Using Next Generation Sequencing to Detect Clinically Relevant Oncogene Mutations in Lung Cancer

by

Minqi Xu

A Practicum submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirement of the degree of

MASTER OF SCIENCE

Department of Pathology
University of Manitoba
Winnipeg

ABSTRACT

Introduction: Modern care of patients with lung cancer requires rapid and accurate diagnosis leading to personalized therapies for individual patients based on molecular characteristics of their tumour. Detecting mutations that predict response to drug quickly and accurately is an essential part of this process. Next generation sequencing (NGS) technologies provide an alternative approach for detecting mutated oncogenes in cancer. We hypothesize that NGS is equal if not superior to standard methods for identifying targetable mutations in the *EGFR* and *ALK* genes in lung cancer.

Methods: DNA and RNA from 38 formalin fixed paraffin embedded lung cancer samples (37 non-small cell lung cancer (NSCLC) and one small cell lung cancer (SCLC)) archived in Diagnostic Services Manitoba were collected and analyzed using gene enrichment methods from Archer Diagnostics followed by sequencing on the Illumina MiSeq NGS machine. Targeted DNA sequencing to detect the *EGFR* mutation was performed on 19 samples while targeted RNA sequencing was applied to 20 samples to identify the *ALK* gene rearrangement. The NGS results were compared with and confirmed by current clinical standard molecular tests for *EGFR* (real-time PCR) and *ALK* (immunohistochemistry and FISH).

Results: Three cases were positive for the *EGFR* mutation and two other samples harbored the *EML4-ALK* fusion genes as determined by NGS. The concordance between NGS and real-time PCR for *EGFR* mutation detection was 88.9%. Additionally, the NGS methodology also

provided profiles of other genes commonly mutated in NSCLC including *KRAS* and *TP53*. The consistency for *ALK* fusion testing was 100% between NGS and FISH.

Conclusion: This study provides support that NGS is a promising diagnostic tool for mutation detection in NSCLC and holds strong potential for an alternative approach to identifying clinically relevant targets such as *EGFR* and *ALK*. Furthermore, NGS provides more information on cancer driven gene mutations than other traditional methods.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and cordial appreciation to Dr. Gefei Qing, Dr. Shantanu Banerji, and Dr. Robert Wightman not only for their outstanding supervision and mentorship throughout the program but also for their great enthusiasm and noble attitude towards scientific research, their expert capacity and profession success in helping patients, and their cordial relations with their colleagues and students. I am very grateful to their priceless contributions to correct and revise this thesis, great help in designing of experiments, stimulating scientific discussions, and most of all the trust placed in my judgment. The experience has been completely rewarding and thoroughly enjoyable.

I would greatly appreciate Dr. Yvonne Myal for participating in my advisory and examining committee. Dr. Myal's support has been invaluable and her many year experience in excellence of graduate education guided me through the program successfully.

Many thanks to Mr. Lance Fuczek for sharing his cutting room, wisdom, and sense of humor with me during my Pathologists' Assistant training.

An extra thank you to Dr. Robert Wightman, Mr. Lance Fuczek, and Dr. Gefei Qing without whom the introduction to the state-of-the-art Pathologists' Assistant program would never have been possible.

I wish to thank Dr. Shantanu Banerji's laboratory (Ms. Cheryl Peltier) at CancerCare Manitoba for essential technical training and support, as well as Pathology/Diagnostic Services Manitoba for specimen contribution. A special thank you and my best wishes to all of the patients whose tissue samples were studied in this research project.

My gratitude also goes out to all of Pathology staff and colleagues for their kind contribution and friendly support throughout my MSc. study. Many thanks are due to the senior and fellow pathologists' assistants at the Health Sciences Centre, St. Boniface Hospital and Grace Hospital with whom two years of PA training and personal interactions have given me a wealth of memories.

I would like to acknowledge Dr. Gabor Fischer, Ms. Michelle Brennan, Ms. Marian Huebner and Ms. Terri Martin for their very helpful administrative and management support.

I am most thankful to my wife, Lucy Ni, who continuously transmits to me her love for life, culture and science. Without her consistent family and spirit support, this work would never have been accomplished. Truly, this work is dedicated to my parents and parents-in-law who have inspired me with their hard work and dedication, as well as to my older brother, my aunt, and the two loved boys of Lucy and mine, Ziming and Zicen, who have all shined my life.

Finally, I would like to express my cordial gratitude to the Diagnostic Services Manitoba Operating Grant, CancerCare Manitoba Foundation, and Dr. Paul H.T. Thorlakson Foundation Grant (University of Manitoba) for their financial support to this research project, as

well as the Faculty of Graduate Studies (University of Manitoba) for supporting and encouraging me in my career pursuit for medical sciences.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A: adenine

AJCC: American Joint Committee on Cancer

ALK: anaplastic lymphoma kinase

AML: acute myelogenous leukemia

BP: base pair

C: cytosine

cDNA: complementary DNA

c-Met: cellular-mesenchymal to epithelial transition factor

DNA: Deoxyribonucleic acid

DSM: Diagnostic Services Manitoba

EGFR: epidermal growth factor receptor

EML4: echinoderm microtubule-associated protein-like 4

ERBB2: erb-b2 receptor tyrosine kinase 2 (human)

ESMO: European Society for Medical Oncology

FFPE: formalin-fixed paraffin-embedded

FISH: fluorescence in situ hybridization

G: guanine

IHC: immunohistochemistry

iPASS: The Iressa Pan-Asia Study

KRAS: V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog

KIF5B: kinesin family 5B gene)

NGS: next generation sequencing

NOS: not otherwise specified

NSCLC: non-small cell lung cancer

PCR: polymerase chain reaction

PIK3CA: phosphatidylinositol-4,5-bisphosphate 3-kinase

RB1: retinoblastoma

RET: rearranged during transfection

RNA: Ribonucleic acid

ROS1: ROS Proto-Oncogene 1

RT: Radiation therapy

SCLC: small cell lung cancer

STK: serine/threonine kinase

T: thymine

TKI: tyrosine kinase inhibitor

TP53: tumor protein p53

TMA: Tissue microarray

1. INTRODUCTION (Part I)

Lung Cancer

1.1. Epidemiology

Lung cancer broadly defines malignant tumours of epithelial origin that develop within the lung parenchyma. Worldwide, lung cancer is the most common cancer among males both in morbidity and mortality (WHO 2014). Among females, lung cancer has the third highest prevalence and is the second leading cause of cancer death following breast cancer (WHO 2014). On average, lung cancer accounts for 19.4% of all cancer deaths globally in 2012, i.e. 1.6 million fatal cases (WHO 2014). In Canada, lung cancer is the most commonly diagnosed cancer along with prostate, breast and colorectal cancers, but is the leading cause of cancer-specific mortality: greater than the next three most lethal cancers combined (**Table 1** and **Table 2**). In 2016, an estimated 26,600 new cases of lung cancer will be diagnosed in Canada, which represents 14% of all new cancer cases. Furthermore, 20,900 Canadian will die from lung cancer and this represents approximate 26% of all cancer mortalities in 2016 (**Table 1** and **Table 2**) (Canadian Cancer Statistics 2016, Alberg *et al.*, 2007). In the United States, the lifetime risk of developing lung cancer is 8% in men and 6% in women (Horn *et al.*, 2015). Recently published statistics data shows that women in Canada have a 1 in 9 chance of dying from lung cancer (Canadian Cancer Statistics 2016). The World Health Organization reported in 2014 that lung cancer is most prevalent in North American, Europe and East Asia with the lowest rates in Africa and South Asia (WHO 2014). Of note, more than one third of new cases in 2012 are in China (Bernard and Christopher, 2014; Zhang *et al.*, 2011). Thus, lung cancer is without doubt the

single most important cause of cancer-related death in industrialized countries and deserves for much medical attention and research focus.

Table 1: Percent distribution of estimated top four of new cancer cases, by sex, Canada, 2016 (Canadian Cancer Statistics 2016)

Males 102,900 New cases		Females 99,500 New cases	
Prostate	21.0%	Breast	25.8%
Colorectal	14.1%	Lung and bronchus	14.1%
Lung and bronchus	14.0%	Colorectal	11.7%
Bladder	6.4%	Body of uterus and uterus NOS	6.6%

Table 2: Percent distribution of estimated top four of cancer deaths, by sex, Canada, 2016 (Canadian Cancer Statistics 2016)

Males (41,700 Deaths)		Females (37,100 Deaths)	
Lung and bronchus	26.1%	Lung and bronchus	26.4%
Colorectal	12.0%	Breast	13.2%
Prostate	9.5%	Colorectal	11.6%
Pancreas	5.8%	Pancreas	6.5%

1.2. Classification

1.2.1. NSCLC and SCLC

Lung cancer is broadly divided into two major histologic categories: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC).

Approximately 85% of new lung cancer cases are NSCLC (Crinò *et al*., 2010). NSCLC itself is a heterogeneous disease with three dominant histologic subtypes based on the microscopic appearance of the cancerous cells: adenocarcinoma, squamous cell, and large cell carcinoma. SCLC accounts for the remainder of new diagnosis. In some cases, there may be a mixed histological pattern with multiple histologic subtypes coexisting within the tumour, for example SCLC in combination with adenocarcinoma. Lifetime exposure to cigarette smoke is considered the greatest risk factor for subsequently developing lung cancer. This observation is particularly true for SCLC and squamous cell carcinomas which are commonly diagnosed in patients who have smoked on average a package of cigarettes daily for 30 years (30 pack-year). By contrast, lung adenocarcinoma is often diagnosed in individuals without a personal history of exposure to cigarette smoke (defined as consuming less than 100 cigarettes in the lifetime). In the last decade, adenocarcinoma arising from the terminal respiratory units of the lung has replaced squamous cell carcinoma arising from the proximal bronchial airways as the most common diagnosed lung cancer in some industrialized countries including Canada. Part of this change is speculated due to a cultural shift towards the consumption of cigarettes with micro-filter causing the deposition of carcinogenic particles in the distant airways (Travis *et al*., 2004). Lung adenocarcinoma is also the most common lung cancer in females (Mitchell *et al.* 2015) and young adults (< 45 year old) (Hsu *et al*., 2012). Besides cigarette smoking, limited evidence

exists to show other environmental risk factors for lung cancer including air pollution, primitive cooking, ionizing radiation and inhalation of asbestos. The exposure to carcinogens found in cigarette smoke and other environmental risk factors causes mutations at cellular level that ultimately leads to lung cancer development.

1.2.2. Clinical subtyping of lung cancer is critical to the management of the disease

SCLC is considered the most aggressive type of lung cancer. These tumours are defined morphologically as having a round to fusiform shapes, scant cytoplasm, finely granular chromatin, and frequent mitotic figures (Kumar *et al*., 9th edition, 2013). At the DNA level, a hallmark of these tumour cells is the almost universal loss of function of the tumour suppressor genes *TP53* and *RB1*, either through mutation or gene deletion (Pfeifer *et al*., 2012). Even when diagnosed at an early clinical stage, all SCLCs are presumed to have metastasized due to the aggressive nature of the disease. Hence, surgery is not typically performed and instead chemotherapy remains the standard of care for all patients. Radiation therapy can be added to some patients where gross disease is confined to a limited region of the chest. Even with optimal therapy, median survival for SCLC is about 6 months (Molina *et al.*, 2008).

By contrast, the treatment of NSCLCs can vary depending of the stage of disease at diagnosis (**Table 3**). Stage I and II disease are considered localized and may be curable by surgical resection alone. Newer therapies for Stage I disease also include stereotactic body radiation therapy which may be used for patients medically ineligible for surgery. Five-year survival for Stage I and II NSCLC with curative intent treatment is approximately 70% and 50%, respectively (Paoletti *et al*., 2011; Howington *et al*., 2013). Stage III disease is usually treated

with a combination of multiple treatment modalities including surgery, radiation, and chemotherapy depending on the clinical scenario. However, even when curative intent treatment is delivered, 5-year survival does not exceed 30% (Wanders *et al.*, 2011). Patients initially diagnosed with Stage I-III disease, may subsequently progress to Stage IV disease even following curative intent therapy. Stage IV or metastatic disease is considered incurable even with aggressive systemic therapy and radiation. Unfortunately 50% of all NSCLC patients present with Stage IV disease at initial diagnosis (Molina *et al.*, 2008). In these patients, all treatment is provided with a palliative intent specifically to reduce disease burden, improve cancer-specific symptoms, improve quality of life, and potentially prolong life. The treatment of Stage IV NSCLC has gone through significant changes since the start of the new millennium, mainly through an improved understanding of the molecular mechanisms underlying NSCLC, and the development of newer therapeutics. The AJCC Cancer Staging Manual including detail of lung cancer staging is available at https://cancerstaging.org/references-tools/deskreferences/Pages/default.aspx.

Table 3: Treatment of NSCLC depending on tumour staging (pre-2017).

