THE MOLECULAR PROFILING OF NON-SMALL CELL LUNG CANCER

by

Marika Forsythe

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
December 2016

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Personalized therapies against lung cancer provide more efficient treatment options than conventional therapies, as they target the genetic makeup. Current treatment options are available against EGFR or ALK genes mutations. 667 cases of surgically resected lung cancer underwent molecular analysis for six possible gene mutations, including EGFR, KRAS, BRAF, PIK3CA, HER2 and ALK. The samples included different types of non-small-cell lung cancer. The highest mutation frequency was the KRAS mutation (24.4%) while the lowest was the HER2 mutation (0%). The significant correlations observed were: against age, where fewer older patients exhibited BRAF mutations; against gender, where females exhibited more EGFR mutations; against cell type, where adenocarcinoma exhibited more KRAS and EGFR mutations; against vascular invasion, where positive individuals had fewer EGFR mutations; against smoking history, where non-smokers exhibited more EGFR mutations. This study provides significant results for better understanding the gene mutation status and their significance in lung cancer patients.

LIST OF ABBREVIATIONS USED

STAT	Signal transducer and activator of transcription
ATF	Activating transcript factors
ATP	Adenosine triphosphate
SH2	Src Homology 2
TKI	Tyrosine kinase inhibitors
BCL2	B-cell lymphoma 2
TRAIL	TNF-related apoptosis-inducing ligand
ER	Endoplasmic reticulum
NF-kB	Nuclear factor kappa-light-chain-enhancer of activated B cells
JNK	c-Jun N-terminal kinase
AP-1	Activator protein 1
RAS	Rat sarcoma
NSCLC	Non-small-cell lung cancer
Trp53	Transformation-related protein 53
RB1	Retinoblastoma 1
ALK	Anaplastic lymphoma kinase
EGFR	Epidermal growth factor receptor
KRAS	V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
PIK3CA	Phosphoinositide-3-kinase, catalytic, alpha polypeptide
BRAF	v-Raf murine sarcoma viral oncogene homolog B1
HER2	Human epidermal growth factor receptor 2
ErbB	Avian erythroblastosis oncogene
RAF	proto-oncogene serine/threonine-protein kinase
MEK	Mitogen-activated protein kinase kinase
ERK	Extracellular signal-regulated kinase
MAPK	Mitogen-activated protein kinase
PI3K	Phosphoinositide 3-kinase
AKT	Protein Kinase B
mTOR	Mechanistic target of rapamycin
JAK	Janus kinase
EML4	Echinoderm microtubule-associated protein-like 4
TFG	TRK-Fused Gene
KIF5	Kinesin superfamily protein 5
PLX4032	Vemurafenib
PCR	Polymerase chain reaction
FFPE	Formalin-fixed, paraffin-embedded
H&E	Haematoxylin and eosin
UV	Ultraviolet
SNP	Single nucleotide polymorphisms
dNTP	Nucleoside triphosphate containing deoxyribose
SAP	Shrimp alkaline phosphatase
IHC	Immunohistochemical assay
FISH	Fluorescent in situ hybridization

SSC	Standard saline citrate
AD	Adenocarcinoma
SQ	Squamous cell carcinoma
LCC	Large cell carcinoma
PLE	Pleomorphic carcinoma
N1	Ipsilateral peribronchial, hilar and intrapulmonary nodes
N2	Ipsilateral mediastinal and subcarinal nodes
NS	Non-significant
OSCC	Oral squamous cell carcinomas
HNSCC	Head and neck squamous cell carcinoma
HPF	Horizon Patient Folder

CHAPTER 1: INTRODUCTION

The term ‘cancer’ is used to describe a disease that causes a cell to proliferate at an abnormal and continual rate without control. These cells have the ability to invade other tissues. Rapid cell division creates a mass of abnormal cells or a ‘tumour’. The tumour can be found in the tissue of origin or it can spread to and invade other parts of the body. The tumour can spread either through invasion or metastasis. Invasion involves direct penetration of the neighbouring tissues by the cancer cells, requiring little migration. Spreading by metastasis means the cancer cells have invaded the lymphatic and/or blood vessels, allowing them to circulate in the body to reach tissues elsewhere. The type of tumours that spread using these methods are called ‘malignant’. They disrupt tissue and organ functions needed for survival[1]. This newly formed tumour is called a metastatic tumour, and it is made from the same cells as the primary tumour. The primary tumour grows where the cancer originated and depicts what type of cancer both tumour types are. For example, if the primary tumour originated in the lung and the metastatic tumour grew in the liver they would both be considered lung cancer due to the origin. Both tumour types share common molecular features, such as specific protein expressions and chromosomal changes[2,3]. Tumours can also be classified as ‘benign’, although they would not be considered cancerous. These types of tumours do not grow continually and spread, however they can still be dangerous as they can press against a vital organ, blood vessel or nerve. If a benign tumour requires treatment, the method employed is generally surgical[4].

There is a large variety of cancer types, all of which are caused by some form of genetic change over time, be it alterations in tumour-suppressor genes, oncogenes or

microRNA genes[5]. While the vast majority of these genetic mutations occur in somatic cells of the body, alternations in the germ-line can cause a heritable cancer within a family. Typically, multiple gene alterations are observed before a tumour arises, implying that a single mutation is not typically sufficient for cancer development[6]. The exact cause of these genetic alterations has never been fully understood, but changes can be triggered by chemicals, radiation or viruses[7].

Tumour suppressor genes code for proteins that inhibit cell proliferation and survival. Therefore, if they are inactivated these regulatory proteins will not be produced in the cells, which will lead to proliferation and tumour development. A common example is the p53 gene, which is mutated in roughly 50% of cancers[1]. The protein it produces functions by binding to DNA, stimulating the production of another protein called p21 which interacts with a cell division-stimulating protein called cdk2. This effectively prevents division from continuing by acting as a 'stop signal'. When the p53 gene is mutated it cannot properly bind to DNA, which will prevent this division process from occurring[8]. This gene is the most frequently observed mutated tumour suppressor in cancers, occurring in 50% of non-small-cell lung cancers, 70% of small-cell lung cancers, and 50% of skin cancers. Its inactivation appears to be essential for the formation of most cancers[9,10].

Almost all oncogenes are derived from proto-oncogenes, which are normal cellular genes involved in growth-controlling pathways. A mutation in a proto-oncogene generally produces an oncogene with a 'gain-of-function'[11]. An example is the RAS gene, which when mutated produces a significant growth-promoting signal. The signal would be continual, resulting in uncontrolled proliferation[12]. Members of the RAS gene family are

the most frequently observed oncogenes in human tumours. They are linked to approximately 20% of human tumours, including 50% of colon cancer and 25% of lung carcinoma[1].

Other types of oncogenes with various cellular functions include the following: overproduction of growth factor receptors on the cell surface, leading to a stronger response to growth factors; activation of the telomerase protein (normally inactive in adults) which prevents the shortening of telomeres and promotes cell growth; creating mutated transcription factors which affects messenger RNA production; production of mutated tyrosine kinases, which results in continual protein phosphorylation and a cascade effect towards cell division and growth. Oncogene amplification is another common occurrence in tumour cells, having a significant role in tumour progression and quicker growth[1,13]. These activated oncogenes can be classified into six large groups, namely transcription factors, chromatin remoldelers, growth factors, growth factor receptors, signal transducers, and apoptosis regulators[14,15,16,17,18,19].

Transcription factors function by interacting with other proteins, potentially increasing or decreasing the expression of genes involved in cell division. The transcription factor genes can be activated by chromosomal translocations which produce fused proteins. These fused proteins can cause the constitutional activation of the transcription factor gene, resulting in continuous expression of genes that affect regulation of cell division[14]. A large number of oncogenes and tumour suppressor genes code for transcription factors. Many of these are inactive under normal conditions, making them ideal target points for developing cancer therapies. There have been three major transcription factor families identified, which are potential targets for further studies due

to their importance in human cancers: the NF-kappaB and AP-1 families, the STAT families and steroid receptors. Additional families that play important roles include the Ets factors, the ATF family, and basic helix-loop-helix transcription factors to name a few[20].

Alteration in the compaction of chromatin can have a significant role in either the expression or repression of genes[15]. This remodeling of chromatin is done by two types of enzymes: ATP-dependent enzymes, which are involved in moving the location of the nucleosomes, and enzymes that alter the N-terminal tails of histones such as histone acetyltransferases, deacetylases, methyltransferases, and kinases[21]. The structure of chromatin and its ability to be transcribed are determined by the interactions of nucleosomes and chromatin-associated proteins, which are further controlled by alterations of the core histones[1]. These alterations in structure caused by the chromatin remodeling proteins determine the future of the cell by exposing or hiding different genes from the transcription process, therefore affecting cell survival and/or proliferation rate[21].

Overexpression of growth factor genes can lead to tumour formation, by causing continuous cell division. They are typically protein or steroid hormones that act by signalling between cells, binding to receptors on the surface of cells and promoting differentiation and division[16,18,22]. Growth factors are important for maintaining homeostasis among cells; an overexpression results in steps necessary for tumour progression. These steps include clonal expansion, invasion across tissue barriers, angiogenesis, and colonization of far-off niches. They are also frequently involved with the development of resistance to cancer treatments, therefore an understanding of the mechanism of action for growth factors may lead to the development of new cancer therapies[1].

Constitutive activation of growth factor receptors will also have the same cellular effect. An example found in many tumours involves mutated EGFR, a transmembrane protein with tyrosine kinase activity (discussed further below). The receptor becomes continually active in the absence of ligand binding due to the deletion of the ligand-binding domain. When activated tyrosine is phosphorylated in the intracellular region of the receptor, forming interaction sites for proteins containing certain binding domains, such as the SH2 domain. Several signalling pathways are regulated through these interactions. Some of the proteins involved also have domains with enzymatic activity or act as activate receptors linked to downstream targets[16,17,23]. The presence of a mutation in receptors such as EGFR have been shown to aid in predicting whether non-small cell lung cancer will respond to tyrosine kinase inhibitors (TKIs); EGFR mutations appear in 15-20% of this type of cancer[22].

Two main groups of signal-transduction pathway components controlled by several oncogenes are non-receptor protein kinases and guanosine-triphosphate-binding proteins. Non-receptor protein kinases can be further split into tyrosine kinases and serine/threonine kinases. If proteins involved in signal transduction exhibit activating mutations they are considered oncogenic[17]. STAT-3 is an example of a signal transducer, causing the activation of transcription-3. It has been associated with inflammation, cellular transformation, survival, proliferation, invasion angiogenesis, and cancer metastasis, mediating these effects through collaboration with other transcription factors. It has been observed as continually active in tumour cells but not normal cells, and is activated by various carcinogens, radiation, viruses, growth factors, oncogenes, and inflammatory

cytokines. Due to STAT-3's important role in cancer development, determining how to inhibit its function is promising for the production of new cancer therapies[25].

There are two main pathways that lead to apoptosis of a cell: the stress pathway and the death-receptor pathway. The former is activated by proteins containing the BCL2 homology 3 domain, which regulates apoptosis by inactivating other proteins normally inhibiting apoptosis; this further activates caspases which in turn leads to apoptosis. The latter involves the binding of Fas ligand, TRAIL, and tumour necrosis factor α to their matching receptors on the cell surface[18,26]. While caspases (proteolytic enzymes) act as main effectors of apoptosis, the process involved in activating this system is less known. There have been two pathways identified upstream of the caspase cascade, one involves death receptors which activate caspase-8, and the other involves mitochondria release of apoptogenic factors which trigger caspase-9[25]. The stressed endoplasmic reticulum (ER) also contributes to apoptosis via the unfolded protein response pathway, producing ER chaperones, and the ER overload response pathway, producing cytokines by NF-kB. There are many other stress-inducing molecules (i.e. p53, JNK, AP-1, etc.), as well as participants in the sphingomyelin pathway that have a significant effect in the process of apoptosis[27].

Three significant mechanisms can cause an alteration in an oncogene or an increase of its expression: chromosomal rearrangement, mutation or gene amplification. Chromosomal rearrangements, inversions and translocations, are commonly observed in cancer cells and have the potential to increase the transcription of the oncogene[28]. Translocations result in the exchange of genetic information between two or more chromosomes. This occurs when a double stranded DNA break occurs in more than one

chromosome, and the broken arms reattach in a reciprocal position. Inversions result from one chromosome being broken in two places and the excised piece reinserting in an inverted position[29]. In certain cancers, fusion genes are created that have a very active promoter fused with another that carries the oncogenic activity[28]. Due to a number of rearrangements having their changes in gene structure and function defined, some general assumptions can be applied to all chromosomal rearrangements in human cancers[30].

Mutations affect the resulting protein produced by the oncogene, by changing its structure and transformation capacity. An example of this is the RAS oncogenes, which become permanently active when there is a mutation in codon 12, 13 or 61; this results in signals being continually transduced, which has an effect on cell proliferation[28]. The signal from the RAS oncogenes pass from protein-to-protein through many different pathways, having effects on a variety of functions including lipid metabolism, DNA synthesis, and cytoskeletal organization. There is optimism for potential cancer treatments for cases expressing a mutation in the RAS oncogene. Attempts are being made to modify RAS using an enzyme, farnesyltransferase, which if inhibited could block the maturation of RAS and eliminate its effect on cell proliferation[12].

While the mechanisms described above are found to occur at tumour initiation or during tumour development, gene amplification is a change that is more likely to occur during tumour progression[28]. Possible causes for gene amplification include chromosomal fragile sites, defects in DNA replication or telomere dysfunction. In terms of clinical significance, amplification is a mechanism of acquired drug resistance. It is also useful in establishing a diagnosis and predicting prognoses[31].

Tumours generally originate from a single cell, however the resulting tissue can display a vast variety of a morphological and physiological characteristics. These can include differing cell receptors, proliferation ability and motility[32]. The differing features effect the survivability of the cells, making certain types better suited to grow and pass on their genes. This creates a group of varying subtypes within the tumour, all with different genetic anomalies and oncogenic driver mutations that establish clinical sensitivities[33]. This tumour heterogeneity influences the effectiveness of therapy treatments. The continual evolution and genetic variance expressed among the different tumour cells allows some to adapt and become resistant to treatment, even though other types might still be susceptible[34]. These new therapy-resistant tumour cells can then metastasize, developing new characteristics and drug resistance. A major reason for therapy failure among cancer patients is drug resistance, increasing the necessity to develop newer therapies (such as multitargeting treatments) to combat these ever-adapting cancer cells[35,36,37]. Many of these mutation types and mechanisms have been associated with the development of lung cancer.

Lung cancer is one of the most frequent causes of cancer-related deaths among men and women in Canada, the United States and many other countries worldwide. One in eight cancer occurrences is lung cancer which leads to the death of 1.1 million people each year with only a 15% five year survival rate[38,39]. 90-95% of the tumours are carcinomas, 5% are carcinoids, and 2-5% are mesenchymal or other miscellaneous neoplasms. Lung carcinomas can be classified into two main types; small cell lung cancer (SCLC) which make up for 15% of cases and non-small-cell lung cancer (NSCLC) which account for 85% of cases. Small cell carcinoma, which is generally found in 14% of males

and 18% of females, is a very malignant epithelial tumour and more often diagnosed after having metastasized[40]. Specific features common to this tumour type include scant cytoplasm, missing or indistinct nucleoli, granular nuclear chromatin which resembles a pattern of salt and pepper, and nuclear molding, i.e. a difficulty in separating adjacent cells due to poorly defined borders[41] (Figure 1). Only 5% of diagnosed patients have a survival rate above 5 years. While mutations in the tumour suppressor genes Trp53 and RB1 have been reported in this cancer type, there is limited knowledge about the causative genetic alterations involved in small cell carcinomas which would help the development of new cancer therapies[42].

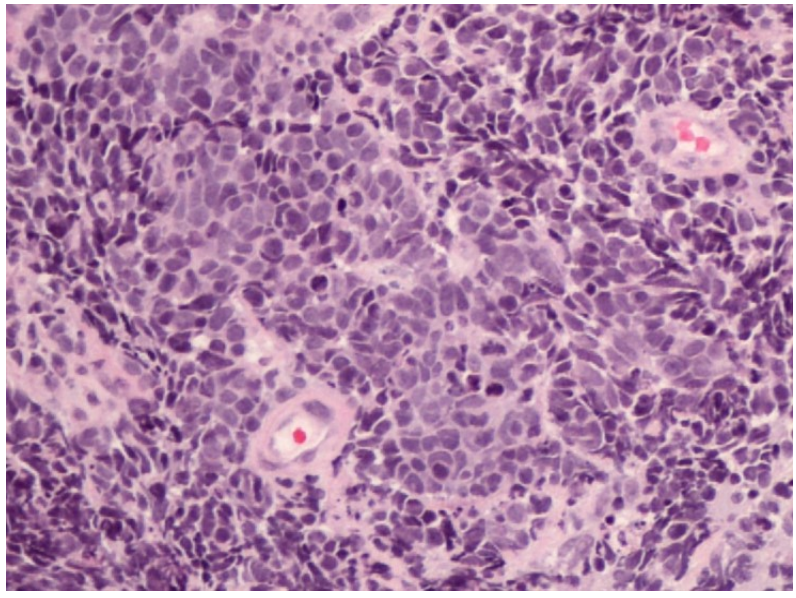

Figure 1. Small cell carcinoma.

