

## ABSTRACT

The human genome contains tens of thousands of gene loci which code for an even greater number of protein and RNA products. The highly complex temporal and spatial expression of these genes makes possible all the biological processes of life. Altered gene expression by mutation or deregulation is fundamental for the development of many human diseases. The ultimate aim of this thesis was to identify gene expression changes relevant to cancer. The advent of genome-wide expression profiling techniques, such as microarrays, has provided powerful new tools to identify such changes and researchers are now faced with an explosion of gene expression data. Processing, comparing and integrating these data present major challenges. I approached these challenges by developing and assessing novel methods for cross-platform analysis of expression data, scalable subspace clustering, and curation of experimental gene regulation data from the published literature. I found that combining results from different expression platforms increases reliability of coexpression predictions. However, I also observed that 'global' correlation between platforms was generally low, and few gene pairs reached reasonable thresholds for 'high-confidence coexpression'. Therefore, I developed a novel subspace clustering algorithm, able to identify coexpressed genes in experimental subsets of very large gene expression datasets. Biological assessment against several metrics indicates that this algorithm performs well. I also developed a novel meta-analysis method to identify consistently reported genes from differential expression studies when raw data are unavailable. This method was applied to thyroid cancer, producing a ranked list of significantly over-represented genes. Tissue microarray analysis of some of these candidates and others identified a number of promising biomarkers for diagnostic and prognostic classification of thyroid cancer. Finally, I present ORegAnno ([www.oreganno.org](http://www.oreganno.org)), a resource for the community-driven curation of experimentally verified regulatory sequences. This resource has proven a great success with ~30,000 sequences entered from over 900 publications by ~50 contributing users. These data, methods and resources contribute to our overall understanding of gene regulation, gene expression, and the changes that occur in cancer. Such an understanding should help identify new cancer mechanisms, potential treatment targets, and have significant diagnostic and prognostic implications.

## BIOGRAPHICAL NOTES

Born: March 7, 1978, Winnipeg, MB, Canada

Academic Studies: B.Sc. (Honours), University of Winnipeg, 2002

## GRADUATE STUDIES

Field of Study: Cancer informatics

### Courses

MEDG505 Genome Analysis  
MEDG520 Advances in Human Molecular Genetics  
MEDG530 Advanced Human Genetics  
MEDG545 Current Topics in Genomics  
MEDG548 Directed Studies  
A  
MBB829 Special Topics in Bioinformatics  
BIOL535 Teaching and Learning in Life Sciences

### Instructors

Drs. S. Jones & P. Hieter  
Dr. R. McMaster  
Dr. J. Friedman  
Dr. A. Brooks-Wilson  
Dr. S. Jones  
Dr. F. Pio  
Drs. E. Rosenberg & J. Berger

## AWARDS

2007 Lloyd Skarsgard Research Excellence Prize  
2006-2008 MSFHR Senior Trainee Award  
2005-2008 CIHR Canada Graduate Scholarship Doctoral Award  
2004 CIHR National Poster Competition Award of Excellence (Silver)  
2004 Albert B. and Mary Steiner Summer Research Award  
2003 UBC Graduate Entrance Scholarship  
2003-2005 MSFHR Junior Trainee Award  
2003-2005 NSERC PGS-A Award

## SELECTED PUBLICATIONS

**Griffith OL\***, Montgomery SB\*, Bernier B, Chu B, Aerts S, Sleumer MC, Bilenky M, Haeussler M, Griffith M, Gallo SM, Giardine B, Mahony S, Hooghe B, Van Loo P, Blanco E, Ticoll A, Lithwick S, Portales-Casamar E, Donaldson IJ, Robertson G, Wadelius C, De Bleser P, Vlieghe D, Halfon MS, Wasserman W, Hardison R, Bergman CM, Jones SJM. 2008. ORegAnno: an open-access community-driven resource for regulatory annotation. *Nucleic Acids Res.* 36:D107-13.  
Wiseman SM, **Griffith OL**, Melck A, Masoudi H, Gown A, Nabi R, Jones SJM. Evaluation of Type 1 Growth Factor Receptor Family Expression in 205 Thyroid Lesions Reveals Diagnostic Utility and Targeted Therapeutic Potential for HER1, HER3, and HER4. *Am J Surg.* Accepted. 17 Dec. 2007.  
Leung S, **Griffith OL**, Masoudi H, Gown A, Jones SJM, Phang T, Wiseman SM. Clinical Utility of Type I Growth Factor Receptor Expression by Colon Cancer. *Am J Surg.* Accepted. 6 Dec. 2007.  
Melck AL, Masoudi H, **Griffith OL**, Rajput A, Wilkins GE, Bugis S, Jones S, Wiseman SM. 2007. Cell Cycle Regulators Show Diagnostic and Prognostic Utility for Differentiated Thyroid Cancer. *Ann Surg Oncol.* 14:3403-3411.

