Gene Expression Biomarkers for Colorectal Neoplasia

by

L. C. LaPointe

B.Sc. (Florida State University) 1991

Department of Medicine

FLINDERS UNIVERSITY OF SOUTH AUSTRALIA

Committee in charge:

Prof. Graeme P. Young, Chair

Dr. Robert A. Dunne
Dr. Peter L. Molloy 2008

Contents

Gene Expression Biomarkers for Colorectal Neoplasia

L. C. LaPointe

Flinders University of South Australia

Department of Medicine

Prof. Graeme P. Young

The aim of this research was to assemble sufficient experimental evidence about candidate gene transcript expression changes between non-neoplastic and neoplastic colorectal tissues to justify future assay development involving promising leads. To achieve this aim, this thesis explores the hypothesis that gene expression-based biomarkers can be used to accurately discriminate colorectal neoplastic tissues from non-neoplastic controls.

This hypothesis was tested by first analysing multiple, large, quality controlled data sets comprising gene expression measurements across colorectal phenotypes to discover potential biomarkers. Candidate biomarkers were then subjected to validation testing using a custom-design oligonucleotide microarray applied to independently derived clinical specimens. A number of novel conclusions are reached based on these data. The most important conclusion is that a defined subset of genes expressed in the colorectal mucosa are reliably differentially expressed in neoplastic tissues. In particular, the apparently high prediction accuracy achieved for single gene transcripts to discriminate hundreds of neoplastic and non-neoplastic tissues provides compelling evidence that the resulting candidate genes are worthy of further biomarker research.

In addition to addressing the central hypothesis, additional contributions are made to the field of colorectal neoplasia gene expression profiling. These contributions include:

The first systematic analysis of gene expression in non-diseased tissues along the colorectum To better understand the range of gene expression in non-diseased tissues, RNA extracts taken from along the longitudinal axis of the large intestine were studied.

The development of quality control methodologies for high dimensional gene expression data Complex data collection platforms such as oligonucleotide microarrays introduce the potential for unrecognized confounding variables. The exploration of quality control parameters across five hundred microarray experiments provided insights about quality control techniques.

The design of a custom microrray comprised of oligonucleotide probesets hybridising to RNA transcripts differentially expressed in neoplastic colorectal specimens A custom design oligonucleotide microarray was designed and tested combining the results of multiple biomarker discovery projects.

Introduction of a method to filter differentially expressed genes during discovery that may improve validation efficiencies of biomarker discovery based on gene expression measurements Differential expression discovery research is typically focused only on quantitative changes in transcript concentration between phenotype contrasts. This work introduces a method for generating hypotheses related to transcripts which may be qualitatively "switched-on" between phenotypes.

Identification of mRNA transcripts which are differentially expressed between colorectal adenomas and colorectal cancer tissues Transcripts differentially expressed between adenomatous and cancerous RNA extracts were discovered and then tested in independent tissues.

In conclusion, these results confirm the hypothesis that gene expression profiling can discriminate colorectal neoplasia (including adenomas) from non-neoplastic controls. These results also establish a foundation for an ongoing biomarker development program.

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief does not contain any material previously or written by another person except where due reference is made in the text.

.....

Lawrence Charles LaPointe

Acknowledgements

Firstly, I would like to thank my supervisors Prof. Graeme Young of Flinders University, Dr. Rob Dunne of CSIRO Mathematical and Information Sciences and Dr. Peter Molloy of CSIRO Molecular Health Technologies. I am indebted to Peter Molloy for reminding me that good science requires precision and careful consideration and that patience is often rewarded. I am grateful to Rob Dunne for teaching me skills that I will use for the rest of my career and for his excellent instruction of complex subject matter. I express my greatest thanks to Graeme Young, without whose guidance I would not have been able to start, conduct, or complete this research.

Collectively, my supervisors' guidance, scientific instruction, and ability to provide insightful criticism made this work possible.

I would like also to thank Clinical Genomics Pty Ltd and Enterix Inc for support of this research, including providing me ample time to dedicate to this study. In particular, I thank Howard Chandler, Max Mawhinney, and Peter Horrobin who have shared my vision that good science makes good business. With their support, I have been able to invest considerable time and energy into this research.

I thank my wife and family for love and support. I especially thank Karen for enduring my absence, inattention, and stress through these years without a single word of objection. Thank you for helping me to make this investment.

Finally, I express my deepest gratitude to the nameless patients and volunteers whose generous gift of clinical specimens forms the cornerstone of this research. To these individuals: your decision to contribute to the benefit of others even while you are confronted by the tragedy of colorectal cancer is inspirational. This thesis is aimed at discovering biomarkers which I hope will help others avoid your pain and I dedicate this work to you.