The three major types of NSCLC include adenocarcinoma (37% of males, 47% of females), squamous cell carcinoma (32% of males, 25% of females), and large cell carcinoma (18% of males, 10% of females)[42]. Non-small-cell lung cancer (NSCLC) traditionally refers to all histological types of lung carcinomas excluding small cell carcinoma.

Adenocarcinoma is a malignant epithelial tumour, expressing glandular differentiation or mucin production by the tumour cells in either lepidic, acinar, papillary, micropapillary, or solid patterns to name a few[41] (Figure 2). Pure lepidic pattern in a tumour is now termed adenocarcinoma-in-situ. Adenocarcinoma is the most common type of lung carcinomas, and the most diverse in histological patterns. Proper diagnosis of adenocarcinoma is sometimes difficult among small biopsies in the absence of the physical structure[43,44]. This type of carcinoma is known to metastasize earlier than others and spreads far throughout the body. The mutation most frequent in adenocarcinoma is KRAS, while EGFR mutations and amplifications are commonly found as well[41].

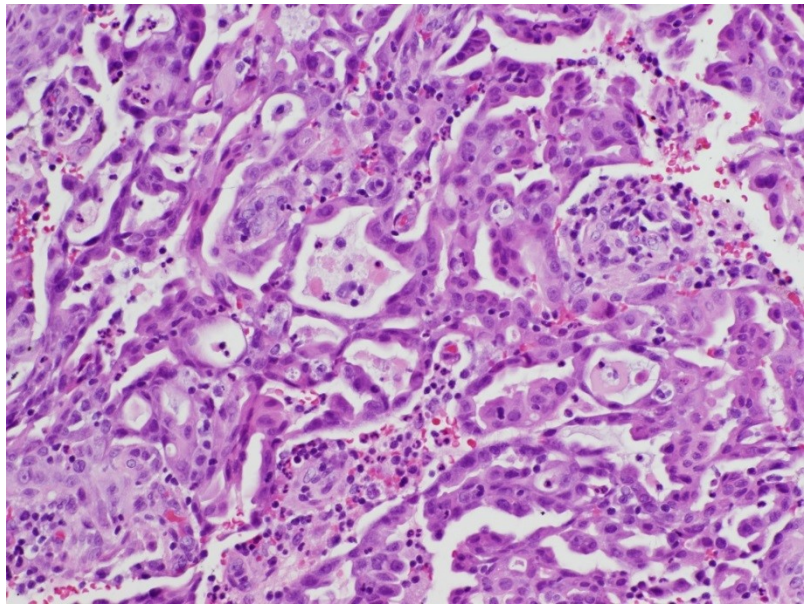

Figure 2: Adenocarcinoma.

Squamous cell carcinoma is defined by the presence of keratinization and/or intercellular bridges, where keratinization can appear in the form of pearls or individual cells containing dense eosinophilic cytoplasm[41] (Figure 3). Squamous cell lung cancer kills approximately 400,000 people per year worldwide. Squamous cell carcinoma has no targeted therapies directed against its genetic alterations[45]. Approximately 80% of

squamous cell tumours exhibit overexpression of EGFR (although rarely a mutation of the gene), while around 30% of these tumours overexpress the HER2 gene[41].

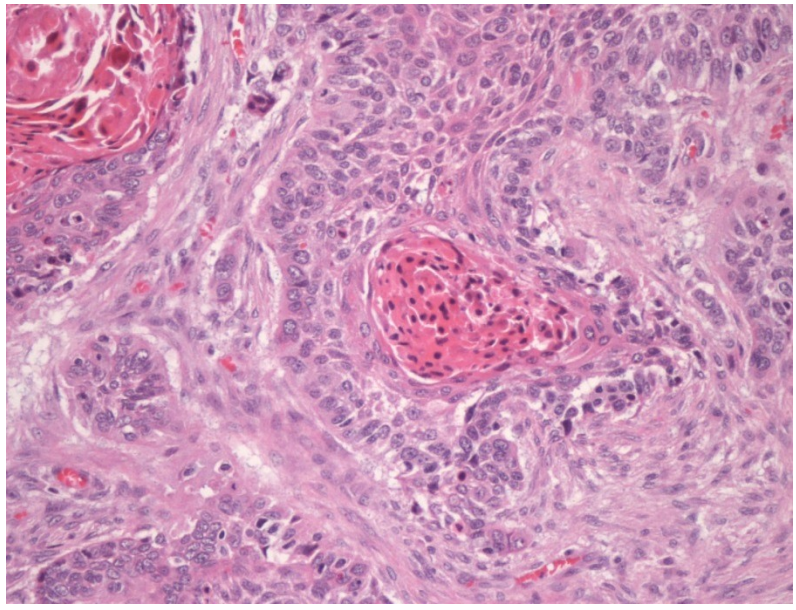

Figure 3: Squamous cell carcinoma.

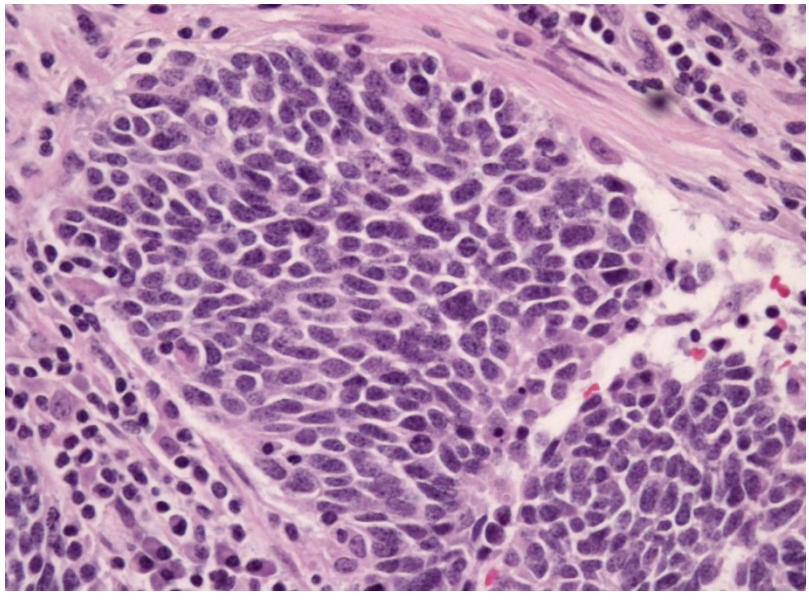

Figure 4: Large cell carcinoma.

Large cell carcinoma is an undifferentiated malignant epithelial tumour, which differs from the other types by the missing glandular and squamous differentiation, as well

as lacking certain cellular characteristics common to small cell carcinomas. The common features of large cell carcinomas are an average amount of cytoplasm with a large nuclei and fairly noticeable nucleoli[41] (Figure 4).

As described above, lung cancer is a group of different cell types dividing involuntarily with different underlying causative mechanisms. This makes the disease difficult to treat. Previous treatment options for advanced stage lung cancer were limited to chemotherapy and radiation which did not offer much 'personalization' based on the clinical data of the patient or pathology of the disease itself, and showed a response rate of around 20%-30%[46]. Surgical resection could be attempted at early stages of the disease, however other treatments such as adjuvant chemotherapy would often be required, either immediately following resection or in later stages of the disease when surgery is no longer an option[47]. More recently a personalized approach is possible, where the genotype of a patient's cancer cells is used to guide therapies. For example, some cancers are closely linked to a specific active kinase and respond better to drugs that inhibit these kinases than to chemotherapy. The response rate is much higher, reaching 75% and patients experience improved quality of life[46].

There are many different gene mutations associated with lung cancer, in particular within adenocarcinoma. Mutations can either be initiators, which cause the cancer to develop by altering key regulatory pathways in the cell, or progressions, which are mutations produced from continual genetic instability, leading to metastasis[47]. Roughly 64% of adenocarcinoma cases contain a driver mutation. The most common driver mutations are found in ALK (anaplastic lymphoma kinase; 6%), EGFR (epidermal growth factor receptor; 23%), KRAS (V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog;

25%), PIK3CA (phosphoinositide-3-kinase, catalytic, alpha polypeptide; 3%), BRAF (v-Raf murine sarcoma viral oncogene homolog B1; 3%), and HER2 (human epidermal growth factor receptor 2; 1%)[46,47]. All of these genes are involved in signalling pathways. Cancers can develop by simple point mutations (ex. KRAS or BRAF), insertions or deletions (ex. EGFR), or inversions/translocations (ex. ALK) whose consequences ultimately result in continuous cell proliferation. These genes are connected by signalling pathways, and therefore a mutation in any produces a similar result: 1) stimulation of the associated protein-tyrosine kinase, causing overproduction of growth factor, 2) continuous DNA synthesis and cell proliferation, and 3) the inhibition of apoptosis.

The cell surface receptors, EGFR and HER2, are part of the ErbB family of receptor tyrosine kinases and serve as mediators for cell signaling within the cell due to their location at the beginning of the pathway[48,49]. The associated tyrosine kinase linked to EGFR causes the biochemical responses initiated by the receptor[50,51]. KRAS and PIK3CA genes are members of other significant pathways, including the RAS-RAF-MEK-ERK and PI3K-AKT-mTOR pathways[52] (Figure 5).

These pathways operate by a cascade of phosphorylation of numerous proteins involved in the regulation of cell growth[53]. They are linked to the EGFR pathway, leading the signaling towards cell proliferation and survival[50,54]. The RAS-RAF-MEK-ERK and PI3K-AKT-mTOR pathways, as well as the surface receptors, serve as ideal targets for cancer treatments due to their effects on cell survival, by either genetic alterations in upstream molecules or genes within the pathway themselves[55]. For instance, a mutation within the EGFR surface receptor in certain domains could invoke dimerization which leads to conformational change of the receptor, exposing the dimerization arm and creating

a dimer interface. This increases tyrosine kinase activity and phosphorylation, leading to RAS activation which then further initiates kinase activities down the pathways[56,57]

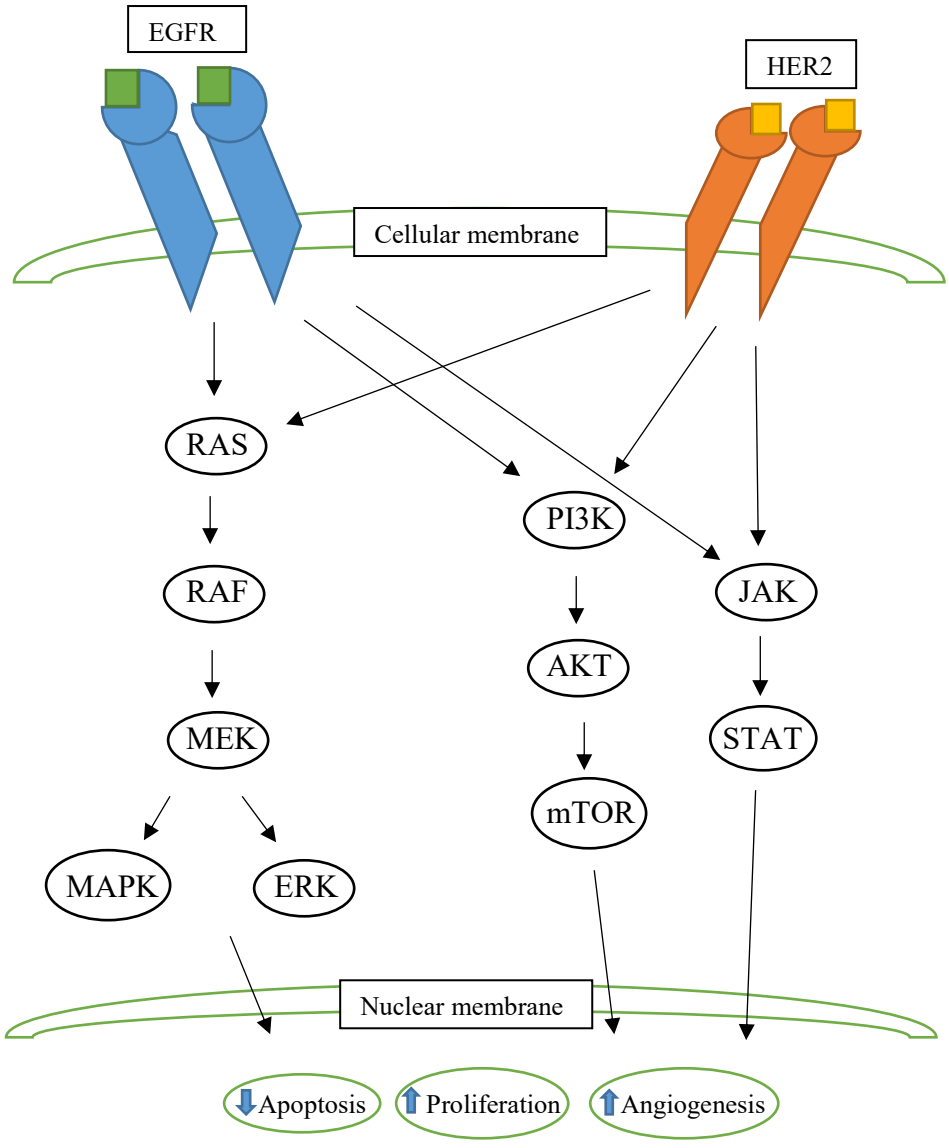

Figure 5: RAS-RAF-MEK-ERK, PI3K-AKT-mTOR and JAK-STAT pathways with associated EGFR and HER2 cell membrane receptors. Activation of these pathways leads to an increase in proliferation, angiogenesis and a decrease in apoptosis.

These types of mutations that increase tyrosine kinase activity could occur at any gene alone the pathway, resulting in its continuous activation. Furthermore, information about the specific genetic change can offer prognostic information on the particular cancer. For example, activation of the PI3K-AKT-mTOR pathway via gene mutation provides resistance to various cancer therapies, indicating a poor prognosis for the cancer[58].

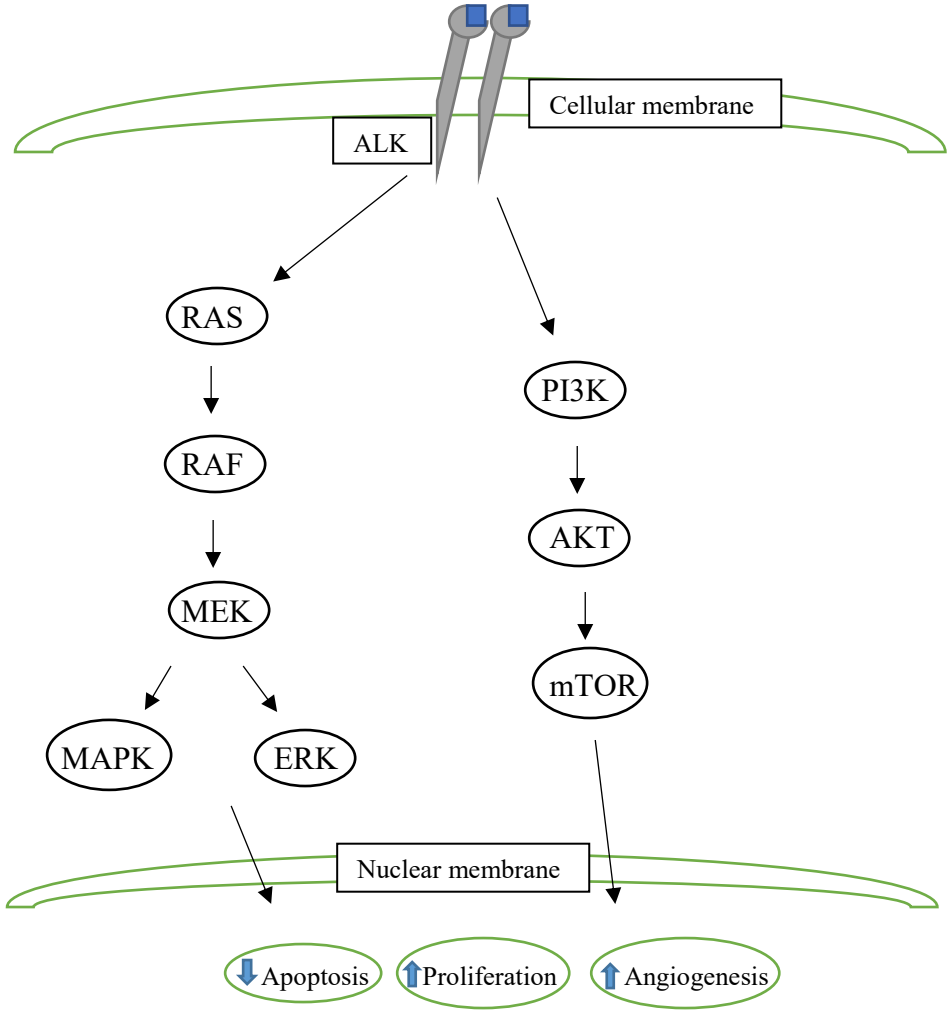

Figure 6: RAS-RAF-MEK-ERK and PI3K-AKT-mTOR pathways with associated ALK tyrosine kinase receptor. Activation of these pathways leads to an increase in proliferation, angiogenesis and a decrease in apoptosis.

