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 Definition

There are many terms used interchangeably with 'genomics' A central focus of cancer 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. Well-known examples include the use of the tyrosine kinase inhibitor imatinib in treating BCR-ABL-translocated chronic myeloid leukemia (CML), anti-HER2 monoclonal antibody trastuzumab in treating HER2/neu-amplified breast cancer, and BRAF inhibitor vemurafenib in treating advanced BRAF-mutated melanoma. Thus, screening for selective driver genes has became standard of care for several cancer types to match patients with specific targeted anti-cancer agents.

Despite the successful applications of targeted therapies, remission was short-lived in a subset of patients due to acquired resistance. response variability within the same tumour types were

associated with increased treatment efficacy, there were reports on treatment response variability and acquired resistance. These were evident that cancer is a complex disease that and heterogenous.

For approximately 25 years,

Due to the complexity of cancer

Cancer initiation and progression are driven by accumulation of genomic and epigenomic alterations that confer selective growth advantages to cells. The advent of MPS technologies has revolutionized cancer genomics research by enabling collective efforts such as TCGA and the ICGC to characterize the genomic architectures of a growing number of tumours. A central recognition from these studies is that cancer is a complex and heterogenous disease, with mutational burdens that vary between tumour types, from 0.28 to 8.15 mutations per megabase in acute myeloid leukemia (AML) and lung squamous carcinoma, respectively. Most cancers typically accumulate an array of mutated genes that interact over time to initiate neoplasia and fuel its progression.

more cancer genes, more than one driver gene

The genetic basis of cancer

Using MPS technologies, collective efforts like TCGA and the ICGC have achieved extraordinary progress in uncovering the genomic architectures of a growing number of tumours, thereby accelerating driver mutation discovery.

In parallel to the rapid characterizing of tumour genomes is the development of a vast spectrum of targeted therapies that

Genomics-driven oncology aims to harness information from host and tumour genomes for patient management and therapeutic intervention. This emerging approach in cancer medicine has demonstrated promising potential in enhancing patient care and is driven by a mature understanding of tumour biology, an expanding spectrum of targeted therapies, and advances in MPS technologies.

The unique features of MPS including its high-throughput nature and ability to provide base-pair resolution of genomic content at reduced cost and turn-around time have led to its rapid adoption in clinical oncology, giving rise to the genomics-driven oncology framework.

MPS platforms can be used to screen for Prior to the genomics era, For approximately 25 years since the late 20th century, dideoxynucleotide However, MPS technologies have boosted the discovery of cancer genes Central to the transition from Sanger is the crucial recognition of cancer as a complex and heterogenous disease that typically involves mutations in a spectrum

The application of genomic information to guide management of cancer holds promising potential for enhancing patient care and improving clinical benefit. While a deeper understanding of tumour biology and vast spectrum clinical feasibility of genomics-driven oncology is driven by

Clinical application of genomic information to guide the can and The advent of MPS technologies has revolutionized cancer genome studies Cancer genome studies have uncovered the genomic architectures of a growing number of tumour types, thereby accelerating driver mutation discovery. These efforts were propelled by the high-throughput nature and declined cost of MPS technologies,

which were also leveraged and adopted by clinical oncology for the management of cancer. The application of MPS to oncology care gave rise These features of MPS technologies were also leveraged and adapted to fit the needs in oncology practice, leading to the transition from morphology-based management of cancers to genomics-driven oncology have propelled these efforts and have been entering clinical practice duthe advantageous in characterizing the complexity of tumour genomes at a global and refined level. The high-throughput nature and ability to identify genomic alterations at base-pair resolution in a cost-effective and timely manner have been leveraged and adopted

Central to the application of genomic information to management of cancer is the deeper understanding of cancer as a complex genomic disease.

Cancer is among the leading causes of mortality, accounting for one in eight deaths globally. been propelled by the extraordinary progress in MPS technologies have propelled cancer genome studies by uncovering the complexity of cancer genomes global and refined, uncovering, complexity Understanding the complexity of the cancer genome was propelled by advances in MPS technologies leading to rapid progress in drive mutation discovery.

Cancer genomics Traditional approaches that categorize cancers based on anatomic origins have shifted towards molecular classification. This transition from a morphology-based to genetic-based disease paradigm of cancer was driven by due to a deeper understanding of the complexity . Additionally, single gene testing has been supplanted by genomic technologies that enabled screening of more target genes, the entire exome, or genome. The application of genomic information to guide patient management and therapeutic intervention holds promising potential for improving patient care in clinical oncology. Central to the emergence of genomics-driven cancer medicine is the deeper understanding of cancer as a complex genomic disease, which was increasingly elucidated by efforts of the TCGA and ICGC

Accelerated drug development, shorter timelines for translation from somatic mutation discovery to clinical actionability

HGP in 2003, The advent of MPS technologies and steep declined in cost paved the opportunity for clinical MPS-based testing for cancer

Clinical use of genomic information to guide patient management and therapeutic intervention has been rapidly adopted in oncology care. Various breakthroughs have created an unparalleled opportunity to test the hypothesis that systematic knowledge of genomic information from individual tumours can improve clinical outcomes for many patients with cancer. These driving factors of the emerging genomics-driven cancer medicine framework include extraordinary progress in tumour biology research, the expanding compendium of anti-cancer targeted drugs, and revolutionary advances in massively parallel sequencing (MPS) technologies.