NSCLC					
Stage I	**Stage II**	**Stage III**	**Stage IV**		
			Adenocarcinoma (EGFR/ALK mutation positive)	**Adenocarcinoma** (EGFR/ALK mutation negative)	**Squamous/Large cell cancer**
•Surgery •Stereotactic radiation	•Surgery •Chemo	•Chemo •RT •Surgery	•EGFR or ALK TKI •Platinum doublet •Pemetrexed •Taxane •Palliative Care	•Platinum doublet •Pemetrexed •EGFR TKI •Taxane •Palliative Care	•Platinum doublet •Taxane •Clinical Trial •Palliative care

TKI: Tyrosine kinase inhibitor, RT: Radiation therapy

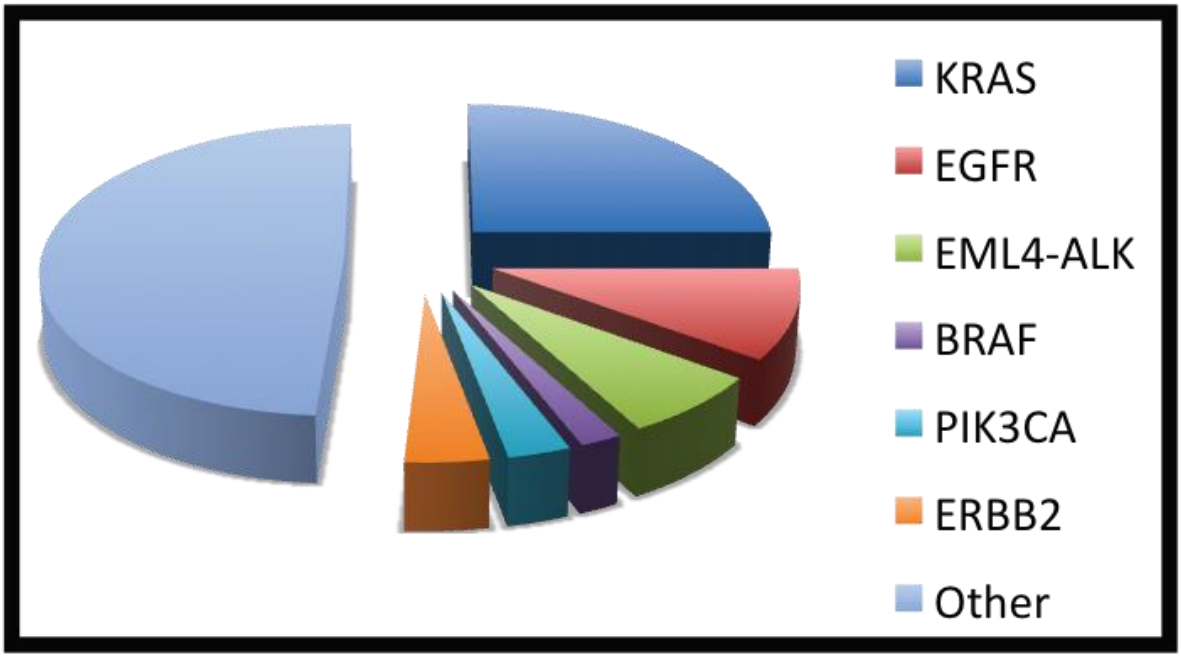
1.3. Somatic Gene Mutations of Lung Cancer

1.3.1 Overview

Cancer arises from gene alterations, or mutations that cause defects in normal cellular function. These mutations are referred to as 'somatic' since they arise in the tumor and are not carried by patients at birth. Genomic studies over the last decade mainly in the most common lung adenocarcinoma subtype have revealed recurrent somatic mutations (**Figure 1**) in the oncogenes *KRAS* and *EGFR*, occurring in 30% and 10% of annual North American cases respectively (Ding *et al*., 2008). *KRAS* mutations are common in smokers while *EGFR* mutations are more common in non-smokers and females. Less common mutations in lung adenocarcinoma include a gene

fusion activating the *ALK* oncogene present in <5% of cases (Soda *et al.*, 2007). Besides *EGFR* mutations, *ALK* fusions are also enriched in non-smokers. *ALK, KRAS* and *EGFR* mutations are typically mutually exclusive and are rarely seen in other NSCLC subtypes. Other potentially targetable mutations in lung adenocarcinoma occur in *BRAF, ERBB2*, and *ROS1*, at frequencies of 1-3% (Imielinski *et al.* 2012). Drugs for each of these targets are at various stages of clinical trials.

Figure 1: Estimated prevalence of common somatic oncogene mutations in lung adenocarcinoma. (Adapted from Kris *et al*. JAMA. 2014. 311(19):1998-2006, courtesy of Dr. Shantanu Banerji)



In 2014, The Cancer Genome Atlas Research Network reported a landmark whole genome analysis of 230 resected lung adenocarcinomas (The Cancer Genome Atlas Research Network, 2014). High rates of somatic mutation (mean 8.9 mutations per megabase) were observed. Of 18 statistically significantly mutated genes noticed, *TP53* (46%) and *KRAS* (33%) are commonly mutated, followed with *STK11* (17%) and *EGFR* (10%). Gene fusions of *ALK* (3 cases) and *ROS1* (4 cases) are also identified. Gender and smoking status correlation with mutation patterns seen in this large scale study is consistent with previous investigations (Ding *et al*. 2008, Weir *et al*., 2007 and Imielinski *et al*., 2012).

1.3.2. EGFR in Lung Cancers

Epidermal Growth Factor Receptor (EGFR) is a transmembrane protein with cytoplasmic kinase activity that transduces important growth factor signaling from the extracellular milieu to the cell (da Cunha Santos *et al*., 2011). Upon activation by its growth factor ligands, EGFR undergoes a transition from an inactive monomeric form to an active homodimer. This configuration change stimulates its intrinsic intracellular tyrosine kinase activity and initiates the downstream signaling cascades, leading to increased proliferation, angiogenesis, metastasis, and decreased apoptosis (Mosesson and Yarden, 2004). Mutations, including increased gene copy number, that cause *EGFR* overexpression or over-activity have been associated with carcinogenesis (Ciardiello and Tortora, 2008; Woodburn, 1999)

Studies show more than 60% of NSCLCs express EGFR. EGFR-activating mutations occur in 47.9% of Asian and 19.2% of Western patients with NSCLC (Dearden *et al*., 2013).

These mutations are usually heterozygous and predominantly associated in *EGFR* exons 18-21, which encodes a portion of the EGFR kinase domain (Soh *et al*., 2009). Approximately 90% of these mutations are short deletions in exon 19 and point mutations of L861Q, G719S and L858R in exon 21 (Ladanyi and Pao 2008, Lynch *et al*., 2004, Paez *et al*., 2004). Mutational data from more than 20,000 NSCLC samples have been reported and are available at the Catalogue of Somatic Mutations in Cancer (http://www.sanger.ac.uk/perl/genetics/CGP/cosmic?action=gene&ln=EGFR) (da Cunha Santos *et al*., 2011). Furthermore, those EGFR-activating mutations are more common in female, never smoker and lung adenocarcinoma. This trend of *EGFR* mutations suggests different pathogenesis underling NSCLC (Sakurada *et al*., 2006).

Almost two-thirds of lung cancers are known to express the EGFR protein (Sherwin *et al*., 1981), this observation led to clinical trials with the oral small molecule inhibitors erlotinib and gefitinib that specifically target EGFR. With both drugs showing evidence of activity, erlotinib was approved for widespread clinical use as second or third line therapy for patients with metastatic lung adenocarcinoma based improved survival compared to placebo (Shephard *et al*., 2005). Response rates were 8.9% however and seemed to be more common in females, non-smokers, and those of East-Asian descent. Subsequent work by several groups revealed the presence of specific point mutations in the *EGFR* gene predicted both sensitivity and resistance to the drugs (Lynch *et al*., 2004; Paez *et al*., 2004; Pao *et al*., 2005 and Mitsudomi *et al*., 2005). Over the last 10 years, there has been considerable clinical trial data supporting the use of anti-EGFR small molecule inhibitors as first line therapy in EGFR-mutant positive patients. The pivotal iPASS trial from East Asia demonstrated a hazard ratio (HR) of 0.48 for progression or

death ($p<0.001$) for mutation positive patients who received gefitinib versus standard carboplatin and paclitaxel chemotherapy (Mok *et al*., 2009). Similarly, the EURTAC trial from Europe comparing erlotinib and platinum based chemotherapy showed a significantly improved progression free survival of 9.4 vs. 5.2 months for erlotinib in *EGFR*-mutant patients (Rosell *et al*., 2012). Overall survival difference again was not significantly different in both studies mainly due to significant crossover in the treatment arms. Both drugs are now approved globally for the first-line treatment of *EGFR*-mutated lung adenocarcinoma. *EGFR* mutation testing is now considered mandatory for all patients with lung adenocarcinoma in Canada (Ellis *et al.*, 2011).

1.3.3. ALK in Lung Cancers

Anaplastic lymphoma kinase (ALK) is a type of tyrosine kinase receptor, which is also a novel oncogene associated with the development of lung adenocarcinoma. In 2007, the Mano laboratory in Japan first reported the presence of oncogenic gene fusions involving the gene *ALK* in lung cancer (Soda *et al*., 2007). *ALK* rearrangements, for example an *EML4-ALK* fusion gene, occur in 2 to 7% of NSCLC. Never smokers with lung adenocarcinoma have a higher rate of *EML4-ALK* fusion mutation (Soda *et al*., 2007; Kwak *et al*., 2010). *EML4-ALK* gene variants are based on distinct breakpoints within *EML4* exons, and the fusion gene encodes a cytoplasmic aberrant protein with constitutive kinase activity (Choi *et al*., 2008; Takeuchi *et al*., 2008 and Kwak *et al*., 2010). Other than NSCLC, the dysregulation of kinase associated with *ALK* somatic mutation has been previously identified in several other types of cancer such as anaplastic large cell lymphoma (Kotok and Aster, 2002), neuroblastoma (George *et al*., 2008; Moose *et al*., 2008) and inflammatory myofibroblastic tumor (Pulford, 2004). In addition, *ALK* fusion genes with

other genes partners beyond *EML4* has been reported in NSCLCs at low frequencies (Takeuchi, 2009).

The presence of *ALK* fusion genes in lung cancers has led to clinical trials using the combined ALK and cMET oral small molecule inhibitor crizotinib (Kwak *et al*., 2010). Trials enrolled only patients with evidence of the *ALK* fusion based on a break-a-part FISH assay. Response rates of over 60% were seen in heavily pretreated patients who typically do not show responses greater than 10% with cytotoxic agents. As reported at the ESMO annual meeting in 2012, evidence of significant progression free survival has now been demonstrated for crizotinib versus standard chemotherapy in the second line setting for lung cancer patients with the *ALK* fusion. The break-a-part FISH assay remains the gold standard and approved assay to confirm the presence of the fusion gene. However, PCR, sequencing, and chromogenic in situ hybridization may also be used (Kim *et al*., 2011).

1.3.4. KRAS in Lung Cancers

V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (*KRAS*) is a member of the RAS family of membrane-associated G proteins and acts downstream of a number of receptor tyrosine kinases including EGFR (Sholl 2014). The activation of KRAS is necessary for signaling in cell growth and proliferation. As a common proto-oncogene, *KRAS* is mutated in 34% of smokers and 6% of never smokers with lung adenocarcinoma (Dogan *et al*., 2012). The high frequency of *KRAS* mutation in smokers strongly suggests its correlation with smoking history. The most common *KRA*S mutation in smokers is a codon 12 G>T transversion which causes a substitution

of Cysteine (C) with Glycine (G) in translation. The single-nucleotide missense alteration of the codons 12 and 13 account for more than 90% of the *KRAS* mutations in lung adenocarcinoma (Dogan *et al.*, 2012 and Kranenburg 2005). However, no KRAS mutation-targeted treatment is approved for clinical use but is an active area of clinical research (Okimoto and Bivona, 2014).

1.4. Personalized Treatment of Lung Cancer

1.4.1 Overview

Traditional cancer therapy is selected based on the organ of tumour origin and then choosing the most appropriate cytotoxic drugs for the disease. Cytotoxic chemotherapy is defined as drugs that interfere with basic functions of cellular replication including DNA synthesis, repair and replication, or interfere with microtubule synthesis. Commonly used cytotoxic drugs for the treatment of NSCLC include platinum agents (DNA cross-linking), etoposide (inhibiting DNA replication), taxanes and vinka alkaloids (affecting normal microtubule function), gemcitabine (nucleoside analogues), and pemetrexed (inhibiting DNA synthesis and cellular metabolism). The non-specific nature of these therapies leads to dose-limiting toxicities through damage to normal tissues including the neuropathy, nephropathy, alopecia, diarrhea, and cytopenias. These toxicities limit the safe delivery of these drugs to patients with poor clinical performance status, multiple comorbidities, and elderly patients (Socinski 2004; Sörenson *et al*., 2001). The identification of oncogenes that drive cancer has led to greater personalization of cancer treatments as new therapy is directed specifically against the protein product altered by the gene mutation. This ‘targeted therapy’ or ‘personalized medicine’ has resulted in treatments that are

both effective against the tumor and less toxic to the patient. This enhances the quality of life of patients diagnosed with an incurable disease.

1.4.2. Targeted Therapy and the Challenge in Manitoba

In Manitoba, targeted therapies are approved for lung adenocarcinoma with *EGFR, ALK*, or *ROS1* mutations. Patients with activating somatic *EGFR* mutations are treated with oral EGFR inhibitors gefitinib or erlotinib as first line therapy with better and longer lasting responses compared to cytotoxic chemotherapy. The utility of these drugs in the first line setting was demonstrated in the iPASS study comparing the use of first line EGFR therapy to standard platinum doublet chemotherapy (Mok *et al*., 2009). This study demonstrated superiority of gefitinib over chemotherapy in patients with a somatic *EGFR* mutation, but also demonstrated inferiority of the same drug in patients without an *EGFR* mutation. Therefore, the current routine use of these agents for first line treatment requires testing all patients with advanced adenocarcinoma for the *EGFR* mutation. At the present time, this test is not available in Manitoba and clinical samples are sent to Calgary Laboratory Services for polymerase chain reaction (PCR)-based detection at a cost of $475/case.