Mutations of the ALK gene in lung cancer are typically inversions that result in a fusion, most commonly with EML4 (echinoderm microtubule-associated protein-like 4). The ALK-EML4 fusion proteins results in a ligand-independent constitutive dimerization of the domain kinase[59.] The interaction between the coiled-coil domain within the EML4 gene of each monomer is continuous, which causes ALK activation[60,61]. TFG and KIF5 are other genes which have been less frequently observed as partners in the ALK fusion genes, though are less common[62,63]. The ALK-EML4 fusion gene causes the continual stimulation of the ALK kinase, which causes a cascade effect throughout its signalling pathway (Figure 6). There are multiple effects as the ALK kinase stimulates different axes, affecting cell survival. However, many gene activations result in cell proliferation which produces the same consequence as with the previously mentioned mutations.

The ALK TK receptor gene is part of the insulin receptor superfamily. It is a transmembrane receptor containing an extracellular domain with an N-terminal signal peptide sequence, as well as binding sites for the activating ligands of ALK, pleiotrophin, and midkine[64]. The ALK rearrangement mutation is a fusion oncogene and is found among 2-7% of non-small-cell lung cancer patients. It is most common among non-smokers or those with limiting smoking history with adenocarcinoma. Currently, advanced lung adenocarcinoma patients and non-small-cell lung cancer with an adenocarcinoma component are tested routinely for ALK rearrangements. Those who test negative for the EGFR mutations will also receiving testing for ALK as the two mutations are generally found to be mutually exclusive[65].

EGFR mutations are observed in patients with advanced adenocarcinoma. In a recent study of EGFR mutations in lung cancer, 16.6% of 2105 patients had a mutation.

90% of EGFR mutations involve exon 19 or exon 21[66,67]. This study, which took into account sex, histological subtype and smoking history, determined that of those who have advanced non-small-cell lung cancer, smokers diagnosed with non-squamous cell carcinoma and all non-smokers of every histological subtype should be tested for the EGFR mutation, due to a high number of patients fitting these criteria exhibiting the mutation[67].

KRAS is one of the most commonly mutated of the RAS genes with mutations, occurring in 30% of adenocarcinomas and 5% of squamous cell carcinomas. The mutations usually occur in codons 12 and 13, and are generally associated with smokers. Specific inhibitors for this mutation have not been found, but the search is still underway[68]. Studies have shown that chemotherapy administered to patients exhibiting KRAS mutations will not exhibit any benefits. Unlike EGFR mutations which are typically found in patients who have never smoked cigarettes, KRAS mutations generally appear in patients who had a large exposure to tobacco[69]. Like with EGFR and ALK mutations, previous studies have shown that EGFR and KRAS appear to be mutually exclusive in lung cancer[70].

HER2 has been found to be over-expressed in 20% of advanced non-small-cell lung cancers, however only 2% of cases have been shown to have an amplification or mutation. HER2 mutations most commonly found in adenocarcinoma cases are among non-smoking women. The mutation of this gene usually occurs as an insertion in the exon 20, which codes for part of the kinase domain; it causes continual receptor activation[71]. Studies have shown that patients with this mutation have a partial response to therapy with

the HER2 inhibitor neratinib, a TKI, in combination with the mTOR (mechanistic target of rapamycin) inhibitor temsirolimus[72].

The BRAF mutation is found in 1-3% of tumours from patients with advanced non-small-cell lung cancer. Interestingly the BRAF mutations is found more frequently in patients with melanoma, occurring in 50% of tumours. Unlike those with EGFR and ALK mutations, this mutation is found throughout smokers and ex-smokers. The effects of specific inhibitors which have proven effective within melanomas containing this mutation, such as PLX4032, are being studied[73]. Dabrafenib is one such example, which acts as a B-Raf enzyme inhibitor that eliminates an abnormal form of the BRAF protein, involved in the regulation of cell growth. It has been used against tumours with BRAF mutations found in melanoma, and results in prolonged survival without tumour progression. Using dabrafenib as a cancer treatment for cancer patients with BRAF mutations is still in the clinical phase, but has the potential of leading to more personalized therapy options[74].

PIK3CA mutations are uncommon in non-small-cell lung cancer, but are seen in approximately 3.7%. However, amplification of this gene has been seen in advanced non-small-cell lung cancer, in particular within squamous cell carcinomas of male smokers. A recent study showed PIK3CA amplification in 8.9% of squamous cell carcinoma cases compared to 2.9% among adenocarcinoma cases[75]. This amplification is not necessarily related to the presence of the mutation however. A study has shown the presence of the mutation and of copy number alterations were mutually exclusive in the patients, which suggests both have equal oncogenic potential[76,77]. The mutations are found in exon 9 among both adenocarcinoma and squamous cell carcinoma cases, and on occasion within

tumours exhibiting an EGFR or KRAS mutation as well. Increased activity in phosphatidylinositol 3-kinase and phosphorylated Akt expression have been observed with PIK3CA alterations. PIK3CA and KRAS have been observed to be one of the most frequently mutated oncogenes among human cancers[78].

Different drugs have been developed which can inhibit the associated kinases to the EGFR and ALK genes and have been proven to be the best option in terms of treatment. Crizotinib (TKI) is effective against ALK gene mutations, while Gefitinib or Erlotinib can eliminate the effects caused by mutations within the EGF pathway acting as TKIs[79]. Another anti-EGFR agent, Cetuximab, has also been developed, which is a chimeric monoclonal antibody that functions against the extracellular domain of the receptor[80]. While Gefitinib and Erlotinib have proven effective in adenocarcinoma cases, they are not effective on patients with squamous cell carcinoma[44]. Cells containing a KRAS mutation have also shown significant resistance to TKIs[70].

In most cases, the tumour recurs after successful initial treatment. For instance, there are cases of EML4-ALK positive non-small-cell lung cancer patients having developed resistance to Crizotinib even after receiving successful treatment for 5 months. This appeared to be due to two new mutations that were discovered within the fusion gene, which allowed resistance to the drug[81]. Other examples are cases with EGFR mutations, such as L858R and exon 19 deletions. Even after treatment with Gefitinib or Erlotinib, some of these cases developed a secondary mutation, T790M, which is resistant to treatment[82]. These occurrences increase the need to continue research and development of drugs directed not only against newer cancer-causing mutations, but also against mutations with previously developed treatments. The latter allows to continually improve treatment

options and aid in the expansion of personalized medicine, which ultimately increases the survival rates of future patients.

Lung cancer is clearly a group of disorders with diverse genetic causes and outcomes. It is important to identify the underlying cause of disease in lung cancer patients for a number of reasons. Previous studies have indicated that the presence of specific mutations predicts prognosis. Furthermore, an understanding of the frequency of these mutations and the mechanisms leading to disease guides efforts to develop directed therapies. The published studies were mainly concentrated in advanced stage of lung cancer patients and little is known about mutations in patients with early stage of the disease and clinical outcome after surgical treatment. Therefore, we studied the mutational status in a group of lung cancer patients treated with surgical resection of the tumours. All samples were submitted with patient consent to access their clinical and pathological history, making it possible to do clinical and pathological correlations.

This study hypothesizes that the determined correlations between molecular and clinical/pathological data for the Nova Scotia population will match those observed in other published studies for different populations across the world. The objectives to test this are fourfold:

1. For each case of lung cancer, the appropriate sample of the tumour must be selected.
2. Molecular profiling is performed on the samples to determine the frequency of the mutation throughout all the lung cancer cases

3. The clinical and pathological features of the patients is collected from various databases and sorted in an excel sheet
4. Accompanying the molecular data of the samples, it is possible to establish any correlations between the molecular and clinical/pathological data of the patients using statistical analysis.

Such correlations provide the preliminary data for future research, both in the development of personalized treatments for new mutations or improvement of treatment options for previously treated mutations. It could also indicate any location-specific correlations, as the samples are obtained from patients within Nova Scotia. These results can guide treatment developments best able to benefit Nova Scotians.

CHAPTER 2: MATERIALS AND METHODS

There are five main procedures involved with specimen analysis: specimen handling and processing, microscopic examination, DNA extractions, SNaPshot genotyping and PCR based gene sizing analysis, and immunohistochemical assay and fluorescent in situ hybridization. The mutations being analyzed differ; some are point mutations where a base is switched, others are insertion/deletion of bases and one is a fusion gene caused by an inversion. Therefore, three different molecular assays must be used for accurate analysis.

2.1 Specimen Identification

The samples used for this study are contained within a tumour bank in the Department of Pathology at the QEII Health Sciences Centre, consisting of cases collected from 2005 until the present. The lung cancer samples are first fixed in formalin and embedded in paraffin (FFPE). A 4-micron section from each tissue block is mounted on glass slide and stained with haematoxylin and eosin (H&E). This allows the pathologist to examine for the amount of tumour present in the block, which will determine its adequacy for further studies. The appropriate tissue blocks were selected based on these results.

2.2 Tumour Identification

Sections of the tissue is reserved for preparation of the H&E stained slides which can be used in microscopic examination of different properties, including: Pathologic type (such as adenocarcinoma, squamous cell carcinoma, large cell carcinoma, pleomorphic carcinoma, carcinoid, etc.), pathologic grade (level of differentiation), pathologic staging, lymphatic/vascular invasion and/or lymph node metastasis.

2.3 Isolation of Tumour from Tissue Blocks

From each selected tumour block 20 µ thick tissue x5 sections were cut and mounted on glass slides. Based on the location of tumour tissue identified on the matching H&E slide, the tumour areas were circled by a lung pathologist. The tissue from the circled area was then scrapped from the slide for DNA extraction. The purpose of circling the tumour area allows for the tumour cells to be concentrated during further analysis. If the tumour area encompasses the entire or near entire H&E slide, the five cut sections were immediately put into a tube for the DNA extraction.

2.4 DNA Extraction

Tissue was extracted from the tumour sections, which had either been placed directly into tubes or were scraped into tubes from the circled areas on the slides using a scalpel. The tissue was suspended in 300 µL Buffer ATL, which is used for tissue lysis and contains a chaotropic salt, and 20 mg/mL of Proteinase K. It was then incubated overnight at 65ºC after being vortexed for 15 s and centrifuged for 20 s. The following day after centrifuging again the tissue lysate layer was removed and put in a new tube, and enough 0.9% saline solution was added to bring the total volume to 400 µL. The DNA was then extracted from the lysed tissue using the MagnaPur Robotic system (Roche) with the MagnaPur2 Compact Nucleic Acid Isolation Kit. This process of isolation used a bead-based kit, where Magnetic Glass Particles (MGPs) were added to the samples, causing nucleic acid to become immobilized on their surfaces. This allowed unbound substances (such as proteins, cell debris, etc.) to be removed, leaving the purified nucleic acids. This

purified DNA can subsequently be used in different molecular analysis procedures, including PCR and sequencing.

2.5 DNA Quantification

The extracted DNA was quantified using the NanoDrop Spectrophotometer instrument from Thermo Scientific. This instrument can determine the quantity and purity of DNA by measuring its absorbance of light in the UV wavelengths. This is done by shining UV light through the sample, and determining how much makes it through vs. how much is absorbed. DNA absorbs at 260 nm, while proteins and other contaminates absorb around 280 nm. To assess the purity of the DNA, the 260/280 ratio is determined, with a higher ratio of ~1.8 indicating a more "pure" sample of DNA.

2.6 Molecular Analysis

2.6.1 Point Mutation

A SNaPshot (multiplexed primer extension assay) was developed to identify the most common point mutations associated with adenocarcinoma, namely, EGFR L858R and T790M, KRAS G12X, BRAF V600E, and PIK3CA E542K and E545K in this study. It had been developed to detect single nucleotide polymorphisms, or SNPs[83]. They represent differences among the nucleotides in DNA, A, T, C and G. This is a highly sensitive and specific assay. Due to its multiplexed nature, the assay can detect up to ten point mutations simultaneously from a very small amount of DNA. This is critical considering the amount of tissue available for testing is often a limiting factor[84]. Throughout the analysis process, three controls were also included in each run. One was a negative cell line, containing none of the mutations. The second was the A549 control,

which had a homogenous KRAS mutation. The third was the H1975 control, which had a heterogeneous EGFR mutation. Primer primers were multiplexed to amplify the seven loci of interest; these are listed as PCR Primers in Table 1. Extension Primers (Table 2) were designed to bind directly adjacent to the SNP of interest and identify the specific SNP. These were purchased from the company Thermo Fisher Scientific.

Table 1: List of PCR primers used for SNaPshot.

Primer # and Sequence	Amplicon size	Target	Stock conc. (µM)	Working conc. (µM)	Final conc. (µM)
215: 5'- ACG TTG GAT GTC ATT ATT TTT ATT ATA AGG CCT GCT G - 3'	186 bp	KRAS ex2	50	0.8571	0.24
216: 5'- ACG TTG GAT GAG AAT GGT CCT GCA CCA GTA A - 3'			50	0.8571	0.24
217: 5'- ACG TTG GAT GTG TTC CCG GAC ATA GTC CAG - 3'	98 bp	EGFR ex20	50	0.2143	0.06
218: 5'- ACG TTG GAT GAT CTG CCT CAC CTC CAC CGT - 3'			50	0.2143	0.06
219: 5'- ACG TTG GAT GGA CAA AGA GCT CAA AGC AA - 3'	108 bp	PIK3CA ex10	50	0.2143	0.06
220: 5'- ACG TTG GAT GTT TAG CAC TTA CCT GTG ACT CCA - 3'			50	0.2143	0.06
221: 5'- ACG TTG GAT GTG CTT GCT CTG ATA GGA AAA TG - 3'	143 bp	BRAF ex15	50	0.2143	0.06
222: 5'- ACG TTG GAT GCT GAT GGG ACC CAC TCC AT - 3'			50	0.2143	0.06
223: 5'- ACG TTG GAT GCC TCC TTC TGC ATG GTA TTC - 3'	101 bp	EGFR ex21	50	0.1875	0.05
224: 5'- ACG TTG GAT GGC AGC ATG TCA AGA TCA CAG - 3'			50	0.1875	0.05

Table 2: List of Extension primers used for SNaPshot.

Primer # and Sequence	Amplicon size	Target	Stock conc. (µM)	Working conc. (µM)	Final conc. (µM)
225: 5'- GAC TGA CTG CTC TTG CCT ACG CCA - 3'	24 bp	KRAS 35 ext R	10	1.49	0.149
226: 5'- CTG ACT GAC TGT GAT TTT GGT CTA GCT ACA G - 3'	31 bp	BRAF 1799 ext F	10	1.22	0.122
227: 5'- GAC TGA CTG ACT GAC TGA CAG ATC ACA GAT TTT GGG C - 3'	37 bp	EGFR 2573 ext F	10	4.00	0.400
228: 5'- CTG ACT GAC TGA CTG ACT GAC TGA CTG ACT AAG GGC ATG AGC TGC - 3'	45 bp	EGFR 2369 ext R	10	2.44	0.244
229: 5'- CTG ACT GAC TGA CTG ACT GAC TGA CTG ACT GAC TTC TCC TGC TCA GTG ATT T - 3'	52 bp	PIK3CA 1624 ext R	10	0.50	0.050
230: 5'- ACT GAC TGA CTG ACT GAC TGA CTG ACT GAC TGA CTG ACT CTT GTG GTA GTT GGA GCT - 3'	57 bp	KRAS 34 ext F	10	0.25	0.025
231: 5'- GAC TGA CTG ACT GAC TGA CTG ACT GAC TGA CTG ACT GAC TGA CTG ATC CTC TCT CTG AAA TCA CT - 3'	65 bp	PIK3CA 1633 ext F	10	0.10	0.010

Their preparation involved adding 1.6 mL of 1/10 dilution of low TE buffer to each tube of primers, which contained quantities of 80,000 pmol, making the concentration 50 µM. Ten 100µL aliquots of the working primers were made, as described in Table 3 and Table 4. Finally, to prepare 20 mM of dNTPs for the first step of PCR, 50 µL of 40mM stock dNTPs were combined with 50 µL of Ultrapure water. The dNTPs were purchased from the company Sigma Aldrich.

Table 3: PCR primer mix for SNaPshot.

