UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN

MASTER OF SCIENCE IN
BIOMEDICAL ENGINEERING

MASTER's THESIS

DEVELOPMENT OF A BIOINFORMATIC PIPELINE FOR THE DETECTION OF SOMATIC MUTATIONS IN LIQUID BIOPSY SAMPLES FROM NON-SMALL CELL LUNG CANCER PATIENTS

VADYM IVANCHUK

2020

MASTER's THESIS

Title . . . . . .

Development of a bioinformatic pipeline for the detection of somatic mutations in liquid biopsy samples from non-small cell lung cancer patients

Author . . . . . .

Vadym Ivanchuk

Supervisor . . . . . .

Atocha Romero Alfonso
Estela Sánchez Herrero

Tenured professor . . . . . .

Milagros Ramos Gómez

Department . . . . . .

Tecnología Fotónica y Bioingeniería

COURT

President . . . . . .

Vocal . . . . . .

Secretary . . . . . .

Substitute . . . . . .

DATE |

GRADE |

UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN

MASTER OF SCIENCE IN
BIOMEDICAL ENGINEERING

MASTER's THESIS

DEVELOPMENT OF A BIOINFORMATIC PIPELINE FOR THE DETECTION OF SOMATIC MUTATIONS IN LIQUID BIOPSY SAMPLES FROM NON-SMALL CELL LUNG CANCER PATIENTS

VADYM IVANCHUK

2020

Abstract

Lung cancer is the most common cancer worldwide, accounting for 2.1 million new cases and 1.8 million deaths in 2018. Therefore, it is one of the most lethal types since more than half of people die within one year of being diagnosed.

Non-small cell lung cancer (NSCLC) is the most prevalent type, accounting for about 80–90%. Besides, more than 50% of these patients are diagnosed in an advanced stage or have gained metastases and thus they have lost the opportunity for surgical treatment. In this way, targeted therapy has gradually been paid more attention to and has become the first-line treatment program.

Currently, optimal management of NSCLC requires tumors to be periodically examined for a variety of predictive and prognostic biomarkers. However, tissue biopsies are an invasive procedure limited to certain locations, not easily acceptable in the clinic, and generally unable to reflect the clonal heterogeneity of the tumor. In this context, liquid biopsy has emerged as a solid alternative to this traditional invasive technique, using different body fluids (peripheral blood, cerebrospinal fluid, etc.) of cancer patients as a source of cancer-derived material.

Over the last decade, the recognition of molecularly defined tumor subsets with unique sensitivities to targeted therapeutics has transformed the management of patients with ALK-mutated lung adenocarcinoma, helping the clinicians to move towards personalized medicine. Molecular therapy is now the standard of care and patient outcomes have greatly improved. However, acquired resistance to targeted agents has become a pivotal issue limiting the long-term benefit of such therapies. Therefore, it becomes evident that real-time detection technologies applied to cancer patients are needed urgently in clinical practice to guide the treatment at the

molecular level after mutations or resistance emergence.

In this context, this project incorporated into standard clinical practice an automatic way to filter and select genetic variants by analyzing and interpreting next-generation sequencing (NGS) data. Specifically, the developed algorithm is capable of identifying clinically relevant deviations from a reference genome in clinical samples analyzed by NGS, allowing clinicians to act more easily on genomic information at the point of patient care. Currently, the analysis of the raw data obtained from the sequencer is not standardized, and before the implementation of the algorithm, it was performed manually, with all the drawbacks that this entailed.

The algorithm design and evaluation were carried out on a study population of 30 patients with advanced NSCLC. After performing a digital PCR confirmation that detected 8 subjects with some type of mutation in the ALK gene, the algorithm showed an accuracy of 82.95% (for a prevalence of ALK mutations of 20%), identifying 7 (87.5%) out of 8 patients as carriers of an ALK alteration. Overall, this study has managed to combine clinical procedures and bioinformatic analysis, obtaining a direct impact at four different levels: (i) for the detection of oncogenic drivers; (ii) for the identification of resistance mutations in patients relapsing on targeted therapies; (iii) for monitoring response to therapies; and (iv) for the prediction of clinical outcomes.

Keywords

Liquid biopsy, circulating tumor DNA (ctDNA), non-small cell lung cancer (NSCLC), targeted therapy, anaplastic lymphoma kinase (ALK), next-generation sequencing (NGS), digital PCR (dPCR), bioinformatics

Resumen

Los tumores pulmonares son el tipo de cáncer más común, contabilizándose en 2018 2.1 millones de casos nuevos y 1.8 millones de muertes. Es además uno de los más letales al fallecer más de la mitad de los pacientes en el término de un año desde que son diagnosticados.

El cáncer de pulmón no microcítico (CPNM) es el tipo más frecuente, representando alrededor del 80–90% de todos los casos. Más del 50% de estos pacientes son diagnosticados en una etapa avanzada o metastásica, imposibilitando un tratamiento quirúrgico eficaz. En este contexto, la terapia dirigida se ha postulado como el tratamiento de primera línea.

Actualmente, el manejo óptimo del CPNM requiere una examinación periódica de los tumores con el fin de detectar biomarcadores predictivos y pronósticos específicos. Sin embargo, las biopsias de tejido son un procedimiento invasivo limitado a ciertas localizaciones, no bien acogidas en la práctica clínica y generalmente incapaces de reflejar la heterogeneidad del tumor. Ante estas limitaciones han emergido las técnicas de biopsia líquida, que emplean diferentes fluidos corporales (sangre periférica, líquido cefalorraquídeo, etc.) de pacientes con cáncer como fuente del material derivado de los tumores.

A lo largo de la última década, la identificación de subconjuntos moleculares con sensibilidades únicas a fármacos específicos ha transformado completamente el tratamiento de los pacientes con adenocarcinoma pulmonar ALK+, favoreciendo los avances hacia una medicina personalizada. La terapia molecular es ahora el estándar en estos pacientes, mejorándose su calidad de vida y supervivencia. No obstante, el beneficio a largo plazo de estas terapias está limitado por la aparición de resistencias

adquiridas a los fármacos administrados. Por ello, se hace evidente la necesidad de tecnologías de seguimiento en tiempo real que guíen el tratamiento a nivel molecular tras el surgimiento de mutaciones de resistencia.

En este contexto, este estudio ha conseguido incorporar a la práctica clínica una forma automática de filtrar y seleccionar variantes genéticas mediante el análisis e interpretación de los datos obtenidos de la secuenciación de nueva generación (NGS). El algoritmo desarrollado ha sido capaz de identificar desviaciones genéticas clínicamente relevantes en muestras de biopsia líquida analizadas por NGS, permitiendo a los oncólogos actuar más decididamente sobre la información genómica de cada caso. Actualmente, el análisis de los datos brutos obtenidos del secuenciador no está estandarizado y antes de la implementación del algoritmo se realizaba manualmente, con todos los inconvenientes que esto conllevaba.

El proceso de diseño y evaluación del algoritmo se realizó en una población de estudio de 30 pacientes con CPNM avanzado. Después de realizarse una confirmación digital por PCR, que identificó a 8 sujetos con mutaciones en el gen ALK, el algoritmo mostró una precisión del 82.95% (para una prevalencia de mutaciones ALK del 20%), siendo capaz de detectar las alteraciones genéticas en 7 (87.5 %) de los 8 pacientes ALK+.

En definitiva, este estudio ha logrado combinar eficazmente los procedimientos clínicos y el análisis bioinformático, teniendo un impacto directo en múltiples niveles: (i) para detectar mutaciones conductoras; (ii) para identificar mutaciones de resistencia; (iii) para monitorizar la respuesta a las terapias; y (iv) para realizar un pronóstico.

Palabras clave

Biopsia líquida, ADN tumoral circulante (ctDNA), cáncer de pulmón no microcítico (CPNM), terapia dirigida, cinasa del linfoma anaplásico (ALK), secuenciación de nueva generación (NGS), PCR digital (dPCR), bioinformática

Glossary

AF	Allele Frequency
AKT	Protein Kinase B
ALCL	Anaplastic Large-Cell Lymphoma
ALK	Anaplastic Lymphoma Kinase
AMP	Association of Molecular Pathologists
BBB	Blood-Brain Barrier
BRAF	V-Raf Murine Sarcoma Viral Oncogene Homolog B1
CAP	College of American Pathologists
cDNA	Complementary DNA
cfDNA	Circulating Free DNA
CNS	Central Nervous System
CNV	Copy-Number Variation
CSF	Cerebrospinal Fluid
CTC	Circulating Tumor Cell
ctDNA	Circulating Tumor DNA
CTLA-4	Cytotoxic T-Lymphocyte-Associated Antigen 4
dbSNP	Database of Single Nucleotide Polymorphisms
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
dPCR	Digital PCR
ECOG	Eastern Cooperative Oncology Group
EGFR	Epidermal Growth Factor Receptor
EMA	European Medicines Agency
EML4	Echinoderm Microtubule-Associated Protein-Like 4
EMT	Epithelial-Mesenchymal Transition
EpCAM	Epithelial Cell Adhesion Molecule
ePCR	Emulsion PCR

ERK	Extracellular Signal-Regulated Kinase
ERRB	Avian Erythroblastosis Oncogene B
FDA	Food and Drug Administration
FGF	Fibroblast Growth Factor
FGFR1	Fibroblast Growth Factor Receptor 1
FISH	Fluorescence in Situ Hybridization
GCO	Global Cancer Observatory
GUI	Graphical User Interface
HER2	Human Epidermal Growth Factor Receptor 2
HMBS	Hydroxymethylbilane Synthase
IASLC	International Association for the Study of Lung Cancer
IGF-1R	Insulin-Like Growth Factor 1 Receptor
IGV	Integrative Genomics Viewer
IHC	Immunohistochemistry
InDel	Insertions and Deletions
ISP	Ion Sphere™ Particle
JAK	Janus Kinase
KRAS	Kirsten Rat Sarcoma
LOD	Limit of Detection
MAF	Mutant Allele Fraction
MAPD	Median of the Absolute values of all Pairwise Differences
MAPK	Mitogen-Activated Protein Kinase
MEK	Mitogen-Activated Protein Kinase Kinase
MET	Mesenchymal-Epithelial Transition
MNP	Multiple-nucleotide polymorphism
mRNA	Messenger RNA
mTOR	Mammalian Target of Rapamycin
NGS	Next-Generation Sequencing
NPM	Nucleophosmin
NRG1	Neuregulin 1
NSCLC	Non-Small Cell Lung Cancer
NTRK1	Neurotrophic Tyrosine Kinase Receptor 1
ORR	Overall Response Rate
OS	Overall Survival
PCR	Polymerase Chain Reaction
PD-1	Programmed Cell Death Protein 1

PD-L1	Programmed Death-Ligand 1
PDGF	Platelet-Derived Growth Factor
PDGFRA	Platelet-Derived Growth Factor Receptor A
PFS	Progression-Free Survival
PI3K	Phosphoinositide 3-Kinase
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
qPCR	Quantitative PCR
RAF	Rapidly Accelerated Fibrosarcoma
RAS	Rat Sarcoma
RET	Rearranged During Transfection
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
ROS1	Reactive Oxygen Species Proto-Oncogene 1
RR	Response Rate
RT-PCR	Reverse Transcription PCR
SCLC	Small Cell Lung Cancer
SNP	Single-Nucleotide Polymorphism
SNV	Single-Nucleotide Variant
STAT	Signal Transducer and Activator of Transcription Proteins
TBP	TATA-Binding Protein
TEP	Tumor-Educated Platelet
TKI	Tyrosine Kinase Inhibitor
TP53	Tumor Protein 53
UMI	Unique Molecular Identifier
VATS	Video-Assisted Thoracoscopic Surgery
VEGF	Vascular Endothelial Growth Factor
VEGFR	Vascular Endothelial Growth Factor Receptor
WT	Wild Type

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Cancer

1.1.1 Background

Cancer is a very broad term that covers more than 200 types of diseases [1]. Each of them can have completely different characteristics from each other and can be considered as independent diseases with their causes, evolution, and specific treatments. However, they all have a common denominator: cancer cells acquire the ability to multiply and spread throughout the body without control.

Cancer arises from the transformation of normal cells into tumor cells in a process that could be divided into at least four consecutive stages: initiation, promotion, progression, and metastasis (Figure 1.1) [2]. Initiation involves the alteration, change, or mutation of genes arising spontaneously or induced by exposure to a carcinogenic agent. The promotion stage is considered to be a relatively lengthy and reversible process in which actively proliferating preneoplastic cells accumulate, while progression is the phase between a premalignant lesion and the development of invasive cancer. Progression is the final stage of neoplastic transformation, where the cells may undergo further genetic and phenotypic changes with invasive and metastatic potential. Finally, metastasis, which is responsible for as much as 90% of cancer-associated mortality, involves the spread of cancer cells from the primary site to other parts of the

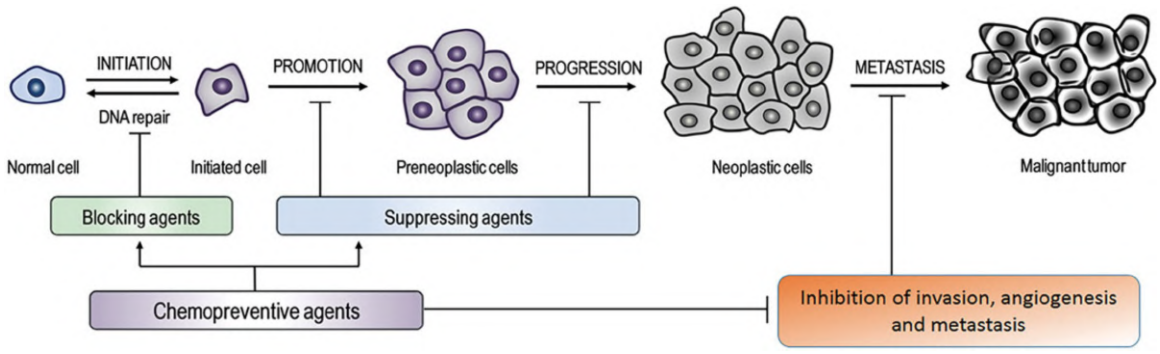


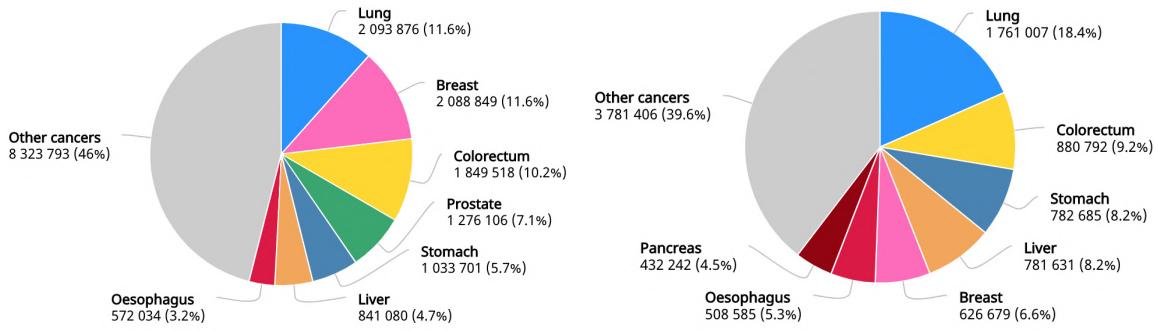
Figure 1.1. Carcinogenesis phases: initiation, promotion, progression, and metastasis. Modified image from [2].

body through the bloodstream or the lymph system [3].