The oncogenic *ALK* gene fusion was discovered in 2007. The fusion results from a chromosomal rearrangement that places the kinase portion of the *ALK* gene adjacent to another activated gene, commonly *EML4*. This leads to the constitutive activation of the ALK protein that in turn signals tumor growth. The clinical gold standard for detecting the fusion gene is using a break-a-part fluorescent in situ hybridization (FISH) assay, after which positive patients

receive the oral inhibitor crizotinib. The test is also unavailable in Manitoba and samples are sent to PhenoPath labs in Seattle at a cost of US$495/case.

Routine clinical testing is not yet approved for *BRAF*, *ERBB2*, and *ROS1* in lung cancer. However, drugs that effectively target these mutations are approved in Canada for other indications. The *ROS1* gene fusion predicts similar response to crizotinib as the *ALK* fusion. *ERBB2* mutations predict response to the pan-EGFR-inhibitor afatinib approved for EGFR-mutated lung cancer. *BRAF* mutated lung tumours respond to vemurafinib which has been approved to treat *BRAF* mutated metastatic melanoma. These mutations are detected either with FISH (*ROS1* fusion) or PCR (*ERBB2* and *BRAF* point mutations) at similar costs per case as *EGFR* and *ALK*. Clinical trials correlating biomarker and respective inhibitor in lung cancer are underway and the expectation of the cancer community is that the drugs will be approved for these indications within a year (Okimoto and Bivona, 2014).

2. INTRODUCTION (Part II)

DNA Sequencing

2.1. Current Methods for Cancer Mutation Testing

As described above in Chapter 1, the majority of mutations amenable to targeted drug therapy are point mutations. The most commonly used method for mutation detection in clinical laboratories is PCR-amplification based sequencing. Multiple variations of this method exist and mainly differ in i) the amount of input DNA, and ii) the sensitivity to detect mutations at a lower allelic fraction. The major first step involves the amplification of a specific region of DNA using gene specific primers. This amplified region is then sequenced using standard capillary sequencing to identify the altered DNA bases. Newer laboratory methods amplify the DNA using customized primers compatible with quantitative PCR, which produce a signal only when the mutation is present. This no longer requires the extra sequencing step and can improve turnaround time for mutation reporting. The limitation of these PCR-amplification methods is that they can only interrogate known mutations and are limited to a single sample per assay. Also, they are often restricted to a single gene, and even then, specific loci within the gene. In general, the cost of testing a clinical sample for mutations ranges from $100 to $500 per case depending on the number of gene loci within the assay.

Gene amplifications and fusion gene detection are most efficiently performed by either PCR or FISH in most laboratories. The current gold standard for detection of *ERBB2* amplification in breast cancer and the *ALK* fusion in NSCLC is FISH. This requires adequate

collection and processing of samples so that nuclear architecture is preserved to provide a satisfactory report. The interpretation of FISH result is subjective and significant variations of FISH results by different assay readers have been reported (Wallander *et al*., 2012). Therefore, properly trained technicians to read the assay manually is essential. Also, FISH is a qualitative molecular testing tool which neither can identify the fusion partners nor distinguish fusion variants. Cost of FISH is typically $300-500 per case. If a particular tumor sample requires both mutation testing and fusion gene testing such as in lung cancer, clinical specimens can become depleted very quickly. In the age of minimally invasive procedures and smaller biopsies, this can often be a significant limitation.

PCR-based testing has a long history of application in genomic study of cancers, particularly the detection of known fusion gene products. This method is widely used for hematologic malignancies where fresh RNA from hematologic cells are readily available. By contrast, most biopsies from solid tumors are stored as formalin fixed paraffin-embedded (FFPE) specimens. RNA in this format unfortunately suffers from increased degradation and cross-linking. Therefore, this method is not currently recommended as a first-line diagnostic tool for *ALK* fusion detection by the guidelines because of concerns of higher risks of assay failure due to false negatives in RNA-based assay in routine FFPE pathology material, and also due to the variability in *EML4-ALK* fusion structure and the existence of other *ALK* fusion partner (Lindeman *et al*., 2013).

2.2. DNA Sequencing

2.2.1. First-generation versus Next-generation Sequencing

The terminology first-generation sequencing typically refers to traditional Sanger sequencing. In brief, this process involves incubating a single-strand of DNA with a DNA primer (specific to the region of interest), DNA polymerase, normal deoxynucleotides, and modified deoxynucleotides usually labeled with a unique fluorescent dye color corresponding to each of the four bases (A, C, G, T) that can be incorporated into the complementary sequence. With incorporation of each new base, this fluorescent dye is activated and can be detected with automated sensors. The chronological sequence of dye being released with each base corresponds to the DNA base order of the region of interest being sequenced (Smith *et al*., 1985). This process overall is fairly efficient and accurate but overall throughput is limited as only one DNA region and one patient can be examined with each analysis.

Next-generation sequencing (NGS) in simplistic terms relies on slight variations of the Sanger sequencing process as its fundamental basis. The key difference is that the process has been adapted and automated so that hundreds of thousands of sequencing reactions can be run in parallel. A variety of methods exist to accomplish this process. **Table 4** summarizes currently available NGS platforms (Goodwin *et al*., 2016). For an extensive description of commonly used platforms, please refer to the review by Metzker (2010). The major accomplishments of next-generation methods have been to increase throughput of sequencing, increase the area of DNA that can be covered, and increase the number of samples that can be processed. More

significantly, these improvements result in a reduction in cost, which makes these platforms increasingly suitable for clinical applications (**Figure 2**).

Table 4: Summary NGS platforms. Reprinted by permission from Macmillan Publishers Ltd: *Nature Reviews Genetics*. June 2016; 17, 333–351. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Copyright (2016).
http://www.nature.com.uml.idm.oclc.org/nrg/journal/v17/n6/full/nrg.2016.49.html

Wildfire	75 (SE)	120 Gb					
	50 (SE)*	160 Gb*					
SOLiD 5500xl	50 (SE)	160 Gb	~1.4 B*	10 d*	≤0.1%, AT bias‡	$251,000‡	$70‡
	75 (SE)	240 Gb					
	50 (SE)*	320 Gb*					
BGISEQ-500 FCS[155]	50–100 (SE/PE)*	8–40 Gb*	NA§	24 h*	≤0.1%, AT bias‡	$250 (REF. 155)	NA§
BGISEQ-500 FCL[155]	50–100 (SE/PE)*	40–200 Gb*	NA§	24 h*	≤0.1%, AT bias‡	$250,000 (REF. 155)	NA§
Sequencing by synthesis: CRT							
Illumina MiniSeq Mid output	150 (SE)*	2.1–2.4 Gb*	14–16 M*	17 h*	<1%, substitution‡	$50,000 (REF. 118)	$200–300 (REF. 118)
Illumina MiniSeq High output	75 (SE)	1.6–1.8 Gb	22–25 M (SE)*	7 h	<1%, substitution‡	$50,000 (REF. 118)	$200–300 (REF. 118)
	75 (PE)	3.3–3.7 Gb	44–50 M (PE)*	13 h			
	150 (PE)*	6.6–7.5 Gb*		24 h*			
Illumina MiSeq v2	36 (SE)	540–610 Mb	12–15 M (SE)	4 h	0.1%, substitution‡	$99,000‡	~$1,000
	25 (PE)	750–850 Mb	24–30 M (PE)*	5.5 h			$996
	150 (PE)	4.5–5.1 Gb		24 h			$212
	250 (PE)*	7.5–8.5 Gb*		39 h*			$142‡
Illumina MiSeq v3	75 (PE)	3.3–3.8 Gb	44–50 M (PE)*	21–56 h*	0.1%, substitution‡	$99,000‡	$250
	300 (PE)*	13.2–15 Gb*					$110‡
Illumina NextSeq 500/550 Mid output	75 (PE)	16–20 Gb	Up to 260 M (PE)*	15 h	<1%, substitution‡	$250‡	$42
	150 (PE)*	32–40 Gb*		26 h*			$40‡
Illumina NextSeq 500/550 High output	75 (SE)	25–30 Gb	400 M (SE)*	11 h	<1%, substitution‡	$250‡	$43
	75 (PE)	50–60 Gb	800 M (PE)*	18 h			$41
	150 (PE)*	100–120 Gb*		29 h*			$33‡
Illumina HiSeq2500 v2 Rapid run	36 (SE)	9–11 Gb	300 M (SE)*	7 h	0.1%, substitution‡	$690‡	$230
	50 (PE)	25–30 Gb	600 M (PE)*	16 h			$90
	100 (PE)	50–60 Gb		27 h			$52
	150 (PE)	75–90 Gb		40 h			$45
	250 (PE)*	125–150 Gb*		60 h*			$40‡
Illumina HiSeq2500 v3	36 (SE)	47–52 Gb	1.5 B (SE)	2 d	0.1%, substitution‡	$690‡	$180
	50 (PE)	135–150 Gb	3 B (PE)*	5.5 d			$78
	100 (PE)*	270–300 Gb		11 d*			$45‡
Illumina HiSeq2500 v4	36 (SE)	64–72 Gb	2 B (SE)	29 h	0.1%, substitution‡	$690‡	$150
	50 (PE)	180–200 Gb	4 B (PE)*	2.5 d			$58
	100 (PE)	360–400 Gb		5 d			$45
	125 (PE)*	450–500 Gb*		6 d*			$30‡
Illumina HiSeq3000/4000	50 (SE)	105–125 Gb	2.5 B (SE)*	1–3.5 d*	0.1%, substitution‡	$740/$900 (REF. 156)	$50
	75 (PE)	325–375 Gb					$31
	150 (PE)*	650–750 Gb*					$22 (REF. 157)

Table 4: Summary NGS platforms (continued)

Sequencing by synthesis: SNA (cont.)							
Illumina HiSeq X	150 (PE)*	800–900 Gb per flow cell*	2.6–3 B (PE)*	<3 d*	0.1%, substitution[9]	$1,000[4,5]	$7.0[9]
Qiagen GeneReader	NA[§]	12 genes; 1,250 mutations[22]	NA[§]	Several days[22]	Similar to other SBS systems[22]	NA[§]	$400–$600 per panel[22]
Sequencing by synthesis: SNA							
454 GS Junior	Up to 600; 400 average (SE, PE)*	35 Mb*	~0.1 M*	10 h*	1%, indel[9]	NA[§]	$40,000[9]
454 GS Junior+	Up to 1,000; 700 average (SE, PE)*	70 Mb*	~0.1 M*	18 h*	1%, indel[9]	$108,000[9]	$19,500[9]
454 GS FLX Titanium XLR70	Up to 600; 450 mode (SE, PE)*	450 Mb*	~1 M*	10 h*	1%, indel[9]	NA[§]	$15,500[9]
454 GS FLX Titanium XL+	Up to 1,000; 700 mode (SE, PE)*	700 Mb*	~1 M*	23 h*	1%, indel[9]	$450,000[9]	$9,500[9]
Ion PGM 314	200 (SE)	30–50	400,000–550,000*	23 h	1%, indel[9]	$49[9]	$25–3,500[9]
	400 (SE)	60–100 Mb*		3.7 h*			
Ion PGM 316	200 (SE)	300–500 Mb	2–3 M*	3 h	1%, indel[9]	$49[9]	$700–1,000[9]
	400 (SE)*	600 Mb–1 Gb*		4.9 h*			
Ion PGM 318	200 (SE)	600 Mb–1 Gb	4–5.5 M*	4 h	1%, indel[9]	$49[9]	$450–800[9]
	400 (SE)*	1–2 Gb*		7.3 h*			
Ion Proton	Up to 200 (SE)	Up to 10 Gb*	60–80 M*	2–4 h*	1%, indel[9]	$224[9]	$80[9]
Ion S5 520	200 (SE)	600 Mb–1 Gb	3–5 M*	2.5 h	1%, indel[9]	$65 (REF. 158)	$2,400*
	400 (SE)*	1.2–2 Gb*		4 h*			$1,200*
Ion S5 530	200 (SE)	3–4 Gb	15–20 M*	2.5 h	1%, indel[9]	$65 (REF. 158)	$950*
	400 (SE)*	6–8 Gb*		4 h*			$475*
Ion S5 540	200 (SE)*	10–15 Gb*	60–80 M*	2.5 h*	1%, indel[9]	$65 (REF. 158)	$300*
Single-molecule real-time long reads							
Pacific BioSciences RS II	~20 Kb	500 Mb–1 Gb*	~55,000*	4 h*	13% single pass, ≤1% circular consensus read, indel[9]	$695[9]	$1,000[9]
Pacific Biosciences Sequel	8–12 Kb[69]	3.5–7 Gb*	~350,000*	0.5–6 h*	NA[§]	$350 (REF. 69)	NA[§]
Oxford Nanopore MK 1 MinION	Up to 200 Kb[159]	Up to 1.5 Gb[159]	>100,000 (REF. 159)	Up to 48 h[160]	~12%, indel[159]	$1,000*	$750*
Oxford Nanopore PromethION	NA[§]	Up to 4 Tb*	NA[§]	NA[§]	NA[§]	$75*	NA[§]
Synthetic long reads							
Illumina Synthetic Long-Read	~100 Kb synthetic length*	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500 (possible barcoding and partitioning errors)	No additional instrument required	~$1,000*
10X Genomics	Up to 100 Kb synthetic length*	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500 (possible barcoding and partitioning errors)	$75 (REFS 72,161)	See HiSeq 2500 +$500 per sample[161]

Approx., approximate; AT, adenine and thymine; B, billion; bp, base pairs; d, days; Gb, gigabase pairs; h, hours; indel, insertions and deletions; Kb, kilobase pairs; M, million; Mb, megabase pairs; NA, not available; PE, paired-end sequencing; SBS, sequencing by synthesis; SE, single-end sequencing; Tb, terabase pairs. *Manufacturer's data. [‡]Rounded from Field Guide to next-generation DNA sequencers[160] and 2014 update. [§]Not available as this instrument will be discontinued or only available as an upgraded version. [‖]As this product has been developed only recently, this information is not available. [¶]Not available as a single instrument.