Primer #	µL of each stock	Stock concentration	Working concentration	Final concentration
215, 216	17.0	50 µM	0.8571 µM	0.24 µM
217, 218	4.3	50 µM	0.2143 µM	0.06 µM
219, 220	4.3	50 µM	0.2143 µM	0.06 µM
221, 222	4.3	50 µM	0.2143 µM	0.06 µM
223, 224	3.7	50 µM	0.1875 µM	0.05 µM
Ultrapure water	932.8			

Table 4: Extension primer mix for SNaPshot.

Primer #	µL of each stock	Stock concentration	Working concentration	Final concentration
225	149	10 µM	1.49 µM	0.149 µM
226	122	10 µM	1.22 µM	0.122 µM
227	400	10 µM	4.00 µM	0.400 µM
228	244	10 µM	2.44 µM	0.244 µM
229	50	10 µM	0.50 µM	0.050 µM
230	25	10 µM	0.25 µM	0.025 µM
231	10	10 µM	0.10 µM	0.010 µM

The initial step involved multiplex PCR to amplify the regions of the DNA that contain the seven mutations of interest, in a reaction mix containing 20 ng of DNA (or 60 np for repeats) from the samples combined with 0.1 µL of Platinum Taq polymerase, 0.6 µL of 50 mM of $MgCl_2$, 0.6 µL of 20 mM of dNTPs, 0.9 µL of dH_2O, 1 µL of 10x PCR Buffer and 2.8 µL of amplification primers 215-224 making a total volume of 10 µL each. The dNTPs used were dideoxy nucleotides, each labelled with a different fluorescence[85]. The reaction mix was heated in the Veriti Thermal Cycler for 8 minutes at 95°C, followed by 40 cycles of 20 seconds at 95°C, 30 seconds at 63°C, 1 minute at 72°C, and one final cycle for 3 minutes at 72°C. 3.3 µL of shrimp alkaline phosphatase (SAP), 4.43 µL of dH_2O, 2.0 µL 10x SAP Buffer and 0.27 µL of exonuclease I (Exo I) were then

added to inactivate excess primers and dNTPs, which was then incubated for 60 minutes at 37°C and 15 minutes at 75°C to inactivate the enzymes.

The following reaction in the SNaPshot assay is primer extension. Extension primers are designed to hybridize directly adjacent to the nucleotide of interest. During the extension reaction, one dideoxy nucleotide was added, which was complementary to the template. The nucleotide added was identified based on the florescent tag. The reaction mix contained 3 µL of PCR product, 2.5 µL of SNaPshot Multiplex Ready Reaction mix (which helped ensure strong analyses of the samples), 3.5 µL of dH_2O and 1 µL of extension primers 225-231, which were added to the PCR products making a total volume of 10 µL for each. It subsequently underwent one cycle in the Thermal Cycler for 30 seconds at 96°C, followed by 25 cycles for 10 seconds at 96°C, 5 seconds at 50°C, 30 seconds at 60°C, and one finale cycle for 3 minutes at 72°C.

After a second treatment with 2.0µL of SAP, 0.1 µL of the products were mixed with 9.7 µL Hi-Di- Formamide and 0.3µL of GeneScan LIZ 120 size standard to obtain a volume of roughly 10 µL each. They were then covered and spun for 1 minute at 1000 rpm. Analysis on the products was then performed on an ABI 3130-XL automated sequencer with the Gene Mapper Analysis Software using the automatic calling parameters to examine for any point mutations. The products were size separated by capillary electrophoresis. Individual extension primers were distinguished by size and the extended nucleotide was identified by fluorescent tag.

2.6.2 Insertion/Deletion Mutation

The Quadruplex fragment analysis genotyping was used to identify insertions and deletions in exons 19 and 20 of EGFR and exon 20 of HER2. Throughout the analysis, three controls are also included in each run. One was a negative cell line, containing none of the mutations. The second was the H1650 control, which had a heterogeneous EGFR deletion mutation. The third was the H1781 control, which had a homogenous HER2 insertion mutation. The primers required for Quadruplex are listed in Table 5, and were purchased from the company Thermo Fisher Scientific.

Table 5: List of primers used for Quadruplex.

Primer # and Sequence	Amplicon size	Target	Stock conc. (µM)	Working conc. (µM)	Final conc. (µM)
207: 5'- NED-GCA CCA TCT CAC AAT TGC CAG TTA - 3'	206 bp	EGFR ex19	100	0.6	0.15
208: 5'- AAA AGG TGG GCC TGA GGT TCA - 3'			100	0.6	0.15
209: 5'- PET-CTC TCA GCG TAC CCT TGT CC - 3'	233 bp	ERRB2	100	1.05	0.2625
210: 5'- CAA AGA GCC CAG GTG CAT AC - 3'			100	1.05	0.2625
211: 5'- VIC-CGA AGC CAC ACT GAC GTG – 3'	246 bp	EGFR ex20	100	1.2	0.30
212: 5'- CCG TAT CTC CCT TCC CTG ATT A - 3'			100	1.2	0.30
213: 5'- 6FAM-GGT GAT CTA TTT TTC CCT TTC TCC - 3'	184 bp	KIT	100	0.9	0.2188
214: 5' GAC ATG GAA AGC CCC TGT T - 3'			100	0.9	0.2188

Their preparation involved adding 800 µL of 1/10 dilution of low TE buffer to each tube of primers, which contained quantities of 80,000 pmol, making the concentration 100 µM. Ten 100 µL aliquots of the working primers were made, as described in Table 6.

Table 6: Primer mix for Quadruplex.

Primer #	**µL of each stock**	**Stock concentration**	**Working concentration**	**Final concentration**
207, 208	3	100 µM	0.6 µM	0.15 µM
209, 210	5.25	100 µM	1.05 µM	0.2625 µM
211, 212	6	100 µM	1.2 µM	0.30 µM
213, 214	4.5	100 µM	0.9 µM	0.2188 µM
Ultrapure water	481.25			

PCR primers designed to hybridize by flanking the common deletion or duplication sites were used in PCR to amplify regions of interest. The initial step of multiplex PCR involved combining 15 ng of DNA (or 60 ng for repeats) from the samples with 0.8 µL of 50 mM of $MgCl_2$, 0.8 µL of 20 mM of dNTPs, 0.2 µL of Platinum Taq, 3.2 µL of dH_2O, and 5.0 µL of fluorescently labelled primers 207-214, creating a total volume of 20 µL each. The tubes are then vortexed for 2 s and centrifuged for 5 s before being loaded in the PCR machine. PCR conditions involved denaturation for 10 minutes at 94°C, followed by 40 cycles of 30 seconds at 94°C, 30 seconds at 55°C, 30 seconds at 72°C, and one final cycle for 10 minutes at 72°C.

The products were mixed with 9.7 µL Hi-Di- Formamide and 0.3 µL of GeneScan LIZ 500 size standard to obtain a volume of roughly 10 µL each. It was then covered and spun for 1 minute at 1000 rpm. Products were size separated using an ABI 3130-XL automated sequencer with the Gene Mapper Analysis Software, which uses the automatic calling parameters to examine for any insertion/deletion mutations. The DNA fragments

were detected as they moved through the sequencer, and placed among a sizing curved generated by the software, based on known size standard fragments. Since each primer was tagged with a different fluorescent marker, individual amplicons were identified based on colour as well as size.

2.6.3. ALK Inversion Mutation

Immunohistochemical assay (IHC) and fluorescent in situ hybridization (FISH) was used to identify ALK gene rearrangements. All cases were first screened for ALK protein overexpression by IHC. Equivocal and positive cases were then confirmed by FISH. 4 μm thick sections are obtained from the FFPE sample blocks and placed on charge slides, and were then dried and melted in an oven for 60 minutes at 45°C. The slides were stained on Ventana's Benchmark Ultra and dewaxed on Ultrausing EZ Prep (dilution 1/10). Ventana's CC1 (Tris/Borate/EDTA, pH 8.0-8.5 buffer) was used for 92 minutes for heat-induced epitope retrieval. Antigen-antibody visualization was performed using the Ultra View DAB detection kit. The first reagent in the kit was 3% hydrogen peroxide, which acted as an inhibitor to endogenous peroxidase. The primary antibody was ALK-5A4 (Leica, # NCL-ALK), which was used at a dilution of 1/10 and incubated for 60 minutes at 37°C. The staining was further enhanced using an amplification kit. Ventana's Hematoxylin and Bluing Regeants were used to counterstain the slides, before they were dehydrated and coverslipped on Ventana's Symphony system. The positive control was from a known ALK IHC positive case confirmed by positive FISH ALK rearrangement, and the negative control was a mouse immunoglobulin G1 serum substitution for the primary antibody.

FISH is an interphase molecular cytogenetic study using an ALK probe on the samples. 4 μm FFPE sections on glass slides were de-paraffinized three times for 10 minutes in xylene, dehydrated twice for 5 minutes in 100% ethylic alcohol, and pre-treated for 120 minutes with 10mM sodium citrate. The sections were then transferred to 2X standard saline citrate (SSC) for 2 minutes at 25°C and washed in de-ionized distilled water. Proteins were first pre-digested in 0.01 N HCl at 25°C and then digested with pepsin solution (75,000 mL in 0.01 N HCl) at 37°C. The slides were again washed with de-ionized distilled water at 25°C, dehydrated in alcohol (70, 85, and 100%) and air dried at room temperature. After appropriate 1:50 dilution of the ALK probe, 10 μL of ALK probe was added to the sections. After denaturation for 5 minutes at 74°C, the slides were placed overnight in a humidified chamber at 37°C to allow for probe hybridization. The following morning the samples were washed with 2X SSC/0.3% NP40 for 2 minutes at 73°C and then 2XSSC for 2 minutes at 60°C before being mounted in 10 μL 4,6-diamidino-2-phenylindole.

ALK rearrangement is considered positive if 15% or more tumour cells observed after a FISH treatment shows a split signal of the fluorescent probes flanking the ALK locus within the nucleus, or 15 of 100 cells analyzed in a sample. A signal is considered split if the red and green signals are separated by ≥2 signal sizes in diameter. The ALK probe is composed of a 3' red and 5' green signal which appear yellow when fused together, such as in normal cells. Abnormal cells, or those with an ALK mutation, exhibit these two separate signals, or a 3' red signal alone. A 5' green signal alone does not indicate an ALK mutation[86].

2.7 Chart Review

All the data of the patients (molecular, pathological, clinical) was obtained from online databases, including Millennium and Horizon Patient Folder (HPF). Data from Millennium was obtained throughout 2013 to 2015, and data from HPF was obtained in July 2015. Such data included patient age, sex, smoking history, survival, cell type, vascular/lymphatic/pleura/lymph node invasion, staging, cancer history, and family cancer history. Survival rate was determined by having a 5 year follow up for all patients to determine their status. This was entered into an excel spreadsheet, which was then used for statistical analysis.

2.8 Statistical Analysis

Statistical analysis was performed using SAS 9.3 software licensed to Dalhousie University (Site 70099642). Frequency cross tabulations of molecular mutation by various pathological/clinical characteristics were generated using the FREQ procedure. Statistical associations between mutation and pathological/clinical characteristics were investigated by means of chi-square tests and logistic regression. The association between molecular mutation and patient survival, controlling for patient and pathological/clinical characteristics was investigated using proportional hazards regression[87], which was done using the PHREG procedure in SAS. The characteristics examined and controlled when found to be independently predictive of survival were age, gender, cancer type, pleural invasion, vascular invasion, lymphatic invasion, lymph node invasion, tumour stage, history of cancer, history of family cancer and smoking history.

CHAPTER 3: RESULTS

Among the 667 cases available, 340 are male and 327 female with the age ranging from 39 to 90 years. There are 438 cases of adenocarcinoma (65.7%), 166 cases of squamous cell carcinoma (24.9%), 43 cases of large cell carcinoma (6.4%), 10 cases of pleomorphic carcinoma (1.5%) and 10 cases carcinoid (1.5%). The age and gender distribution data are listed in Table 7.

Table 7: Age and gender distribution of the total 667 lung cases.

Age Groups	*Gender*		
	Male	**Female**	**Total**
39-59 years	68	79	147
60-74 years	193	180	373
75-90 years	79	68	147
Total	340	327	667

The vast majority of the patients are in early stage (stage IA-IIA) at the time of surgical treatment. Detailed staging is presented in Table 8. The table shows the observed frequency in each cell, the expected frequency under the null hypothesis of no association, the contribution of each cell to the overall chi-square statistic, and the row percentage. Given that the degrees of freedom of the overall chi-square statistic is equal to the (R-1)(C-1) where R and C are the numbers of rows and columns, respectively, the contribution of each cell to the overall chi-square statistic is a chi-square with slightly less than 1 degree of freedom. A value in excess of 3.841, this being the 95th percentile of a chi-square with

one degree of freedom, would be reasonable evidence that the observed frequency deviates from what is expected if there is no association. None of the chi-square values in the table exceeds 3.841. Roughly 33.1% of the tumour tissues subjected to molecular genotyping involved one of the six selected gene mutations. KRAS was the most frequent mutation (24.4%), followed by EGFR (6.4%), BRAF (0.9%), PIK3CA (1.2%) and ALK gene rearrangement (0.2%). HER2 mutations were not present at all in this cohort (Figure 7). Among 43 cases of EGFR mutations 21 cases were exon 19 deletion (3.1%), 2 cases were exon 20 insertion (0.3%), and 20 cases were L858R mutation (3.0%), and among 8 cases of PIK3CA 4 cases were PIK3CA E452K (0.6%) and 4 cases were PIK3CA E454K (0.6%).

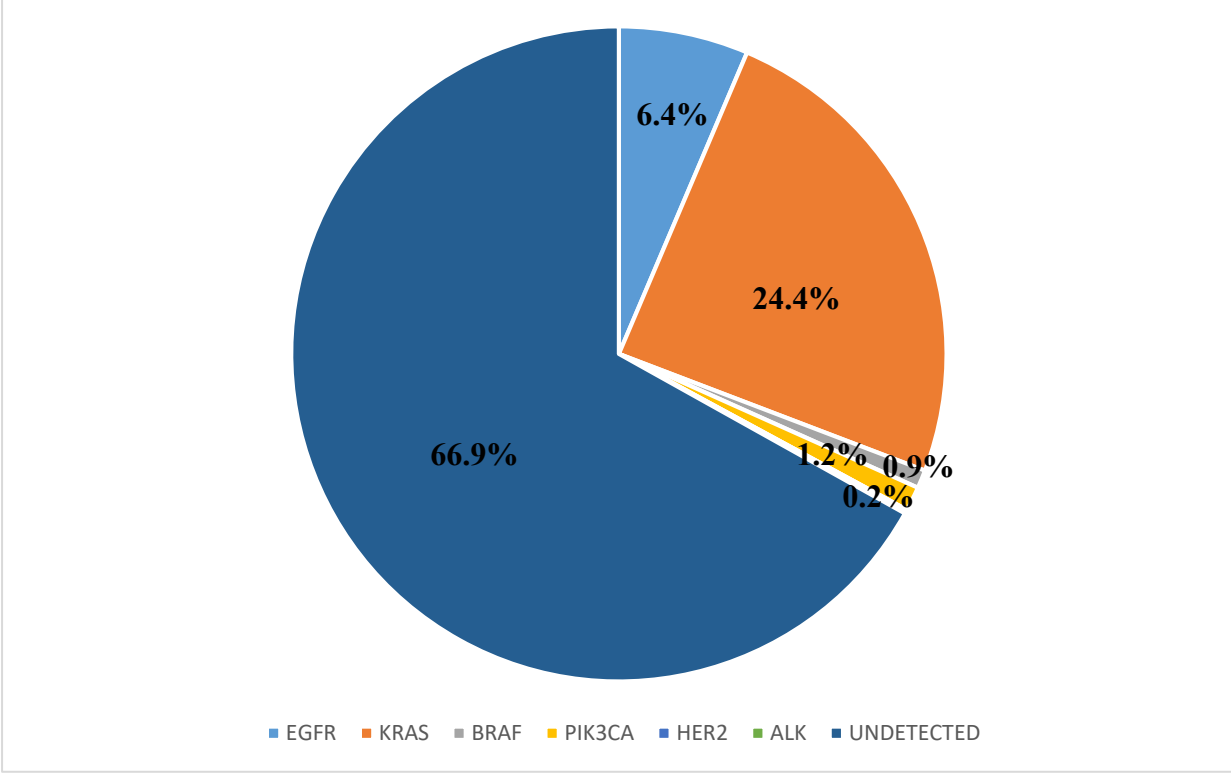

Figure 7: Frequency of the mutations within the 667 samples tested. 6.4% of the observed mutations were EGFR, 24.4% were KRAS, 0.9% were BRAF, 1.2% were PIK3CA, and 0.2% were ALK. No HER2 mutations were observed in this group of samples.

Table 8: Statistics table of tumour stage by mutation. The chi-square test yielded a p-value of 0.9527. Forty-two of the cases did not have this information unavailable.