Somatically acquired genomic alterations play a significant role in driving malignant transformation through conferring selective growth advantages to cells. Through unveiling oncogenic mechanisms underlying driver mutations, anti-cancer drugs can be customized against tumour-defining

mutations. A well-known example is the treatment of BCR-ABL-translocated chronic myeloid leukemia (CML) with the targeted agent imatinib. The aberrant chromosomal translocation involving the BCR and ABL genes in CML was first discovered through karyotyping in 1960 with ensuing efforts in elucidating downstream mechanisms of the BCR-ABL fusion protein. After 40 years of investigative research followed by drug development and clinical trials, the tyrosine kinase inhibitor imatinib was approved by the Food and Drug Administration (FDA) as a targeted therapy against the constitutively active ABL1 kinase in BCR-ABL-translocated CML. The drastic improvement in clinical outcome of BCR-ABL-translocated CML demonstrated that genetic profiling of tumours holds great potential in enhancing patient care through molecularly targeted therapies.

The notable success of imatinib in BCR-ABL-translocated CML also revealed that basic research in tumour biology must go hand-in-hand with drug discovery and development. Although necessary, functional understanding of tumorigenic mechanisms governed by somatic genomic alterations is not sufficient for clinical practicality of genomics-driven oncology. The accelerated progress in the pharmaceutical development pipeline was demonstrated by the shorter duration required for translation of driver mutation discovery to clinical proof-of-concept. For instance, it only took three years for the ALK inhibitor crizotinib to enter Phase II clinical trials since the discovery of the chromosomal rearrangement of ALK in non-small cell lung carcinoma (NSCLC) compared to the 41-year-timeline of imatinib in BCR-ABL-translocated CML. In addition, the spectrum of targeted therapeutics is expanding with the National Cancer Institute listing 19 targeted therapeutics that have entered clinical practice in 2012 and nearly 150 compounds listed as clinical candidates in a public database. Aside from imatinib, examples of mechanistic findings of driver mutations that resulted in clinical deployment of targeted therapies include the monoclonal antibody trastuzumab in HER2/neu positive breast cancer, PARP inhibitors in BRCA1/2 breast cancer, and BRAF inhibitors in melanoma (Table 1). Hence, emergence of the genomics-driven oncology approach is feasible because screening for clinical biomarkers would facilitate treatment choice and enrollment to clinical trials based on predictive tumour genetic biomarkers. A successful example is treatment of CKIT-positive gastrointestinal stromal tumours (GIST) with imatinib which was previously FDA-approved for CML and also inhibit the activity of CKIT.

Despite the promising results of imatinib in early 2000s, several key realizations were made in the proceeding decade including that human cancers are The notable success of imatinib in treating BCR-ABL-translocated CML was ensued by many promising examples wherein elucidating oncogenic driver pathways gave rise to clinical deployment of molecular targeted therapies (Table 1). This expanding compendium of targeted anti-cancer therapies has enabled the translation of cancer genomics to Paired with the deeper understanding of For instance, have ensued since the notable success of imatinib in treating BCR-ABL-translocated CML. In addition to the promising examples wherein elucidating oncogenic driver pathways gave rise to clinical deployment of molecular targeted therapies (Table 1), advances in massively parallel sequencing (MPS) technologies have

enabled genome sequencing to be clinically feasible. For approximately 25 years since the late twentieth century, Sanger sequencing, also known as dideoxynucleotide chain termination sequencing, was the most widespread method for DNA sequencing. However, the many advantages of MPS compared to Sanger sequencing was the Sanger method, also known as the dideoxynucleotide chain termination method. At present, MPS technology through reduced cost, time, labour, and DNA input compared to traditional Sanger sequencing.

Aside from screening for known

- not one gene is responsible for tumour progression - CML is an outlier - drug resistance - resistance to imatinib - different type of biomarkers

In the era of genomics, the molecular classification of cancer continues to expand and in turn has the potential to facilitate the development of novel biomarkers.