Figure 2: Improvement in DNA sequencing equipment (courtesy of Dr. Yvonne Myal)

Then (first generation sequencing): 33 years to sequence one human genome on one machine

Now (next generation sequencing): 2 weeks to sequence one human genome on one machine

Next-generation DNA sequencing efforts are typically divided into whole-genome and sub-whole-genome projects. Whole-genome sequencing aims to cover each base in the genome on average 30 times so that subsequent analytical tools can call the presence of a mutation with sufficient confidence. This method is useful for point mutation discovery, detection of fusion genes, and for examining non-coding regions of the genome. The main limitation to whole-genome sequencing is the substantial cost associated with covering the vast genomic territory. This can become even more problematic in polyploid cancer samples. Also, the functional relevance of the vast amount of data obtained from whole-genome sequencing can be limited. Sub-whole genome sequencing restricts the DNA territory sequenced so that the cost of sequencing can be reduced. This is accomplished through a process of targeted sequencing. This can be accomplished either by targeted amplification of specific DNA regions (for example via PCR) and then subsequently sequencing only the amplified DNA on next-generation machines. Another solution is to design oligonucleotides 'baits' that are complementary to a DNA region of interest to 'capture' only this specific territory for sequencing (Gnirke *et al*., 2009). Typical sub-genome methods used are whole-exome sequencing where oligonucleotides against the whole coding region of the genome (or exome) are used as baits. All functional mutations to date are known to reside within the exome that account for 1% of the whole genome. Mutations that alter the protein structure may be the most relevant to cancer biology and actionable targets for future therapy. Also, by sequencing a smaller territory, there is increased sensitivity of detecting mutations that may be present in sub-clonal populations of the tumour. Restricting sub-genome sequencing to an even smaller territory, for example clinically relevant genes, makes next-generation methods more suitable for every day clinical applications.

2.2.2. Next-generation Sequencing (NGS) Improves Cancer Gene Mutation Discovery

The year 2001 saw the first published reports of the complete human genome sequence, approximately 6 gigabases in total (Lander *et al*., 2001 and Venter *et al*., 2001). These initial efforts relied on the step-by-step assembly of individual sequences of human DNA through a collaboration of multiple centers across the globe, a laborious and costly endeavor. Not until 2009 was the first complete cancer genome published. The major advance leading to this accomplishment was the development of next-generation sequencing technologies that allowed the parallel sequencing of thousands of DNA fragments simultaneously (Wheeler *et al*., 2008). This resulted in both improved speed of sequencing and reduction in the amount of input DNA from an individual case. The first cancer sequencing effort focused on the acute myelogenous leukemia (AML) genome of a single patient (Ley *et al*., 2008). The complete genome sequence of a breast tumor was the first for an epithelial cancer (Shah *et al*., 2009). In 2008, two lung cancer cell lines were sequenced (Campbell *et al*., 2008). Two years later, Lee et al. reported the whole genome sequencing of lung adenocarcinoma and normal lung tissue from a same patient. Of the 530 somatic single nucleotide variants found, 393 were detected in coding regions including in the *KRAS* proto-oncogene. In addition, 53 structural variations were identified (Lee *et al*., 2010). The massive parallel sequencing a SCLC cell line was first reported in 2010 (Pleasance *et al*., 2010). Ju's research team published the whole genome study of paired NSCLC and normal lung tissue from a never smoking patient, and a novel fusion gene, *KIF5B-RET*, was discovered (Ju *et al*., 2012). More recently, The Cancer Genome Atlas Research Network published comprehensive molecular profiling of 230 resected lung adenocarcinomas using messenger RNA, microRNA and shotgun DNA sequencing (2014). While many somatic mutations were observed in the individual lung cancer, the functional significance of any

individual mutation remained unknown. Increasing the number of samples sequenced was proposed to improve the sensitivity of detecting significantly mutated genes to better understand cancer pathogenesis.

The significant cost of genome sequencing (>$100,000 per sample) initially limited expansion of sequencing efforts. Improvements in both sequencing technology and analytic methods resulted in further cost reductions. For a detailed discussion of these technical improvements, please refer to the review by Meyerson *et al* (2010). Over the last two years, dozens of cancer genome sequencing papers have been published each involving larger numbers of patient samples. The more recent efforts have involved over a hundred cases per project. These studies have revealed that the cancer genome is complex and that many potential targets for therapy, both point mutations and fusion genes, may exist in individual patients. Unfortunately, many of these new targets occur at frequencies under 10% within individual cancer types. Thus efficient methods to detect these mutations and enroll patients in clinical trials for targeted therapy must be implemented.

2.2.3. Bringing Next-generation Technology to Clinical Laboratory

The value of next-generation sequencing methods for detecting rare mutations within tumours has been widely demonstrated. The key next step in the process is to bring the technology to the clinical laboratories where pathologic specimens from patients are collected and examined on a daily basis. Instruments that can serve this function have three major requirements: i) accuracy of sequencing, ii) ease of use, and iii) modest physical space needs. Multiple instruments are

currently entering clinical use that meet these requirements led by the Ion Torrent PGM Sequencer and the Illumina MiSeq Personal Sequencer. These machines occupy less space than most flow cytometry instruments and can perform the majority of sequencing steps including analysis in an automated fashion. Also, these instruments are compatible with a number of commercial methods for targeting sub-genomic regions.

For the analysis of genes relevant to cancer, targeted methods can concentrate on sequencing only the handful of clinically important genes. This can be limited to just the few that currently guide targeted therapies, like those described earlier. Or they can be more expansive to include additional genes that are potential targets for newer drugs currently in early phase clinical trials. A major restriction to analysis of cancer samples is the availability of sufficient DNA. With more widespread use of minimally invasive procedures, many biopsies are of limited size. Once these specimens are processed for histology and routine immunohistochemistry, limited material remains for subsequent DNA analysis. Platforms like the Ion Torrent PGM Sequencer can function with DNA input as little as 10ng. Thus, DNA obtained from as few as 3,000 cells may be sufficient. Recently, success in next generation sequencing in lung adenocarcinoma specimens of fine needle aspiration has been reported by Qiu *et al* (2015) who applied 10 ng of pooled amplicons for the study. However, if there is still insufficient DNA, repeat biopsies may be necessary to make the correct molecular diagnosis.

Tumor Heterogeneity: Of note, a major issue now being realized in epithelial cancers is the existence of intra-tumor heterogeneity (Gerlinger *et al.*, 2012). In practical terms, this means

that a single biopsy of any given sample may not contain all genetic abnormalities within the tumor and may confound current efforts to personalize patient treatments.

2.2.4. NGS in Manitoba

A NGS platform was established at the University of Manitoba in 2012 for research use. Diagnostic Services Manitoba (DSM), the main molecular pathology provider in the province, has established a clinical NGS platform. At present, its main purpose is to provide germline mutation analysis for heritable diseases. Our research project expands on these efforts to apply the next-generation sequencing technology with both research and clinical applications to the clinical challenge of somatic mutation detection to guide cancer therapy.

3. RATIONAL, HYPOTHESIS AND RESEARCH DESIGN

3.1. Rational and Hypothesis

The trend in oncology practice is to personalized therapy for individual patients based on molecular characteristics of their tumour. Targeted therapies for lung cancer are already in routine clinical use while others are progressing rapidly through pre-clinical and clinical development. Detecting the genetic abnormality targeted by these drugs quickly and accurately is essential for appropriate patient selection. As described, the ability to perform routine lung cancer gene testing in Manitoba is limited. The next-generation sequencing provides an alternative approach for detecting mutated oncogenes in lung cancers. The investment in NGS technology by DSM represents an opportunity to erase the current cancer gene testing deficit and position Manitoba as a national leader in the field. Current commercially available NGS-based cancer assays are robust for point mutation detection. Fusion gene sequencing requires a different NGS workflow that traditionally runs separate from point mutation detection. We aim to develop an NGS-based gene assay including *ALK* fusion and *EGFR* mutation for lung cancer. We hypothesize that this approach is likely equivalent in sensitivity to standard methods and significantly more cost-effective.

3.2 Research Design

DNA and RNA from 38 formalin-fixed paraffin embedded NSCLC samples archived in Diagnostic Services Manitoba will be extracted and analyzed using gene enrichment methods from Archer Diagnostics followed by sequencing on the Illumina MiSeq NGS machine. Targeted

DNA sequencing to detect the *EGFR* mutation will be performed on 19 samples while targeted RNA sequencing will be applied to 20 samples to identify the *ALK* gene rearrangement. The NGS results will be compared with and confirmed with current clinical standard molecular tests for EGFR (real-time PCR) and ALK (immunohistochemistry and FISH).

4. MATERIALS & METHODS

4.1. Patients and Sample Selection

Archived formalin-fixed paraffin-embedded (FFPE) tissue blocks were collected from 38 patients diagnosed with lung carcinoma from 1999 to 2015 at the Diagnostic Services of Manitoba, Canada. Selection of the patients for the study was not random. The selected patients were those with NSCLC histology. Based on the WHO 2004 classification, 37 of the patients were diagnosed with adenocarcinoma and 1 with small cell carcinoma. The ratio of male and female is 0.81 (male: 17, female: 21). The mean age of the 17 male patient is 63.9 years old, ranging from 47 to 83 years. Mean age is 67.2 years old for the 21 female patients (49 to 81 years). The tumor tissue percentage of the specimens was above 15% as assessed by a pathologist (Dr. Gefei Qing). Of those specimens, 33 are lung resection samples and 5 are from needle biopsy (lymph nodes, pleura or vertebra).

Among 18 patients, the *EGFR* gene mutation and *ALK* fusion status have been previously tested either with real-time PCR or FISH, respectively. The smoking history of most cases is unknown. Dr. Gefei Qing at the Diagnostic Services Manitoba is the only individual in our research team who knows the mutation status of individual samples. All other personnel are blinded until all analysis is complete.

4.2. ALK fusion detection

4.2.1 Immunohistochemistry

A widely accepted method to screen patients for *ALK* fusion is immunohistochemistry (IHC) with ALK-specific antibodies that are sensitive and have a high negative predictive value. In 2013, using a 40-case TMA, Dr. Gefei Qing and DSM tested the IHC approach. Independent sections were stained with the ALK 5A4 (Novacastra) and D5F3 (Cell Signaling) antibodies. Three cases (including a positive control) stained positive with both antibodies. Positive cases and 3 random negative cases were sent to PhenoPath Laboratories (Seattle, USA) for FISH confirmation. The 3 IHC positive staining cases were confirmed FISH positive. IHC is now used routinely to screen for the *ALK* fusion gene in Manitoba. Approximately 180 advanced Stage NSCLC are tested in DSM laboratories annually.

4.2.2. Fluorescence In Situ Hybridization

All IHC positive or equivocal staining cases are sent to PhenoPath Laboratories (http://phenopath.com/, Seattle, USA) for FISH confirmation. The confirmation test was performed using the Vysis® ALK Break Apart FISH Probe IVD Kit (Abbott Molecular, Chicago, IL) following the IVD probe package insert protocol with modifications. The modifications include use of a Metasystems image analysis system to enhance scoring of tumor cells and an optimized digestion, pretreatment, and wash protocol comparable to the IVD kit protocol. These modifications have been validated by PhenoPath and approved for clinical use. Each *ALK* test performed is evaluated by a pathologist including analysis of an H&E stained section for tumor cell adequacy. Probed slides are scanned in areas selected by a pathologist using the MetaSystemsTM imaging platform. Tumor cells are then scored and each case is

resulted according to the IVD kit protocol. For each case 50 tumor cells are evaluated. If more than 50% of the tumor cells are positive then the case is positive. If less than 10% of the tumor cells are positive then the case is negative. If there is between 10% to 50% positivity, 50 more cells are evaluated by a second analyst. After the second count, a sample is positive if 15% or more of the 100 cells counted in total by the 2 analysts are positive. The performance characteristics of this test have been validated by Phenopath Laboratories.

4.3. Real-Time Polymerase Chain Reaction for *EGFR* Testing

Samples negative for *ALK* fusion were sent to Calgary Laboratory Services (Calgary, Canada) for RT-PCR based *EGFR* mutation assessment. According to the assessment reports provided, *EGFR* mutation is tested by RT-PCR utilizing the Qiagen EGFR RGQ PCR Kit. The PCR Kit screens for the following mutations: deletions in exon 19 (detects the presence of any of the common deletions but does not distinguish between them), T790M (This mutation has been associated with resistance to anti-EGFR therapy), L858R, L861Q, G719X (detects the presence of G719S, G719A, or G719C but does not distinguish between them), S768I, and 3 insertions in exon 20 (detects the presence of any of 3 insertions but does not distinguish between them). A result of "not detected" refers to those mutations screened for in this kit but not found to be mutated.