Tumour stage	*Mutation*							
	ALK	**BRAF**	**EGFR**	**KRAS**	**PIK3CA**	**HER2**	**OTHERS**	**Total**
IA								
Frequency	0	4	14	56	1	0	131	206
Expected	0.3296	1.9776	12.854	49.44	2.6368	0	138.76	
Chi-square	0.3296	2.0682	0.1021	0.8704	1.016	0	0.4341	
Row %	0.00	1.94	6.80	27.18	0.49	0.00	63.59	
IB								
Frequency	1	0	9	29	1	0	84	124
Expected	0.1984	0.1904	7.7376	29.76	1.5872	0	83.526	
Chi-square	3.2387	0.1904	0.206	0.0194	0.2172	0	0.0027	
Row %	0.81	0.00	7.26	23.39	0.81	0.00	67.74	
IIA								
Frequency	0	2	5	26	4	0	91	128
Expected	0.2048	1.2288	7.9872	30.72	1.6384	0	86.221	
Chi-square	0.2048	0.484	1,1172	0.7252	3.404	0	0.2649	
Row %	0.00	1.56	3.91	20.31	3.13	0.00	71.09	
IIB								
Frequency	0	0	1	6	0	0	24	31
Expected	0.0496	0.2976	1.9344	7.44	0.3968	0	20.882	
Chi-square	0.0496	0.2976	0.4514	0.2787	0.3968	0	0.4657	
Row %	0.00	0.00	3.23	19.35	0.00	0.00	77.42	
IIIA								
Frequency	0	0	7	24	1	0	64	96
Expected	0.1536	0.9216	5.9904	23.04	1.2288	0	64.666	
Chi-square	0.1536	0.9216	0.1702	0.04	0.0426	0	0.0069	
Row %	0.00	0.00	7.29	25.00	1.04	0.00	66.67	
IIIB								
Frequency	0	0	0	2	0	0	3	5
Expected	0.008	0.048	0.312	1.2	0.064	0	3.368	
Chi-square	0.008	0.048	0.312	0.5333	0.064	0	0.0402	
Row %	0.00	0.00	0.00	40.00	0.00	0.00	60.00	
IV								
Frequency	0	0	3	7	1	0	23	34
Expected	0.0544	0.3264	2.1216	8.16	0.4352	0	22.902	
Chi-square	0.0544	0.3264	0.3637	0.1649	0.733	0	0.0004	
Row %	0.00	0.00	8.82	20.59	2.94	0.00	67.65	
Total	1	6	39	150	8	0	421	625

Further examining the individual cancer types revealed that 209 samples of the 438 adenocarcinoma cases had a mutation (47.7%), 3 samples of the 43 large cell carcinoma cases (7.0%), 8 samples of the 166 squamous cell carcinoma cases (4.8%), 1 sample of the 10 pleomorphic carcinoma cases (10.0%) and 0 samples of the 10 carcinoid cases (0%) (Table 9).

Table 9: Statistics table of cell type by mutation. The chi-square test yielded a p-value of <0.0001.

Cell Type	*Mutation*							
	ALK	**BRAF**	**EGFR**	**KRAS**	**PIK3CA**	**HER2**	**OTHERS**	**Total**
AD								
Frequency	1	6	**42**	**156**	4	0	**229**	438
Expected	0.6567	3.94	**28.237**	**107.04**	5.2534	0	**292.88**	
Chi-square	0.1795	1.077	**6.7084**	**22.397**	0.299	0	**13.931**	
Row %	0.23	1.37	**9.59**	**35.62**	0.91	0.00	**52.28**	
SQ								
Frequency	0	0	**1**	**3**	4	0	**158**	166
Expected	0.2489	1.4933	**10.702**	**40.567**	1.991	0	**111**	
Chi-square	0.2489	1.4933	**8.7951**	**34.789**	2.0271	0	**19.902**	
Row %	0.00	0.00	**0.60**	**1.81**	2.41	0.00	**95.18**	
LCC								
Frequency	0	0	0	**3**	0	0	**40**	43
Expected	0.0645	0.3868	2.7721	**10.508**	0.5157	0	**28.753**	
Chi-square	0.0645	0.3868	2.7721	**5.3647**	0.5157	0	**4.3997**	
Row %	0.00	0.00	0.00	**6.98**	0.00	0.00	**93.02**	
PLE								
Frequency	0	0	0	1	0	0	9	10
Expected	0.015	0.09	0.6447	2.4438	0.1199	0	6.6867	
Chi-square	0.015	0.09	0.6447	0.853	0.1199	0	0.8003	
Row %	0.00	0.00	0.00	10.00	0.00	0.00	90.00	
Carcinoid								
Frequency	0	0	0	0	0	0	10	10
Expected	0.015	0.09	0.6447	2.4438	0.1199	0	6.6867	
Chi-square	0.015	0.09	0.6447	2.4438	0.1199	0	1.6418	
Row %	0.00	0.00	0.00	0.00	0.00	0.00	100.00	
Total	1	6	43	163	8	0	445	667

When examining the molecular profiling among each cancer type, the following frequencies were found. 52.3% of the adenocarcinoma samples exhibited no detected

mutations, with the KRAS mutation being the most frequently observed (35.6%), followed by EGFR (9.6%), BRAF (1.4%), PIK3CA (0.9%) and ALK (0.2%) (Figure 8). 95.2% of the squamous cell carcinoma samples exhibited no detected mutations, with the PIK3CA mutation being the most frequently observed (2.4%), followed by KRAS (1.8%) and EGFR (0.6%) (Figure 9).

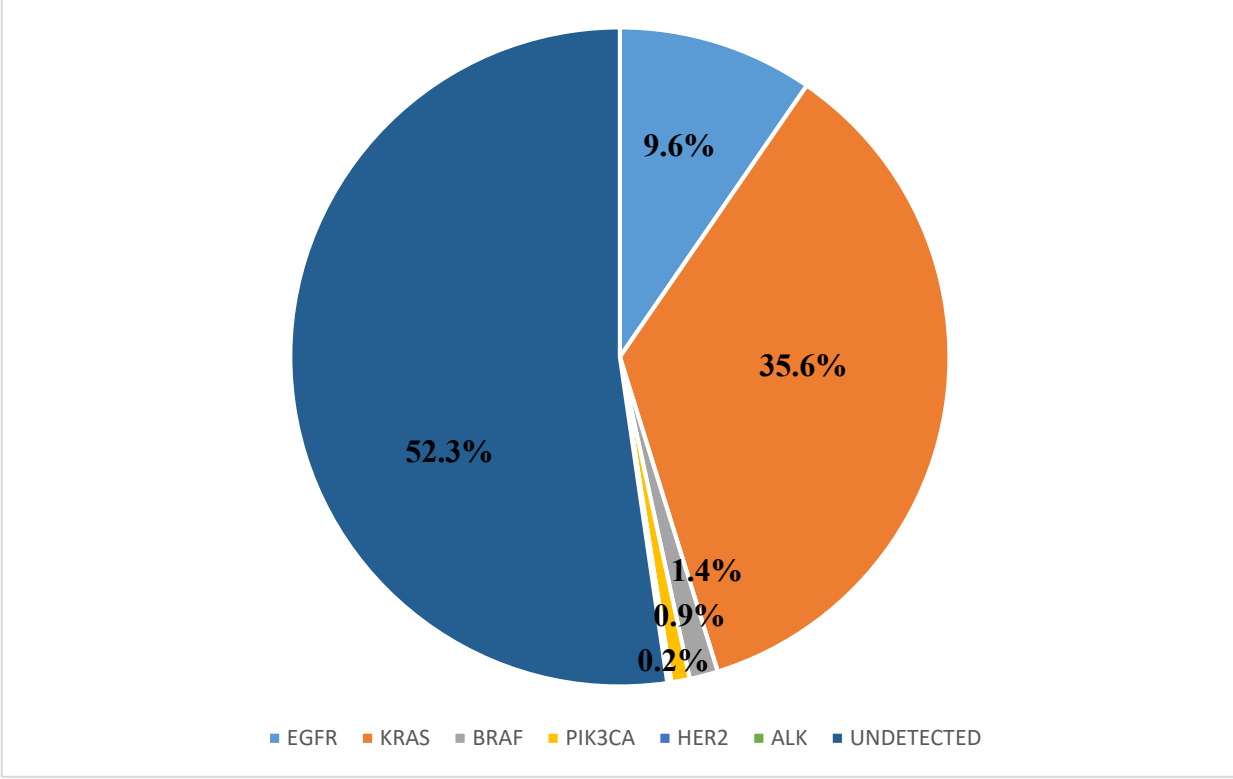

Figure 8: Frequency of the mutations within the 438 adenocarcinoma tested. 9.6% of the observed mutations were EGFR, 35.6% were KRAS, 1.4% were BRAF, 0.9% were PIK3CA and 0.2% were ALK. No HER2 mutations were observed in this group of samples.

26 cases had two primary lung tumours with 13 of them showing no detected mutations. Of the remaining 13, six individuals had a mutation in one tumour and not the other, five individual had the same mutation (KRAS) in both tumours, and two individuals had a different mutation (EGFR and KRAS; KRAS and BRAF) in both tumours. There were also eight cases where the patient had three primary lung tumours, three of which

had no detected mutations, one individual who had the same mutation (KRAS) in all three tumours, two individuals who only had a mutation (KRAS) in one of their tumours, one individual with a (different) mutation (EGFR and KRAS) in only two of their tumours, and one individual with the same mutation (KRAS) in two of their tumours and a different one (BRAF) on the third.

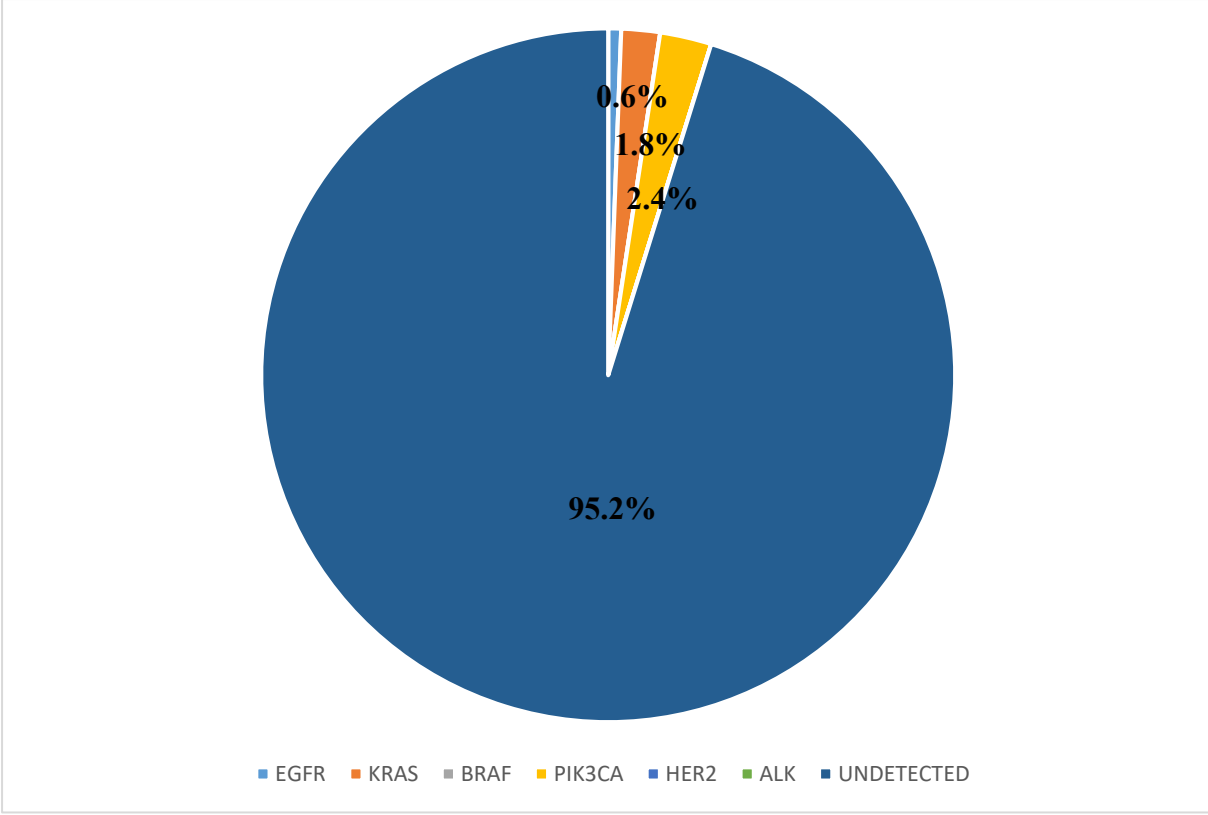

Figure 9: Frequency of the mutations within the 166 squamous cell carcinoma samples tested. 0.6% of the observed mutations were EGFR, 1.8% were KRAS, and 2.4% were PIK3CA. No BRAF, HER2 or ALK mutations were observed in this group of samples.

Among the 667 patients being analyzed, 202 of them are now deceased. All living patients also receive a follow-up accession five years after initial diagnosis. Of the 465 still alive, and 169 of them (36.3%) had a mutation (Figure 10), while 52 of the 202 deceased patients (25.7%) demonstrated a mutation (Figure 11).

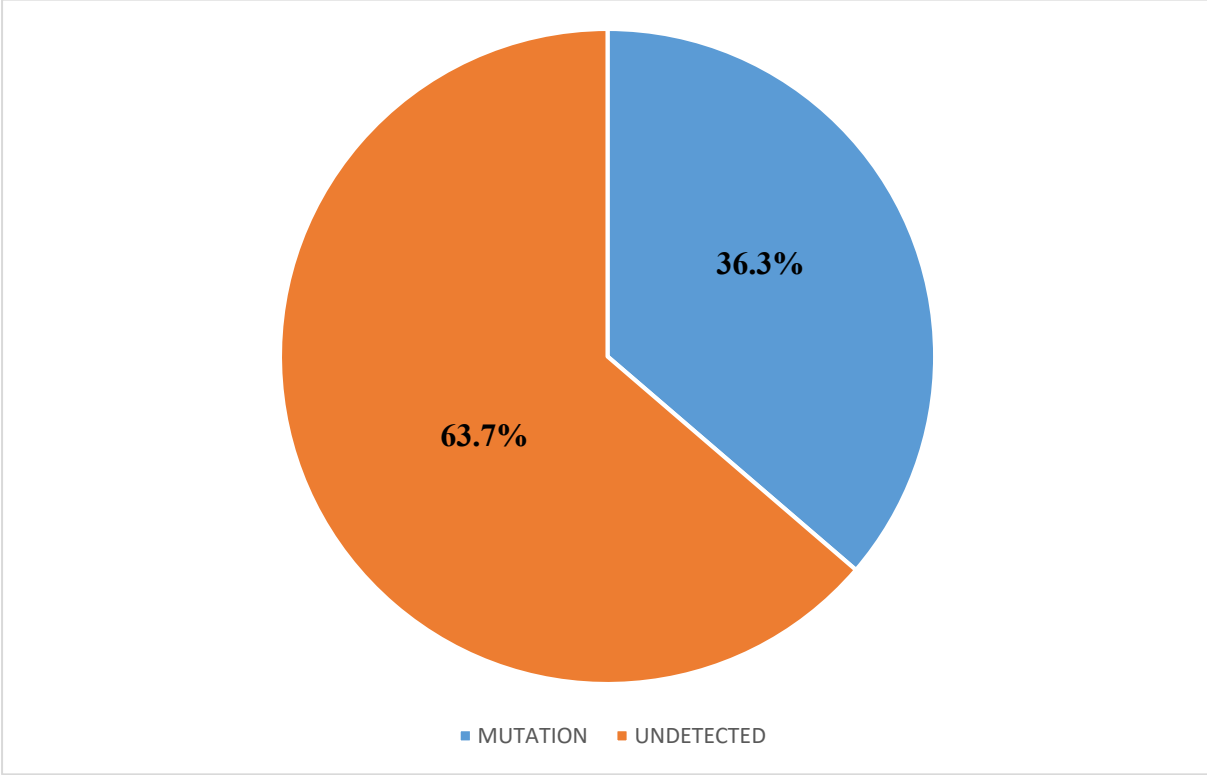

Figure 10: Mutations frequency vs undetected mutations among the 465 living patients who were tested. 36.3% of the patients exhibited one of the six mutations while 63.7 % of the patients did not.

Statistical analysis performed on the patient data yield a variety of correlations. When examining the survival data, a different statistics test had to be used for the analysis than for the others, as certain variables, including gender and age needed to be controlled for to avoid diluting the data or creating a bias. Therefore, a regression analysis model was employed to do this, while the remaining used a chi-square test. As well, to avoid further diluting the results only individuals with the EGFR, KRAS and 'other' mutations were used in the analysis. This is because there were too few individuals exhibiting the BRAF, PIK3CA and ALK mutations, therefore they were combined in the 'other' category with the individuals containing an unidentified mutation.

