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

Current research in cancer genomics primarily focus on somatically acquired mutations that drive malignant transformation through conferring selective growth advantages to cells. These efforts are demonstrated by formation of large-scale collaborations such as the The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC), which aim to characterize and catalog the genomic landscapes of diverse tumour types. Understanding oncogenic mechanisms underlying driver somatic mutations have led to the development of targeted therapies, which resulted in improved clinical outcomes for various cancer subtypes. However, germline genetic variants can also influence cancer treatment by affecting drug targets and disposition, thereby causing interpatient differences in drug response. These germline variants, known as pharmacogenomic (PGx) variants, can assist with treatment selection, optimal drug dosing, and identifying toxicity risk to reduce cancer therapeutics-associated morbidities.

Advances in massively parallel sequencing (MPS) technologies have revolutionized genetic testing in clinical oncology through enabling surveillance of increased genomic depth and breadth with less DNA in a cost-effective and timely manner. Nevertheless, clinical application of MPS approaches to cancer medicine still encounter several challenges and financial barriers. One of these challenges is caused by formalin fixation of tumour biopsies. Tumour biopsies are routinely formalin-fixed paraffin-embedded (FFPE) to preserve morphology and cellular characteristics for histologic examination. Moreover, most clinical laboratories prefer storage of FFPE blocks at ambient temperature to avoid cost inflicted by maintaining fresh-frozen specimens. However, formalin fixation causes DNA fragmentation and base transition artifacts, which could result in false-negative or false-positive variant calls. These sequence artifacts are particularly concerning in a clinical setting because failure to detect or inaccurate detection of cancer biomarkers could have devastating consequences for patients and their families.

Another challenge in clinical MPS-based testing in oncology practice is the lack of matched normal DNA, which is not commonly collected in the clinic due to increased cost and logistical difficulties. Without matched normal DNA, determining the somatic or germline nature of the variant calls, which is essential for translating MPS data into clinically actionable information, rely heavily on filtering and interpretation using databases such as dbSNP, ExAC, and COSMIC. The bottleneck of MPS data generation to interpretation for clinical use is yet another limiting factor of clinical genomic sequencing. Despite the ability of MPS approaches to screen increased genomic content, these methods lead to higher rates of detecting variants of uncertain significance (VUS) that lack evidence of clinical utility. Conversely, incidental findings with medical value to patient care may arise while there are ethical controversies and very few guidelines on the management of this category of variants.

The main objective of this thesis is to investigate whether germline PGx variants can be accurately and sensitively detected in FFPE tumour DNA sequenced by a clinical targeted MPS panel. To achieve this objective, key challenges in clinical genomic sequencing that were briefly described above were addressed. This introductory chapter is organized into five sections to provide the necessary background knowledge: (1) Describes driving forces that led to emergence of genomics-driven oncology; (2) Introduces different applications of MPS to provide an overview of technologies behind sequence data generation; (3) Introduces bioinformatics pipeline for variant calling, which generated input data analyzed in this thesis; (4) Expands on key challenges in clinical genomic sequencing and potential solutions; (5) Emphasizes on the importance of germline PGx testing in oncology care.

1.2 Genomics-Driven Oncology

1.2.1 Definitions

Genomics-driven oncology is defined as the use of genomic information to provide guidance for disease management and therapeutic intervention in oncologic care. The application of NGS to oncology, or genomics-driven cancer medicine, is conceptually logical and simple: First, the genome of a patients tumor is sequenced, and all genetic differences from the standard human reference genome are identified. Because all human beings have many normal genetic variants that differ from the reference genome, the tumor sequence is compared with the patients constitutional (germline) genome to determine which alterations in the tumor are somatic (and therefore potentially pathogenic) and which are germline (and probably not cancer-related). Next, the somatic mutation list is filtered through a database of mutations that may render tumors sensitive to established and emerging anticancer drugs. Finally, an annotated list is provided to the treating physician to be used in clinical decision making and clinical research design. However, several technical and ethical challenges must be addressed before real-time application of NGS can become a reality in cancer medicine.

1.2.2 Genomic Alterations in Cancer Pathogenesis

1.2.3 Clinical Deployment of Targeted Cancer Therapies

1.2.4 Advances in DNA Sequencing Technologies

The first human reference genome was established in 2003 through completion of the Human Genome Project, which instigated major developments in DNA sequencing technologies and computational tools for large-scale genomic data analysis. As a result, MPS technologies have emerged with increased throughput, sensitivity, and cost efficiency, leading to the genomic characterization of a growing number of tumours. As the application of MPS technologies in cancer genomic studies continued to accelerate the progress of driver gene discoveries and drug development, the clinically feasible features of MPS have led to its rapid integration in oncology practice, giving rise to the genomics-driven oncology framework.

Genomics-driven oncology is defined as the use of genomic information to provide guidance for disease management and therapeutic intervention in oncologic care. One of the driving forces of this emerging approach is the expanding knowledge in tumour biology. A central focus of tumour biology research is elucidating oncogenic mechanisms driven by somatic mutations that confer selective growth advantages to cells. Translation of these findings into targeted therapies have demonstrated pronounced improvement in clinical outcomes, leading to the transition from morphology-based to genetic-based management of cancer. A well-known example is the treatment of BCR-ABL-translocated chronic myeloid leukemia (CML) with the tyrosine kinase inhibitor imatinib, which targets the constitutively active ABL1 kinase as a result of the BCR-ABL fusion gene.