4.4. Next Generation Sequencing

4.4.1. DNA isolation

Two 10 µm FFPE sections were cut for each sample using a microtome. Sections underwent DNA extraction using the QIAamp® DNA FFPE Tissue isolation kit (QIAGEN) according to manufacture instructions. Briefly, one 10 µm FFPE tissue section is placed in 1.5 ml micro-centrifuge tube to which 1 ml xylene is added to the sample before centrifuging at full speed for 2 minutes at room temperature to remove paraffin. One ml of 99% ethanol is used to extract residual xylene and discarded after centrifugation. After all residual ethanol has evaporated, sample is lysed under denaturing condition (incubating at 56 degrees Celsius for one hour) with proteinase K. To reverse formalin crosslinking, the sample is then incubated at 90 degrees Celsius for one hour with the buffer provided. Sample DNA binds to the membrane provided and contaminations flow through by centrifugation. Residual contaminants are removed by repeated washing and centrifugation cycles. Pure and concentrated DNA is eluted from the membrane and stored at -20 degrees Celsius until use.

4.4.2. RNA isolation

RNA extraction was completed using AllPrep® DNA/RNA FFPE kit (QIAGEN) according to manufacture instructions. In Brief, one 10 µm FFPE tissue section from each sample is deparaffinised using xylene before proteinase K is added and incubated at 56 degrees Celsius for 15 minutes to release RNA molecules from crosslinked protein molecules. Centrifuge for 15 minutes at 20,000 x *g* to collect RNA containing supernatant. The supernatant is then incubated at 80 degrees Celsius for 15 minutes to reverse formaldehyde modification. Next, the sample is

treated with DNase to remove DNA contamination. Sample RNA binding to the column is washed and centrifuged for purification. Extracted RNA is eluted from the column and stored at -20 degrees Celsius until use.

4.4.3. DNA Quantification

DNA quality was assessed with the QuantiFluor™ dsDNA system (Promega) according to manufacturer's guide. In brief, the Fluorometer is calibrated with blank and 500ng/ml DNA standard. Then the concentration of the unknown DNA sample is calculated by comparing and plotting to the standard DNA concentration curve.

4.4.4. RNA quantification

RNA quality was assessed with the QuantiFluor™ RNA system (Promega) according to manufacturer's guide. Briefly, the Fluorometer is calibrated with blank and 2,000ng/ml RNA standard. Then the concentration of the unknown RNA sample is calculated by comparing and plotting to the standard RNA concentration curve.

4.4.5. Gene Enrichment and Library Construction

4.4.5.1. Archer VariantPlex CTL Kit for Illumina (http://archerdx.com/variantplex/ctl-dna?mid=nav, ArcherDX, Boulder, CO) was applied to capture all exons of *EGFR*, *KRAS*, *ERBB2*, and *BRAF* combined additional gene territories as described (**Table 5**). This represents a total genomic territory of ~56 kb. Fifty ng of DNA from 19 samples were digested into 150-200

bp fragments prior to DNA library construction according to manufacturer's instructions. Illumina-specific barcoded adapters with total length of 121 bases were attached to each fragment to generate sequencing libraries. Pooled libraries were then hybridized to the biotinylated baits. Baits are then captured using streptavidin coated magnetic beads. After elution and washing, the DNA representing only regions of interest was ready for sequencing. DNA quality is assessed both before and after library construction to assess suitability for sequencing.

4.4.5.2. Archer FusionPlex CTL Kit for Illumina (http://archerdx.com/fusionplex-assays/ctl-rna?mid=nav, ArcherDX, Boulder, CO) was ordered to capture exons 19, (intron 19), 20, 21, 22, 23 of ALK and its fusion partners (Table 6). Following the manufacture's protocol, 20 ng RNA of 20 samples were loaded for cDNA synthesis prior to cDNA fragmentation and library construction. Once the cDNA synthesis was accomplished, the remaining steps are highly similar as described above in the section of 4.4.3.1.

Table 5: Gene mutations and or copy number variations (CNV) detected by Archer VariantPlex CTL Kit

BRAF	***CCND1***	***CTNNB1***	***DDR2***	***EGFR***	***EIF1AX***
ERBB2	***FGFR1***	***FGFR2***	***FGFR3***	***GNAS***	***HRAS***
IDH1	***IDH2***	***KIT***	***KRAS***	***MAP2K1***	***MDM2***
MET	***NRAS***	***PDGFRA***	***PIK3CA***	***PTEN***	***RET***
ROS1	***STK11***	***TERT***	***TP53***	***TSHR***	

4.4.6. DNA (cDNA) sequencing

Purified libraries were sequenced on the Illumina MiSeq machine (Harisendy *et al*., 2011). The MiSeq captures sequential images of fluorescent-labeled nucleotides added simultaneously to ~$15x10^6$ clusters isolated in parallel on a sequencing chip. Each cluster represents a million copies of an individual DNA strand. Images are stacked to determine the specific DNA sequence of each cluster. The MiSeq sequences 5000 Megabases (Mb) of DNA in a single run. Up to 20 samples can be sequenced on each run: this represents the maximum number of samples that can sequenced in a single run while achieving at least 200 fold sequencing depth of coverage for each base in the region of interest. Paired-end DNA sequencing of 200 bp length were performed.

4.4.7. Data Analysis

Publicly available software Archer Analysis (http://archerdx.com/software/analysis) was used to perform the analysis. Briefly, after filtering reads and alignment to the reference genome, paired-end reads were analyzed for the presence of an *ALK* fusion (**Figure 3**) and other common gene mutations in lung cancers including *EGFR*, *KRAS*, *ROS1* and *TP53*.

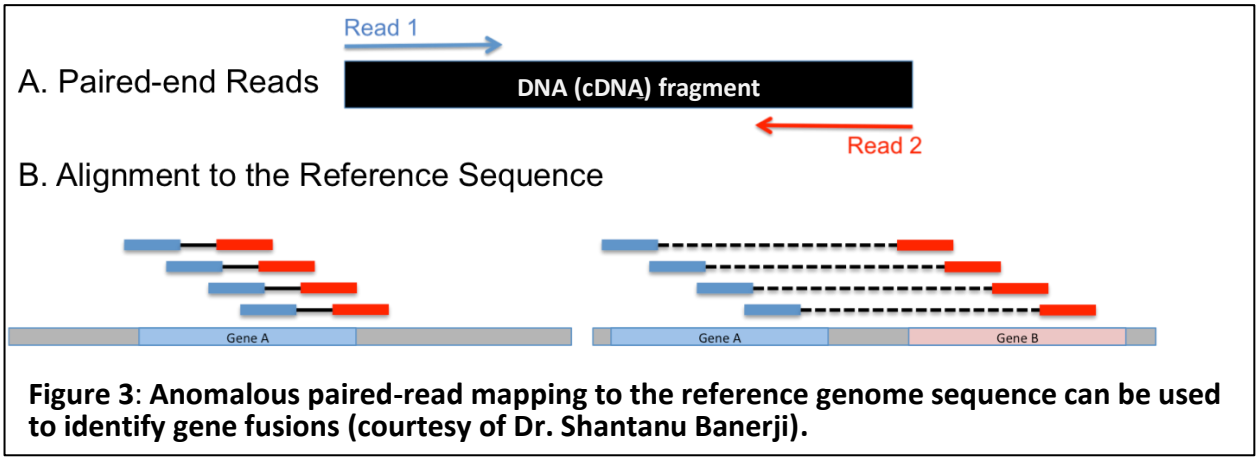


Figure 3: **Anomalous paired-read mapping to the reference genome sequence can be used to identify gene fusions (courtesy of Dr. Shantanu Banerji).**

5. RESULTS

5.1. Targeted DNA Sequencing for Gene Mutation Detection

Of those 19 cases subjected to targeted DNA sequencing, 7 cases have been previously tested for *EGFR* mutation by RT-PCR with 2 cases being positive for the L858R mutation and other 5 cases being negative. The DNA sequencing results are consistent with 6 out of those 7 cases except for one discordant case that is RT-PCR positive but DNA sequencing negative for *EGFR* mutation (**Table 6**). Among the remaining 12 cases with unknown status for *EGFR* mutation, targeted sequencing detected a L858R mutation in 2 cases that were subsequently confirmed with RT-PCR tests (**Table 7**). The overall concordant rate between targeted DNA sequencing and RT-PCR in *EGFR* mutation detection in this study is 88.9% (8 out of 9). Furthermore, the DNA sequencing revealed 3 cases having a *KRAS* mutation in codon G12C, 2 cases harboring a *KRAS* mutation in codon G12V, and 1 case with *TP53* (Y220C) mutation (**Tables 6** and **7**). No single case holds two or more common lung cancer oncogenic mutations simultaneously as per the DNA sequencing results.

5.2. Targeted RNA Sequencing for ALK Fusion Detection

Targeted RNA sequencing for *ALK* fusion detection was performed on 20 cases. Among those cases, 11 cases have been pre-screened for *ALK* fusion using IHC assay with 3 positive, 1 equivocal and 7 negative stains observed. All ALK-IHC positive or equivocal cases were further tested with ALK-FISH methods, and 2 IHC positive cases were confirmed to harbor the *EML4-*

ALK fusion gene. The two FISH-confirmed *ALK* fusion positive cases were also identified by the targeted RNA sequencing (**Figure 4** and **Figure 5**). The results between RNA sequencing and IHC/FISH are 100% concordant (**Table 8**). No *ALK* fusion was revealed by the sequencing among those 9 cases whose *ALK* fusion conditions were unknown prior to sequencing, as shown in **Table 9**.

Table 6: NGS of 7 cases with known *EGFR* mutation status pre-assessed by RT-PCR.

Sample Code	EGFR mutation status (RT-PCR)	NGS Mutation Call	Codon of mutation	Total NGS reads
VP10	L858R	EGFR	L858R	1,496,233
VP7	L858R	_	_	1,551,457
VP3	Negative	KRAS	G12C	1,505,662
VP4	Negative	_	_	1,704,674
VP4 repeat	Negative	_	_	1,509,067
VP5	Negative	_	_	971,178
VP6	Negative	_	_	1,999,338
VP9	Negative	KRAS	G12V	1,233,455
VP9 repeat	Negative	KRAS	G12V	1,758,788

Table 7: NGS of 12 cases with unknown *EGFR* mutation status. (* The EGFR-L858R mutations of V17 and V19 were confirmed with RT-PCR afterwards)

Sample Code	EGFR mutation status (unknown)	NGS Mutation Call	Codon of mutation	Total NGS reads
VP1	Unknown	_	_	2,234,547
VP13	Unknown	TP53	Y220C	1,529,724
VP8	Unknown	KRAS	G12C	842,383
VP8 repeat	Unknown	KRAS	G12C	1,355,976
VP11	Unknown	_	_	821,298
VP12	Unknown	_	_	390,299
VP14	Unknown	_	_	904,321
VP17*	**Unknown**	**EGFR**	**L858R**	**1,766,884**
VP18	Unknown	KRAS	G12V	1,274,231
VP19*	**Unknown**	**EFGR**	**L858R**	**2,184,210**
VP21	Unknown	KRAS	G12C	1,832,898
VP22	Unknown	_	_	626,113
VP20	Unknown	_	_	Failed

Table 8: NGS of 11 cases with known *ALK* fusion status pre-determined by FISH and IHC.

Sample Code	ALK-IHC	ALK-FISH	NGS Total Reads	NGS Gene Fusion Call
FP3	Negative	-	1,492,711	0
FP10	Equivocal	Negative	1,829,497	0
FP13	Positive	Negative	1,167,394	0
FP1	**Positive**	**Positive**	**273,152**	**EML4-ALK**
FP23	Negative	-	1,064,785	0
FP2	**Positive**	**Positive**	**208,279**	**EML4-ALK**
FP5	Negative	-	960,224	0
FP6	Negative	-	190,192	0
FP7	Negative	-	101,514	0
FP9	Negative	-	110,666	0
FP15	Negative	-	183,848	0

Table 9: NGS of 9 cases with unknown *ALK* fusion status.

Sample Code	ALK-IHC	ALK-FISH	NGS Total Reads	NGS Gene Fusion Call
FP8	Unknown	Unknown	980,353	0
FP17	Unknown	Unknown	687,747	0
FP18	Unknown	Unknown	1,188,151	0
FP19	Unknown	Unknown	1,105,897	0
FP20	Unknown	Unknown	1,090,132	0
FP11	Unknown	Unknown	935,418	0
FP14	Unknown	Unknown	870,559	0
FP21	Unknown	Unknown	1,395,113	0
FP22	Unknown	Unknown	876,248	0

Figure 4: Detection of *EML4-ALK* fusion in lung adenocarcinoma case FP1 by targeted RNA sequencing.