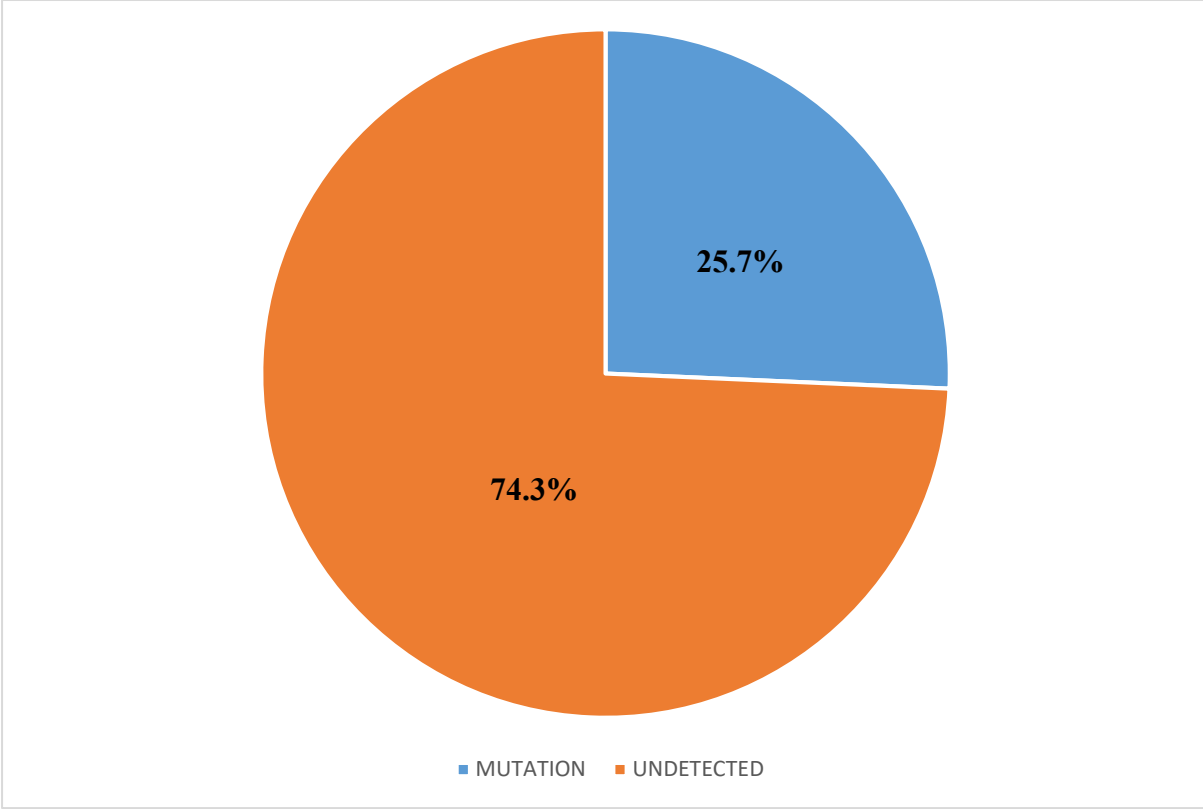

Figure 11: Mutations frequency vs undetected mutations among the 202 deceased patients who were tested. 25.7% of the patients exhibited one of the six mutations while 74.3% of the patients did not.

The patients were divided into three groups based on their age, one group of individuals' ages 39-59, a second group of individuals ages 60-74, and a third group of individuals ages 75-90. Analysis of presence of a mutation among these three age groups yielded a p-value of 0.0689, which implies weak evidence that different mutations are associated with age. Specifically, it was determined that fewer individuals in the oldest age group exhibited the BRAF mutation than expected (Table 10).

Table 10: Statistics table of age by mutation. The chi-square test yielded a p-value of 0.0689.

Age	*Mutation*							
	ALK	**BRAF**	**EGFR**	**KRAS**	**PIK3CA**	**HER2**	**OTHERS**	**Total**
39-59								
Frequency	0	1	13	45	2	0	86	147
Expected	0.2204	1.3223	9.4768	35.924	1.7631	0	98.294	
Chi-square	0.2204	0.786	1.3099	2.2933	0.0318	0	1.5376	
Row %	0.00	0.68	0.68	30.61	1.36	0.00	58.50	
60-74								
Frequency	1	1	18	89	4	0	260	373
Expected	0.5592	3.3553	24.046	91.153	4.4738	0	249.41	
Chi-square	0.3474	1.6534	1.5204	0.0508	0.0502	0	0.4495	
Row %	0.27	0.27	4.83	23.86	1.07	0.00	69.71	
75-90								
Frequency	0	**4**	12	29	2	0	100	147
Expected	0.2204	**1.3223**	9.4768	35.924	1.7631	0	98.294	
Chi-square	0.2204	**5.4221**	0.6718	1.3344	0.0318	0	0.0296	
Row %	0.00	**2.72**	8.16	19.73	1.36	0.00	68.03	
Total	1	6	43	163	8	0	446	667

Examining presence of a mutation between the two genders yielded a p-value of 0.0070, which implies reasonable evidence that different mutations are associated with gender. Specifically, it was determine that more females exhibited the EGFR mutation than expected, while fewer males did (Table 11).

Table 11: Statistics table of gender by mutation. The chi-square test yielded a p-value of 0.0070.

Gender	*Mutation*							
	ALK	**BRAF**	**EGFR**	**KRAS**	**PIK3CA**	**HER2**	**OTHERS**	**Total**
Female								
Frequency	1	3	**31**	89	3	0	200	327
Expected	0.4902	2.9415	**21.081**	79.912	3.9220	0	218.16	
Chi-square	0.5301	0.0012	**4.6670**	1.0335	0.2167	0	1.5117	
Row %	0.31	0.92	**9.51**	27.30	0.92	0.00	61.04	
Male								
Frequency	0	3	**12**	74	5	0	246	340
Expected	0.5105	3.0631	**21.952**	83.213	4.0841	0	227.18	
Chi-square	0.5105	0.0013	**4.5117**	1.0201	0.2054	0	1.5596	
Row %	0.00	0.88	**3.53**	21.76	1.47	0.00	72.35	
Total	1	6	43	163	8	0	445	667

Cancer type has been divided into adenocarcinoma, squamous cell carcinoma, large cell carcinoma, pleomorphic carcinoma and carcinoid (miscellaneous types). Analysis of the presence of a mutation among each type yielded a p-value of <0.0001, which implies strong evidence that different mutations are associated with different cancer types. Specifically, a higher number of KRAS and EGFR mutations were observed among adenocarcinoma cases, while the tested panel of mutations were uncommon in squamous cell carcinoma (Table 9). In certain cases, invasion of the tumour cells into different areas of the body occurred. These invasions involve the pleura (Table 12), vasculature (Table 13) and lymphatics (Table 14). Examining presence of a mutation against individuals exhibiting pleural, vascular or lymphatic invasion and those who do not yielded p-values of 0.5070, 0.0072 and 0.1163, respectively, which implies no evidence that different mutations are associated with pleura or lymphatic invasion, but reasonable evidence with vascular invasion. Specifically, fewer individuals with the EGFR mutation exhibited vascular invasion than expected.

Table 12: Statistics table of pleural invasion by mutation. The chi-square test yielded a p-value of 0.5070. Five of the cases did not have this information unavailable.

Pleural invasion	*Mutation*							
	ALK	**BRAF**	**EGFR**	**KRAS**	**PIK3CA**	**HER2**	**OTHERS**	**Total**
No								
Frequency	1	5	33	128	4	0	341	512
Expected	0.7734	4.6405	33.257	124.52	6.1873	0	342.62	
Chi-square	0.0664	0.0279	0.002	0.0973	0.7732	0	0.007	
Row %	0.20	0.98	6.45	25.00	0.78	0.00	66.60	
Yes								
Frequency	0	1	10	33	4	0	102	150
Expected	0.2266	1.3595	9.7432	36.48	1.8127	0	100.38	
Chi-square	0.2266	0.0951	0.0068	0.332	2.6394	0	0.0262	
Row %	0.00	0.67	6.67	22.00	2.67	0.00	68.00	
Total	1	6	43	161	8	0	443	662

Table 13: Statistics table of vascular invasion by mutation. The chi-square test yielded a p-value of 0.0072. Ten of the cases did not have this information unavailable.

Vascular invasion	*Mutation*							
	ALK	**BRAF**	**EGFR**	**KRAS**	**PIK3CA**	**HER2**	**OTHERS**	**Total**
No								
Frequency	0	4	**35**	89	3	0	247	378
Expected	0.5753	2.8767	**24.164**	93.205	4.6027	0	252.58	
Chi-square	0.5753	0.4386	**4.8588**	0.1898	0.5581	0	0.1231	
Row %	0.00	1.06	**9.26**	23.54	0.79	0.00	65.34	
Yes								
Frequency	1	1	**7**	73	5	0	192	279
Expected	0.4247	2.1233	**17.836**	68.795	2.2973	0	186.42	
Chi-square	0.7795	0.5943	**6.5829**	0.2571	0.7561	0	0.1667	
Row %	0.36	0.36	**2.51**	26.16	1.79	0.00	68.82	
Total	1	5	42	162	8	0	449	657

Table 14: Statistics table of lymphatic invasion by mutation. The chi-square test yielded a p-value of 0.1163. Eleven of the cases did not have this information unavailable.

Lymphatic invasion	*Mutation*							
	ALK	**BRAF**	**EGFR**	**KRAS**	**PIK3CA**	**HER2**	**OTHERS**	**Total**
No								
Frequency	1	5	35	111	4	0	293	449
Expected	0.6845	3.4223	28.747	110.88	5.4756	0	299.79	
Chi-square	0.1455	0.7274	1.3602	0.0001	0.3977	0	0.1538	
Row %	0.22	1.11	7.80	24.72	0.89	0.00	65.26	
Yes								
Frequency	0	0	7	51	4	0	145	207
Expected	0.3155	1.5777	13.253	51.119	2.5244	0	138.21	
Chi-square	0.3155	1.5777	2.9503	0.0003	0.8626	0	0.3335	
Row %	0.00	0.00	3.38	24.64	1.93	0.00	70.05	
Total	1	5	42	162	8	0	438	656

Invasion can also occur in the lymph nodes, which are divided into two groups based on their location in the body. N1 are the ipsilateral peribronchial, hilar and intrapulmonary nodes, and N2 are the ipsilateral mediastinal and subcarinal nodes. Examining presence of a mutation against individuals exhibiting invasion in one of these groups, or none at all, yielded a p-value of 0.9108, which implies no evidence that different mutations are associated with lymph node invasion (Table 15).

Table 15: Statistics table of lymph node invasion by mutation. The chi-square test yielded a p-value of 0.9108. Fifty of the cases did not have this information unavailable.

Lymph node invasion	*Mutation*							
	ALK	**BRAF**	**EGFR**	**KRAS**	**PIK3CA**	**HER2**	**OTHERS**	**Total**
None								
Frequency	1	4	25	96	4	0	261	391
Expected	0.6337	3.1686	24.715	95.057	5.0697	0	262.36	
Chi-square	0.2117	0.2182	0.003	0.0094	0.2257	0	0.007	
Row %	0.26	1.02	6.39	24.55	1.02	0.00	66.75	
N1/not N2								
Frequency	0	1	8	32	4	0	102	147
Expected	0.2383	1.1912	9.2917	35.737	1.906	0	98.635	
Chi-square	0.2383	0.0307	0.1796	0.3909	2.3006	0	0.1148	
Row %	0.00	0.68	5.44	21.77	2.72	0.00	69.39	
N2/not N1								
Frequency	0	0	3	6	0	0	13	22
Expected	0.0357	0.1783	1.3906	5.3485	0.2853	0	14.762	
Chi-square	0.0357	0.1783	1.8626	0.0794	0.2853	0	0.2103	
Row %	0.00	0.00	13.64	27.27	0.00	0.00	59.09	
N1 and N2								
Frequency	0	0	3	16	0	0	38	57
Expected	0.0924	0.4619	3.6029	13.857	0.7391	0	38.246	
Chi-square	0.0924	0.4619	0.1009	0.3313	0.7391	0	0.0016	
Row %	0.00	0.00	5.26	28.07	0.00	0.00	66.67	

All tumours are designated a stage based on their size, involvement of nodes and progression of metastasis. The seven stages are IA, IB, IIA, IIB, IIIA, IIIB, and IV. Analysis of presence of a mutation among these stages yielded a p-value of 0.9527, which implies no evidence that different mutations are associated with differing tumour stages (Table 8).

A number of patients had previous cases of cancer in other parts of the body, unrelated to the present lung cancer. Analysis of presence of a mutation against cancer history (previous cases of cancer or none) yielded a p-value of 0.9471, which implies no evidence that different mutations are associated with cancer history (Table 16).

Table 16: Statistics table of history of cancer by mutation. The chi-square test yielded a p-value of 0.9471. 135 of the cases did not have this information unavailable.

History of cancer	*Mutation*							
	ALK	**BRAF**	**EGFR**	**KRAS**	**PIK3CA**	**HER2**	**OTHERS**	**Total**
No								
Frequency	1	3	28	87	5	0	286	410
Expected	0.7707	3.0827	26.974	90.169	5.3947	0	283.61	
Chi-square	0.0682	0.0022	0.0391	0.1114	0.0289	0	0.0202	
Row %	0.24	0.73	6.83	21.22	1.22	0.00	69.76	
Yes								
Frequency	0	1	7	30	2	0	82	122
Expected	0.2293	0.9173	8.0263	26.831	1.6053	0	84.391	
Chi-square	0.2293	0.0075	0.1312	0.3743	0.0971	0	0.0677	
Row %	0.00	0.82	5.74	24.59	1.64	0.00	67.21	
Total	1	4	35	117	7	0	368	532

A number of patients had a history of cancer among their family members as well. Analysis of presence of a mutation against history of familial cancer (previous cases of cancer or none) yielded a p-value of 0.2862, which implies no evidence that different mutations are associated with history of familial cancer (Table 17).

Table 17: Statistics table of history of family cancer by mutation. The chi-square test yielded a p-value of 0.2862. 145 of the cases did not have this information unavailable.

History of family cancer	*Mutation*							
	ALK	**BRAF**	**EGFR**	**KRAS**	**PIK3CA**	**HER2**	**OTHERS**	**Total**
No								
Frequency	1	2	28	87	6	0	298	422
Expected	0.8084	4.0421	28.295	88.927	5.659	0	294.27	
Chi-square	0.0454	1.0317	0.0031	0.0418	0.0205	0	0.0473	
Row %	0.24	0.47	6.64	20.62	1.42	0.00	70.62	
Yes								
Frequency	0	3	7	23	1	0	66	100
Expected	0.1916	0.9579	6.705	21.073	1.341	0	69.732	
Chi-square	0.1916	4.3539	0.013	0.1763	0.0867	0	0.1997	
Row %	0.00	3.00	7.00	23.00	1.00	0.00	66.00	
Total	1	5	35	110	7	0	364	522

Smoking history of the patients has been classified as Current/Past smoker versus Non-smoker, or more precisely, ever smoked versus never smoked. Investigation of the type of mutation in relation to smoking history yielded a p-value of <0.0001, which is very strong evidence of an association. Specifically, a higher number of individuals with the EGFR mutation than expected were non-smokers (Table 18). This result was further investigated and confirmed using the LOGISTIC and CATMOD procedures in SAS, adjusting for age and gender. Since neither age nor gender nor their interaction effect were significant in these analyses, they were dropped from the respective models. As a result, we were left merely with investigating the association between a binary smoking history variable and a binary mutation variable (EGFR versus other). The resulting odds ratio estimate is 10.5 with 95% Wald confidence limits of (4.5, 24.1). In other words, the odds of never smoking was 10.5 times higher in the group of patients with an EGFR mutation than in the group with other mutations. It is unlikely that this magnitude of association in a 2x2 table with cell frequencies of 11, 21, 24 and 479 is merely an artifact of one of the cells having an expected frequency (2.0935) that is less than 5.

Table 18: Statistics table of smoking history by mutation. The chi-square test yielded a p-value of <0.0001. 132 of the cases did not have this information unavailable.

Smoking history	*Mutation*							
	ALK	**BRAF**	**EGFR**	**KRAS**	**PIK3CA**	**HER2**	**OTHERS**	**Total**
Never								
Frequency	0	0	**11**	4	0	0	17	32
Expected	0.0598	0.2991	**2.0935**	7.1178	0.4187	0	22.011	
Chi-square	0.0598	0.2991	**37.893**	1.3657	0.4187	0	1.1409	
Row %	0.00	0.00	**34.38**	12.50	0.00	0.00	53.13	
Ever								
Frequency	1	5	24	115	7	0	351	503
Expected	0.9492	4.7009	32.907	111.88	6.5813	0	345.99	
Chi-square	0.0038	0.019	2.4107	0.0869	0.0266	0	0.0726	
Row %	0.20	0.99	4.77	22.86	1.39	0.00	69.78	
Total	1	5	35	119	7	0	368	535

The survival analysis was conducted in two stages. First, a univariate proportional hazards model was fitted with the sole predictor being molecular mutation categorized as either EGFR, KRAS or OTHER. The likelihood ratio test for equality of survival across these three categories yielded a statistically significant p-value of 0.0327. Specifically, the mortality rate among those with the EGFR mutation was less than in the other two categories (Table 19).

