GERMLINE PHARMACOGENOMICS TESTING IN FORMALIN-FIXED PARAFFIN-EMBEDDED TUMOURS

by

Shyong Quin Yap

B.Sc. (Hons), Trent University, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Experimental Medicine Program)

The University of British Columbia
(Vancouver)

January 2017

Abstract

This document provides brief instructions for using the `ubcdiss` class to write a UBC-conformant dissertation in LaTeX. This document is itself written using the `ubcdiss` class and is intended to serve as an example of writing a dissertation in LaTeX. This document has embedded Unique Resource Locators (URLS) and is intended to be viewed using a computer-based Portable Document Format (PDF) reader.

Note: Abstracts should generally try to avoid using acronyms.

Note: at University of British Columbia (UBC), both the Graduate and Postdoctoral Studies (GPS) Ph.D. defence programme and the Library's online submission system restricts abstracts to 350 words.

Preface

At UBC, a preface may be required. Be sure to check the GPS guidelines as they may have specific content to be included.

Table of Contents

List of Tables

List of Figures

Glossary

This glossary uses the handy `acroynym` package to automatically maintain the glossary. It uses the package's `printonlyused` option to include only those acronyms explicitly referenced in the LaTeX source.

GPS Graduate and Postdoctoral Studies

PDF Portable Document Format

URL Unique Resource Locator, used to describe a means for obtaining some resource on the world wide web

Acknowledgments

Although this thesis only bears one name, its completion would be impossible without the contribution of many individuals. First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Aly Karsan, for the opportunity to work with his team of diverse talents as well as his patience, guidance, and extensive knowledge in clinical informatics.

I would also like to thank my labmates from the bioinformatics team, Kieran and Rod, and members from the Centre of Clinical Genomics, Liz and Jill, for their insightful comments and help throughout my time in the lab. I would like to extend my gratitude to my supervisory committee members, Dr. Ryan Morin and Dr. Martin Hirst, for their knowledgeable feedback and continuous effort in asking me difficult questions which motivated me to widen my research perspective.

My sincere thanks also goes to my friends, who supported me and lifted my spirits through the tough times. Last but not least, I would like to thank my family: mom and dad, for always encouraging my interest in science and listening to my endless science talks and my sisters, for believing in my ability even when I doubt myself. This thesis is yours as much as it is mine.

Chapter 1

Introduction

1.1 Objectives

Genetic variants in genes encoding drug metabolizing enzymes, drug transporters, and drug targets can cause interpatient differences in drug response. Pharmacogenomics (PGx) applies genomic approaches to evaluate the association of genetic variants with drug efficacy and toxicity. If a PGx variant demonstrates clinical utility, screening for this biomarker would guide treatment selection as well as optimization of treatment dosage and duration, thereby improving therapeutic effectiveness and safety. In the context of cancer, both somatic mutations in the tumour genome and germline genetic variants can influence a patient's treatment choice. While somatic mutations typically affect activity or expression of drug targets, germline variants present in normal tissues affect drug disposition which includes absorption, distribution, metabolism, and excretion of the administered drug. Although somatic mutations have been implicated in cancer PGx, this thesis will focus on germline PGx variants with significant impacts on treatment response to cytotoxic chemotherapy.

Clinical use of genomic information has been rapidly adopted by oncology practice to arrive at more informed decision with regards to patient management and therapeutic intervention. This emerging framework of genomics-driven cancer medicine has been revolutionized by advances in next-generation sequencing (NGS) technologies, declined cost in genome sequencing, and development of bioinformatics analytic tools. There are various genomic approaches that interrogate different proportions of the genome. At present, targeted NGS panels that screen for genomic alterations in a subset of genes are the most practical in the clinic based on clinical actionability, cost-efficiency, and short turnaround time. Other comprehensive approaches such as whole exome sequencing (WES) which survey only coding regions in the genome and whole genome sequencing (WGS) are also translated for clinical use. However, limitations to these comprehensive approaches exist due to incomplete annotation of the human genome and challenges in interpreting genomic data for medical decision-making.

Despite the many promising applications of NGS-based testing in cancer medicine, several challenges are yet to be resolved. One of the challenges is sequence artifacts caused by formalin fixation. For decades, formalin fixation has been the standard procedure to preserve morphology and cellular characteristics of tissue biopsies for histologic examination. Compared to storing fresh-frozen tissues, long-term storage of formalin-fixed paraffin-embedded (FFPE) blocks at room temperature is more cost-effective and therefore, frequently used for molecular analysis. However, formalin fixation causes DNA fragmentation and base transition artifacts (i.e. C>T and G>A), which could result in false-positive or false-negative variations in the genome. This is particularly concerning in the clinic because this genomic information is used to inform therapeutic intervention of a patient. Secondly, the lack of matched normal DNA is a challenge faced by present-day clinical NGS-based testing. Processing of matched normal DNA is not a routine practice in the clinical setting due to logistical difficulties in obtaining a blood or saliva sample, increased cost, and an underappreciation of the potential value of the matched normal. Without matched normal DNA, detection of clinically relevant germline variations (e.g. germline PGx variants) from somatic mutations in the tumour genome rely heavily on filtering with existing variant databases such as dbSNP and ExAC. Additional challenges in clinical implementation of NGS-based testing include lack of methods to interrogate low-complexity regions of the genome, a consensus standard for variant calling and annotation, and methods for functional prediction of novel genomic variants. Nevertheless, the scope of this thesis will only address the first two challenges described through analyzing data from a targeted NGS panel for solid tumours.