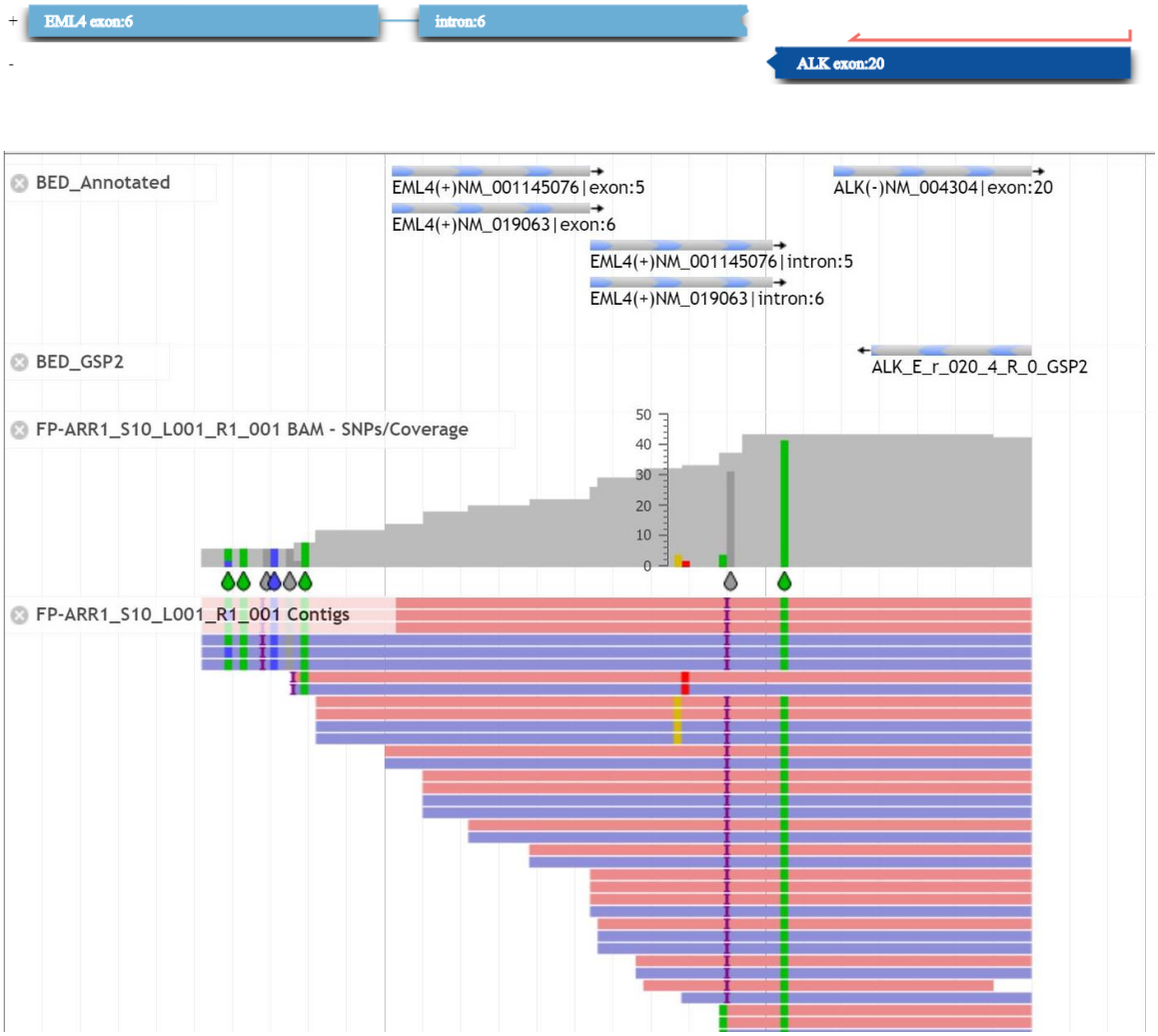
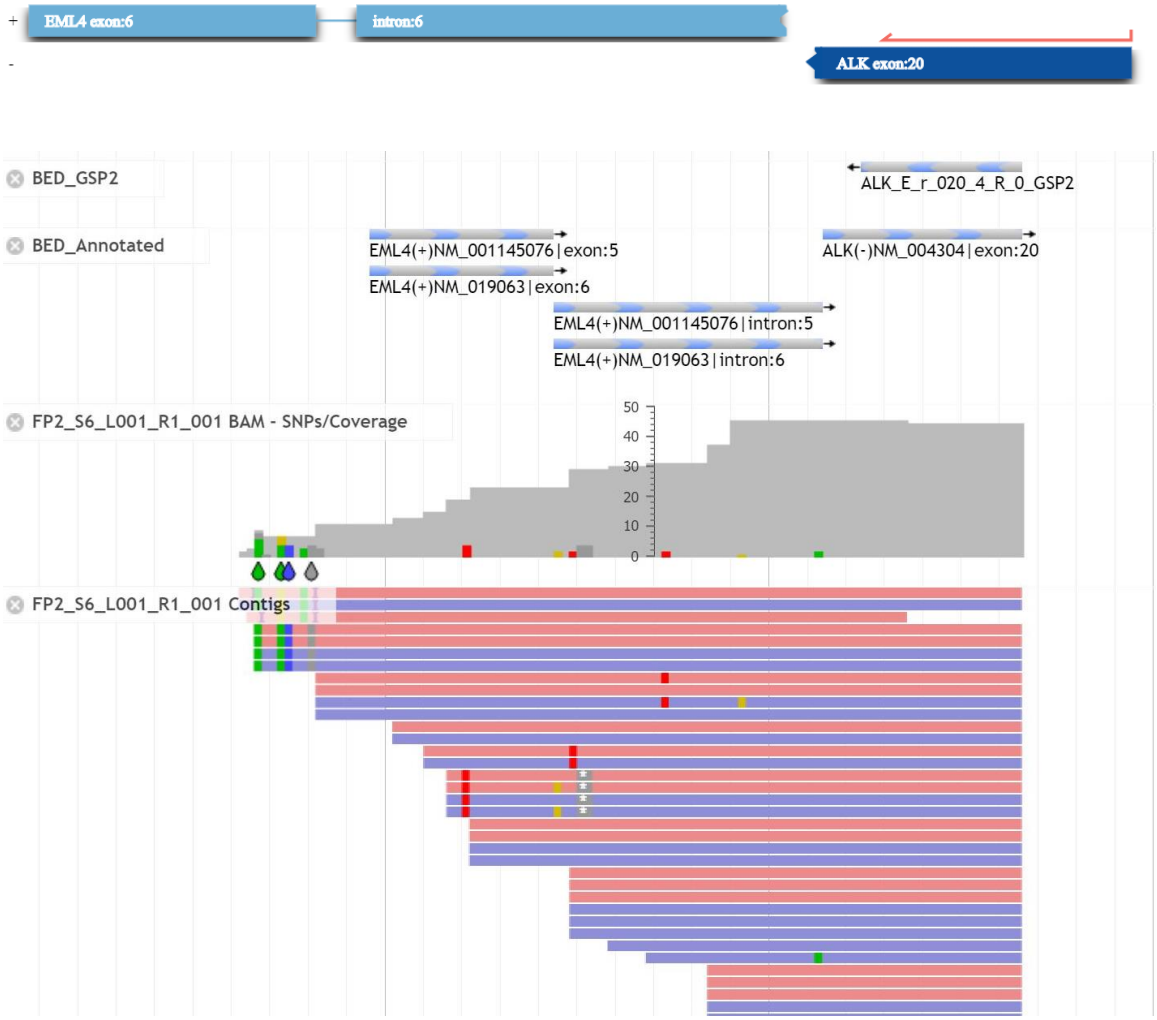


Figure 5: Detection of *EML4-ALK* fusion in lung adenocarcinoma case FP2 by targeted RNA sequencing.



6. DISCUSSION & CONCLUSION

Traditional cancer therapy is selected based on the organ of tumor origin and then choosing the most appropriate cytotoxic drugs that poisoned proliferating cells. The non-specific nature of these therapies led to dose-limiting toxicities. The identification of oncogenes that drive cancer has led to greater personalization of cancer treatments as now therapy is directed specifically against the protein product altered by the gene mutation. This 'targeted therapy' has resulted in treatments that are both effective and less toxic to the patient. To identify a qualified candidate for targeted therapy, a highly sensitive and specific testing for targeted gene mutation is mandatory.

Three targeted therapies are approved in Manitoba to treat lung adenocarcinoma including anti-EGFR (gefitinib or erlotinib) and anti-ALK (crizotinib). Clinical use of these agents requires testing patients for the *EGFR* mutation and *ALK* rearrangement. The 'gold standard' or recommended molecular tests, PCR and FISH, are not currently available in Manitoba and clinical samples are sent out of province for testing at a cost of $475/case and $495/case, respectively. Furthermore, these 'low-throughput' methods are often limited to screen a single gene with known variations. As a qualitative assay, FISH detects gene fusions but cannot distinguish between the fusion partners and fusion variants. The interpretation of FISH results is subjective and, therefore requires highly trained individuals to read the tests manually.

In the past decade, next generation sequencing has been developed and become the cutting edge platform for molecular testing of multiple cancers. A major merit of NGS is the capacity to

analyze a spectrum ranging from the whole genome or limited to selection of multiple genes specific genes depending on the indication. This enables oncologists to achieve comprehensive knowledge of genetic information of cancer patients and hence to optimize personalized treatment to increase the standard of cancer care. After the first NGS-based cancer sequencing was completed in an acute myelogenous leukemia case in 2008 (Campbell *et al*. 2008), the sequencing of lung cancers were conducted and reported by several research teams (Shah *et al*., 2009; Lee *et al*., 2010; Pleasance *et al*., 2010 and Ju *et al*., 2011). Recently, NGS studies are being focused on clinical cases for the detection and validation of the gene alterations for targeted therapy, such as *ALK* fusion and *EGFR* mutation, in individual cases.

Break-apart FISH assay is the current gold standard test for *ALK* fusion gene detection. As discussed previously, FISH methods provides very limited information on fusion partners and fusion variants. A threshold (usually splitting signals in 15% or more of 50 evaluated tumor nuclei) is required for a positive A*LK* FISH result. Therefore, reading FISH signals can be challenging and subjective, particularly when the break-apart signal patterns are subtle. This may result in variations among individual FISH readers or between different laboratories (Wallander *et al*., 2012). In addition, the FISH assay is relatively expensive and labor intensive.

Considering the limitations of the FISH assay as above, alternative and or complimentary diagnostic tools have been explored, for example NGS and IHC. Attributing to the optimization of antibodies and application of amplification technology, IHC has been developed into a highly sensitive and specific test for *ALK* fusion detection in lung cancers (Mino-Kenudson *et al.*, 2010, Marchetti *et al*., 2016, Savic *et al.*, 2013, Ying *et al.*, 2013, Lantuejoul *et al.*, 2015 and Pekar-

Zlotin *et al*., 2015). Compared to FISH, IHC is cheaper, faster, easy accessible, less labor intensive and more readily available for reading results. The D5F3 antibody based Ventana automated IHC assay has been approved by the US FDA for detecting the *ALK* gene rearrangements. Of note, as a molecular approach, by targeting protein expression, IHC offers only indirect diagnostic information on the presence of gene fusions or amplifications. Like FISH, IHC cannot distinguish fusion sites and fusion partners (Dagogo-Jack and Shaw, 2016).

Targeted sequencing is a promising molecular test for *ALK* rearrangement detection. In 2014, Abel *et al.* employed NGS to analyze a series of FFPE samples from seven *ALK* rearranged cancers (six lung adenocarcinomas and one anaplastic large cell lymphoma), six *KMT2A* rearranged leukemias, and 77 *ALK/KMT2A* rearrangement-negative cancers, as previously tested by FISH. The researchers reported that *ALK* rearrangements can be detected from targeted gene panel-based NGS with sensitivity and specificity equivalent to that of FISH, while NGS provides finer-scale information and increased efficiency for molecular oncology testing, for example identifying translocation partners without specifically targeting any of these partners (Abel *et al*., 2014). Tuononen *et al*. (2012) assessed the *ALK* fusion status for 95 FFPE tumor tissue samples from 87 patients with NSCLC by FISH and real-time RT-PCR, for 57 specimens from 56 patients by targeted resequencing, and for 14 specimens from 14 patients by IHC. The results of NGS correlated significantly with those from FISH, real-time RT-PCR, and IHC. Compared with other methods, targeted sequencing can identify all variations of *ALK* fusions but is time-consuming as ten days were needed to complete the analysis. FISH, IHC and RT-PCR *ALK* assays all can be done in 2 or 3 days (Tuononen *et al*., 2012).

Consistent with the two studies reported above, our study demonstrates a 100% correlation between targeted sequencing and FISH method for *ALK* fusion assessment among 11 cases with lung adenocarcinoma. Moreover, we observed discordance in one case (case number FP13, **Table 8**) that is IHC positive but FISH negative. In favor of the FISH result, the sequencing did not identify any *EML4-ALK* fusion in this single case. The discordant results between *ALK* IHC and *ALK* FISH have been previously reported by multiple research groups (Cabillic *et al*., 2014, Pekar-Zlotin *et al*., 2015, Marchetti *et al*., 2016, Rosoux *et al.*, 2016 and Peled *et al*., 2012). In Cabillic's study, *ALK* fusion mutation were identified in 150 of 3244 NSCLC cases. Interestingly, the consistency of FISH and IHC results presents in only 80 of 150 specimens. The remaining specimens were either FISH positive/IHC negative (36 cases) or FISH negative/IHC positive (19 cases). A primary monoclonal ALK antibody (mAb) from Abcam (clone 5A4; Abcam, Cambridge, United Kingdom) was used for the IHC study reported by Cabillic *et al*. Although the definite mechanism underlying the discordance is not clear, several factors have been proposed, including low expression level of ALK protein in lung cancers (Mino-Kenudson *et al*., 2010), variable properties and stabilities among different chimeric ALK proteins (Bargalet *et al*., 2011), silenced transcription/translation of ALK proteins (Cabillic *et al*. 2014), and *ALK* fusion variants undetectable for FISH (Wallander *et al*., 2012). Even poor technical skills and artifacts may lead to the discordance (Cabillic *et al*., 2014; Ilie *et al*., 2012). In our study, only one case was discordant (IHC positive, but FISH and NGS negative for *ALK* fusion). A possible explanation is that the *ALK* expression may be associated with other forms of gene mutations rather than rearrangement, such as *ALK* amplifications or activating point mutations as reported in neuroblastoma, renal clear cell adenocarcinoma, and myofibroblastic tumors (Ogura *et al.*, 2012; Butrynski *et al.*, 2010; Sukov *et al.*, 2012). The higher false positive rates that contributes

to the high sensitivity but lower specificity of ALK IHC has been accepted by the academic community as a cost-effective screening test for the presence of the gene fusion. In the clinic, IHC discordant cases still receive the anti-ALK therapy and have been reported to have treatment responses. However, in this particular case, the associated clinical data was not retrieved as part of the study.