During this multi-step development of tumors, cancer cells acquire some biological capabilities that are considered the hallmarks of cancer [4]. They include sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, activating invasion and metastasis, reprogramming of energy metabolism, and evading immune destruction. Furthermore, underlying these hallmarks are genome instability, which generates the genetic diversity that expedites their acquisition and inflammation, which fosters multiple hallmark functions. All these characteristics constitute an organizing principle to rationalize the complexities of this neoplastic disease and are widely recognized and applicable in cancer treatments, especially in targeted therapies.

These biological changes in normal cells are primarily the result of the interaction between a person's genetic factors and 3 categories of external agents, including physical carcinogens (ultraviolet and ionizing radiation), chemical carcinogens (asbestos, components of tobacco smoke, etc.), and biological carcinogens (infections from certain viruses, bacteria or parasites) [1]. Aging is another fundamental factor for the development of cancer since its incidence rises dramatically with age [5].

Finally, the magnitude of this problem is reflected by the Global Cancer Observatory (GCO) in the fact that this pathology is the leading cause of death worldwide, accounting for an estimated 9.6 million deaths in 2018 [5]. Furthermore, every year about 18.1 million new cases are diagnosed, being the lung and breast cancers the most



(a) Number of new cases.

(b) Number of deaths.

Figure 1.2. Cancer incidence and mortality statistics worldwide in 2018. Images from [6].

common types with a joint incidence of 23.2% (Figure 1.2a). However, when it comes to mortality, lung cancer stands out above the rest with 1.76 million deaths, doubling even the numbers of the second most lethal cancer, the colorectal (Figure 1.2b).

1.1.2 Lung Cancer

Lung cancer is a group of diseases resulting from the malignant growth of cells of the respiratory tract, in particular of the lining epithelium of the bronchi, bronchioles, or alveoli. In the last century, it has progressed from an uncommon disease to the most prevalent cancer in the world and the most common cause of death from cancer. Its mortality is very close to its incidence since only 20.5% of people are still alive 5 years after being diagnosed in the United States [7]. Furthermore, according to a population-based study of lung cancer survival and stage at diagnosis worldwide, only 15–20% of cases are diagnosed when the tumor can be removed by surgical resection [8]. In 20-25% of people, the neoplasm is detected when it has a regional extension and can still be cured by combined chemotherapy and radiation therapy techniques. However, in 55–65% of cases, this type of cancer is diagnosed at an advanced stage when only palliative care treatments or participation in clinical trials are considered, which are focused on treating the symptoms rather than cancer itself.

This progressive tumor development, which dramatically decreases patients' survival, can spread and proliferate in three main ways [7]:

- **Local growth**, which can eventually lead to an invasion through the lung walls of some of the surrounding structures, such as the heart, great vessels, esophagus, or vertebral bodies, depending on the initial location of the neoplasm.

- **Lymphatic dissemination**. It is done through the lymph. The presence of tumors in the middle and lower third of the lungs tend to proliferate mainly to the mediastinal lymph nodes, whereas tumors in the upper third are more likely to affect the supraclavicular lymph nodes.

- **Hematogenous dissemination**. The propagation of tumor cells is carried out through the blood vessels and can potentially affect the liver, adrenal glands, brain, and bones.

To diagnose lung cancer and determine whether it has spread to other parts of the body, many different tests are used and some of them can also help to decide which treatments might work best. The steps to make an effective diagnosis usually include studying the patients' medical history, laboratory tests, imaging tests, biopsies, and biomarker tests. However, currently, only a biopsy can provide a definite diagnosis of lung cancer.

In this way and as a previous step to treatment, it is essential to make an accurate histological diagnosis of the tumor as each one of its types presents a different natural evolution. The most common forms of lung cancer are classified according to the characteristics of the cells from which they are derived [9], distinguishing two large groups:

- **Small cell lung cancer (SCLC)**. It is a very aggressive malignant neoplasm that accounts for about 15% of lung cancer cases and is characterized by a rapid doubling time, high growth ratio, and a greater propensity for early development of widespread metastases [10]. The origin of SCLC tumor cells has not been formally identified, although they are commonly thought to arise from neuroendocrine cells in the bronchus [11]. Finally, this neoplasm has been described mainly in the elderly with a history of long exposure to tobacco.

- **Non-small cell lung cancer (NSCLC)**. It is the most common type of lung cancer, accounting for approximately 85% of all cases, and it is the most common form in women, young adults, and people who haven't smoked [12]. Furthermore, this histological subtype of lung cancer is not as lethal as SCLC, since it usually grows and spreads more slowly. In this sense, depending on the staging, this kind of patients are often eligible for a wider variety of treatments ranging from surgery to radiation therapy, chemotherapy, immunotherapy, and targeted therapy.

Finally, despite the severity of this disease, it is largely preventable since there is a close relationship between the amount of exposure to tobacco carcinogens and the subsequent appearance of lung cancers several years or decades later [13]. It is estimated that tobacco smoking explains almost 90% of lung cancer risk in men and 70–80% in women [14].

1.2 Non-Small Cell Lung Cancer (NSCLC)

1.2.1 Histological Classification

There are three major types of non-small cell lung cancer:

- **Squamous-cell carcinoma**, which comprises 25–30% of all NSCLC cases [12]. It arises from early versions of squamous cells in the airway epithelial cells in the central part of the lungs, that is, in the bronchi. Therefore, squamous cell carcinoma tends to cause symptoms earlier since it affects the larger airways of the lungs. On the other hand, although this type of cancer is generally slow-growing, it has the potential to spread to multiple sites due to its initial location, including the brain, spine and other bones, adrenal glands, and liver [15]. In addition, this subtype is strongly correlated with cigarette smoking, more than any other form of NSCLC [13, 14].

- **Adenocarcinoma**. It is the most common type of lung cancer, accounting for approximately 40% of NSCLC [12]. Compared to the other forms, it tends

to grow more slowly and has a greater chance of being found before spreading outside the lungs. It usually starts in glandular cells, which secrete mucus and other substances, and tends to develop in smaller airways, such as alveoli. Additionally, it is normally located more along the outer edges of the lungs and as a result, it frequently affects the pleura and chest wall [16]. Currently, it is not only the most common type in smokers, but it is also the most common form of lung cancer in women, Asians, and people under the age of 45 [17]. Finally, in recent years, this histological variant has gained special interest with the discovery of a series of molecular alterations in a subgroup of patients that is allowing them to be treated with targeted therapies [18].

- **Large cell (undifferentiated) carcinoma**. It accounts for 5–10% of lung cancers [12]. However, as more exact ways of diagnosing lung cancer have come into use, many tumors previously labeled as undifferentiated large cell carcinoma can now be classified more appropriately as poorly differentiated adenocarcinoma or squamous cell carcinoma [19]. For this reason, the incidence of this type of tumor continues to decrease and it is often diagnosed by default through the exclusion of other possibilities and where is no evidence of squamous or glandular maturation. It usually begins in the epithelial cells on the outer edges of the lungs and tends to grow and spread rapidly [20], which can make it harder to treat. Furthermore, and as in the other types, it is concluded that cigarette smoking is the predominant cause of large cell lung cancer [21].

Classifying NSCLC tumors into histological subtypes was the first step towards a better distinction of the molecular, phenotypical, and prognostic features of NSCLC and became the first approach to develop personalized treatment strategies based on the tumor characteristics. However, the development of molecular biology techniques applied to the study of cancer revealed the complexity of genomic and epigenetic differences between these histological subtypes, unraveling the existence of potent oncogenic drivers in lung adenocarcinomas and giving rise to the era of targeted therapies for this disease [22].

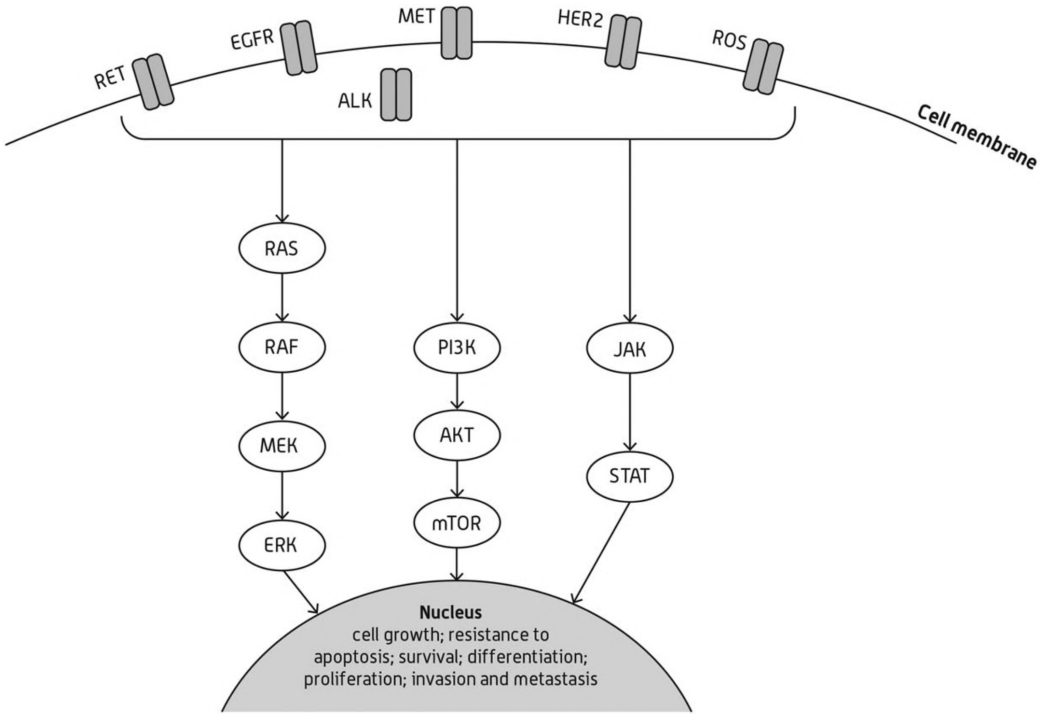


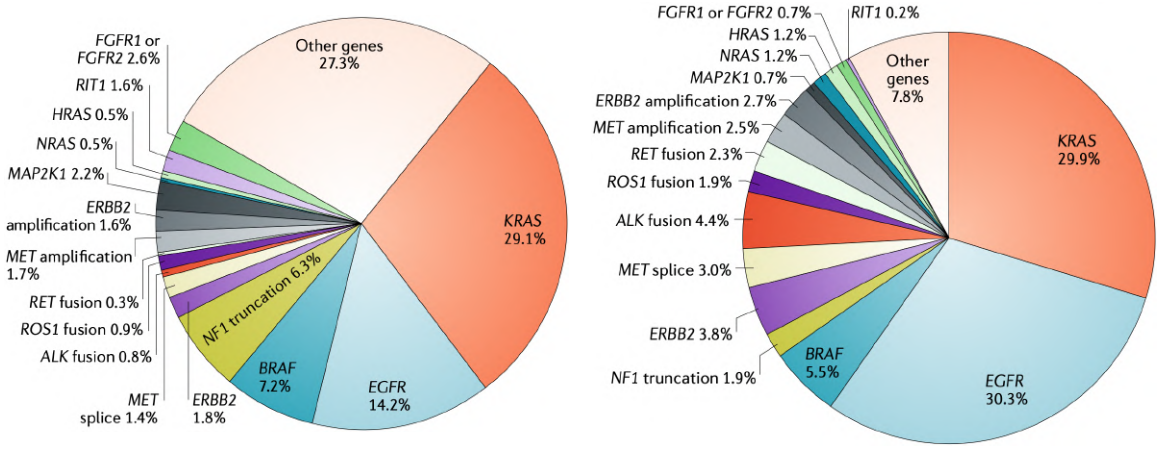
Figure 1.3. Molecular pathways of non-small cell lung cancer. AKT, protein kinase B; ALK, anaplastic lymphoma kinase; EGFR, epidermal growth factor receptor; ERK, extracellular signal-regulated kinase; HER2, human epidermal growth factor receptor 2; JAK, janus kinase; MEK, mitogen-activated protein kinase kinase; MET, mesenchymal-to-epithelial transition; mTOR, mammalian target of rapamycin; PI3K, phosphoinositide 3-kinase; RAF, rapidly accelerated fibrosarcoma; RAS, rat sarcoma; RET, rearranged during transfection; ROS, reactive oxygen species; STAT, signal transducer and activator of transcription proteins. Modified image from [25].

1.2.2 Molecular Classification of Lung Adenocarcinomas

Lung adenocarcinomas are classified according to the presence of well-identified and characterized molecular alterations that drive cancer initiation and progression. This genomic classification relies on the detection of fusions, copy-number variations, insertions and deletions, single-nucleotide variants, and multiple-nucleotide polymorphisms in key genes. Ultimately, these alterations translate at the functional level into the activation of oncogenes and the inactivation of tumor suppressor genes [23,24], which play a crucial role in signal transduction, cellular proliferation, differentiation, and other regulatory mechanisms (Figure 1.3).

The most relevant driver mutations leading to the classification of genomic sub-

types of lung adenocarcinoma are EGFR, ALK, ROS1, KRAS, BRAF, RET, NTRK1, NRG1, HER2 (also named ERBB2), and MET. The prevalence of each of these mutations, as well as the notable differences between the oncogenic paradigm of the early and more advanced stages of this disease, are reflected in Figure 1.4. It must also be taken into consideration that this distribution is not uniform worldwide and that there are significant changes depending on the different geographical areas [22].



(a) Early-stage adenocarcinoma.

(b) Metastatic-stage adenocarcinoma.

Figure 1.4. Oncogenic driver paradigm of lung adenocarcinoma molecular classification. Modified images from [26].

Up to 50% of all lung cancers patients harbor a genetic alteration and 20% have a targetable mutation [27]. Therefore, whenever feasible, patients with advanced NSCLC should have tumor assessed for the presence of a driver alteration [28]. However, guidelines from the College of American Pathologists (CAP), the International Association for the Study of Lung Cancer (IASLC) and the Association of Molecular Pathologists (AMP) recommend analysis of either the primary tumor or of metastasis for EGFR and ALK, which are mutually exclusive, for all patients whose tumor contains an element of adenocarcinoma and regardless of the clinical characteristics of the subjects [28,29]. The reason is that certain genetic changes, such as EGFR mutations or ALK translocations are strongly associated with a positive targeted therapeutic response. Additionally, there are also drugs approved for mutations in the ROS1, MET, RET, BRAF (V600E mutation), and NTRK genes. For other prevalent alterations such as KRAS, there is currently no approved treatment and the clinical utility

has yet to be established [30].

