TARGETED NEXT-GENERATION SEQUENCING FOR CLINICAL DIAGNOSTICS OF SOLID 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 inter-individual 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-drive cancer medicine has been driven 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 collection 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 coding regions in the genomie and whole genome sequencing (WGS) are also translated for clinical use. However, limitations to these comprehensive approaches exist due to due to the incomplete annotation of the human genome and difficulty in interpreting genomic data for medical decision-making.

Although there are many promising applications of clinical NGS testing in cancer medicine, several challenges are yet to be resolved. One of the challenges is the sequencing of DNA from formalin-fixed paraffin-embedded (FFPE) tumours. Additional challenges include xx, but this thesis will be addressing formalin artifacts in DNA and disadvantages in tumour profiling without matched normal DNA.

This emerging framework of genomics-driven cancer medicine involves screening for genomic alterations that have diagnostic, prognostic, predictive or pharmacogenomic (PGx) clinical utility.

Diagnostic biomarkers facilitate tumour type or subtype classification whereas prognostic biomarkers provide information on disease outcome based on overall and disease-free survival rates. Moreover, predictive biomarkers would inform treatment selection while PGx biomarkers assist with reducing drug toxicity through optimization of treatment dosage and duration.

There are various genome sequencing strategies used in the clinic to screen and facilitate the discovery of clinical biomarkers. These strategies differ in the proportion of the genome they interrogate. There are various strategies that differ in the proportion of the genome

patient care, treatment outcome, treatment response, risk stratification, susceptibility, toxicity, adverse drug events Key words: pharmacogenomic, germline variants, targeted amplicon-based MPS panel, formalin artifacts, tumour-only genomic profiling

1.2 Pharmacogenomics in Clinical Oncology

1.3 The Evolution of Next-Generation Sequencing

1.4 Bioinformatics Tools for Variant Calling

1.4.1 Types of Genomic Alterations

1.4.2 Variant Calling Pipelines

1.4.3 Variant Calling Algorithms

1.5 Challenges in Clinical Next-Generation Sequencing Testing

1.5.1 Formalin Artifacts in DNA

1.5.2 Tumour-only Profiling

1.6 Summary

Chapter 2

Variant Calling in Formalin-fixed Paraffin-embedded Tumours

2.1 Overview

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 fi

xation 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.

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.