Chan SK, **Griffith OL**, Tai IT, Jones SJM. Meta-analysis of Colorectal Cancer Gene Expression Profiling Studies Identifies Consistently Reported Candidate Biomarkers. *Cancer Epidemiol Biomarkers Prev*. Accepted. 6 Nov. 2007.

**Griffith OL**, Melck A, Jones SJM, Wiseman SM. Thyroid Cancer: Identification of Gene Expression Markers for Diagnosis. In *Methods of Cancer Diagnosis, Therapy and Prognosis*. Hayat MA, editor. Springer Publishing Company. New York, NY. Accepted. 5 Jul. 2007.

Wiseman SM, **Griffith OL**, Deen S, Rajput A, Masoudi H, Gilks B, Goldstein L, Gown A, Jones SJM. 2007. Identification of Molecular Markers Altered During Transformation of Differentiated Into Anaplastic Thyroid Carcinoma. *Arch Surg*. 142:717-729.

**Griffith OL**, Melck A, Jones SJM, Wiseman SM. 2006. A Meta-analysis and Meta-review of Thyroid Cancer Gene Expression Profiling Studies Identifies Important Diagnostic Biomarkers. *J Clin Oncol*. 24:5043-5051.

Gao BJ, **Griffith OL**, Ester M, Jones SJ. 2006. Discovering significant OPSM subspace clusters in massive gene expression data. In *Proceedings of the 12th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining* (Philadelphia, PA, USA, August 20 - 23, 2006). KDD '06. ACM Press, New York, NY, 922-928.

Montgomery SB\*, **Griffith OL\***, Bilenky M, Pleasance ED, Prychyna Y, Sleumer MC, Zhang X, Jones SJM. 2006. ORegAnno: An open access database and curation system for literature-derived promoters, transcription factor binding sites, and regulatory variations. *Bioinformatics*. 22:637-640.

**Griffith OL**, Pleasance ED, Fulton DL, Oveisi M, Ester M, Siddiqui AS, Jones SJM. 2005. Assessment and Integration of Publicly Available SAGE, cDNA Microarray, and Oligonucleotide Microarray Expression Data for Global Coexpression Analyses. *Genomics*. 86:476-488.

\*These authors contributed equally to this work.

## SELECTED PRESENTATIONS

**Griffith OL**, Montgomery SB, Bergman CM, Bilenky M, Chu B, Pleasance ED, Prychyna Y, Sleumer MC, Zhang X, Jones SJM. ORegAnno: A Community-Based Annotation System for Literature-Derived Regulatory sequences. Oral presentation. AGBT. Marco Island, Florida, USA. Feb. 2007.

**Griffith OL**. Annotation standards in Oreganno. Oral presentation. The RegCreative Jamboree, Ghent University, Ghent, Belgium. Nov. 2006.

**Griffith OL**, Pleasance E, Fulton D, Bilenky M, Robertson G, Oveisi M, Pan YJ, Ester M, Siddiqui AS, Jones SJM. Gene expression platforms for global coexpression analyses – assessment and integration for study of gene deregulation in cancer. Invited oral and poster presentation. International Workshop on Encoding Information in DNA Sequence. Okinawa, Japan. Feb. 2005.

## SUPERVISORY COMMITTEE

Dr. Steven Jones, Research Supervisor (Medical Genetics)

Dr. Marco Marra (Medical Genetics)

Dr. Angela Brooks-Wilson (Medical Genetics)

Dr. Isabella Tai (Medicine, Gastroenterology division)

Dr. Francis Ouellette (Medical Genetics)



**PROGRAMME**

The Final Oral Examination  
For the Degree of

DOCTOR OF PHILOSOPHY  
(Medical Genetics)

**OBI L. GRIFFITH**

B.Sc. (Honours), University of Winnipeg, 2002

Wednesday, April 2, 2008, 9:00am  
Room 200, Graduate Student Centre

**“Identification of Gene Expression Changes in Human Cancer Using  
Bioinformatic Approaches”**

**EXAMINING COMMITTEE**

Chair:

To be completed by FoGS

Supervisory Committee:

Dr. Steven Jones, Research Supervisor (Medical Genetics)  
Dr. Isabella Tai (Medicine, Gastroenterology division)

University Examiners:

Dr. George Mackie (Biochemistry and Molecular Biology)  
Dr. Michael Murphy (Microbiology and Immunology)

External Examiner:

To be completed by FoGS