Regarding co-occurring genomic alterations, they have emerged as core determinants of the molecular and clinical heterogeneity in different NSCLC subgroups, including those driven by EGFR mutations and ALK translocations. Specifically, these subgroups are dominated by co-occurring TP53 alterations, PIK3CA mutations, and FGFR1 amplifications [26].

Epidermal Growth Factor Receptor (EGFR)

Driver mutations in the EGFR gene (chromosome 7) can be found in 10–20% of the Caucasian population diagnosed with NSCLC, but are more prevalent in Asians (35–45%) [31, 32]. Patients who are women, nonsmokers, and diagnosed with a pulmonary adenocarcinoma are more likely to harbor an EGFR mutation, but it can also be found in subjects without these characteristics [22].

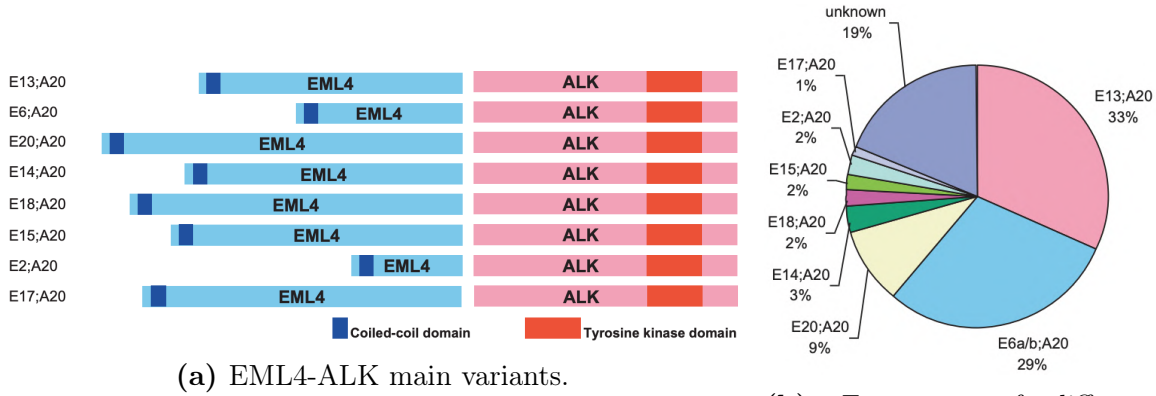
EGFR is one of the four members of the family of membrane receptors with tyrosine kinase activity known as ERBB. Activation of this receptor leads to both increased cell proliferation and slowed apoptosis by stimulating oncogenic pathways such as MAPK/ERK, also known as RAS/RAF/MEK/ERK, and PI3K/AKT/mTOR (Figure 1.3).

This type of EGFR activating mutations in NSCLC, and discovered in 2004, is located in exons 18–21, which encode part of the tyrosine kinase domain. However, 90% of these alterations are exon 19 in-frame deletions of amino acids 747–750 (45%) and exon 21 mutations resulting in L858R substitutions (40–45%) [33]. The remaining 10% of mutations involve exons 18 and 20.

Anaplastic Lymphoma Kinase (ALK)

ALK rearrangements are oncogenic drivers that occur in 3–7% of patients with NSCLC and are more common among younger patients with a light smoking history, adenocarcinoma histology, and absence of EGFR and KRAS mutations [25, 31].

ALK promotes the downstream activation of proliferative and anti-apoptotic signals via intracellular pathways [34, 35], including Ras/ERK, PI3K/AKT, and JAK/



(a) EML4-ALK main variants.

(b) Frequency of different EML4-ALK variants.

Figure 1.5. Different variants of EML4-ALK. The most common variants are E13;A20 (variant 1), E20;A20 (variant 2) and E6a/b;A20 (variant 3). The nomenclature refers to the exon in EML4 translocated to the exon in ALK. Modified images from [38].

STAT (Figure 1.3), which eventually culminate in oncogenesis.

In ALK-activated NSCLC, the predominant molecular event leading to ALK activation is the juxtaposition of the N-terminal portion of the protein encoded by the echinoderm microtubule-associated protein-like 4 (EML4) gene with the intracellular domain of the ALK tyrosine kinase [36]. This somatic rearrangement was first described as a molecular driver in a small cohort of Japanese NSCLC patients in 2007 [36]. Ever since 15 distinct variants of EML4-ALK have been identified with some of them being expressed as multiple isoforms. Variants 1, 2 and 3a/b are the most common (Figure 1.5b), accounting for around 90% of all cases [37], and as described previously, all of them include the kinase domain of ALK, encoded by exons 20–29, but differ in size based on the EML4 breakage point (Figure 1.5a).

1.2.3 International Staging System: the TNM Classification

Cancer staging is a critical step in the diagnostic process. Its objectives are:

- Aid treatment planning.
- Provide an indication of prognosis.
- Assist in the evaluation of treatment results.

- Facilitate the exchange of information between treatment centers.
- Contribute to continuing investigations of human malignancies.

The fundamental structure of stage classification is the TNM system, which since its introduction in the 1970s has undergone significant revisions with the latest, 8th edition, being effective internationally from 2018 [39]. It describes the anatomical extent of the disease and, using alphanumeric codes, NSCLC is typically given a clinical stage based on the results of a physical exam, biopsy, and imaging tests.

Tumor Staging (T)

The T stage (Table 1.1) is determined by the size of the primary tumor in the long axis measured in multiplanar reconstruction and its involvement with the adjacent structures.

Tx	Primary tumor cannot be assessed or tumor is proven by the presence of malignant cells in sputum or bronchial washings but not visualized by imaging or bronchoscopy
T0	No evidence of primary tumor
Tis	Carcinoma in situ
T1	Tumor ≤ 3 *cm* in greatest dimension, surrounded by lung or visceral pleura, without bronchoscopic evidence of invasion more proximal than the lobar bronchus (i.e., not in the main bronchus)
T1mi	Minimally invasive adenocarcinoma: solitary adenocarcinoma, ≤ 3 *cm* with a predominantly lepidic pattern and ≤ 5 mm invasion in greatest dimension
T1a	Tumor ≤ 1 *cm* in greatest dimension
T1b	Tumor > 1 *cm* but ≤ 2 *cm* in greatest dimension
T1c	Tumor > 2 *cm* but ≤ 3 *cm* in greatest dimension
T2	Tumor > 3 *cm* but ≤ 5 *cm* or having any of the following features: - Involves main bronchus regardless of distance from the carina, but without the involvement of the carina - Invades visceral pleura - Associated with atelectasis or obstructive pneumonitis that extends to the hilar region, involving part or all of the lung
T2a	Tumor > 3 *cm* but ≤ 4 *cm* in greatest dimension
T2b	Tumor > 4 *cm* but ≤ 5 *cm* in greatest dimension
T3	Tumor > 5 *cm* but ≤ 7 *cm* in greatest dimension or associated with separate tumor nodule(s) in the same lobe as the primary tumor or directly invades any of the following structures: chest wall (including the parietal pleura and superior sulcus tumors), phrenic nerve and parietal pericardium
T4	Tumor > 7 *cm* in greatest dimension or associated with separate tumor nodule(s) in a different ipsilateral lobe than that of the primary tumor or invades any of the following structures: diaphragm, mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body and carina

Table 1.1. T-descriptors of the TNM classification for lung cancer [40].

Nodal Staging (N)

The nodal staging (Table 1.2) assesses tumor burden in the regional hilar and mediastinal nodes.

Nx	Regional lymph nodes cannot be assessed
N0	No regional node metastasis
N1	Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes, including involvement by direct extension
N2	Metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s)
N3	Metastasis in the contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene or supraclavicular lymph node(s)

Table 1.2. N-descriptors of the TNM classification for lung cancer [40].

Metastasis Staging (M)

M staging (Table 1.3) is defined by the presence of metastasis beyond the regional lymph nodes.

M0	No distant metastasis
M1	Distant metastasis
M1a	Separate tumor nodule(s) in a contralateral lobe tumor; tumor with pleural or pericardial nodules or malignant pleural or pericardial effusion
M1b	Single extrathoracic metastasis in a single organ and involvement of a single nonregional lymph node
M1c	Multiple extrathoracic metastases in a single or multiple organs

Table 1.3. M-descriptors of the TNM classification for lung cancer [40].

Stage Grouping

All possible combinations of the T, N, and M categories are used to create TNM subsets (Table 1.4), which are then combined into stage groupings that share similar prognoses, ranging from one to four (I through IV).

1.2.4 Current Treatment Options

The treatment of pulmonary adenocarcinoma depends on several factors including stage, resectability, performance status, histology, and genomic alterations acquired

Occult carcinoma	TX	N0	M0
Stage 0	Tis	N0	M0
Stage IA1	T1mi	N0	M0
	T1a	N0	M0
Stage IA2	T1b	N0	M0
Stage IA3	T1c	N0	M0
Stage IB	T2a	N0	M0
Stage IIA	T2b	N0	M0
Stage IIB	T1a-c	N1	M0
	T2a	N1	M0
	T2b	N1	M0
	T3	N0	M0
Stage IIIA	T1a-c	N2	M0
	T2a-b	N2	M0
	T3	N1	M0
	T4	N0	M0
	T4	N1	M0
Stage IIIB	T1a-c	N3	M0
	T2-ab	N3	M0
	T3	N2	M0
	T4	N2	M0
Stage IIIC	T3	N3	M0
	T4	N3	M0
Stage IVA	Any T	Any N	M1a
	Any T	Any N	M1b
Stage IVB	Any T	Any N	M1c

Table 1.4. Lung cancer stage grouping [40].

by the individual tumor [12]. As in most types of cancer, treatment approaches can be divided into 5 categories: surgery, chemotherapy, radiotherapy, immunotherapy, and targeted therapy.

Surgical Resection

Surgery is the most effective therapy for stages I to II and selected cases of stage IIIA NSCLC [41]. To determine if the tumor is resectable, imaging studies and biopsies are completed as well as an evaluation of the patient to verify if the surgery will be tolerable. In this procedure, surgeons can perform a lobectomy if the tumor is located only in one lobe, or a pneumonectomy if it affects more than one lobe or the main bronchus. This latter intervention carries significant operative mortality and long-term morbidity, although outcomes are currently improving for resectable early-stage lung cancer owing to advances in minimally invasive thoracic surgery technologies such as video-assisted thoracoscopic surgery (VATS) [42]. However, despite the progress

being made in this field, a high percentage of tumors will recur, with 5-year overall survival ranging from 83% for stage IA to 36% for stage IIIA disease [40].

On the other hand, some patients who have undergone resection surgery may benefit from adjuvant therapy. It may include radiation, chemotherapy, and targeted therapy and is associated with improved survival in patients with resected stage II and IIIA NSCLC [43].

Chemotherapy

Chemotherapy has been the mainstay of treatment in patients with advanced (stage IV) and unresectable lung tumors for approximately 3 decades and it is currently given with palliative intent. The first-line therapy is platinum-based doublet chemotherapy, combining cisplatin or carboplatin with another cytotoxic agent and the regimens might be influenced by the patients' histology, age, comorbidity, performance status, response and molecular genetic features of cancer [41].

Approximately 40% of newly diagnosed lung cancer patients are stage IV and their median overall survival (OS) is 4.5 months when no chemotherapy is given [44]. However, the 1-year OS improves from 10–20% up to 30–50% with chemotherapy, reducing disease-related adverse events [45].

Radiotherapy

Radiation therapy delivers high-energy ionizing radiation to destroy cancer cells by damaging their DNA. Therefore, it can help control or eliminate tumors at specific sites in the body. In the treatment of stage I and stage II NSCLC, this treatment alone is considered only when the tumor is unresectable, being associated with a 3-year cancer-specific survival rate approaching 55% in early-stage lung cancer [46]. However, it also can be part of palliative care to improve quality of life in NSCLC patients who do not respond to surgery or chemotherapy or as adjuvant therapy for patients who have undergone a resection surgery to reduce the risk of lung cancer relapse.

Immunotherapy

Immunotherapy is a breakthrough treatment in oncology that uses the patients' immune system to fight off cancer. Since some tumor cells share characteristics with healthy cells, the goal is to stimulate the host's immune system so that it can target cancer cells, stopping or slowing their growth and preventing their spread to other parts of the body.

Therapeutic approaches that modulate the immune system in patients with lung cancer have traditionally concentrated on vaccines and have generally been ineffective. However, most recent approaches focused on a series of ligands and receptors that inhibit or stimulate the immunological synapse have been approved for the treatment of patients with metastatic NSCLC and have been shown in several randomized trials to lead to better outcomes compared to standard chemotherapy [47]. Specifically, these new strategies are targeting immune checkpoint pathways, which include the blockade of the inhibitory receptors cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) and programmed cell death 1 (PD-1) and its ligand, PD-L1.

Targeted Therapy

The identification of targetable gene alterations has dramatically transformed the management of malignant tumors, allowing individualized therapy, and leading to remarkable responses in selected patients treated with matched tyrosine kinase inhibitors (TKIs). Therefore, it is a successful strategy and results in improved OS of patients with a driver mutation as compared to patients with a driver mutation that do not receive targeted therapy [48]. In addition, the toxicity profile, in this case, is milder compared to conventional chemotherapy.

Targeted therapies using EGFR TKIs against tumors with EGFR mutations, and ALK TKIs against tumors with ALK fusions thus have become an increasingly standard treatment in clinical settings. Currently, almost all guidelines recommend the use of EGFR TKIs, such as Gefitinib (Iressa®), Erlotinib (Tarceva®) or Afatinib (Gilotrif®), and ALK TKIs, such as Crizotinib (Xalkori®) and Alectinib

(Alecensa®), as first-line treatments [22, 34, 41]. Through genomic testing, other molecular changes have been found including gene rearrangements of ROS1 and RET, amplification of MET, and activating mutations in BRAF, HER2, and KRAS genes (Table 1.5). However, most lung adenocarcinomas either lack an identifiable driver oncogene or harbor mutations that are more responsive to chemotherapy due to the lack of truly effective drugs [49].

Gene	Most Common Alteration	Prevalence	Treatment
EGFR	Exon 19 deletion and exon 21 activating mutation	15–35%	Erlotinib (Tarceva®), Afatinib (Gilotrif®), Gefitinib (Iressa®), Osimertinib (Tagrisso®), Dacomitinib (Vizimpro®)
ALK	Fusion of partner gene with exon 20 of ALK	3–7%	Crizotinib (Xalkori®), Ceritinib (Zykadia®), Alectinib (Alecensa®), Brigatinib (Alunbrig®), Lorlatinib (Lorbrena®)
KRAS	Codon 12 mutation	25–30%	AMG 510, MRTX849
BRAF	V600E mutation	3–5%	Dabrafenib (Tafinlar®) + Trametinib (Mekinist®), Vemurafenib (Zelboraf®)
ROS1	Translocation	∼ 2%	Crizotinib (Xalkori®), Ceritinib (Zykadia®), Lorlatinib (Lorbrena®), Entrectinib (Rozlytrek®)
RET	Translocation	∼ 2%	Multitarget TKIs such as Cabozantinib (Cabometyx®), Vandetanib (Caprelsa®) and Alectinib (Alecensa®), and selective RET inhibitors such as LOXO-292 (Selpercatinib®) and BLU667 (Pralsetinib®)
MET	Amplification and exon 14 skipping mutation	∼ 3%	Crizotinib (Xalkori®), Glesatinib, Capmatinib
HER2	Activating exon 20 in-frame insertion	2–3%	Poziotinib, Afatinib (Gilotrif®), Trastuzumab-emtansine (Kadcyla®)

Table 1.5. Overview of driver genetic alterations potentially amenable to targeted treatment [35, 41, 45, 50–55]. The most common alterations as well as their prevalence are shown along with the main treatments, including both those approved and marketed, and those that are undergoing clinical trials.

