies: B. Sc. University of British Columbia, 2008

Current Position: Ph.D. candidate, UBC (BC Cancer Research Centre)

GRADUATE STUDIES

Field of Study: Bioinformatics

Courses (500 level and above) for example:		**Instructors**
BIOF 501A	Special Topics in Bioinformatics	Drs. R. Brinkman and F. Pio
CPSC 545	Algorithms in Bioinformatics	Dr. I. Meyer
MEDG 505	Genome Analysis	Drs. P. Heiter and A. Rose
STAT 548G	Directed Studies	Dr. R. Balshaw
BIOF 520	Problem-Based Learning in Bioinformatics	Dr. F. Brinkman
CPSC 540	Machine Learning	Dr. N. de Freitas

AWARDS

Lloyd Skarsgard Graduate Research Excellence Award (2012)
NSERC PGS (2010-2013)
Bioinformatics Training Program CIHR/MSFHR Graduate Scholarship (2008-2010)

SELECTED PUBLICATIONS

Roth, A., Khattra, J., Yap, D., Laks, E., **Ha, G**., Aparicio, S, Bouchard-Cote, A., Shah, SP. (2014) PyClone: Statistical inference of clonal population structure in cancer from deep digital sequencing. *Nature Methods*, 11:396–398.

Bashashati, A.*, **Ha, G**.*, Tone, A.*, Ding, J., Prentice, L. M., Roth, A., Rosner, J., Shumansky, K., Kalloger, S., Senz, J., Yang, W., McConechy, M., Melnyk, N., Anglesio, M., Luk, M. T. Y., Tse, K., Zeng, T., Moore, R., Zhao, Y., Marra, M. A., Gilks, B., Yip, S., Huntsman, D. G., McAlpine, J. N., and Shah, S. P. (2013). Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *Journal of Pathology*, 231:21–34.

Ha, G., Roth, A., Lai, D., Bashashati, A., Ding, J., Goya, R., Giuliany, R., Rosner, J., Oloumi, A., Shumansky, K., Chin, S.-F., Turashvili, G., Hirst, M., Caldas, C., Marra, M. A., Aparicio, S., and Shah, S. P. (2012). Integrative analysis of genome-wide loss of heterozygosity and mono-allelic expression at nucleotide resolution reveals disrupted pathways in triple negative breast cancer. *Genome Research*, 22(10):1995–2007.

Shah, S. P., Roth, A.*, Goya, R., Oloumi, A.*, **Ha, G**.*, Zhao, Y.*, Turashvili, G.*, Ding, J.*, Tse, K.*, Haffari, G.*, Bashashati, A.*, Prentice, L. M., Khattra, J., Burleigh, A., Yap, D., Bernard, V., McPherson, A., Shumansky, K., Crisan, A., Giuliany, R., Heravi-Moussavi, A., Rosner, J., Lai, D., Birol, I., Varhol, R., Tam, A., Dhalla, N., Zeng, T., Ma, K., Chan, S. K., Griffith, M., Moradian, A., Cheng, S. W., Morin, G. B., Watson, P., Gelmon, K., Chia, S., Chin, S. F., Curtis, C., Rueda, O. M., Pharoah, P. D., Damaraju, S., Mackey, J., Hoon,

K., Harkins, T., Tadigotla, V., Sigaroudinia, M., Gascard, P., Tlsty, T., Costello, J. F., Meyer, I. M., Eaves, C. J., Wasserman, W. W., Jones, S., Huntsman, D., Hirst, M., Caldas, C., Marra, M. A., and Aparicio, S. (2012). The clonal and mutational evolution spectrum of primary triple- negative breast cancers. *Nature*, 486(7403):395–399.

Curtis, C., Shah, S. P., Chin, S. F., Turashvili, G., Rueda, O. M., Dunning, M. J., Speed, D., Lynch, A. G., Samarajiwa, S., Yuan, Y., Gräf, S., **Ha, G**., Haffari, G., Bashashati, A., Russell, R., McKinney, S., METABRIC Group, Langerod, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Borresen-Dale, A. L., Brenton, J. D., Tavaré, S., Caldas, C., and Aparicio, S. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, 486(7403):346–352.

PRESENTATIONS

"Profiling the Subclonal Copy Number Architecture from Whole Genome Sequencing of Heterogeneous Tumours" Seminars in Quantitative Biology, Cancer Research UK, Cambridge, UK. January 27, 2014.

"Probabilistic inference of subclonal copy number alterations and LOH in whole genome sequencing of tumours" 21st Annual International Conference on Intelligent Systems for Molecular Biology, High Throughput Sequencing Analysis and Algorithms (HiTSeq) Special Interest Group. Berlin, Germany. July 20, 2013.

"Profiling copy number aberrations and loss of heterozygosity mutational landscapes in cancer genomes" Research Seminar Series. BC Cancer Research Centre, Vancouver. Feb. 25, 2013.

"APOLLOH: copy number aware approach to detect loss of heterozygosity in tumour genome sequence data". 19th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB), High Throughput Sequencing Analysis and Algorithms (HiTSeq) Special Interest Group. Vienna, Austria. July 16, 2011.

SUPERVISORY COMMITTEE

Dr. Sohrab P Shah
Dr. Samuel Aparicio
Dr. Steven Jones
Dr. Raymond Ng
Dr. Wyeth Wasserman

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)

GAVIN HA

B. Sc., University of British Columbia, 2008

Example: Friday, June 27 2014, 12:30 pm
Room 203, Graduate Student Centre
Latecomers will not be admitted

"Probabilistic Approaches for Profiling Copy Number Aberrations and Loss of Heterozygosity Landscapes in Cancer Genomes"

EXAMINING COMMITTEE

Chair:
Dr. Keith Humphries (Medical Genetics)

Supervisory Committee:
Dr. Sohrab P Shah, Research Co-supervisor (Pathology and Laboratory)
Dr. Samuel Aparicio, Research Co-supervisor (Pathology and Laboratory)

University Examiners:
Dr. Jenny Bryan (Statistics)
Dr. Phil Heiter (Medical Genetics)

External Examiner:

Dr. Michael Hallett	Dr. Jens Lagergren
Centre for Bioinformatics	School of Computer Science
McGill University	Karolinska Institutet Science Park
Montreal, Quebec	Solna, Sweden