The primary objective of my research is to determine whether a clinical targeted NGS panel that only sequence FFPE tumour DNA can be harnessed to report high quality germline PGx variants that would inform chemotherapeutic intervention. To address this objective, I evaluated the concordance of germline PGx variants between FFPE tumour DNA and the gold standard peripheral blood DNA. In addition, I assessed the extent of formalin induced DNA damage and modification by comparing quality metrics of sequencing data and formalin artifacts between FFPE tumour and peripheral blood.

This introductory chapter will be organized into four sections to provide sufficient background information for the work presented in this thesis. The first section will introduce the adoption of PGx findings in oncology care by providing examples of promising applications with an emphasis on germline PGx variants that affect response to cytotoxic chemotherapy. The second chapter will focus on the evolution of NGS technologies and its contribution to the emergence of genomics-driven cancer medicine. Genomic approaches that interrogate different content of the genome will also be described and examples of their clinical use will be provided. In the third chapter, I will provide an overview on a variant calling pipeline and introduce different types of genomic alterations. I will also describe different variant calling algorithm and methods for variant annotation and interpretation. Finally, the fourth section will focus on challenges in the clinical application of

NGS-based testing, expanding on formalin fixation artifacts and the lack of matched normal samples. These are problems addressed in this thesis which aims to investigate whether germline PGx variants can be detected with precision in a clinical targeted NGS panel that only sequence DNA from FFPE tumours.

1.2 Pharmacogenomics in Clinical Oncology

Cancer pharmacogenomics takes into account tumour-associated somatic mutations and germline variants. Somatic mutations in driver gen

Pharmacogenomics (PGx) applies genomic approaches to evaluate the interaction of genetic variants with drug response. These variations affect genes encoding drug targets as well as drug disposition proteins involved in drug absorption, distribution, metabolism, and excretion (ADME). The goals of PGx studies are to elucidate biological mechanisms underlying interpatient variability in drug efficacy and toxicity as well as identify PGx biomarkers with clinical utility, which would guide selection of treatment type, optimal dosage, and duration.

Cancer PGx takes into account tumour-associated somatic mutations and germline variants. Somatic mutations in driver genes promote malignant transformation through conferring selective growth advantage to the cells. Characterization of somatic driver mutations has provided an avenue for development of molecularly targeted drugs against specific tumour-defining somatic mutations. Hence, screening for these specific

somatic mutations serving as genomic predictors of tumour response and providing new leads for drug development germline variants optimize cancer drug dosing and predict the susceptibility of patients to the adverse side effects of these drugs - knowledge that can be used to improve benefit:risk ratio of cancer treatment for individual patients

1.2.1 Somatic Mutations and Targeted Therapies

1.2.2 Germline Variants and Chemotherapy-Associated Morbidity

1.3 Applications of Massively Parallel Sequencing

1.3.1 The Evolution of Massively Parallel Sequencing

1.3.2 Targeted Sequencing

1.3.3 Whole Exome Sequencing

1.3.4 Whole Genome Sequencing

1.3.5 Summary

1.4 Bioinformatics Tools for Variant Calling

1.4.1 Types of Genomic Alterations

1.4.2 Variant Calling Algorithms

1.4.3 Variant Annotation and Interpretation

1.5 Challenges in Clinical Next-Generation Sequencing Testing

1.5.1 Sequence Artifacts by Formalin Fixation

- Transition vs. transversion

1.5.2 Lack of Matched Normal DNA

1.6 Summary

Chapter 2

Germline Pharmacogenomics Testing in Formalin-Fixed Paraffin-Embedded Tumours

2.1 Overview

Genomics-driven cancer medicine is driven by the advances in massively parallel sequencing (MPS) technologies, the reduced cost of genome sequencing, and development of bioinformatics analytic tools. This emerging framework in the oncology care aims to use genomics information to inform

The Oncopanel is a clinical targeted sequencing panel for solid tumours provided by the CCG at the BCCA. In addition to somatic mutations, it screens for germline variants in PGx genes such as DPYD, GSTP1, MTHFR, TYMP, TYMS, and UGT1A1 (Table 1). Detection of germline PGx variants is essential for chemotherapy selection and optimization of treatment dosage and duration. The Oncopanel is also delivered as a single sample clinical assay in which genetic variants are detected in DNA from FFPE tumours. However, formalin fixation causes DNA fragmentation and base transition artifacts (i.e. C>T and G>A). Hence, I investigated whether germline PGx variants could be detected with high sensitivity and precision in FFPE tumour DNA compared to blood DNA which is the gold standard for germline variant calling.

2.2 Methods

2.3 Results and Discussion

Bibliography

Appendix A

Supporting Materials

This would be any supporting material not central to the dissertation. For example:

- additional details of methodology and/or data;
- diagrams of specialized equipment developed.;
- copies of questionnaires and survey instruments.