The presence of genetic alterations, which leads to the use of targeted agents as first-line therapy, is associated with improved progression-free survival (PFS), 10 versus 7.1 months, and OS, 16.5 versus 11.8 months, according to a nationwide French study [27]. Similarly, in the United States, the Lung Cancer Mutation Consortium conducted a study reflecting that the median survival was 3.5 years for patients with an oncogenic driver and genotype-directed therapy, compared to 2.4 years for subjects with any oncogenic drivers and therefore not eligible for targeted therapy [48].

However, despite the initial efficacy of these inhibitors, patients eventually develop acquired drug resistance within 1 to 2 years from the start of TKI treatment, regardless of the line of therapy [56]. In this way, efforts are currently focusing on identifying

new therapeutic biomarkers, generating next-generation drugs to overcome drug resistance, and developing methods to diagnose driver gene mutations using less invasive biopsy techniques, such as liquid biopsy.

In this thesis, only the identification of point mutations at the ALK locus in NSCLC patients with adenocarcinoma histology will be addressed.

1.3 ALK-Positive Non-Small Cell Lung Cancer

As mentioned in section 1.2.2, EML4-ALK variants are the most common ALK rearrangement seen in NSCLC patients.

Anaplastic lymphoma kinase is a 1,620 amino acid transmembrane protein and is a member of the insulin receptor tyrosine kinases. It was originally identified in 1994 in anaplastic large-cell lymphoma (ALCL) as a fusion to a part of the nucleophosmin (NPM) protein. On the other hand, echinoderm microtubule-associated protein-like 4 (EML4) is a cytoplasmic protein essential for the formation of microtubules and microtubule-binding protein.

EML4 and ALK are closely located genes situated only 12.7 megabases apart on the short arm of chromosome 2, 2p21 and 2p23, respectively. Inversion of this part of the short arm of chromosome 2 [Inv(2)(p21p23)], that juxtaposed the 5' end of the EML4 gene with the 3' end of the ALK gene, creates a fusion between these two genes by joining exons 1–13 of EML4 to exons 20–29 of ALK (Figure 1.5a).

As described in section 1.2.2 and section 1.2.4, since its discovery in lung adenocarcinomas in 2007 and the subsequent identification of at least 15 different variants through reverse transcription polymerase chain reaction (RT-PCR) or next-generation sequencing (NGS) [37], there has been a revolution in molecular-targeted therapy that has transformed the outlook for NSCLC patients.

1.3.1 Current Therapeutic Agents: Targeted ALK Inhibitors

Advanced NSCLC associated with the ALK fusion oncogene is highly sensitive to ALK tyrosine kinase inhibitors (TKIs), which represent a newer group of medications

ALK inhibitor	Approval Status	Clinical Trial	Median PFS (Months)	Targeted ALK Mutations
Crizotinib	FDA approved, first-line	PROFILE 1014 (phase III) Crizotinib vs. chemotherapy	10.9	L1198F
Ceritinib	FDA approved, first-line	ASCEND-4 (phase III) Ceritinib vs. crizotinib	16.6	I1171T/N L1196M S1206C/Y G1269A/S
Alectinib	FDA approved, first-line	ALEX (phase III) Alectinib vs. crizotinib	25.7	L1152P/R C1156Y/T F1174C/L/V L1196M S1206C/Y G1269A/S
Brigatinib	Accelerated FDA approval, second-line	ALTA-1L (phase III) Brigatinib vs. crizotinib	Not reached	I1151Tins L1152P/R C1156Y/T F1174C/L/V L1196M G1202R G1269A/S
Lorlatinib	Accelerated FDA approval, second- and third-line	NCT01970865 (phase I/II) Single-arm trial	11.4	I1151Tins L1152P/R C1156Y/T I1171T/N/S F1174C/L/V L1196M G1202R S1206C/Y E1210K G1269A/S

Table 1.6. ALK inhibitors and main clinical trials that are currently being carried out [57]. The progression-free survival (PFS) for each of them is also shown, as well as the main mutations against which they are directed.

that prevents intracellular signaling from the ALK alteration, leading to decreased cellular proliferation. However, due to the acquisition of secondary mutations and the activation of alternative pathways, it is common for patients who initially responded positively to treatment to relapse within a year post-treatment [41].

Currently, the main clinical options are crizotinib (Xalkori®), ceritinib (Zykadia®), alectinib (Alecensa®), brigatinib (Alunbrig®) and lorlatinib (Lorbrena®). Their main characteristics are summarized in Table 1.6.

First-Generation ALK Inhibitor: Crizotinib

Crizotinib was the first ALK TKI approved for treating NSCLC patients, and although it was originally described as a dual MET/ALK kinase inhibitor, it was later found to be active on ROS1-driven tumors.

It was initially approved in the United States by the Food and Drug Administration (FDA) in 2011 based on the initial results from the single-arm phase I (PROFILE 1001) and II trials (PROFILE 1005), which reported a response rate (RR) of more than 60% and a median progression-free survival (PFS) of 8–10 months among participants with previous treatment experience [58]. Subsequent randomized controlled trials, such as PROFILE 1014 and PROFILE 1007, compared crizotinib treatment with chemotherapy in naive and progressing after one prior platinum-based treatment patients, respectively. Improvements in PFS were reported but there was no corresponding improvement in overall survival (OS) since a high proportion of patients who experienced disease progression on chemotherapy crossed over to crizotinib treatment. During the PROFILE 1007 clinical trial, crizotinib was associated with a RR of 65% and a median PFS of 7.7 months while chemotherapy showed a RR of 20% and a median PFS of 3.0 months [59]. On the other hand, in the crizotinib group of the PROFILE 1014 study median PFS was 10.9 months and RR was 74%, while in chemotherapy PFS was 7.0, and RR was 45% [41]. Based on these data, crizotinib was approved both as first-line and as subsequent therapy in patients with ALK-positive NSCLC.

Second-Generation ALK Inhibitors: Ceritinib, Alectinib, and Brigatinib

ALK-positive NSCLC treated with crizotinib usually becomes resistant within the first year of treatment, being the central nervous system (CNS) the most common site of progression [25,41,60]. Therefore, second-generation ALK-inhibitors have been developed to overcome this acquired resistance and have been investigated both in crizotinib-refractory and in crizotinib-naive settings. However, currently, no trial has compared them head-to-head [61].

Ceritinib was the first second-generation ALK inhibitor to obtain FDA approval and was intended for patients with ALK-positive metastatic NSCLC who had become resistant to crizotinib or whose disease had progressed. In the ASCEND-4 study, it was compared with platinum-based chemotherapy in naive patients. The median PFS was 16.6 months in the ceritinib group and 8.1 months in the chemotherapy group [62].

Alectinib was the next to obtain FDA approval as a first-line treatment option after the results from the global randomized phase III ALEX trial, which demonstrated superior PFS with alectinib compared to crizotinib (25.7 versus 10.4 months [25]) in patients with previously untreated ALK-positive NSCLC. In addition, the use of alectinib showed lower toxicity and delayed CNS disease progression [63].

Brigatinib is another second-generation ALK TKI approved for second-line treatment of patients with ALK-rearranged NSCLC that progressed on crizotinib. Like alectinib, it has also demonstrated better PFS compared to crizotinib through the phase III ALTA-1L randomized trial. The one-year PFS was 67% in patients receiving brigatinib and 43% in patients receiving crizotinib [64].

Finally, despite being effective against the more common crizotinib-resistant mutations, such as the L1196M gatekeeper mutation, resistance to second-generation ALK inhibitors has also been reported (e.g., G1202R and 1151Tins mutations) [65]. Therefore, additional treatment options are necessary and these generally include third-generation ALK inhibitors such as lorlatinib.

Third-Generation ALK Inhibitor: Lorlatinib

Despite the efficacy of second-generation ALK inhibitors, and as with crizotinib, patients end up relapsing. Almost 56% of patients who progress with second-generation ALK inhibitors will end up developing ALK resistance mutations (ceritinib 54%, alectinib 53%, and brigatinib 71%), while in crizotinib-resistant patients, such secondary mutations occur only in 20% of subjects [66]. Therefore, lorlatinib might be an effective solution for patients with ALK-positive NSCLC who have become resistant to currently available TKIs.

Lorlatinib is a third-generation ALK TKI approved by the FDA in November 2018 and by the European Medicines Agency (EMA) in February 2019 that offers several advantages over second-generation TKIs. It has activity against most of the known ALK resistance mutations, including G1202R and 1151Tins, and can penetrate the blood-brain barrier (BBB) to achieve therapeutic CNS drug concentrations [60,64,65]. Data from the phase I portion of an ongoing phase I/II study (NCT01970865) have proven that lorlatinib is more potent and selective than other known ALK inhibitors, including second-generation inhibitors [67]. The overall response rate (ORR) with lorlatinib was 46% and the median PFS was 11.4 months, although the majority of patients had received two or more prior ALK TKIs. In addition, this study demonstrated the ability of lorlatinib to decrease the size of brain metastases. However, despite its effectiveness, resistance to lorlatinib also emerges [68].

1.3.2 Molecular Mechanisms of ALK-TKI Resistance

As described previously, despite treatment with crizotinib or with second- or third-generation ALK TKIs, all patients with ALK-positive NSCLC end up developing acquired resistance. To assess them, pre- and post-TKI paired biopsies are critical for the identification of specific therapeutic targets in resistant ALK-positive patients.

There are two major types of ALK TKI resistance mechanism:

- **Primary resistance**. In the PROFILE 1001 crizotinib study, approximately 10% of patients treated with crizotinib presented primary resistance, which is the lack of response to ALK-inhibitors [58]. However, currently, the mechanisms underlying primary resistance are not clearly defined.

- **Secondary resistance**. It is usually classified into two distinct groups:

  - **ALK-dependent**, which includes secondary mutations in the ALK tyrosine kinase domain or amplification of the ALK fusion gene, remaining the tumor dependent on ALK signaling.

– **ALK-independent**, including activation of bypass signaling pathways that cause tumor cells to lose dependency on ALK.

ALK-dependent Resistance: Secondary Mutations in the ALK Tyrosine Kinase Domain

Secondary mutations in kinases are a common mechanism of pharmacological resistance to kinase inhibitors, accounting for approximately 25% of all cases [56,69]. The first reported mutation was the substitution of leucine by methionine at the 1196 position (L1196M gatekeeper mutation) in the ALK tyrosine kinase domain, which eventually prevents the crizotinib binding. The remaining not-gatekeeper ALK resistance mutations include ATP-binding pocket (G1269A), N-terminus lobe (C1156Y, L1152R, and I1151Tins), C-terminus of the αC helix (F1174C/L/V) and solvent front mutations (G1202R, G1202del, D1203N and S1206Y/C) [60].

In general, new-generation ALK inhibitors tend to overcome some of these resistance processes (Table 1.6), extending the median time to progression to more than 2 years [37]. However, as patients are treated with sequential ALK TKIs, compound resistances are emerging and adding another layer of complexity, such as C1156Y/ I1171N after progression on sequential crizotinib, ceritinib, and alectinib treatment, and E1210K/D1203N after sequential crizotinib and brigatinib [60].

ALK-Dependent Resistance: Amplification of ALK

A gain of ALK gene fusion copy number appears to be sufficient to confer resistance to crizotinib, although it occurs less frequently than secondary mutations, accounting for approximately 15% of acquired resistance [70]. However, it has not yet been detected in patients using second- and third-generation TKIs, making it vulnerable to more potent ALK inhibitors [60]. Furthermore, mutations and amplification may be found simultaneously in the same patient.

ALK-Independent Resistance: Activation of Bypass Signaling Pathways

Activation of alternative downstream signaling pathways contributes to acquired resistance in up to 20% of ALK-positive patients [70, 71]. These include activation of alternative proliferation pathways such as EGFR, HER2, KRAS, insulin-like growth factor 1 receptor (IGF-1R), MEK/ERK, or PI3K/AKT/mTOR [60]. To overcome them, combined treatment is necessary to inhibit both the ALK and the bypass pathways. However, no combination therapy has yet been clinically approved, nor has a multiple kinase inhibitor been developed.

1.3.3 Methods of Detection

Fluorescence in situ hybridization (FISH) is currently the reference detection method for ALK fusions [37]. It uses two specific DNA probes coupled to a fluorescent marker that covers the ALK region to be studied. In this way, it is possible to measure the distance between these two signals and, therefore, detect possible translocations. The strengths of FISH lie in its sensitivity and specificity.

Immunohistochemistry (IHC) is another method for detecting ALK rearrangements in lung cancer by measuring protein expression using especially labeled antibodies that bind to the proteins of interest. The advantages of IHC are mainly its low cost, it does not require special equipment and it is generally accurate [64]. However, despite being a reliable screening tool, FISH confirmation is required in positive and even in some cases for negative IHC in patients susceptible to ALK rearrangements (younger age, light smokers and testing negative for other mutations, such as EGFR and KRAS, which usually do not coexist with ALK) [72].

Recently, non-in situ testing approaches, including NGS and quantitative PCR (qPCR) assays, as well as RT-PCR for transcription, are emerging as independent methods to identify driver genes [37]. This highly specific techniques offer the additional advantage of identifying the ALK-associated fusion gene and the specific variants involved, although they require quality DNA samples. Additionally, NGS can detect isolated resistance mutations, allowing the patients' targeted therapy to

be changed at disease progression. However, due to the diversity of resistance mechanisms, it is necessary to perform repeated biopsies at progression. This re-biopsying is often not available in most cases since lung tumor biopsy procedures are highly invasive and patients are quite fragile. Therefore, a noninvasive approach such as liquid biopsy is an advantage and in some cases the only viable option.

1.4 Liquid Biopsy for ALK-Positive Non-Small Cell Lung Cancer

One of the central goals of oncology research is to find minimally invasive methods to assess cancer. In the past decade, owing to technological advances, molecular diagnostics performed on tumor-derived material in the circulation and other biofluids, including urine, saliva, cerebrospinal fluid (CSF), or pleural effusions, rather than solid tumor tissue, have gained great interest [25]. In this way, liquid biopsies provide the opportunity to obtain dynamic genomic information over the course of disease in a less invasive and less expensive manner, helping to predict and monitor treatment responses, assess the emergence of drug resistance, quantify minimal residual disease and capture the molecular heterogeneity of a tumor better than a conventional tissue biopsy [65, 73].