Recent evidence indicates a significant *ALK* false negative rate for FISH exists. A published retrospective study shows that 4 out of 5 lung adenocarcinoma cases with IHC positive/FISH negative were positive for *ALK* rearrangement based on NGS assay (Pekar-Zlotin *et al*., 2015). More recently, Ali *et al*. reported that 11 of 31 NGS *ALK* positive NSCLC cases were FISH negative. The majority of these NGS-positive/FISH-negative cases were responsive to crizotinib. This implies the sensitivity of FISH is inferior to that of NGS in the assessment of the ALK rearrangement (Ali *et al*., 2016). Volckmar *et al*. applied NGS to facilitate detection of the classic E13-A20 *EML4-ALK* fusion in an *ALK* FISH/IHC inconclusive biopsy of a stage IV lung cancer patient. As proposed by the researchers, the negative FISH and IHC results might be caused by tissue artifacts introduced by cauterization and surgical instruments. However, NGS yielded a clear positive result in the biopsy sample and was confirmed with further molecular testing on the surgical resection specimen. Volckmar *et al.* suggest that NGS is particularly advisable in a case where FISH and IHC *ALK* results are inconclusive or tissue artifacts are suspect (Volckmar *et al*. 2016).

Addition to *ALK* fusion, our NGS study investigated other common gene mutations found in lung cancer including *EGFR, KRAS* and *TP53*. Our data shows a high concordance (8 out of 9

cases, **Table 6** and **Table 7**) of the results for *EGFR* mutation detection between the targeted NGS and RT-PCR methods. Only one case (case number VP7, **Table 6**) was *EGFR* mutation positive revealed by the real-time PCR but not confirmed by the NGS. In a study reported by Tuononen *et al*. (2013), FFPE tumor tissue specimens from 81 NSCLC patients were analyzed for *EGFR*, *KRAS*, and *BRAF* mutations using targeted NGS and real-time PCR. The percentage of concordance for *EGFR* mutation between two approaches is 96.3%. In Tuononen's study, the *EGFR* mutations of 3 cases were detected by real-time PCR but not by NGS, similar to the observed rate in our study. The authors suggested that the NGS method used appears not to be as sensitive as the real-time PCR for lower allelic fraction, in which the sensitivity is set at 1% of allelic frequency. Another potential explanation the authors provided for the discordance is that the hybridization-based enrichment method of the NGS favored the enrichment of wild-type allele, which may cause the chance of sequencing of wild-type allele higher than that of a lower abundance variant allele. The subsequent Sanger sequencing confirmed that all three discrepant cases are negative for *EGFR* mutation and, hence, supports the NGS results. However, the author frankly pointed out that the confirmation of Sanger sequencing is problematic as Sanger sequencing is less sensitive than real-time PCR in gene mutation detection. Indeed, the mechanism underlying the inconsistency between targeted NGS and real-time PCR is not fully understood (Tuononen *et al*. 2013).

In a number of studies recently published (Shao *et al*., 2016, Xu *et al*., 2016 and Gao *et al*., 2016) noted a significant rate of concordance up to 100% were observed between targeted NGS and conventional testing tools including real-time PCR for mutation profiling of *EGFR* in NSCLC. Based on the results acquired, all of those independent research groups tend to agree

that the sensitivity of targeted NGS may be modestly inferior to real-time PCR but this sequencing technology provides more genetic information such as detection of multiple mutations simultaneously and uncovering non-hotspot mutations. Thus, the most feasible explanation for the discordant *EGFR* mutation result in our study is the unequal sensitivity between targeted NGS and real-time PCR. The *EGFR* mutation frequency of this discordant case appears to fall in a "grey zone" which can be detected by real-time PCR but is out of the range of targeted NGS. However, other possibilities for the discordance cannot be definitely ruled out, for example, DNA quality/quantity flaw of FFPE sample achieved and relatively low amount of tumor component of tissue sample. Reexamining the tissue sample and retesting of NGS and real-time PCR can be an option for further investigation. Again, obtaining clinical information on outcome of targeted therapy will be helpful in determining the *EGFR* mutation status for this case in our study. Finally, tumor heterogeneity may also have played a role where the section sampled in the current NGS study did not contain the mutation, whereas the one for RT-PCR did contain the mutation. This can be confirmed by going back and examining the block used for testing carefully.

Of note, in those studies reported above, targeted NGS was mainly used as a validation approach to compare with and confirm with the mutation status of *EGFR* pre-determined by conventional diagnostic tools. In our study, 9 cases with unknown condition of EGFR mutation was first assessed with targeted NGS. 2 out of 9 cases are positive for *EGFR L858R* mutation. Subsequently, the 2 positive cases were confirmed with real-time PCR assay ordered. This supports the value of NGS as an alternative tool for *EGFR* mutation diagnosis in lung cancer.

Besides the identification of *EGFR* mutations, our NGS data demonstrated other common tumor driven gene mutations for lung cancers including 5 cases with *KRAS* mutations (3 cases in codon G12C and 2 cases in codon G12V) and 1 cases harboring *TP53* mutation in codon Y220C. Cancer driven gene mutations of lung cancer are typically mutually excluded. This exclusion is also noticed in our study. No single case of our study harbors two or more of those common oncogenic mutations associated with lung cancer including *ALK*, *EGFR* and *KRAS*. Neither IHC/FISH nor PCR assays as designed are capable of detecting these additional gene mutations in a single run assay. The essence of NGS screening as many as thousands of genes simultaneously makes NGS a unique and strong tool to profile gene mutations in lung cancer and optimized for the current era of personalized medicine. This high-throughput sequencing not only helps to identify known cancer driven genes but also benefits the searching of candidate or novel gene mutations for carcinogenesis and potential target therapy. For example, Ilumina plans to develop a pan-cancer screening protocol to sequence tens of thousands of genomes using circulating tumor DNA (Goodwin *et al*., 2016; https://www.illumina.com/company/news-center/press-releases/press-release-details.html?newsid=2127903).

To the best of our knowledge, our single-blinded study is the first such study in the Province of Manitoba and among the first efforts across Canada to explore the clinical application of NGS to detect therapeutically relevant oncogene mutations in lung cancer. All results acquired in this study are consistent with literature and current evidence. Our data is in favor of the high accuracy, sensitivity and specificity of NGS in *ALK*-fusion and *EGFR* mutation detections in lung cancer. Our study supports that NGS has the potential to become one of the standard initial or confirmation test for molecular genotyping of NSCLC. One minor limitation

of our study is the limited correlation with clinical information such as smoking status and clinical treatment response of most patients. This prevents us from correlating presence of gene mutations to smoking status and compare to the published literature. The other minor limitation is the non-random selection of tumor cases. This results in the reported occurrence rate of oncogene mutations observed in our study not to reflect the mutation frequency of lung cancer in the population. Also the relatively small scale of research subjects makes the statistical analysis a challenge.

To date, the ability to perform lung cancer gene testing in Manitoba is limited. A NGS platform was established at the University of Manitoba in 2012 for research use. Diagnostic Services Manitoba (DSM), the main molecular pathology provider in the province, has established a clinical NGS platform to serve a variety of needs including germline mutation analysis for heritable diseases and somatic mutation detection in cancer. Our study provides support to developing this NGS platform for cancer gene testing in Manitoba, starting with lung cancer.

7. FUTURE DIRECTIONS

In the last decade, rather than being an add-on or complimentary approach, next generation sequencing has been developed and commercialized into a mainstay of genetic tool in biological research, laboratory diagnosis and clinical management, such as detection of clinically relevant oncogene mutations as conducted in our study. The advent of rapid, high-throughput and low-cost sequencing is prompting to bridge genomic testing to clinically practicable measurements, exemplified with optimizing personalized treatment in cancer patients. Thus, the trend to incorporation of NGS to clinical use is increasing enormously.

Scholars, research teams and institutes at the University of Manitoba / CancerCare Manitoba including Dr. Shantanu Banerji and Dr. Gefei Qing are focusing on NGS-based translation medicine particularly in Oncology. For this research study as part of this group, we have successfully employed NGS for the detection of *ALK* and *EGFR* mutations in a group of lung cancer cases, which is the first such study in Manitoba. Meanwhile, the study has explored a practical protocol for local use of NGS in lung cancer management. Our preliminary data supports the routine use or at least optional use of NGS to investigate clinically relevant oncogenes.

As of future directions, a cohort with expanded sample number could be considered for further validation of NGS in analysis of *ALK* and *EGFR* mutations. This can also be extended to other gene mutations with clinical significance such as *ROS-1* and *KRAS*. A close follow-up of clinical outcome of anti-EFGR and anti-ALK therapy in selected patients will be essential to

clinically confirm mutation status revealed by those molecular diagnostic assays including NGS. Determine the lower threshold of *ALK* fusion and *EGFR* mutation detection to optimize the protocol through sensitivity control. Combine methods for *ALK* and *ROS-1* fusion gene detection with NGS-based *EGFR*, *KRAS*, *ERBB2*, and *BRAF* point mutation detection to produce a comprehensive lung cancer test. Besides lung cancer, the significance of NGS platforms in molecular profiling of other malignancies is worth exploring, like breast cancer, leukemia and colorectal tumor. The economics of NGS investment is another interesting topic to look at in the future.

8. REFERENCES

Abel HJ, Al-Kateb H, Cottrell CE, *et al.* Detection of gene rearrangements in targeted clinical next-generation sequencing. *J Mol Diagn.* 2014 Jul;16(4):405-17.

Alberg AJ, Ford JG, Samet JM. Epidemiology of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest* 2007;132:29S-55S.

Ali SM, Hensing T, Schrock AB, *et al.* Comprehensive Genomic Profiling Identifies a Subset of Crizotinib-Responsive ALK-Rearranged Non-Small Cell Lung Cancer Not Detected by Fluorescence In Situ Hybridization. *Oncologist.* 2016 Jun;21(6):762-70.

Bernard WS, and Christopher PW. *World cancer report 2014.* IARC Press. pp. 350-352.

Biesalski HK, Bueno de Mesquita B, Chesson A, *et al.* (1998). "European Consensus Statement on Lung Cancer: risk factors and prevention. Lung Cancer Panel". *CA Cancer J Clin.* 48 (3): 167–176; discussion 164–166.

Butrynski JE, D'Adamo DR, Hornick JL, *et al.* Crizotinib in ALK-rearranged inflammatory myofibroblastic tumor. *N Engl J Med*. 2010 Oct 28;363(18):1727-33.

Cabillic F, Gros A, Dugay F, *et al.* Parallel FISH and immunohistochemical studies of ALK status in 3244 non-small-cell lung cancers reveal major discordances. *J Thorac Oncol.* 2014 Mar;9(3):295-306.

Campbell PJ, Stephens PJ, Pleasance ED, *et al*. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet*. 2008 Jun;40(6):722-9.

Canadian Cancer Statistics 2016. Canadian Cancer Society. https://www.cancer.ca/~/media/cancer.ca/CW/cancer%20information/cancer%20101/Canadian%20cancer%20statistics/Canadian-Cancer-Statistics-2016-EN.pdf?la=en

Ciardiello F, Tortora G. 2008. EGFR antagonists in cancer treatment. *N. Engl. J. Med.* 358:1160–74

Choi H, Mazzone P. Radon and lung cancer: assessing and mitigating the risk. *Cleveland Clinic Journal of Medicine*. 2014 Sept; 81 (9): 567–575.

Choi YL, Takeuchi K, Soda M, *et al*. Identification of novel isoforms of the EML4-ALK transforming gene in non-small cell lung cancer. *Cancer Res*. 2008 Jul 1;68(13):4971-6

Crinò L, Weder W, van Meerbeeck J, Felip E. Early stage and locally advanced (non-metastatic) non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *ESMO Guidelines Working Group Ann Oncol*. 2010 May; 21 Suppl 5:v103-15.

da Cunha Santos G, Shepherd FA, Tsao MS. EGFR mutations and lung cancer. *Annu Rev Pathol*. 2011;6:49-69. Review.

Dagogo-Jack I, Shaw AT. Screening for ALK Rearrangements in Lung Cancer: Time for a New Generation of Diagnostics? *Oncologist*. 2016 Jun;21(6):662-3.

Dearden S, Stevens J, Wu YL, Blowers D. Mutation incidence and coincidence in non small-cell lung cancer: meta-analyses by ethnicity and histology (mutMap). *Ann Oncol*. 2013;24:2371–2376

Ding L, Getz G, Wheeler DA, *et al*. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 2008;455:1069-75.

Dogan S, Shen R, Ang DC, *et al*. Molecular epidemiology of EGFR and KRAS mutations in 3,026 lung adenocarcinomas: higher susceptibility of women to smoking-related KRAS-mutant cancers. *Clin Cancer Res*. 2012;18(22):6169–6177.

Ellis PM, Blais N, Soulieres D, *et al*. A systematic review and Canadian consensus recommendations on the use of biomarkers in the treatment of non-small cell lung cancer. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* 2011;6:1379-91.

Gao SP, Chang Q, Mao N, *et al*. JAK2 inhibition sensitizes resistant EGFR-mutant lung adenocarcinoma to tyrosine kinase inhibitors. *Sci Signal*. 2016 Mar 29;9(421):ra33.

George RE, Sanda T, Hanna M, *et al*. Activating mutations in ALK provide a therapeutic target in neuroblastoma. *Nature*. 2008 Oct 16;455(7215):975-8.

Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 2012;366:883-92.

Gnirke A, Melnikov A, Maguire J, *et al*. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol.* 2009 Feb;27(2):182-9.

Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016 May 17;17(6):333-51

Harismendy O, Schwab RB, Bao L, *et al.* Detection of low prevalence somatic mutations in solid tumors with ultra-deep targeted sequencing. *Genome biology* 2011;12:R124.

Hecht, SS (2012). "Lung carcinogenesis by tobacco smoke". International Journal of Cancer. 131 (12): 2724–2732.

Horn L, Lovly CM, Johnson, DH. Chapter 107: Neoplasms of the lung. In Kasper DL, Hauser SL, Jameson JL, et al.'s *Principles of Internal Medicine* (19th ed.). McGraw-Hill. 2015.

Howington JA, Blum MG, Chang AC, Balekian AA, Murthy SC. Treatment of stage I and II non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines.
Chest. 2013 May;143(5 Suppl):e278S-313S

Hsu CL, Chen KY, Shih JY, *et al*. Advanced non-small cell lung cancer in patients aged 45 years or younger: outcomes and prognostic factors. *BMC Cancer*. 2012 Jun 13;12:241.

Ilie M., Long E, Butori, C *et al.* ALK-gene rearrangement: a comparative analysis on circulating tumour cells and tumour tissue from patients with lung adenocarcinoma
Annals of Oncology 23: 2907–2913, 2012

Imielinski M, Berger AH, Hammerman PS, *et al*. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 2012;150:1107-20.

Ju YS, Lee WC, Shin JY, *et al*. A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. *Genome Res.* 2012 Mar;22(3):436-45.

Kim H, Yoo SB, Choe JY, *et al*. Detection of ALK gene rearrangement in non-small cell lung cancer: a comparison of fluorescence in situ hybridization and chromogenic in situ hybridization with correlation of ALK protein expression. *J Thorac Oncol*. 2011 Aug;6(8):1359-66.

Kranenburg O. The KRAS oncogene: past, present, and future. *Biochim Biophys Acta.* 2005;1756(2):81–82

Kumar V, Abbas AK, Aster JC. Chapter 5 . *Robbins Basic Pathology* (9th ed.). 2013, Elsevier Saunders. p. 199.

Kutok JL, Aster JC. Molecular biology of anaplastic lymphoma kinase-positive anaplastic large-cell lymphoma. *J Clin Oncol.* 2002 Sep 1;20(17):3691-702. Review.

Kwak EL, Bang YJ, Camidge DR, *et al*. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med* 2010;363:1693-703.

Ladanyi M, Pao W. Lung adenocarcinoma: guiding EGFR-targeted therapy and beyond. *Mod Pathol*. 2008 May;21 Suppl 2:S16-22. Review.

Lander ES, Linton LM, Birren B et al. Initial sequencing and analysis of the human genome. *Nature*. 2001 Feb 15;409(6822):860-921.

Lantuejoul S, Rouquette I, Blons H, *et al.* French multicentric validation of ALK rearrangement diagnostic in 547 lung adenocarcinomas. *Eur Respir J.* 2015 Jul;46(1):207-18.

Lee W, Jiang Z, Liu J, Haverty PM, *et al.* The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature.* 2010 May 27;465(7297):473-7.

Ley TJ, Mardis ER, Ding L, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 2008;456:66-72.

Lindeman NI, Cagle PT, Beasley MB, *et al*Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. *J Thorac Oncol.* 2013 Jul;8(7):823-59

Lynch TJ, Bell DW, Sordella R, et al. 2004. Activating mutations in the epidermal growth factor receptor underlying responsiveness of nonsmall-cell lung cancer to gefitinib. *N. Engl. J. Med.* 350:2129–39 8, 9.

Marchetti A, Di Lorito A, Pace MV, *et al.* ALK Protein Analysis by IHC Staining after Recent Regulatory Changes: A Comparison of Two Widely Used Approaches, Revision of the Literature, and a New Testing Algorithm. *J Thorac Oncol*. 2016 Apr;11(4):487-95.

Merck Manual Professional Edition, Online edition. Lung Carcinoma: Tumors of the Lungs. Retrieved 15 August 2007

Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. Nature reviews Genetics 2010;11:685-96.

Metzker ML. Sequencing technologies - the next generation. Nature reviews Genetics 2010;11:31-46.

Mino-Kenudson M, Chirieac LR, Law K, *et al.* A novel, highly sensitive antibody allows for the routine detection of ALK-rearranged lung adenocarcinomas by standard immunohistochemistry. *Clin Cancer Res.* 2010 Mar 1;16(5):1561-71.

Mitchell RS, Kumar V, Abbas AK, Fausto N. Chapter 13, box on morphology of adenocarcinoma. *Robbins Basic Pathology* (8th ed.). Philadelphia: Saunders.

Mitsudomi T, Kosaka T, *et al*. Mutations of the epidermal growth factor receptor gene predict prolonged survival after gefitinib treatment in patients with non-small-cell lung cancer with postoperative recurrence. *J Clin Oncol.* 2005;23:2513–5220

Mok TS, Wu YL, Thongprasert S, *et al*. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *The New England journal of medicine* 2009;361:947-57.

Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc*. 2008 May;83(5):584-94. Review.

Mosesson Y, Yarden Y. 2004. Oncogenic growth factor receptors: implications for signal transduction therapy. *Semin. Cancer Biol.* 14:262–70. Review.

Mossé YP, Laudenslager M, Longo L, *et al*. Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature*. 2008 Oct 16;455(7215):930-5.

Ogura T, Hiyama E, Kamei N, Kamimatsuse A, Ueda Y, Ogura K. Clinical feature of anaplastic lymphoma kinase-mutated neuroblastoma. *J Pediatr Surg.* 2012 Oct;47(10):1789-96.

Okimoto RA, Bivona TG. Recent advances in personalized lung cancer medicine. *Per Med.* 2014;11(3):309-321

Paez JG, Janne PA, Lee JC, *et al*. 2004. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science.* 2007 June 4; 304:1497–500

Pao W, Miller VA, Politi KA, *et al*. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med*. 2005 Mar;2(3):e73.

Paoletti L, Pastis NJ, Denlinger CE, Silvestri GA. A decade of advances in treatment of early-stage lung cancer. *Clin Chest Med*. 2011 Dec;32(4):827-38. Review.

Peifer M, Fernández-Cuesta L, Sos ML et al. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nat Genet.* 2012 Oct;44(10):1104-10.

Pekar-Zlotin M, Hirsch FR, Soussan-Gutman L, *et al.* Fluorescence in situ hybridization, immunohistochemistry, and next-generation sequencing for detection of EML4-ALK rearrangement in lung cancer. *Oncologist*. 2015 Mar;20(3):316-22.

Peled N, Palmer G, Hirsch FR, *et al.* Next-generation sequencing identifies and immunohistochemistry confirms a novel crizotinib-sensitive ALK rearrangement in a patient with metastatic non-small-cell lung cancer. *J Thorac Oncol*. 2012 Sep;7(9):e14-6.

Pleasance ED, Stephens PJ, O'Meara S, *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 2010 Jan 14;463(7278):184-90.

Qiu T, Guo H, Zhao H, Wang L, Zhang Z. Next-generation sequencing for molecular diagnosis of lung adenocarcinoma specimens obtained by fine needle aspiration cytology. *Sci Rep*. 2015 Jun 11;5:11317.

Rosell R, Carcereny E, Gervais R, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol* 2012;13:239-46.

Sakurada A, Shepherd FA, Tsao MS. Epidermal growth factor receptor tyrosine kinase inhibitors in lung cancer: impact of primary or secondary mutations. *Clin Lung Cancer*. 2006 May;7 Suppl 4:S138-44. Review.

Rosoux A, Pauwels P, Duplaquet F *et al.* Effectiveness of Crizotinib in a Patient With ALK IHC-positive/FISH-negative Metastatic Lung Adenocarcinoma. *Lung Cancer* 98, 118-121. 2016 Jun 03

Savic S, Bode B, Diebold J, *et al.* Detection of ALK-positive non-small-cell lung cancers on cytological specimens: high accuracy of immunocytochemistry with the 5A4 clone. *J Thorac Oncol*. 2013 Aug;8(8):1004-11.

Shah SP, Morin RD, Khattra J, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* 2009;461:809-13.

Shao D, Lin Y, Liu J, Wan L, *et al.* A targeted next-generation sequencing method for identifying clinically relevant mutation profiles in lung adenocarcinoma. *Sci Rep.* 2016 Mar 3;6:22338.

Shepherd FA, Rodrigues Pereira J, Ciuleanu T, *et al.* Erlotinib in previously treated non-small-cell lung cancer. *N Engl J Med*. 2005 Jul 14;353(2):123-32.

Sherwin SA, Minna JD, Gazdar AF, Todaro GJ. Expression of epidermal and nerve growth factor receptors and soft agar growth factor production by human lung cancer cells. *Cancer Res.* 1981 Sep;41(9 Pt 1):3538–3542.

Sholl LM. Biomarkers in lung adenocarcinoma: a decade of progress. *Arch Pathol Lab Med.* 2015 Apr;139(4):469-80. Review.

Smith LM, Fung S, Hunkapiller MW, Hunkapiller TJ, Hood LE. The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis. *Nucleic Acids Res*. 1985 Apr 11;13(7):2399-412.

Socinski MA. Clinical issues in the management of non-small-cell lung cancer and the role of platinum-based therapy. *Clin Lung Cancer*. 2004 Mar;5(5):274-89. Review.

Soda M, Choi YL, Enomoto M, *et al.* Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 2007;448:561-6.

Soh J, Okumura N, Lockwood WW, *et al*. Oncogene mutations, copy number gains and mutant allele specific imbalance (MASI) frequently occur together in tumor cells. *PLoS One*. 2009 Oct 14;4(10):e7464

Sörenson S, Glimelius B, Nygren P; SBU-group. Swedish Council of Technology Assessment in Health Care. A systematic overview of chemotherapy effects in non-small cell lung cancer. *Acta Oncol.* 2001;40(2-3):327-39. Review.

Sukov WR, Hodge JC, Lohse CM, *et al.* ALK alterations in adult renal cell carcinoma: frequency, clinicopathologic features and outcome in a large series of consecutively treated patients. *Mod Pathol.* 2012 Nov;25(11):1516-25.

Takeuchi K, Choi YL, Soda M, *et al*. Multiplex reverse transcription-PCR screening for EML4-ALK fusion transcripts. *Clin Cancer Res*. 2008 Oct 15;14(20):6618-24

The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014; 511:543–50

Travis WD, Brambilla E, Müller-Hermelink HK, Harris CC. Pathology & Genetics Tumours of the Lung, Pleura. *World Health Organization Classification of Tumours* (2004) (http://www.iarc.fr/en/publications/pdfs-online/pat-gen/bb10/BB10.pdf)

Tuononen K, Kaur Sarhadi V, Wirtanen A, *et al.* Targeted Resequencing Reveals ALK Fusions in Non-Small Cell Lung Carcinomas Detected by FISH, Immunohistochemistry, and Real-Time RT-PCR: A Comparison of Four Methods. *BioMed Research International,* Volume 2013, Article ID 757490, 9 pages

Venter JC, Adams MD, Myers EW, *et al*. The sequence of the human genome. *Science*. 2001 Feb 16;291(5507):1304-51.

Volckmar AL, Endris V, Bozorgmehr F, *et al.* Next-generation sequencing facilitates detection of the classic E13-A20 EML4-ALK fusion in an ALK-FISH/IHC inconclusive biopsy of a stage IV lung cancer patient: a case report. *Diagn Pathol*. 2016 Nov 18;11(1):133.

Wallander ML, Geiersbach KB, Tripp SR, Layfield LJ.Comparison of reverse transcription-polymerase chain reaction, immunohistochemistry, and fluorescence in situ hybridization methodologies for detection of echinoderm microtubule-associated proteinlike 4-anaplastic lymphoma kinase fusion-positive non-small cell lung carcinoma: implications for optimal clinical testing. *Arch Pathol Lab Med*. 2012 Jul;136(7):796-803.

Wanders R, Steevens J, Botterweck A, *et al*. Treatment with curative intent of stage III non-small cell lung cancer patients of 75 years: a prospective population-based study. *Eur J Cancer*. 2011 Dec;47(18):2691-7.

Weir BA, Woo MS, Getz G, *et al*. Characterizing the cancer genome in lung adenocarcinoma. *Nature* 2007 Dec 6;450(7171):893-8.

Wheeler DA, Srinivasan M, Egholm M, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 2008;452:872-6.

Woodburn JR. The epidermal growth factor receptor and its inhibition in cancer therapy. *Pharmacol. Ther*. 1999; 82:241–50

World Health Organization. World Cancer Report 2014. 2014. pp. Chapter 5.1.

World Health Organization. World Cancer Report 2014. 2014. pp. Chapter 1.1.

Xu J, Jin B, Chu T, *et al.* EGFR tyrosine kinase inhibitor (TKI) in patients with advanced non-small cell lung cancer (NSCLC) harboring uncommon EGFR mutations: A real-world study in China. *Lung Cancer*. 2016 Jun;96:87-92.

Ying J, Guo L, Qiu T, *et al*. Diagnostic value of a novel fully automated immunochemistry assay for detection of ALK rearrangement in primary lung adenocarcinoma. *Ann Oncol* (2013) 24 (10): 2589-2593.

Zhang J, Ou JX, Bai CX. Tobacco smoking in China: prevalence, disease burden, challenges and future strategies. *Respirology*. 2011 Nov;16(8):1165-72.