Although the development of imatinib as a therapeutic for a genomic alteration heralded the era of molecularly targeted agents, several lessons would emerge over the ensuing decade. First, unlikeCML,most cancers are not homogeneously propelled by a single genomic driver alteration; instead, they comprise rare disease subsets with a variety of genomic alterations. Second, single- agent therapies against a single genomic target have not been as successful in achieving cures or long-term survival as was imatinib in CML. Thus, CMLhas been the exception and not the rule, which highlights the importance of developing rational combination therapies and elucidating mechanisms of drug resistance to single agents.

BCR-ABL fusion protein as a result of chromosomal translocation in chronic myeloid leukemia (CML) patients of elucidation of the discovery of constitutively activated ABL kinase the treatment with the ABL1 kinase inhibitor imatinib, which the drastic improvement in disease outcome for BCR-ABL-translocated chronic myeloid leukemia patients notable success in the promising treatment of Traditional approaches have focused on classifying cancers based on anatomic sites and histopathologic features. However, the notable success in targeted therapy treating BCR-ABL-translocated chronic myeloid leukemia (CML) with the ABL1 kinase inhibitor imatinib has led to the transition from morphology-based to molecular classification of cancers.

With the completion of the Human Genome Project in 2003 and advances in massively parallel sequencing technologies, collaborative efforts such as The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC) are working towards characterizing the genomic landscapes of diverse cancer types.

Developing a personalized therapy strategy to ensure an optimal outcome for individual cancer patients is possible given the dramatic progress in basic cancer research at the molecular and cellular levels, the rapid advancement of new technologies that enable fast and cost-effective comprehensive characterizations of tumors at the molecular level, and an expanding compendium of targeted cancer therapeutics.

Convergence in biological discoveries, MPS technologies, and drug development

Advances in massively parallel sequencing technologies,

- Cancer classification has evolved from tissue of origin to molecular classification. - CML and imatinib - success led to emphasis on molecular classification - Genomics era - expansion of characterizing genomic landscape and cataloguing driver mutations - Different types of biomarkers - Standard of care genetic screening - breast/HER2, NSCLC/EGFR/KRAS/BRAF, Melanoma/BRAF, GIST/KRAS - Facilitate discovery, enrollment to clinical trials, guide precision medicine

1.3 Massively Parallel Sequencing

1.3.1 Massively Parallel Sequencing Technologies

1.3.2 Applications of Massively Parallel Sequencing

Targeted Sequencing

1.3.3 Whole Exome Sequencing

1.3.4 Whole Genome Sequencing

1.4 Bioinformatics Tools for Variant Calling

1.4.1 Types of Genomic Alterations

There are different types of genomic alterations.

1.4.2 Variant Calling Pipeline

1.4.3 Variant Calling Algorithms

1.4.4 Variant Curation and Interpretation

1.5 Challenges in Clinical Genomics

1.5.1 DNA Damage by Formalin Fixation

- Fragmentation - Transition vs. transversion

1.5.2 Lack of Matched Normal DNA

1.5.3 Variant of Unknown Significance

1.5.4 Incidental Findings

1.6 Pharmacogenomics in Clinical Oncology

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

1.6.1 Targeted Therapies

Tamoxifen etc.

1.6.2 Chemotherapy-Associated Morbidities

DPYD, MTHFR, GSTP1, TYMP, TYMS, UGT1A1

1.7 Summary

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

Chapter 2

Germline Pharmacogenomics Testing in Formalin-Fixed Paraffin-Embedded Tumours

2.1 Overview

Application of genome information to guide patient management and therapeutic intervention holds great potential in improving oncology care. One of the driving forces that led to clinical feasibility of genomic sequencing is the advent of massively parallel sequencing (MPS) technologies, which enabled sensitive and accurate sequencing of more target genes with less DNA in a cost-effective and timely manner. At present, various MPS approaches are entering, or have entered the clinic such as targeted sequencing panels, whole exome sequencing, and whole genome sequencing, which create the opportunity to further develop novel clinical biomarkers in addition to screening for biomarkers with established clinical utility.

In the context of cancer, Clinical biomarkers can be classified as diagnostic, prognostic, predictive, and pharmacogenomic (PGx). In the context of cancer, both somatic Germline variants that affect influence drug response are PGx biomarkers.

In the context of cancer, PGx biomarkers are

Clinical molecular laboratories are rapidly adopting and leveraging the advances in massively parallel sequencing (MPS) technologies for germline and tumour profiling. have driven the clinical use of genomic information to guide patient management and therapeutic intervention in oncology care. The ability of MPS to sensitively and accurately sequence more target genes with less DNA in a cost-effective and timely manner perfectly meet the clinical reality which is to do more with less.

reduced cost, time and laborhigh throughput nature, decreased sequencing cost, and increased sensitivity, and ability to Genomics-driven cancer medicine aims is driven by the advances in mas-

sively 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

Chapter 3

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.