Currently, in the follow-up of ALK-positive patients who are treated with TKIs, liquid biopsies are the standard of care. In fact, the US FDA has recognized liquid biopsy as the first diagnostic tool for the identification of NSCLC patients with driver mutations [73].

1.4.1 Currently Available Blood-Based Tumor Biomarkers

The most common biomarkers of liquid biopsy are circulating tumor cells (CTCs), circulating tumor DNA (ctDNA), exosomes, and tumor-educated platelets (TEPs), which can be used to determine the tumors molecular profile and capture the heterogeneity of metastatic cancers (Figure 1.6).

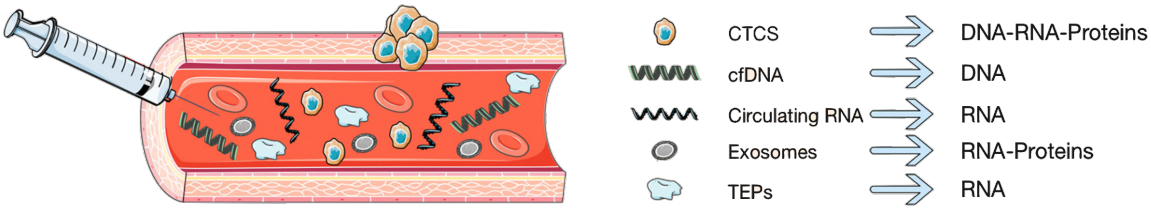


Figure 1.6. Tumor-derived components used as liquid biopsy biomarkers. Modified image from [74].

Circulating Tumor Cells (CTCs)

CTCs are cancer cells that detach from the primary lesions or tumor metastases and then enter into the blood circulation. Specifically, these cells typically undergo an epithelial-mesenchymal transition (EMT) process, which allows them to enter the circulation and reach another remote target, where they undergo a reverse process called mesenchymal-epithelial transition (MET). Ultimately, this results in further cancer proliferation and therefore metastasis formation.

Although extremely rare, since only one CTC can be detected in an average of 10^6–10^7 leukocytes, CTCs are found in around 50% of patients affected by metastatic epithelial tumors and are used for the detection and monitoring of non-hematologic cancers [75, 76]. However, CTCs are fragile and thus require more complex laboratory procedures for isolation and enumeration, and the technology to characterize them is still evolving. Furthermore, the clinical significance of CTCs enumeration as a reliable diagnostic biomarker has yet to be established, although a high CTC count generally correlates with worse prognosis and metastatic spread of cancer [75].

The CellSearch® system is the most known U.S. FDA-approved CTC detection technology and it is capable of isolating CTCs with the use of anti-epithelial cell adhesion molecule (EpCAM) ferromagnetic beads.

Circulating Tumor DNA (ctDNA)

ctDNA is a single strand or double-stranded circulating free DNA (cfDNA) portion derived from apoptotic, necrotic, or living tumor cells that actively release DNA in the bloodstream by mechanisms not yet fully understood [77].

Currently, several studies are investigating the association of ctDNA levels with tumor burden, tumor response, and survival outcome of lung cancer patients [77, 78]. However, since the absolute amount of ctDNA as a diagnostic tool is not significant, the attention has focused on real-time monitoring of tumor-specific genetic alterations to track the response to targeted therapies and drug resistance. In this context, novel and highly sensitive blood-based assays such as NGS and polymerase chain reaction-based techniques (PCR), including digital PCR (dPCR), have emerged and are managing to overcome the problems of tumor DNA isolation and sequencing, which traditionally were slow and laborious processes. In fact, at present, it is possible to detect ctDNA even if it represents less than 1.0% of the total cfDNA [78, 79].

ctDNA is now the most widely used biomarker for detecting genetic abnormalities in ALK-positive NSCLC patients. In this context, a positive finding of an actionable mutation in a cfDNA sample using PCR-based methods is enough to start or change a targeted treatment [80].

Exosomes

Exosomes are cell-derived vesicles of endocytic origin with a diameter of 40–100 nm that are released into the extracellular space and a variety of body fluids. They contain proteins as well as a range of nucleic acids, including DNA, mRNAs, and miRNAs, which are transferred to target cells, modulating their activities. In cancer they have a key role in metastatic niche preparation and cell-cell communication, favoring tumor growth, progression, and drug resistance [81]. In addition, exosomes have a bilayer lipid membrane that protects nucleic acids, particularly RNAs, from ribonuclease degradation and extreme pH when found in body fluids. Therefore, there is a growing interest in these vesicles because their content could potentially serve as a biomarker in diagnosis, prognosis, and monitoring of the response to treatment [80,82]. However, its isolation can be challenging.

Several techniques have been used for exosome isolation, such as density-gradient centrifugation, ultracentrifugation, and immunoaffinity methods based on the presence of specific protein markers, including CD63, CD9, and CD81 [81]. Once isolated,

they can be utilized to characterize the RNA and thus genetic alterations by using different techniques, including RT-PCR, nucleic acid sequencing, ELISE, or Western Blot. In summary, it is a promising source for analyzing EML4-ALK fusions, although its use in clinical practice has not yet been standardized. [82].

Tumor-Educated Platelets (TEPs)

Platelets are the second most abundant cell type in peripheral blood and are known for their role in hemostasis. However, when they interact with tumor cells, they can undergo modifications, which will eventually affect tumor progression, dissemination, and establishment of an ideal environment for the proliferation of cancer [3,83]. Specifically, this interaction induces the transference of tumor-associated biomolecules, including multiple RNA types and proteins, to platelets, resulting in the creation of TEPs.

When isolated, these platelets constitute a source of tumor RNA for general molecular traces for cancer surveillance [82, 84]. Several methodologies have been used for platelet isolation and RNA profiling, although, as with exosomes, the complexities of these procedures are slowing their implementation in clinical routine. Currently, RNA transcripts could be quantified using microarray hybridization techniques, RT-PCR, or NGS. Therefore, in addition to exosomes, TEPs seem to have the potential to be used as liquid biopsy biomarkers for a variety of clinical and research applications, including the identification of mutations [82].

Chapter 2

Research Goals

Diagnosis and targeting of ALK-positive NSCLC now form part of clinical routine, and the results obtained with ALK inhibitors are unprecedented considering the PFS and OS. However, although the development of novel drugs will continue to be a priority of cancer treatment, there is a need for better prognostic tools to assess therapeutic response, monitor tumor progression, and predict clinical outcomes.

One of these tools is a next-generation sequencing (NGS). It provides unparalleled sequencing information and is becoming the main mechanism for both understanding tumor heterogeneity and biomarker discovery in cancer research. However, current clinical laboratory practice guidelines for NGS do not provide definitive guidance on filtering and confirming NGS variants.

Furthermore, since the detection of low-frequency mutant alleles may not always be called confidently in NGS, an independent and more targeted and sensitive assay like dPCR is often used. This secondary confirmation is essential for clinical decision making, where the validation of the mutation has real-world implications.

In this context, the main objective of this thesis is to develop a new tool that addresses the standardization of filtering of sequenced data to minimize the probability of false-positive and false-negative results, which will prevent unnecessary dPCR runs, optimizing time and costs. In short, this bioinformatic pipeline will be able to fully automate the filtering of mutations that is carried out before final confirmation by dPCR, enabling a better stratification of patients and improving the current analysis

of sequenced data from liquid biopsy samples. Additionally, it will allow oncologists to have more reliable information on the patients' tumor genotype, favoring the introduction of more individualized and optimal treatments in clinical practice without the risks associated with tissue biopsies.

Finally, this goal has led to the use in this study of non-invasive methods of describing tumors molecular profile to filter driver mutations in the ALK gene, including ctDNA and exosomes. Specifically, with the combination of molecular techniques and bioinformatics, this master's thesis aims to:

- Perform a clinical study of mutations from circulating DNA and exosomes in a cohort of patients with advanced NSCLC.
- Develop an automatic algorithm for filtering ALK-positive NSCLC somatic mutations.
- Estimate the utility of NGS and the developed algorithm for:
  - Detecting oncogenic drivers.
  - Identifying resistance mutations.
  - Monitoring therapy response.
  - Predicting clinical outcomes.
- Assess the utility of liquid biopsy for the study of somatic mutations in the ALK gene from free circulating DNA and exosomes.

Chapter 3

Materials and Methods

3.1 Patients Characterization

This study is part of the research project entitled "Clinical utility of liquid biopsy in non-small cell lung cancer patients with EML4-ALK translocation". Its objective is to determine the best strategy for the identification of EML4-ALK translocations and to assess the clinical utility of periodic tumor monitoring in ALK-positive NSCLC patients from liquid biopsies. In this context, between June 2015 and July 2019, several ALK-positive NSCLC advanced subjects were recruited from 6 hospitals across Spain, including the Hospital Universitario Puerta de Hierro, the Complejo Hospitalario Universitario A Coruña, the Hospital Universitario Fundación Jiménez Díaz, the Hospital Clínic Barcelona, the Hospital General Universitario de Alicante, and the Hospital General Universitario de Valencia.

The primary study was conducted under the precepts of the Helsinki Declaration and was approved by the Hospital Puerta de Hierro Ethics Committee (internal code 79-18). Patients were eligible if they consented to allow their clinical information to be used in the previously mentioned research project. Their clinical history was queried for information on the age of diagnosis, sex, smoking status, histology, stage, Eastern Cooperative Oncology Group (ECOG) performance status, and current therapies.

Eligible patients had histologically confirmed diagnosis of stage III–IV NSCLC that was ALK-positive, a measurable disease according to the response evaluation

criteria in solid tumors (RECIST, version 1.1), were candidates to be treated with an ALK inhibitor, and were 18 years old or older. Exclusion criteria included the impossibility of frequent venipuncture or evidence of any other major clinical disorder or finding that would have made it undesirable for the patient to participate in the study.

Finally, for this project, a total of 39 liquid biopsies (37 plasma samples and 2 CSF samples) obtained from 30 patients who have met the above restrictions were used.

3.2 Laboratory Procedures

In order to identify the required parameters to design and implement the automatic algorithm for filtering genetic variants, several clinical procedures were carried out in the Liquid Biopsy Laboratory of the Hospital Universitario Puerta de Hierro, including the collection and conditioning of the ALK-positive NSCLC samples, their sequencing and analysis, and the subsequent confirmation of the identified mutations.

3.2.1 Sample Collection and Processing

Once an ALK gene mutation is diagnosed in a patient with NSCLC according to an anatomic pathology report (FISH and/or IHC), peripheral blood is drawn into a 9 mL Streck Cell-Free DNA BCT® (Streck, La Vista, NE, USA) tube and transported to the Liquid Biopsy Laboratory, where it is preserved until its processing. Within the context of the aforementioned project, blood samples were extracted from patients in the pre-treatment stage, at months 2, 4, 6, 8, 12, 15 and 18, and at progression.

In this study, isolation of cfDNA and exosomes from the cellular fraction was achieved by two consecutive centrifugation processes at room temperature, the first at 1500 g for 10 min and the second at 5000 g for 20 min. After discarding the supernatant containing the clean plasma, the last step of the cfDNA and exosome RNA isolation was performed using the QIAamp® Circulating Nucleic Acid Kit (QIAgen, Valencia, CA, USA) and the exoRNeasy® Serum/Plasma Maxi Kit (QIAgen, Valen-

cia, CA, USA) respectively, following the manufacturer's instructions. Until library preparation for sequencing, processed samples were stored at -80 °C.

Regarding CSF samples, their processing was identical to that described for plasma. These were first collected in a 9 mL Streck Cell-Free DNA BCT® (Streck, La Vista, NE, USA) tube, and then the CSF nucleic acids were isolated using the QIAamp® Circulating Nucleic Acid Kit (QIAgen, Valencia, CA, USA).

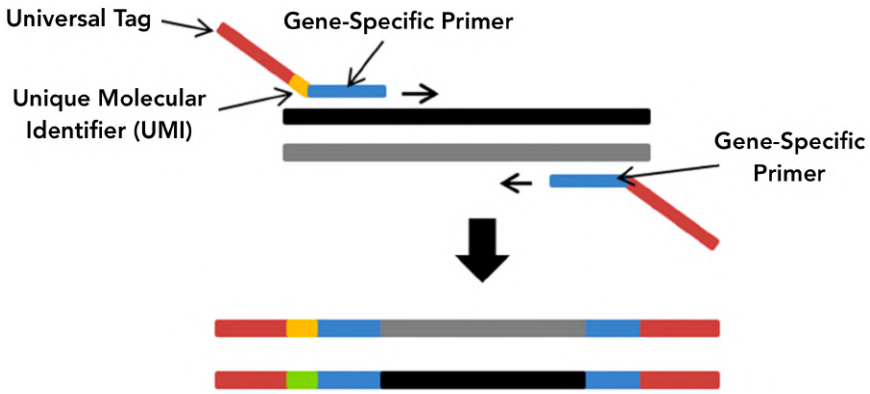
3.2.2 Library Preparation

NGS libraries were prepared from 10.4 μL of cfDNA or exosome RNA using the Oncomine™ Pan-Cancer Cell-Free Assay (Thermo Fisher, Palo Alto, CA, USA) according to the manufacturer's instructions. This multi-biomarker amplicon-based assay is optimized to detect more than 900 hotspots in 52 key genes from liquid biopsy samples containing tumor-derived DNA and RNA. Additionally, it enables a limit of detection (LOD) down to 0.1%, allowing the study of single-nucleotide variants (SNVs), short insertions and deletions (InDels), copy-number variations (CNVs), and fusions.

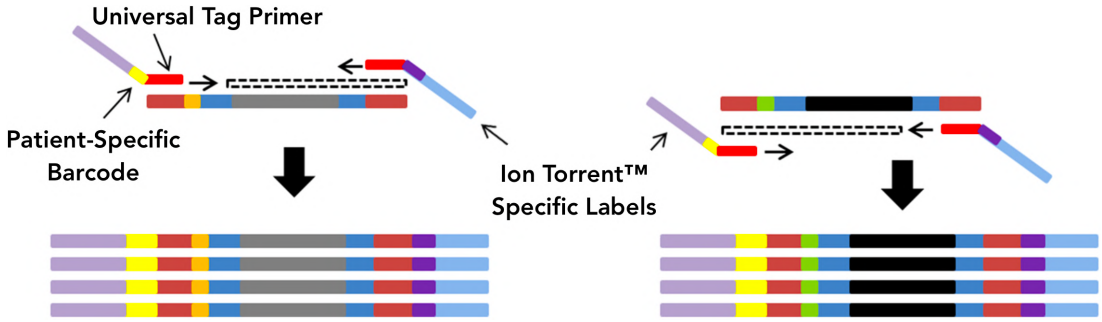
The steps that were followed to prepare the NGS libraries are described below:

1. **Reverse transcription of cell-free nucleic acids**. After isolating the free-circulating DNA and exosome RNA, complementary DNA (cDNA) was synthesized in this step. Specifically, the reverse transcription of exosome RNA was performed using the PrimeScript™ RT Reagent Kit (TaKaRa, Japan).