Table 19: Univariate analysis of survival by mutation. Individuals with ALK, BRAF and PIK3CA mutations were grouped with Other due to the small sample sizes.

Mutation	**Number of patients**	**Number of deaths**	**Hazard Ratio (vs Other)**	**P-value**
EGFR	43	6	0.376	**0.0189**
KRAS	163	40	0.787	0.1777
Other	461	156	-	-
EGFR vs KRAS vs Other	-	-	-	**0.0327**

A multivariate proportional hazards analysis was then performed controlling for potential confounders. A number of factors were independently predictive of survival (Table 20). Older patients experienced shorter survival than younger patients. Males had shorter survival than females. The statistical significance of the age*gender interaction is due to younger females having slightly shorter survival times than predicted by the model with just age and gender main effects. Individuals with tumours at a more advanced stage had shorter survival than those with less advanced staging. Those with pleural invasion tended to have shorter survival than those without pleural invasion. Similarly, the greater the nodal invasion, the shorter the survival. The remaining factors were not significantly associated with survival. In particular, no association was found between any specific mutation and survival when controlling for the aforementioned variables.

Table 20: Multivariate analysis of survival. Hazard ratios are expressed per unit increment in the risk category. For example, the hazard ratio for age 60-74 versus <60 and for age 75+ versus 60-74 was 1.642. The hazard ratio for males versus females was 1.529. NS denotes a non-significant p-value using a cutoff of 0.1.

Effect	**Hazard Ratio**	**P-Value**
Age (<60, 60-74, 75+)	1.642 (p<0.0001)*	<0.0001
Gender (F, M)	1.529 (p=0.0047)*	0.0026
Age*Gender interaction	-	0.0325
Staging (I, II, III, IV)	1.644	<0.0001
Pleural involvement (No, Yes)	1.335	0.0900
Nodal involvement (0, 1, 2)	1.270	0.0365
Vascular involvement (No, Yes)	-	NS
Lymphatic involvement (No, Yes)	-	NS
Cancer history (0, 1)	-	NS
Smoking history (0,1)	-	NS
History of family cancer	-	NS
Celltype (Adeno, Carc, LCC, PLE, Sq)	-	NS
Mutation (EGFR, KRAS, Other)	**-**	**NS**

* These two hazard ratios and their associated p-values were obtained from a model without the age*gender interaction term. All other hazard ratios and p-values were obtained from a model that included the age*gender interaction. Neither the model with interaction nor the model without interaction included the variables with $p > 0.1$ (NS).

CHAPTER 4: DISCUSSION

SNaPshot analysis has proven to be an effective method for determining the molecular data of tissue samples. Firstly, it has high analytical sensitivity, performance rate, mutation detection rate and relatively inexpensive[84]. It can be done rapidly with a minimal amount of DNA unlike with direct sequencing, which requires a much higher amount[88]. The ability to analyze for multiple mutations at a time is another benefit to using SNaPshot. Not only does it make it more cost effective, eliminating the need to run multiple different analyses on one sample, analysing for multiple mutations provides preliminary data to guide further research on treatment options. In previous studies, therapies that target the EGFR mutation, which have been developed, rarely have an effect on the KRAS mutation[89]. KRAS mutations are more commonly found in tumours than EGFR mutations, which increases the need to develop personalized treatments. Such therapies are currently underway and have currently reached the clinical trials level[90]. By analyzing for the mutations without current therapies, and even those that do, allows researchers to continue working towards developing these personalized therapies beyond standard chemoradiation therapy.

Gene mutations can be broken down further into specific exon mutations within the genes. Certain exon mutations are more commonly observed than others for specific genes, which aided in determining which mutations to analyze for in this study in order to maximize results. For instance, exon 9 mutations (E545K or E542K) within PIK3CA are some of the most commonly observed[91]. Among the EGFR gene, the most prevalent mutation types are the exon 19 deletion and L858R point mutation (exon 21) which account for approximately 90% of all the EGFR mutations[92,93]. The exon 20 insertion is a

significantly rarer EGFR mutation, accounting for 9% of EGFR mutations[93]. However, this particular mutation has proven resistant to developed TKI treatments (Erlotinib and Gefitinib), requiring the development of new treatments[94,95].

The frequencies of mutations, especially EGFR and ALK, in our data are lower than those found in other studies in the literature, while KRAS is similar to those published by Sequist, *et al.* However, this study did not distinguish which exon mutations of each gene were analyzed for, which would suggest they only tested for the most common ones.[46] The exact reason for the difference in results is unclear. One of the possible explanations is that in other studies, the examined tissues were mainly obtained from advanced stage lung cancers while in our study the cases were relatively early stage lung cancers. In addition, the cases reported in the literature were often using small biopsy specimens whereas in our study surgically resected cancer tissue were used. Although the statistical analysis of mutations in different tumour stages in our cohort produced non-significant results, it may be related to the sample size since most of our cases are at relatively early stage of the disease (Table 20). It is also worth noting that when observing mutation frequencies among adenocarcinoma cases, which is the most frequently observed subtype, the results were lower than those observed in previous studies except the KRAS mutation, which was 10% higher. These results even contradict those from a similar study done at the Queen Elizabeth II Health Sciences Centre, led by another cohort. Their study examined only adenocarcinoma cases and observed mutation frequencies similar to those in previously published studies[46]. This further implies the smaller sample size in this study has influenced the results.

As expected from the literature, only a small percentage of the squamous cell carcinoma exhibited any mutations we screened for (~5%)[96]. As well, of the seven cases exhibiting a PIK3CA mutation four of them were squamous cell carcinoma (Figure 9) and the remaining three were adenocarcinoma (Figure 8). This was expected as these mutations are equally as common among both cancer types[97]. This specific mutation has more commonly been observed in oral squamous cell carcinoma (OSCC), being the oral cavity, pharynx, and larynx, although typically in advanced stages. This would suggest the mutation has a larger role in tumour progression as opposed to its initiation[98]. A broader examination of head and neck squamous cell carcinoma (HNSCC) reported a higher frequency of the PIK3CA mutation as well, although the highest concentration was found in pharyngeal cancers[99]. The EGFR mutation has also been observed in lung squamous cell carcinoma cases at a relatively high frequency[100]. It is possible the results differed in this study due to a smaller sample size of squamous cell carcinoma and early stages of the tumours.

Although reasonable, the evidence that a specific mutation affects survival has not always been consistent with previous studies. KRAS mutations in particular has not exhibited a negative effect on survival rate of patients as it has in other studies, which would have suggested a poor prognosis[101]. EGFR mutations have been shown in previous research to have a higher survival rate, which was also observed in this study[46]. However, this correlation did not control for other effects on the patient data, including age and gender, as when a regression model was used which incorporated other important variables there was no significant correlation between the observed frequencies. This could indicate

that the clinical and pathological data of the patients have a larger impact on survival rate than the specific mutations themselves.

The gene mutations tested in this study, especially EGFR, ALK, and KRAS, are thought to be mutually exclusive[102]. There are instances, however, where multiple mutations have been observed in one tumour. In a previous study involving 552 samples, the resulting analysis showed 134 tumours with a KRAS mutation and 73 with an EGFR mutation. However, five of the tumours with a KRAS mutation and two of the tumours with an EGFR mutation also exhibited a PIK3CA mutation[46]. It has also been observed in a previous study that 70% of PIK3CA mutations among lung cancer, specifically adenocarcinoma, are exhibited simultaneously with an EGFR, KRAS or ALK mutation and are linked to a poor prognosis for the patients[103]. Three patients exhibiting both a PIK3CA E542K and EGFR L858R mutation were observed to be resistant to EGFR tyrosine-kinase inhibitor (TKI) treatments, further supporting the need for development of new personalized treatments[104].

The very strong evidence linking smoking history to mutation status was reported in the literature. Studies have shown relationships between smoking and the mutations, with a significant correlation between cigarette smoking and the KRAS mutation. The same results were found between genders, stating there was a higher association between females and the KRAS mutation[105,106,107]. Neither of these correlations were observed in this study. However, the opposite trend for smoking history was observed with the EGFR mutation, in that if the individual had this mutation there was a higher likelihood that they had never smoked. Previous studies suggest they make up a distinct subset of lung cancers[108,109]. The “never smoking” lifestyle of the patient is not an effective predictor to

whether the patient will benefit from tyrosine-kinase inhibitors however, making molecular profiling still more reliable[110].

Although none of the cases in this study exhibited the T790M EGFR mutation, it has been observed in another series of cases. Typically, this mutation is viewed as a "second-site mutation", appearing in more than 50% of EGFR-mutated lung cancer patients who have previously received Erlotinib or Gefitinib. The tumours that develop the T790M mutation have acquired resistance to these treatments[111,112]. It has been noted, however, that after a period of time without either treatment this resistant mutation may disappear on its own, allowing for the previous treatment to be resumed effectively[46]. The lung study cases in our centre exhibiting this mutation have not been previously treated before developing it, meaning it would have occurred during initial tumour formation. This is a rare occurrence, as in previous studies this mutation is found in less than 5% of untreated EGFR-mutated tumours[113]. In these cases, the T790M EGFR mutation usually occurs in conjunction with another EGFR mutation, resulting in decreased sensitivity to treatments[114]. At our centre for example, the T790M EGFR was observed with the L858R EGFR mutation within the same tumour.

An important limitation to observe throughout this study is the number of individuals exhibiting these mutations. While there was a significant number of individuals who exhibited the KRAS mutation, 163 in total, only one individual exhibited the ALK mutation. The small sample sizes have the potential to skew the overall data, making it necessary to recognize this constraint when drawing conclusions. Overall patient survival was also measured from disease onset which occasionally predated the recorded date of diagnosis, whereas the clinical/pathological characteristics of the patient

would have only been recorded at diagnosis. Date of onset was used instead of date of diagnosis because the latter was not available for all patients used in the study. Finally, patients who reached the point for a five-year follow up could not always be reached. Therefore it was impossible to tell if they were actually living or not. If it could not be confirmed that the patient was deceased they were listed as living in the study, which may not have been true.

Another limitation to the study included the inability to evaluate all the various types of mutations found among each gene type. For instance, while we analyzed for the EGFR L858R and T790M point mutations, and exon 19 deletion and exon 20 insertion, there are other types of EGFR mutations not included in this study. These include point mutations G719C/S/A in exon 18, V765A AND T783A in exon 20, and L861Q in exon 21, to name a few[115]. Due to certain constraints (i.e. funds, resources, etc.), it was impossible to analyze for all possible mutation types in these genes, and therefore only the most significant mutation types were chosen for this study.

This study has provided an analysis of lung cancer tumours at early stages, indicating the frequency of mutations and specific correlations with patient data. Evidence also suggests that these mutation status may have a potential influence on the five year survival rate of the patients. This not only provides preliminary data on the study of the individuals mutations, but aids in guiding future research in personalized therapies which can lead to an increased rate of successful cancer treatments. Due to the results differing from those in previously published data also causes the hypothesis to be rejected, and could lead future cancer-related studies for Nova Scotians in a different direction.

BIBLIOGRAPHY

1. Cooper G. M. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; 2000.

2. Metastatic Cancer. *National Cancer Institute at the National Institutes of Health.* March 28, 2013. http://www.cancer.gov/about-cancer/understanding/what-is-cancer/metastatic-fact-sheet Accessed June 5, 2016.

3. What are primary and secondary cancers? *Macmillan Cancer Support.* April 1, 2014. http://www.macmillan.org.uk/cancerinformation/cancertypes/unknownprimary/aboutcup/primarysecondarycancers.aspx#DynamicJumpMenuManager_6_Anchor_2 Accessed June 5, 2016.

4. Benign tumors. *MedlinePlus.* April 13, 2016. https://www.nlm.nih.gov/medlineplus/benigntumors.html Accessed May 2, 2016.

5. What is Cancer? *National Cancer Institute at the National Institutes of Health.* March 7, 2014. http://www.cancer.gov/cancertopics/cancerlibrary/what-is-cancer Accessed February 3, 2015

6. Croce, C. M. Oncogenes and Cancer. *N Engl J Med*. 2008; 358(5): 502-11.

7. Bertram, J. The molecular biology of cancer. *Mol. Aspects Med.* 2000; 21 (6): 167–223.

8. The p53 tumour suppressor protein. *Genes and Disease [Internet]*. http://www.ncbi.nlm.nih.gov/books/NBK22268/ Accessed June 4, 2016.

9. Vogelstein, B., Sur, S., and Prives, C. p53 : The Most Frequently Altered Gene in Human Cancers. *Nature Education.* 2010; 3(9): 16.

10. Rivlin, N., Brosh, R., Oren, M., *et al.* Mutations in the p53 Tumour Suppressor Gene. *Genes Cancer*. 2011; 2(4): 466-74.

11. Lodish, H., Berk, A., Zipursky, S. L., *et al.* Molecular Cell Biology. 4th edition. New York: W. J. Freeman; 2000.

12. Goodsell, D. S. The Molecular Perspective: The *ras* Oncogene. *The Oncologist.* 1999; 4(3): 263-4.

13. Lloyd, E. Examples of Oncogenes. *Bright Hub.* January 7, 2010. http://www.brighthub.com/science/genetics/articles/76140.aspx Accessed June 5, 2016.

14. Introna, M. and Golay, J. How can oncogenic transcription factors cause cancer: a critical review of the myb story. *Leukemia.* 1999; 13(9): 1301-6.

15. Wolfe, A. Chromatin remodeling: why it is important in cancer. *Oncogene.* 2001; 20(24): 2988-90.

16. Goustin, A. S., Leof, E. B., Shipley, G. D., *et al.* Growth Factors and Cancer. *Cancer Research.* 1986; 46(3): 1015-29.

17. Mosesson, Y. and Yarden, Y. Oncogenic growth factor receptors: implications for signal transduction therapy. *Semin Cancer Bio.* 2004; 14(4): 262-70.

18. Chauhan, A. K. and Varma, A. A Textbook of Molecular Biotechnology. First edition. New Delhi, India: I.K. International Publishing House Pvt. Ltd.; 2009.

19. Kromer, G. The proto-oncogene Bcl-2 and its role in regulating apoptosis. *Nature medicine.* 1997; 3(6): 614-20.

20. Libermann, T. A. and Zerbini, L. F. Targeting transcription factors for cancer gene therapy. *Curr Gene Ther.* 2006; 6(1): 17-33.

21. Davis, P. K. and Brackmann, R. K. Chromatin Remodeling and cancer. *Cancer Biol Ther.* 2003; 2(1): 22-9.

22. Witsch, E., Sela, M., and Yarden, Y. Roles for growth factors in cancer progression. *Physiology (Bethesda)*. 2010; 25(2): 85-101.

23. Sierke, S. L., Longo, G. M., and Koland, J. G. Structural basis of interactions between epidermal growth factor receptor and SH2 domain proteins. *Biochem Biophys Res Commum.* 1993; 191(1): 45-54.

24. Sgambato, A., Casaluce, F., Maione, P., *et al.* The role of EGFR tyrosine kinase inhibitors in the first-line treatment of advanced non small cell lung cancer patients harboring EGFR mutation. *Curr Med Chem.* 2012; 19(20): 3337-52.

25. Aggarwal, B. B., Kunnumakkara, A. B., Harikumar, K. B., *et al.* Signal transducer and activator of transcription-3, inflammation, and cancer: how intimate is the relationship? *Ann N Y Acad Sci.* 2008; 1171: 59-76.

26. Guicciardi, M. E. and Gores, G. J. Life and death by death receptors. *FASEB J.* 2009; 23(6): 1625-37.

27. Herr, I. and Debatin, K. M. Cellular stress response and apoptosis in cancer therapy. *Blood.* 2001; 98(9): 2603-14.

28. Kufe, D. W., Pollock, R. E., Weichselbaum, R. R., *et al.* Holland-Frei Cancer Medicine. Sixth edition. Hamilton, ON: BC Decker; 2003.

29. Griffiths, A. J. F., Gelbart, W. M., Miller, J. H., *et al.* Modern Genetic Analysis. First edition. New York: W. H. Freeman; 1999.

30. Tsai, A. G. and Lieber, M. R. Mechanisms of chromosomal rearrangement in the human genome. *BMC Genomics.* 2010; 11(Suppl 1): S1.

31. Albertson, D. G. Gene amplification in cancer. *Trends Genet.* 2006; 22(8): 447-55.

32. Nicolson G. L. Generation of phenotypic diversity and progression in metastatic tumor cells. *Cancer Metastasis Rev.*1984; 3(1): 25–42.

33. Luo, S. Y. and Lam, D. C. L. Oncogenic driver mutations in lung cancer. *Translational Respiratory Medicine.* 2013; 1: 6.

34. Calderwood, S. K. Tumor heterogeneity, clonal evolution, and therapy resistance: an opportunity for multitargeting therapy. *Discov Med*. 2013; 15(82): 188-94.