Several successful applications of targeted therapy ensued the example of imatinib and BCR-ABL-translocated CML such as the use of anti-HER2 monoclonal antibody trastuzumab in treating HER2/neu-amplified breast cancer and BRAF inhibitor vemurafenib in treating advanced BRAF-mutated melanoma. The promising potential of targeted anti-cancer agents accelerated the progress of drug discovery and development as evident by the drastically decreasing timelines between driver mutation discovery and clinical proof-of-concept. For instance, it only took three years for the ALK inhibitor crizotinib to enter Phase II clinical trials since identification of ALK translocations in non-small cell lung carcinoma (NSCLC) whereas the Food and Drug Administration (FDA) approval of imatinib for treatment of BCR-ABL-translocated CML took 41 years since discovery of the Philadelphia chromosome. Consequently, there is an extensive compendium of targeted therapeutics with 19 listed as clinically approved by the National Cancer Institute in 2012 while approximately 150 compounds were listed as clinical candidates.

The enhanced understanding of oncogenic pathways and growing spectrum of targeted therapies have created the perfect opportunity for clinical screening of driver mutations to match patients with targeted treatments. Conversely, patients without specific mutations could be spared treatment-

associated toxicities. For example, screening for KRAS mutations in codon 12 or 13 could prevent treating colorectal cancer (CRC) patients with anti-EGFR monoclonal antibody or EGFR inhibitors, which are associated with toxicity risk, as these patients are known to respond poorly. Despite the initial efficacy of targeted treatments, tumours could develop resistant mechanisms causing cancer relapse. One of the crucial realization from proceeding studies is that cancer is a heterogenous disease, in that a tumour can consist of multiple subclonal populations and resistant cancer cells may already exist at an early stage. Hence, treatment resistance occurs after the dominant clone has been wiped out allowing the resistant subclone to proliferate and metastasize. Complexities derived from the vast mutational profiles of tumours and intratumoural heterogeneity revealed that surveillance of multiple cancer genes with increased coverage depth during the course of a disease is essential for positive clinical outcomes. To achieve this, single gene assays using the Sanger method, also known as the dideoxynucleotide chain termination method, are not feasible due to time, labour, and cost constraints.

With the advent of MPS technologies, genome sequencing can be accomplished at a reduced cost of less than $5000 per genome within days. To put this into perspective, the first human reference genome sequenced using the Sanger method was completed at a cost of $2.7 billion over 13 years. Advances in MPS technologies have revolutionized cancer genomics by enabling international consortia such as TCGA and the ICGC to uncover the complex genomic architectures of various tumour types, thereby shedding insights into drug resistant mechanisms and potentiating therapeutic strategies against cancer relapse. The high-throughput nature of MPS and its ability to generate robust genomic information in a timely and cost-efficient manner are also capable of overcoming the limitations of single gene assays in the clinic. Thus, various MPS approaches, most commonly targeted gene panels, have been rapidly adopted in clinical oncology to inform medical decision-making based on a patient's genomic make-up. Although the path to genomics-driven oncology was paved by a deeper mechanistic understanding of oncogenic pathways and the accelerated progress in targeted therapeutics development, the emergence of MPS technologies played a significant role in providing detailed insights into the cancer genome as well as a feasible method to generate genomic information for clinical use.

1.3   Massively Parallel Sequencing Technologies

The Human Genome Project, which assembled the first human reference genome, was completed in 2003 at an expense of $2.7 billion within 13 years. This was a cost and turn-around time that would not be feasible for routine usage in research and clinical settings, thereby stimulating the advancement of DNA sequencing technologies. As a result, various MPS technologies have emerged and have been adapted to fit the needs of scientific research as well as clinical applications.

1.3.1   Sequencing by Ligation

1.3.2   Sequencing by Synthesis

1.4   Applications of Massively Parallel Sequencing

1.4.1   Targeted Sequencing

1.4.2   Whole Exome Sequencing

1.4.3   Whole Genome Sequencing

1.5   Bioinformatics Tools for Variant Calling

1.5.1   Types of Genomic Alterations

There are different types of genomic alterations.

1.5.2   Variant Calling Pipeline

1.5.3   Variant Calling Algorithms

1.5.4   Variant Curation and Interpretation

1.6   Challenges in Clinical Genomics

1.6.1   DNA Damage by Formalin Fixation

- Fragmentation - Transition vs. transversion

1.6.2 Lack of Matched Normal DNA

1.6.3 Variant of Unknown Significance

1.6.4 Incidental Findings

1.7 Pharmacogenomics in Clinical Oncology

Cancer biomarkers, which are central to the genomics-driven oncology approach to medical decision-making, can be classified as diagnostic, prognostic, predictive, and pharmacogenomics (PGx). PGx markers are germline genetic variants that affect genes encoding drug targets as well as drug disposition proteins involved in absorption, distribution, metabolism, and excretion (ADME). Pharmacogenomics (PGx) applies genomic approaches to evaluate the interaction of genetic variants with drug response. These variations affect . 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

1.7.1 Targeted Therapies

Tamoxifen etc.

1.7.2 Chemotherapy-Associated Morbidities

DPYD, MTHFR, GSTP1, TYMP, TYMS, UGT1A1

1.8 Summary

The advent of MPS technologies has refined analysis of the cancer genome at base-pair resolution,

Chapter 2

Materials and methods

2.1   Patient Samples

2.2   DNA Extraction

2.3   Library Preparation and Illumina MiSeq Sequencing

2.4   Data Processing

2.5   Data Analysis and Visualization

Chapter 3

Results

Chapter 4

Discussion and conclusion

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.