2. **Target amplification**. After the fragmentation of the isolated and synthesized DNA using enzymatic methods, 2 cycles of tagging PCR were performed producing labeled DNA amplicons within a specific size range (Figure 3.1a). In this sense, prior to target enrichment and library amplification, each original DNA molecule was assigned a random unique molecular identifier (UMI). This technology is aimed to detect, quantify, and sequence unique DNA fragments with high-resolution, allowing the identification and removal of amplification artifacts arising from library preparation. In addition, this ligated adapter also contains the first sample index and a universal tag.



(a) Incorporation of UMIs in two DNA fragments. This first PCR step assigns a random molecular barcode (UMI) to each of the DNA sequences, including gene-specific primers and a universal tag.



(b) Amplification of the target amplicons. This second PCR uses custom primers to amplify the barcoded molecules with a universal tag sequence, adding patient-specific barcodes along with NGS-specific sequences.

Figure 3.1. Schemes of the amplification and tagging technology used in in the Oncomine™ Pan-Cancer Cell-Free Assay. True hotspot variants can be distinguished from workflow errors by the frequency of variants in families that have the same UMI.

3. **Target amplicons purification**. Sample purification is a critical step for obtaining accurate NGS data. Therefore, a specific reagent with magnetic beads was used in this study to remove short primers, unincorporated deoxyribonucleotide triphosphates (dNTPs), enzymes, short-failed PCR products, and salts from the PCR reactions.

4. **Amplification of the target amplicons with barcode adapted primers**. In this step, specific DNA adapter sequences were annealed to the 5' and 3' ends of the amplicon DNA, enabling the analysis of multiple samples simultaneously. Specifically, this second PCR (18 cycles) was performed to add

patient-specific barcodes along with particular labels to create Ion Torrent™ sequencer-compatible libraries (Figure 3.1b).

5. **Barcoded library purification**. To guarantee reliable sequencing results, and in a process similar to the previous one, a purification of the barcoded amplicons was carried out.

6. **Size selection**. To isolate the DNA fragment sizes of interest for efficient and high-quality DNA sequencing, magnetic beads with varying concentrations of buffers were used for size selection.

7. **Library quantification**. To estimate the DNA concentration and to ensure equal representation of the indexed libraries, each of the samples was quantified with a qPCR system (40 cycles) and compared to the standard curve. Finally, the definitive barcoded libraries were pooled and adjusted to a final concentration of 50 pM.

AMPure XP® magnetic beads (Beckman Coulter, Inc., Brea, CA, USA) were used to carry out the purification and size selection steps, while the Ion Library TaqMan™ Quantitation Kit (Thermo Fisher, Palo Alto, CA, USA) and the StepOnePlus™ Real-Time PCR System (Thermo Fisher, Palo Alto, CA, USA) were used for the quantification.

3.2.3 Template Preparation and Chip Loading

Template preparation is based on an emulsion PCR (ePCR) method (Figure 3.2). It starts with the denaturation of the fragmented and ligated DNA, which is then mixed with beads. Each bead has thousands of copies of an oligonucleotide, which is complementary to one of the adapters and to which the corresponding DNA fragments will be attached by clonal amplification. Empty beads are discarded, remaining the rest of them in the chip wells (one bead per well).

For both template preparation and chip loading, the Ion Chef™ System (Thermo Fisher, Palo Alto, CA, USA) was used. It automates these processes, loading up to

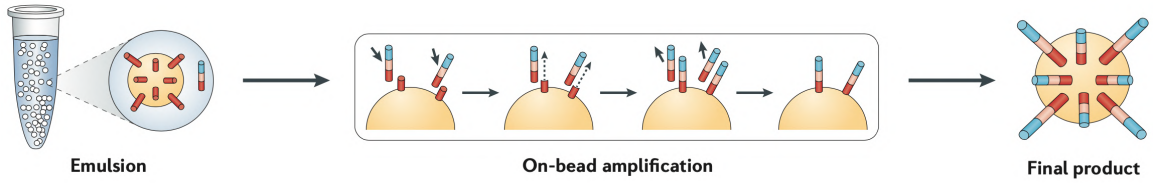


Figure 3.2. Ion Torrent™ template preparation technique. Micelle droplets are loaded with primers, complementary adapters, dNTPs, and DNA polymerase. Then PCR is carried out within the micelle, covering each bead (100–200 million of beads) with thousands of copies of the same DNA sequence. Modified image from [85].

8 samples on a single Ion 550™ Chip (Thermo Fisher, Palo Alto, CA, USA), which is subsequently used for sequencing.

3.2.4 Sequencing

The previously loaded chips were sequenced on an Ion GeneStudio™ S5 Sequencer (Thermo Fisher, Palo Alto, CA, USA). It is a semiconductor-based NGS platform that decodes the template DNA sequence by detecting H^+ ions that are released upon the incorporation of nucleotides to the DNA fragment attached to its corresponding bead.

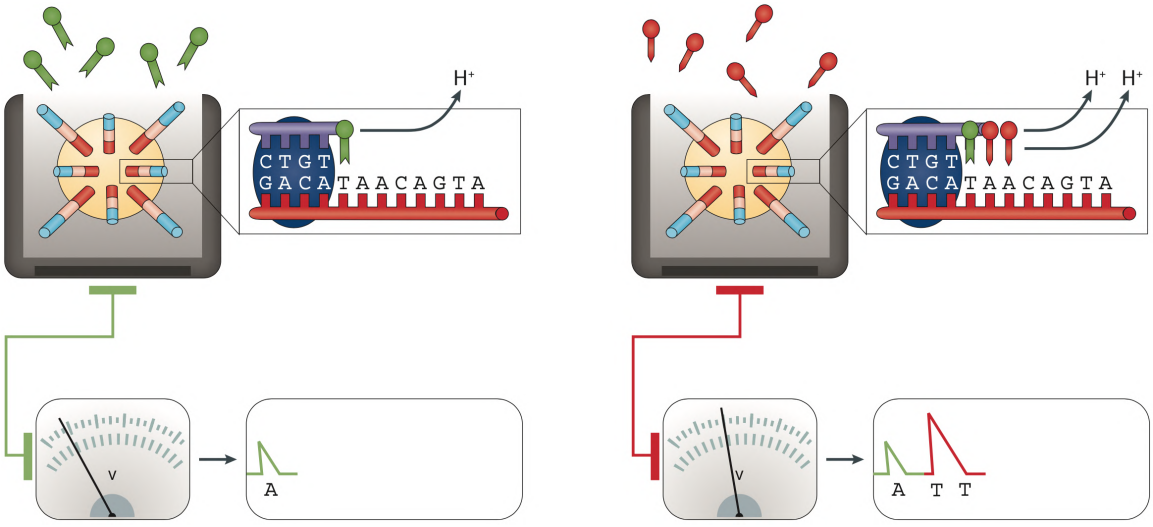


Figure 3.3. Ion Torrent™ measurement of nucleotide addition: incorporation of an adenine (A) followed by the addition of two thymines (T). Modified image from [85].

This process occurs simultaneously in millions of wells, which contain template-attached beads incubated with DNA polymerase and a particular type of dNTP.

Therefore, if that specific dNTP matches the growing template strand, the DNA polymerase adds it, resulting in a discharge of a H^+ ion that will change the pH of the solution in the well. This shift is recorded and converted to voltage, indicating that the nucleotide was incorporated and the base was called (Figure 3.3).

3.2.5 Sequenced Data Analysis

Data analysis was completed in two consecutive steps. First, the primary analysis was performed to study the overall quality of the sequenced data. Subsequently, the identified genetic variants were analyzed and verified using several software tools. In this last step, data of interest was also collected for the development of the algorithm.

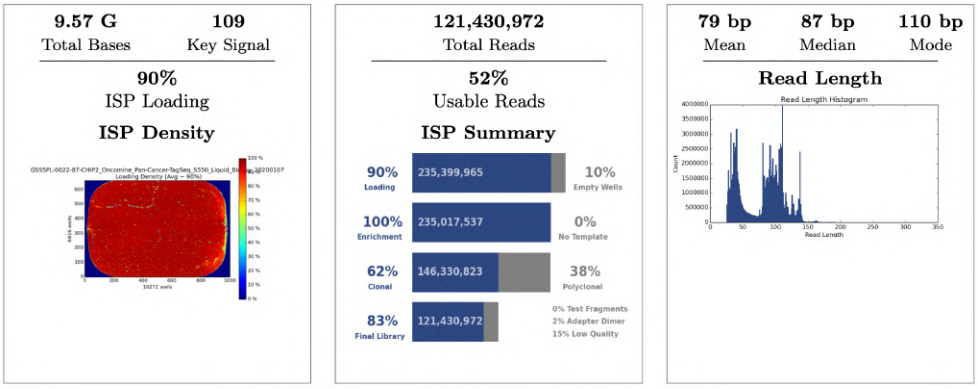
Primary Analysis

The sequencing runs were scheduled with the Torrent Suite™ v5.12 software, generating at the end of the process a file with the main quality specifications (Figure 3.4a), which include:

- **ISP loading density**. Percentage of chip wells that contain an Ion Sphere™ Particle (ISP).
- **Total reads**. The total number of reads associated with a specific barcode.
- **Read length (*bp*)**. Length of the called reads.

Additionally, for each of the samples analyzed in the run (8 samples), the server generates an additional report with the following parameters (Figure 3.4b):

- **Bases**. The number of base additions for each barcode.
- **$\geq$ Q20**. The number of reads that have a predicted quality score of Q20 or better.
- **Reads**. The total number of filtered and trimmed library reads.
- **Mean read length (*bp*)**. Average read length.



(a) Quality overview of a specific NGS run. The sequencing quality is directly proportional to the value of the ISP loading parameter.

Barcode Name	Sample	Bases	$\geq Q20$	Reads	Mean Read Length	Read Length Histogram
TagSequencing_12	ODA 20180320	2,825,685,525	2,514,260,390	28,933,171	97 bp	
TagSequencing_13	ATC 20180321	1,400,011,194	1,237,381,913	14,736,586	95 bp	
TagSequencing_14	MCPA 20180410	1,507,632,040	1,335,900,783	15,938,654	94 bp	
TagSequencing_15	MESR 20181003	1,079,195,152	952,654,081	11,936,405	90 bp	

(b) Quality parameters of 4 of the 8 sequenced samples. A Gaussian base pair distribution in each of the histograms is indicative of well-prepared libraries.

Figure 3.4. Ion Torrent™ NGS run summary (NGS run ID: 20200108).

This raw data was processed automatically on the Torrent Server™ v5.12 and then aligned to the reference hg19 genome. In this context, to study the sequence coverage for target genomic regions, the Coverage Analysis Plugin v5.12.0.0 was used, obtaining the following parameters (Figure 3.5):

- **Mapped reads**. The total number of reads that were mapped to the reference.
- **On target (%)**. Percentage of reads that were mapped to any targeted region.
- **Mean depth**. The average number of reads that align to known reference bases.
- **Uniformity (%)**. Percentage of bases, in all targeted regions, that is covered by at least 20% of the average base coverage depth reads.

Library type: Tag Sequencing
Target regions: Oncomine_PANCAN_cfNA_v5.Designed

Barcode Name	Sample	Mapped Reads	On Target	Mean Depth	Uniformity
TagSequencing_12	ODA 20180320	27,865,285	97.12%	103,068	99.03%
TagSequencing_13	ATC 20180321	13,965,158	96.23%	50,204	99.99%
TagSequencing_14	MCPA 20180410	15,101,783	96.48%	54,070	99.21%
TagSequencing_15	MESR 20181003	10,796,593	95.47%	37,660	99.68%

2 | 4 items per page | 4 - 8 of 8 items

Figure 3.5. Coverage Analysis Plugin output parameters of 4 of the 8 sequenced samples (NGS run ID: 20200108). Successful sequencing shows a percentage of reads mapped to target regions and uniformity values close to 100%.

Secondary Analysis

After ensuring that the sequencing had been performed as expected, the resulting data of the quality control passing samples were uploaded in BAM format to the Ion Reporter™ v5.12. Variant calling, annotation, and filtering were performed on this platform using the Oncomine™ TagSeq Pan-Cancer Liquid Biopsy v2.1 workflow. This pipeline analysis also included the mapping of the sequenced reads to defined target regions (Oncomine™ Pan-Cancer DNA Regions v1.0) and the variant calling using the Oncomine™ Variant Annotator v2.3 plugin. For each of the samples the following parameters were obtained:

- **Median read coverage**. Median coverage across targets.
- **Median molecular coverage**. The median number of individual interrogated DNA molecules across targets.
- **Limit of Detection (LOD)**. Numeric range (%) where the minimum represents the median value across all targets and the maximum represents the LOD for the 80^{th} percentile targets.

For sensitive variant detection down to 0.1% frequency, optimal results are obtained when targeting a median read coverage > 25000, median molecular coverage > 2500, and both numbers of the LOD segment are $\leq 0.1\%$ [86].

Furthermore, with the Ion Reporter™ platform, it was possible to identify and analyze each of the deviations from a reference genome based on their biological and clinical relevance, obtaining information about the gene, the location, the allele frequency, and the specific type of each variant. The main parameters of interest for each of these mutations were:

- **Molecular depth**. The number of interrogated DNA molecules containing target.
- **Molecular counts**. The number of detected DNA molecules containing a variant allele.

On the other hand, it is possible to download all this raw data in a *.tsv* file, which includes an even more detailed analysis of all the variants. This file is precisely the one that was used to design the bioinformatic pipeline, along with the parameters described above.

In addition, all candidate mutations were manually reviewed using the Integrative Genomics Viewer (IGV) v2.3.40 to be subsequently confirmed by dPCR. The polymorphisms were also filtered out based on the following population databases: 1000 Genome Project, NHLBI GO Exome Sequencing Project (ESP), database of single nucleotide polymorphisms (dbSNP), and ExAC (genomAD). Finally, the clinical significance of somatic variants was determined according to the Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer [87].

3.2.6 Confirmation of the Sequenced Results

Digital PCR is a tool that is used to improve the efficiency of the NGS workflow and for the verification of sequencing data. In this study, the previously identified mutations (section 3.2.5) in the cfDNA or cDNA samples were confirmed by the QuantStudio® 3D Digital PCR System (Applied Biosystems, South San Francisco, CA, USA) according to the manufacturer's specifications. The reverse transcription of exosome RNA to synthesize cDNA was performed using the PrimeScript™ RT Reagent Kit (TaKaRa, Japan).