35. Dexter, D. L. and Leith, J. T. Tumor heterogeneity and drug resistance. *J Clin Oncol.* 1986; 4(2): 244-57.

36. Heppner, G. H. and Miller B. E. Therapeutic implications of tumor heterogeneity. *Semin Oncol.* 1989; 16(2): 91-105.

37. Somasundaram, R., Villanueva, J., and Herlyn, M. Intratumoral heterogeneity as a therapy resistance mechanism: role of melanoma subpopulations. *Adv Pharmacol.* 2012; 65: 335-59.

38. Lung Cancer Staging. *American Joint Committee on Cancer.* Atlanta, GA: Atlanta Cancer Society; 2009.

39. Cancer Facts and Figures 2007. *American Cancer Society.* Atlanta, GA: American Cancer Society; 2007.

40. Schiller, J. H., Harrington, D., Belani, C. P., *et al.* Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *N Engl J Med.* 2002; 346(2): 92-8.

41. Kumar, V., Abbas, A. K., Fausto, N., *et al.* Robbins and Cotran Pathologic Basis of Disease. Eighth edition. Philadelphia, PA: Saunders/Elsevier; 2010.

42. Dooley, A. L., Winslow, M. M., Chiang, D. Y., *et al.* Nuclear factor I/B is an oncogene in small cell lung cancer. *Genes Dev.* 2011; 25(14): 1470-5.

43. Afrose, R., Akram, M., Karimi, A. M., *et al.* Correlation of age and gender with different histological subtypes of primary lung cancer. *Med J DY Patil Univ.* 2015; 8(4): 447-451.

44. Srigley, J. R. Benign mimickers of prostatic adenocarcinoma. *Modern Pathology.* 2004; 17: 328-48.

45. Cooney, E. Cracking Squamous Cell Lung Cancers. *Harvard Medical School.* September 10, 2012. http://hms.harvard.edu/news/cracking-squamous-cell-lung-cancers-9-10-12 Accessed September 29, 2014.

46. Sequist, L. V., Heist, R. S., Shaw, A. T., *et al.* Implementing multiplexed genotyping of non-small-cell lung cancers into routine clinical practice. *Annals of Oncology.* 2011; 22(12): 2616-24.

47. Nomura, D. (2013). Chemical Carcinogenesis: Initiation, Promotion and Progression [PowerPoint Slides]. Retrieved from http://nature.berkeley.edu/~dnomura/pdf/Lecture7Carcinogenesis.pptx

48. Cheng, L., Alexander, R. E., MacLennan, G. T., *et al.* Molecular pathology of lung cancer: key to personalized medicine. *Modern Pathology.* 2012; 25(3): 347-69.

49. Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol*. 2001; 2(2): 127–37.

50. Sharma S. V., Bell D. W., Settleman J., Haber D. A. Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer* 2007; 7(3): 169–81.

51. Lemmon, M. A. and Schlessinger, J. Cell signalling by receptor-tyrosine kinases. *Cell*. 2010; 141(7): 1117-34.

52. Leto, S. M. and Trusolino, L. Primary and acquired resistance to EGFR-targeted therapies in colorectal cancer: impact on future treatment strategies. *J Mol Med (Berl).* 2014; 92(7): 709-22.

53. Hilger, R. A., Scheulen, M. E., and Strumberg, D. The Ras-Raf-MEJ-ERK Pathway in the Treatment of Cancer. *Onkologie.* 2002; 25(6): 511-8.

54. Zenonos, K. and Kyprianou, K. RAS signalling pathways, mutations and their role in colorectal cancer. *World J Gastrointest Oncol.* 2013; 5(5): 97-101.

55. Chappell, W. H., Steelman, L. S., Long, J. M., *et al.* Ras/Raf/MEK/ERK and PI3K/PTEN/Akt/mTOR inhibitors: rationale and importance to inhibiting these pathways in human health. *Oncotarget.* 2011; 2(3): 135-64.

56. Siegelin, M. D. and Borczuk, A. C. Epidermal growth factor receptor mutations in lung adenocarcinoma. *Laboratory Investigation*. 2014; 94(2): 129-37.

57. Dawson, J. P., Berger, M. B., Lin, C., *et al.* Epidermal Growth Factor Receptor Dimerization and Activation Require Ligand-Induced Conformational Changes in the Dimer Interface. *Mol Cell Biol.* 2005; 25(17): 7734-42.

58. LoPiccolo, J., Blumenthal, G. M., Bernstein, W. B., *et al.* Targeting the PI3K/Akt/mTOR pathway: Effective combinations and clinical considerations. *Drug Resistance Updates.* 2008; 11(1-2): 32-50.

59. Shaw, A. T., and Solomon, B. Anaplastic lymphoma kinase (ALK) fusion oncogene positive non-small cell lung cancer. *Clin Cancer Res.* 2011; 17(8): 2081.

60. Doebele, R. C., Pilling, A. B., Aisner, D. L., *et al.* Mechanisms of Resistance to Crizotinib in Patients with ALK Gene Rearranged Non-Small Cell Lung Cancer. *Clin Cancer Res.* 2012; 18(5): 1472-82.

61. Soda, M., Choi, Y. L., Enomoto, M., *et al*. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature*. 2007; 448(7153): 561-6.

62. Hernádez, L., Pinyol, M., Hernádez, S., *et al.* TRK-fused gene (TFG) is a new partner of ALK in anaplastic large cell lymphoma producing two structurally different TFG-ALK translocations. *Blood.* 1999; 94(9): 3265-8.

63. Takeuchi, K. Choi, Y. L., Togashi, Y., *et al.* KIF5B-ALK, a novel fusion oncokinase identified by an immunohistochemistry-based diagnostic system for ALK-positive lung cancer. *Clin Cancer Res.* 2009; 15(9): 3143-9.

64. Grande, E., Bolós, M., and Arriola, E. Targeting Oncogenic ALK: A Promising Strategy for Cancer Treatment. *Mol Cancer Ther*. 2011; 10(4): 569-79.

65. Gainor, J. F., Varghese, A. M., Ou, S. H., *et al.* ALK rearrangements are mutually exclusive with mutations in EGFR or KRAS: an analysis of 1,638 patients with non-small cell lung cancer. *Clin Cancer Res.* 2013; 19(15): 4273-81.

66. Pao, W., and Miller, V. Epidermal growth factor receptor mutations, small-molecule kinase inhibitors, and non-small-cell lung cancer: current knowledge and future directions. *J Clin Oncol*. 2005; 10(23): 2556–68.

67. Zhang, A., Stiegler, A. L., Boggon, T. J. EGFR-mutated lung cancer: a paradigm of molecular oncology. *Oncotarget.* 2010; 1(7): 497-514.

68. Westcott, P. M. K., and To, M. D. The genetics and biology of *KRAS* in lung cancer. *Chin J Cancer I.* 2013; 32(2): 63-70.

69. Suzuki, M., Shigematsu, H., Iizasa, T., *et al.* Exclusive mutation in epidermal growth factor receptor gene, HER-2, and KRAS, and synchronous methylation of nonsmall cell lung cancer. *Cancer*. 2006; 106(10): 2200-7.

70. Pao W., Wang T. Y., Riely G. J., *et al.* KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. *PLoS Med* 2005; 2(1): e17.

71. Heinmöller, P., Gross, C., Beyser, K., *et al.* HER2 status in non-small cell lung cancer: results from patient screening for enrollment to a phase II study of herceptin. *Clin Cancer Res*. 2003; 9(14): 5238-43.

72. Freudlsperger, C., Burnett, J. R., Friedman, J. A., *et al.* EGFR-PI3K-AKT-mTOR signaling in head and neck squamous cell carcinomas: attractive targets for molecular-oriented therapy. *Expert Opin Ther Targets*. 2011; 15(1): 63–74.

73. Tsai, J., Lee, J. T., Wang, W., *et al.* Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc Natl Acad Sci USA.* 2008; 105(8): 3041-46.

74. Falchook, G. S., Long, G. V., Kurzrock, R., *et al*. Dabrafenib in patients with melanoma, untreated brain metastases, and other solid tumours: a phase 1 dose-escalation trial. *Lancet*. 2012; 379(9829): 1893-901.

75. Angulo, B., Suarez-Gauthier, A., Lopez-Rios, F., *et al.* Expression signatures in lung cancer reveal a profile for EGFR-mutant tumours and identify selective PIK3CA overexpression by gene amplification. *J Pathol*. 2008; 214(3): 347-56.

76. Okudela, K., Suzuki, M., Kageyama, S., *et al. PIK3CA* mutation and amplification in human lung cancer. *Pathol Int.* 2007; 57(10): 664-71.

77. Kawano, O., Saskia, H., Okuda, K., *et al. PIK3CA* gene amplification in Japanese non-small cell lung cancer. *Lung Cancer.* 2007; 58(1): 159-60.

78. Yamamoto, H., Shigematsu, H., Nomura, M., *et al. PIK3CA* Mutations and Copy Number Gains in Human Lung Cancers. *Cancer Res.* 2008; 68(17): 6913-21.

79. Cataldo, V. D., Gibbons, D. L., Pérez-Soler, R., *et al.* Treatment of Non-Small-Cell Lung Cancer with Erlotinib or Gefitinib. *N Engl J Med.* 2011; 364(10): 947-55.

80. Kim, E. S., Khuri, F. R., and Herbst, R. S. Epidermal growth factor receptor biology (IMC-C225). *Curr Opin Oncol.* 2001; 13(6): 506-13.

81. Shaw, A. T., Kim, D. W., Nakagawa, K., *et al.* Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med.* 2013; 468(25): 2385-94.

82. Stewart, E. L., Tan, S. Z., Liu, G., *et al.* Known and putative mechanisms of resistance to EGFR targeted therapies in NSCLC patients with EGFR mutations – review. *Transl Lung Cancer Res.* 2015; 4(1): 67-81.

83. Lindblad-Toh, K., Winchester, E., Daly, M. J., *et al.* Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat Genet.* 2002; (4): 381-6.

84. Wu, C. C., Lu, Y. C., Chen, P. J., *et al.* Application of SNaPshot multiplex assays for simultaneous multigene mutation screening in patients with idiopathic sensorineural hearing impairment. *Laryngoscope*. 2009; 119(12): 2441-6.

85. Cristofaro, J. D., Silvy, M., Chiaroni, J., *et al.* Single PCR Multiplex SNaPshot Reaction for Detection of Eleven Blood Group Nucleotide Polymorphisms. *J Mol Diagn.* 2010; 12(4) 453-60.

86. 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. *Ann Oncol.* 2012; 23(11) 2907-13.

87. Cox, D. R. Regression Models and Life Tables. *JRSS(B)*. 1972; 34(2): 187-220.

88. Su, Z., Dias-Santagata, D., Duke, M., *et al.* A Platform for Rapid Detection of Multiple Oncogenic Mutations With Relevance to Targeted Therapy in Non-Small-Cell Lung Cancer. *J Mol Diagn.* 2001; 13(1): 74-84.

89. Kang, S., Bader, A. G., and Vogt, P. K. Phosphatidylinositol 3-kinase mutations identified in human cancer are oncogenic. *Proc Natl Acad Sci U S A.* 2005; 102(3): 802-7.

90. Pao, W., Miller, V., Zakowski, M., *et al.* EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumours to gefitinib and erlotinib. *Proc Natl Acad Saci U S A.* 2004; 101(36): 13306-11.

91. Arcila, M. A., Nafa, K., Chaft, J. E. *et al. EGFR* exon 20 insertion mutations in lung adenocarcinomas: Prevalence, molecular heterogeneity, and clinicopathologic characteristics. *Mol Cancer Ther.* 2013; 12(2): 220-9.

92. Parsons, B. L. and Myers, M. B. Personalized cancer treatment and the myth of KRAS wild-type colon tumors. *Discov Med.* 2013; 15(83): 259-67.

93. Jänne, P. A., Shaw, A. T., Pereira, J. R., *et al.* Selumetinib plus docetaxel for KRAS-mutant advanced non-small-cell lung cancer: a randomised, multicentre, placebo-controlled, phase 2 study. *Lancet Oncol.* 2013; 14(1): 38-47.

94. Engelman, J. A., Zejnullahu, K., Gale, C. M., *et al.* PF00299804, an irreversible pan-ERBB2 inhibitor, is effective in lung cancer models with EGFR and ERBB2 mutations that are resistant to geftinib. *Cancer Res.* 2007; 67(24): 11924-32.

95. Greulich, H., Chen, T. H., Feng, W., *et al.* Oncogenic transformation by inhibitor-sensitive and resistant EGFR mutants. *PLoS Med*. 2005; 2:e313.

96. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature.* 2012; 489(7417): 519-25.

97. Kawano, O., Saskia, H., Endo, K., *et al.* PIK3CA mutation status in Japanese lung cancer patients. *Lung Cancer.* 2006; 54(2): 209-15.

98. Kozaki, K., Imoto, I., Pimkhaokham, A., *et al. PIK3CA* mutation is an oncogenic aberration at advanced stages of oral squamous cell carcinoma. *Cancer Science*. 2006; 97(12): 1351-58.

99. Qiu, W., Schöneleben, F., Li, X., *et al. PIK3CA* Mutations in Head and Neck Squamous Cell Carcinoma. *Clin Cancer Res.* 2006; 12(5): 1441-6.

100. Hendler, F. J. and Ozanne, B. W. Human squamous cell lung cancers express increased epidermal growth factor receptors. *J Clin Invest.* 1984; 74(2): 647-51.

101. Guan, J. L., Zhong, W. Z., An, S. J., *et al.* KRAS mutation in patients with lung cancer: a predictor of poor prognosis but not for EGFR-TKIs or chemotherapy. *Ann Surg Oncol.* 2013; 20(4): 1381-8.

102. Hofman, P., Ilie, M., Hofman, V., *et al.* Immunohistochemistry to identify *EGFR* mutations or *ALK* rearrangements in patients with lung adenocarcinoma. *Annals of Oncology.* Oxford, UK: Oxford University Press; 2011.

103. Chaft, J. E., Arcila, M. E., Paik, P. K.*, et al.* Coexistence of PIK3CA and other oncogene mutations in lung adenocarcinoma-rationale for comprehensive mutation profiling. *Mol Cancer Ther.* 2012; 11(2): 485-91.

104. Han, J-Y, Kim, S. H., Lee, S. Y., *et al.* Comparison of targeted next-generation sequencing with conventional sequencing for predicting the responsiveness to epidermal growth factor receptor-tyrosine kinase inhibitor (EGFR-TKI) therapy in never-smokers with lung adenocarcinoma. *Lung Cancer.* 2014; 85(2): 161-7.

105. Nelson, H. H., Christiani, D. C., Mark, E. J., *et al.* Implications and prognostic value of *K-ras* mutation for early-stage lung cancer in women. *J Natl Cancer Inst.* 1999; 91(23): 2032–8.

106. Nelson, H. H., Christiani, D. C., Wiencke, J. K., *et al. K-ras* mutation and occupational asbestos exposure in lung adenocarcinoma: asbestos-related cancer without asbestosis. *Cancer Res.* 1999; 59(18): 4570–3.

107. Husgafvel-Pursiainen, K., Karjalainen, A., Kannio, A., *et al.* Lung cancer and past occupational exposure to asbestos: role of *p53* and *K-ras* mutations. *Am J Respir Cell Mol Biol.* 1999; 20(4): 667–74.

108. Shigematsu, H., Lin, L., Takahashi, T., *et al.* Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst.* 2005; 97(5): 339-46.

109. Gazdar, A. F. Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. *Oncogene.* 2009; (Suppl 1): S24-31.

110. Jackman, D.M., Miller, V.A., Cioffredi, L.A., *et al*. Impact of epidermal growth factor receptor and *KRAS* mutations on clinical outcomes in previously untreated non-small cell lung cancer patients: Results of an online tumor registry of clinical trials. *Clin Cancer Res*. 2009; 15(16): 5267-73.

111. Kobayashi, S., Boggon, T. J., Dayaram, T., *et al.* EGFR mutation and resistance of non-small-cell lung cancer to geftinib. *N Engl J Med*. 2005; 352(8): 786-92.

112. Pao, W., Miller, V. A., Politi, K. A., *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; 2(3): e73.

113. Inukai, M., Toyooka, S., Ito, S., *et al.* Presence of epidermal growth factor receptor gene T790M mutation as a minor clone in a non-small cell lung cancer. *Cancer Res.* 2006; 66(16): 7854-8.

114. Wu, J. Y., Yu, C. J., Chang, Y. C., *et al.* Effectiveness of tyrosine kinase inhibitors on “uncommon” epidermal growth factor receptor mutations of unknown clinical significance in non-small cell lung cancer. *Clin Cancer Res*. 2011; 17(11): 3812-21.

115. Reddi, H. V. Mutations in the EGFR Pathway. *AACC.* October 1, 2013. https://www.aacc.org/publications/cln/articles/2013/october/egfr-mutations Accessed June 5, 2016.