Briefly, dPCR is based on the fact that the random distribution of molecules in many partitions follows a Poisson distribution. Each partition, which contains a specific labeled DNA sequence, acts as an individual and isolated PCR reaction. After amplification, the augmented target sequences are detected by fluorescence, which is sufficient to determine their concentration. In this study, this process was carried out following the next steps:

1. **Set up of the dPCR reaction**. It was performed by mixing a specific sample with the master mix and the assays, obtaining a final volume of 18 μL. Specifically, this reaction included 8.55 μL of template cfDNA or cDNA, 9 μL of 20X QuantStudio™ 3D Master Mix and 0.45 μL of 40X TaqMan™ assays (predesigned TaqMan™ Liquid Biopsy dPCR assay and custom TaqMan™ assay). These assays include specific primers and labeled probes that have a fluorescent reporter dye (VIC™ or FAM™) attached to its 5' end, and a quencher dye at its 3' end (ROX™). In all assays, probes complementary to mutated and wild type (WT) DNA sequences were labeled with the FAM™ and VIC™ dye respectively.

2. **Chip loading** (Figure 3.6). 14.5 μL of the dPCR reaction was loaded using the QuantStudio™ 3D Digital PCR Chip Loader into the QuantStudio™ 3D Digital PCR 20K Chip v2, which has 20000 wells containing a single DNA molecule that will be amplified independently in the next phase. In addition, positive and negative controls were included in each dPCR run to ensure that the primers have attached to the DNA strand and that there has been no contamination.

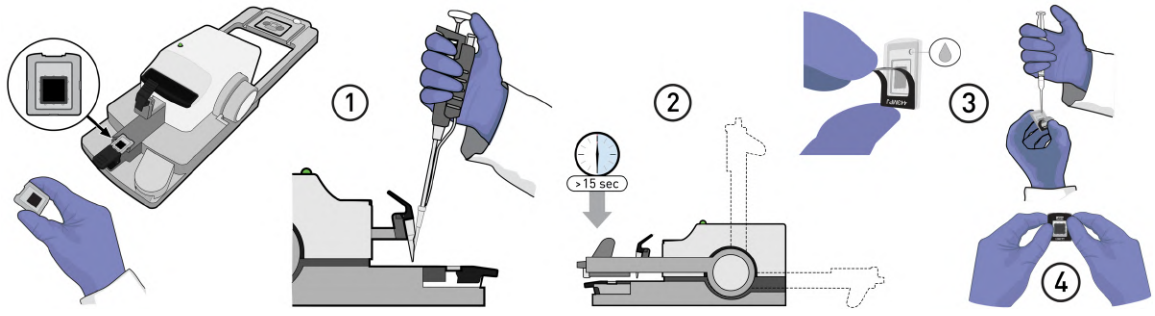


Figure 3.6. Chip loading: (1) charging of the dPCR reaction into the QuantStudio™ 3D Digital PCR 20K Chip v2; (2) lid application; (3) filling of the assembly with Immersion Fluid; and (4) sealing of the loading port. Modified image from [88].

3. **Thermocycling**. This process, which was divided into 3 consecutive stages, was carried out with the QuantStudio® 3D Digital PCR System (Applied Biosystems, South San Francisco, CA, USA). The temperature, duration, and number of cycles for each of the stages are summarized in Table 3.1. During the amplification stage, the Taq polymerase enzyme cleaves the aforementioned FAM™ and VIC™ bound probes, generating fluorescent signals that will be recorded and associated with specific wells on the chip in the next step.

Stage 1	**Stage 2**		**Stage 3**	
96 °C	56 °C	98 °C	72 °C	22 °C
10 *min*	2 *min*	30 *s*	10 *min*	30 *min*
1 cycle (hold)	40 cycles		1 cycle (hold)	

Table 3.1. Digital PCR reaction conditions used in this study for DNA denaturation, annealing, and elongation stages.

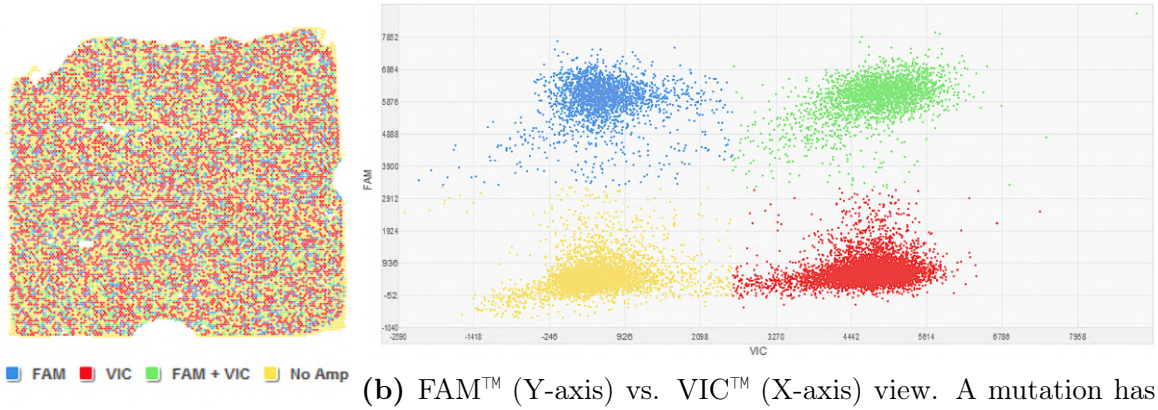
4. **Chip reading**. After completing the thermocycling process, the chips were loaded into the QuantStudio™ 3D Digital PCR Instrument to read their fluorescence. This reading was performed twice to ensure that the subsequent data review is accurate and reliable.

5. **Results interpretation**. The previously processed data was visualized and analyzed using the QuantStudio™ 3D Analysis Suite™ Cloud Software (Figure 3.7). In this context, samples with a mutant allele fraction (MAF) equal to or higher than 0.1% were considered positive (Equation 3.1). Additionally, the LOD was assessed for all individual assays, being lower than 0.1% in all cases.

$$MAF = \frac{FAM_{copies}/\mu L}{FAM_{copies}/\mu L \ + \ VIC_{copies}/\mu L} \tag{3.1}$$

where $FAM_{copies} \equiv$ number of reads of the mutated sequences

$VIC_{copies} \equiv$ number of reads of the WT sequences



(a) Chip view.

(b) FAM™ (Y-axis) vs. VIC™ (X-axis) view. A mutation has been detected if the blue cluster is identifiable in the dispersion chart.

Figure 3.7. Quality and calls review of the dPCR. The red, blue, and green colors correspond to the VIC™ (fluorescence of the WT sequences), FAM™ (fluorescence of the mutated sequences), and VIC™+FAM™ groups respectively, while the yellow represents the wells without amplification.

3.3 Algorithm Specifications and Requirements

The main objective of the developed algorithm is to fully automate the mutation filtering that is carried out before its ultimate confirmation by dPCR. In this way, this solution aims to simplify and automate the analysis of the NGS data described in section 3.2.5. Eventually, this will save time, remove the human factor, and therefore improve the quality and efficiency of the entire process.

The different parameters of the proposed algorithm were identified according to the confirmed variants, both positive and negative, obtained from the liquid biopsy samples used in this study. To find specific patterns, each type of these clinically relevant deviations from a reference genome were studied individually using the obtained specifications described in section 3.2.5 and the *non-filtered-oncomine.tsv* output file, which is generated after each sequencing. It contains both variants that have passed the Ion Reporter™ internal filter (Oncomine™ Variants v5.12) and initially discarded mutations. Therefore, apart from discerning whether a reading of an allele fraction represents a true mutation or is an artifact that should be ignored, the developed algorithm in this study will also be able to identify samples that have been initially discarded by the filters of the Ion Reporter™ platform but are liable to be false-

negatives.

This analysis and identification of the filtering parameters were carried out with respect to each type of variant found during the study, including:

- Fusions.
- Copy-number variations (CNVs).
- Single-nucleotide polymorphisms (SNPs).
- Insertions and deletions (InDels).
- Multiple-nucleotide polymorphisms (MNPs).

For each of them, minimum standards were established to consider them valid and proceed with their confirmation by dPCR. Specifically, parameters such as the overall error of the NGS assay, the limit of detection (LOD), the coverage depth, the percentage of targeted bases sequenced at that coverage depth, the total number of target reads covering a variant region, the number of reads supporting a specific variant, and the clinical significance, among others, were taken into account for the selection of the different thresholds.

Regarding the technical requirements, the programming language used in this study was R (v3.6.3), a free software environment for statistical computing and graphics. Furthermore, its capabilities were extended through additional packages:

- **Tcl/Tk package**. Tcl is a general-purpose multi-paradigm system scripting language that provides the ability for applications to communicate with each other, while Tk is a cross-platform widget toolkit used for building graphical user interfaces (GUIs). It is included with the installation of R.

  In this thesis, the Tcl/Tk package was used to offer the end-user an intuitive graphical interface to carry out the entire filtering process, including loading the *non-filtered-oncomine.tsv* files, selecting the type of variants to filter by, and saving mutations susceptible to being positive.

- **Scales (v1.1.1) package**. It contains functions that convert data values to perceptual properties.

  In this particular case, it was used to transform the raw data obtained from the *non-filtered-oncomine.tsv* files into interpretable values for the end-user.

Finally, to make programming more user-friendly through a graphical interface for R, the integrated development environment RStudio® Desktop v1.2.5033 was used.

Chapter 4

Results

4.1 Study Cohort

This study included a cohort of 30 patients with advanced NSCLC from whom a total of 39 samples were extracted, 37 of plasma and 2 of cerebrospinal fluid (CSF). Within the plasma samples, 24 (64.9%) were cfDNA-based liquid biopsies and 13 (35.1%) were drawn for exosome isolation. All samples were collected upon disease progression as ascertained according to RECIST v1.1 criteria.

Patients' clinicopathological information at the time of the liquid biopsies sampling is summarized in Table 4.1. The median age was 54 years (range: 36–81), with a uniform distribution of females (N=17) and males (N=13). Most of the patients were never smokers (60%), which is in accordance with the ALK-positive NSCLC characteristics, 8 were former smokers (26.7%), and 4 were current smokers (13.3%). All the individuals underwent a successful histological examination, which revealed that the majority presented an adenocarcinoma (93.3%), and the remaining 2 cases a neuroendocrine carcinoma (6.7%). Furthermore, most of them were diagnosed with stage IV (76.7 %) and stage III (20 %) disease, although at the time the samples were collected and this study was performed, all of them had metastatic cancer (stage IV). Treatment procedures were performed in 6 different hospitals across Spain using the main ALK inhibitors, except for 2 cases that were on chemotherapy and 2 that had no associated treatment. In this context, 16 patients were receiving second-

Clinical Feature	Grouping	N	%
Age of diagnosis	Median	54	-
Sex	Female	17	56.7%
	Male	13	43.3%
Smoking status	Never smokers	18	60%
	Former smokers	8	26.7%
	Current smokers	4	13.3%
ECOG performance status	0	14	46.7%
	1	13	43.3%
	2	3	10%
Histology	Adenocarcinoma	28	93.3%
	Neuroendocrine carcinoma	2	6.7%
Initial clinical stage	IV	23	76.7%
	III	6	20%
	II	1	3.3%
Treatment	Alectinib	11	36.7%
	Crizotinib	9	30%
	Ceritinib	4	13.3%
	Chemotherapy	2	6.7%
	No treatment	2	6.7%
	Brigatinib	1	3.3%
	Lorlatinib	1	3.3%

Table 4.1. Clinicopathological characteristics of the study population (N=30) at the time of the sample collection.

generation ALK inhibitors such as alectinib (36.7%), ceritinib (13.3%), and brigatinib (3.3%), while 9 were administered the first-generation ALK inhibitor crizotinib (30%). The remaining patient was prescribed the third-generation inhibitor lorlatinib (3.3%). Regarding their ECOG performance status at diagnosis, it ranged from 0 (N=14) to 1 (N=13).

Finally, it should be mentioned that in this study a subject was considered as two different patients if progression to a specific treatment was observed in that patient, that is if another sample was obtained and sequenced with the patient receiving different treatment than the initial one. On the other hand, patients with multiple sequenced samples but with the same treatment were counted as a single one. In this context, this study had 27 unique individuals, 3 of them progressing to an ALK inhibitor.

4.2 Next-Generation Sequencing Performance

The liquid biopsy samples from each of the patients were analyzed using NGS, which identified genetic mutations according to the wide range of hotspots in key genes addressed by the Oncomine™ Pan-Cancer Cell-Free Assay.

For this study, 39 samples were sequenced using 9 different NGS runs. On average, the percentage of chip wells that contained an ISP was 85.0%, resulting in 53.4% of usable sequences (library ISPs that passed the polyclonal, low quality, and primer dimer filters). These called reads had an average length of 76.8 *bp*. The average of mapped reads per sample was about 9.1 million, while its mean sequencing depth was 25050 reads. Regarding the coverage uniformity, it was higher than 74.0% in all the analyzed samples.

Under these conditions, more than 170 potentially malignant mutations in genes such as ALK, EGFR, KRAS, TP53, PIK3CA, FGFR1, or PDGFRA, among others, were initially accounted. From this set, 84 of the alterations involved the ALK gene. Each of these possible mutations was manually filtered as described in section 3.2.5. As a result of this assessment, 67.9% of the initial ALK variants were excluded from further analysis. In this context, 27 potential ALK alterations were finally identified from 18 different samples (17 patients), which were subsequently verified by digital PCR.

On the other hand, taking into account the clinical interest of pathological mutations, 7 additional samples associated with 12 potential alterations from 6 different patients were also confirmed by dPCR since their characteristics (molecular coverage, read counts, etc.) made them susceptible to being positive. These samples included mutations in the EGFR, KRAS, TP53, and PIK3CA genes.

Finally, it should be noted that all the selected samples had a MAF above the limit of detection (0.1%).

4.3 Mutational Spectrum

Mutations of genes of clinical significance were confirmed by dPCR in order to collect the necessary data, especially the false positive rate, to define the variant filtering criteria according to that described in section 3.2.5.

Among the 39 potential mutations identified from 25 samples, 14 alterations (35.9%) were confirmed as positive by dPCR. 9 of them were ALK variants, of which 8 were unique identifications associated with a specific patient. The remaining ALK mutation was confirmed in two different samples from the same patient. Regarding the other 5 mutations, 2 were found in the TP53 gene, 2 in PIK3CA, and 1 in KRAS.

On the other hand, 40 alterations reported by NGS, but without a clear initial representation, were also reaffirmed by digital PCR and included in the final mutational spectrum. Of these, 39 were in the TP53 gene (only on one occasion the same variant was identified in two different samples from the same patient) and 1 in FGFR1.

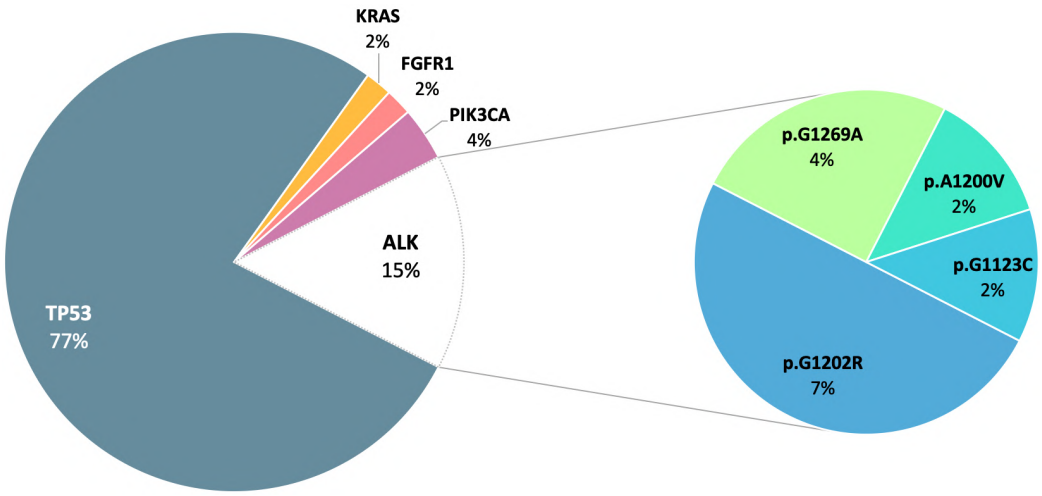


Figure 4.1. Frequency of the mutations identified in the study population, as well as the different types of ALK missense mutations.

In this context, 52 unique mutations were finally reported. As shown in Figure 4.1, the most widely represented mutations occurred in the TP53 gene, with 40 cases. It was common to find multiple alterations of this type in the same sample and coexisting with ALK mutations, which are the target of this study. On the other hand,

Gene	Genomic Alteration	Nucleotide Change	Number of Cases
TP53	p.P92A	c.274C>G	13
	p.A88fs	c.261_262insC	4
	p.R282W	c.844C>T	3
	p.V197fs	c.590delT	1
	p.M133T	c.398T>C	1
	p.H380fs	c.1140delT	1
	p.P359A	c.1075C>G	1
	p.R248Q	c.743G>A	1
	p.Y220C	c.659A>G	1
	p.S90fs	c.267_268insC	1
	p.G105C	c.313G>T	1
	p.S15*	c.42_43insT	1
	p.S149fs	c.445delT	1
	p.Q16fs	c.45delT	1
	p.V157F	c.469G>T	1
	p.R158C	c.472C>T	1
	p.M160I	c.480G>T	1
	p.C176Y	c.527G>A	1
	p.S241A	c.721T>G	1
	p.P250L	c.749C>T	1
	p.G262R	c.784insC	1
	p.R273Gly	c.817C>G	1
	p.?	c.919+10C>T	1
PIK3CA	p.E545K	c.1633G>A	1
	p.E545A	c.1634A>C	1
KRAS	p.G12C	c.34G>T	1
FGFR1	p.T174fs	520delA	1

Table 4.2. TP53, PIK3CA, KRAS, and FGFR1 gene mutations identified in the study population and their prevalence.

alterations were also detected in the PIK3CA, KRAS, and FGFR1 genes, appearing both individually and concurring with the rest of the detected variants. These mutations are found in more detail in Table 4.2. It is worth noting the KRAS G12C alteration, which was categorized as a variant with strong clinical significance since it is sensitive to an experimental KRAS inhibitor (AMG 510) [89]. Additionally, the PIK3CA E545K and E545A mutations were classified as variants with potential clinical significance.

The average number of variants per patient was 1.7. However, 7 patients did not present any mutation and in 22 of them, no ALK alteration was found. Furthermore,

there were no significant differences in the amount of cfDNA input between patients with and without an associated mutation. In this sense, only 1 subject had a single ALK mutation, while the remaining 7 patients had it coexisting with alterations in the TP53 gene.

The ALK variants confirmed by dPCR are represented in Table 4.3. As can also be seen in Figure 4.1, the most prevalent mutation was G1202R, identified in 4 different patients who had progressed on alectinib (N=3) and ceritinib (N=1). The G1123C mutation (N=1) also arose as a result of the failure of a second-generation inhibitor. The remaining G1269A (N=2) and A1200V (N=1) alterations were detected upon progression to crizotinib. Furthermore, these mutations were detected at a median MAF of 0.4%.

Patient ID	**Nucleotide Change**	**Amino Acid Change**	**Progression to**	**MAF (dPCR)**
ARS_1	p.G1269A	c.3806G>C	Crizotinib	0.42%
ARS_2	p.G1202R	c.3604G>A	Ceritinib	2.12%
AHT	p.A1200V	c.3599C>T	Crizotinib	-
AMLO	p.G1269A	c.3806G>C	Crizotinib	2.80%
GCP	p.G1123C	c.3367G>T	Alectinib	-
MCPA	p.G1202R	c.3604G>A	Alectinib	0.40%
MESR	p.G1202R	c.3604G>A	Alectinib	0.30%
MSCS	p.G1202R	c.3604G>A	Alectinib	0.01%

Table 4.3. Characteristics of the ALK mutations identified in this study.

On the other hand, in 12 patients only TP53 mutations could be confirmed, in 1 individual concurrence of alterations in PIK3CA and TP53 was found, in another only 1 variant of PIK3CA was detected, and the last patient had mutations in the FGFR1, KRAS, and TP53 genes.

Regarding the different types of the verified variants, they are summarized at a cohort level in Figure 4.2 along with the main clinical characteristics of the patients. In this context, 14 subjects (46.7%) were identified with multiple alterations. The most frequent mutation type was missense SNPs, detected in 40 different cases (29 in TP35, 8 in ALK, 2 in PIK3CA, and 1 in KRAS). Additionally, 12 InDels were also identified (11 in TP35 and 1 in FGFR1).

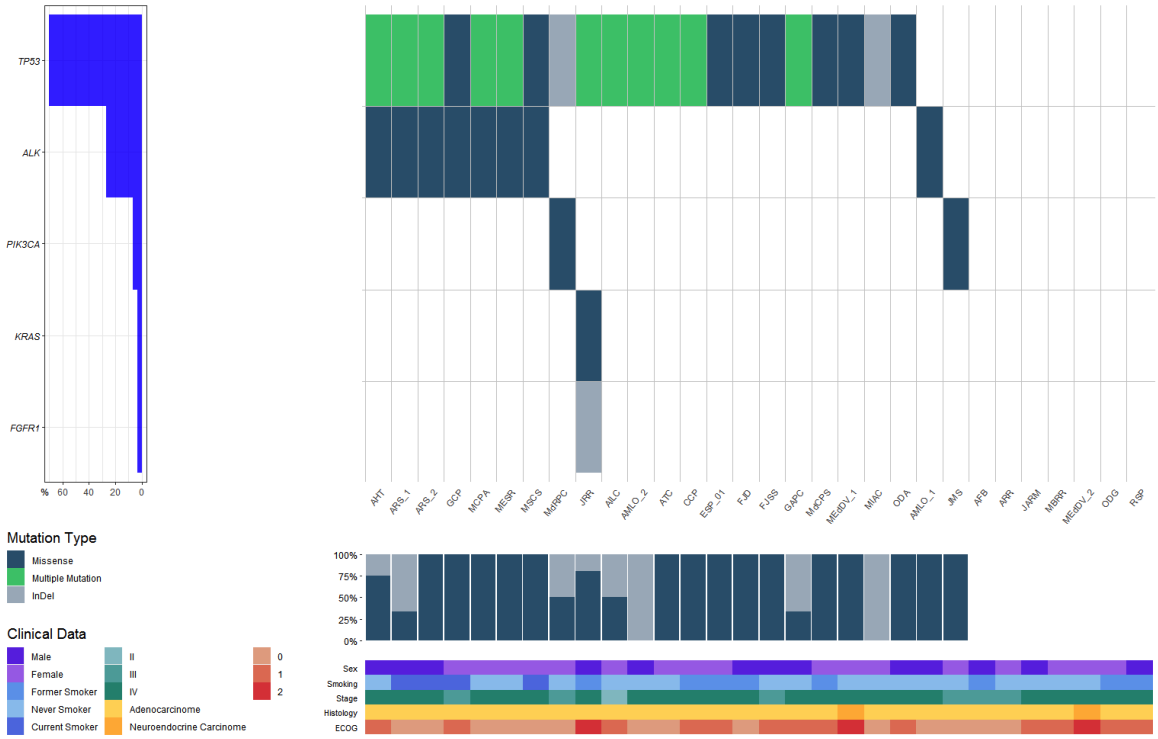


Figure 4.2. Co-mutation plot of the subjects of this study and their corresponding somatic mutations. The clinical features of each of the patients (gender, smoking status, diagnostic stage, histology, and ECOG performance status) and the characteristics of their mutations (type, prevalence, and associated gene) are shown. The multiple alteration class appears broken down into the subtypes that compose it.

Finally, all these results, and especially the confirmed positive (N=8) and negative (N=19) variants associated with the ALK mutations, were used to adjust the different parameters of the implemented algorithm.

4.4 Implemented Pipeline

The development of the bioinformatic pipeline was divided into two steps: the implementation of the filtering algorithm itself; and that of a graphical interface to simplify the process of selecting parameters and saving the output variants along with their properties in a *.csv* file.

4.4.1 Algorithm Characterization

In order to detect all variants at the ALK gene locus, specific conditions for each of them were established from a set of previously validated samples. The flowchart in Figure 4.3 shows the basic structure of the developed pipeline and the selection criteria based on certain variables as presented in the *non-filtered-oncomine.tsv* file.

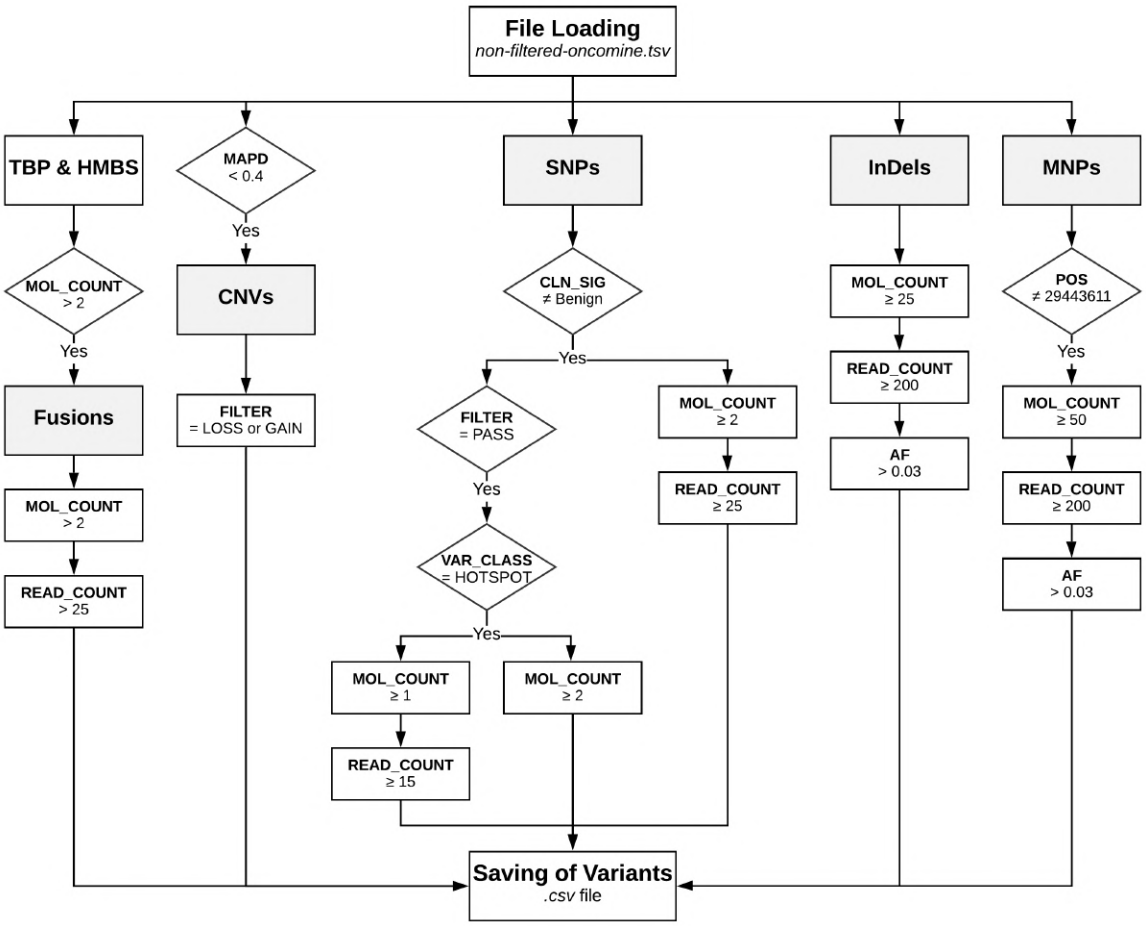


Figure 4.3. Flowchart of the bioinformatic pipeline optimized for the processing and assessment of variants at the ALK gene locus. MOL_COUNT, molecular coverage; READ_COUNT, read coverage; MAPD, median of the absolute values of all pairwise differences; FILTER: Ion Reporter™ internal filter (Oncomine™ Variants v5.12); CLN_SIG, clinical significance; VAR_CLASS: Oncomine™ Variant Class; POS: variant position; AF, allele frequency.

Fusions Filtering

The fusions filter selects translocations at the ALK locus with molecular coverage > 2, and fusion reads > 25. These thresholds were established according to the clinical study data and the Oncomine™ Pan-Cancer Cell-Free Assay recommendations.

Additionally, two control genes were included in each sequencing process to measure the transcript abundance, TATA-binding protein (TBP) and hydroxymethylbilane synthase (HMBS). Both must have molecular coverage > 2 to validate the filtered fusion variants by ensuring their correct amplification.

Copy-Number Variations Filtering

Following the Ion Reporter™ recommendations, to make a CNV call the median of the absolute values of all pairwise differences (MAPD) must be < 0.4. MAPD measures the absolute difference between the log_2 copy-number ratios of adjacent amplicons and then calculates the median across all wells (Equation 4.1).

$$MAPD = median(\mid x_{i+1} - x_i \mid) \tag{4.1}$$
$$\text{where } x_i \equiv log_2 \text{ ratio for marker i}$$

Larger MAPD values indicate lower coverage uniformity and greater noise, resulting in a higher probability of erroneous CNV calls. Therefore, only samples showing an MAPD < 0.4 were considered in further analysis, which consisted of selecting ALK gene copy-number gains or losses.

Single-Nucleotide Polymorphisms Filtering

Taking into account that false positives of this type of variants are not common, the proposed algorithm makes a call as long as any of the following conditions is met, discarding variants with a benign or likely benign clinical significance:

- SNPs in hotspot regions that have passed the Oncomine™ Variants v5.12 filter and that have been detected in at least 1 molecular count with ≥ 15 reads.

- SNPs in hotspot regions that have passed the Oncomine™ Variants v5.12 filter and that have been detected in at least 2 molecular counts.
- SNPs that have been detected in at least 2 molecular counts with ≥ 25 reads.

Insertions and Deletions Filtering

The sequencing results usually present doubtful data regarding InDels. Therefore, the restrictions are more severe for this filter, which selects only variants that have been detected in at least 25 molecular counts with ≥ 200 reads and an allele frequency (AF) ≥ 0.03.

Multiple-Nucleotide Polymorphisms Filtering

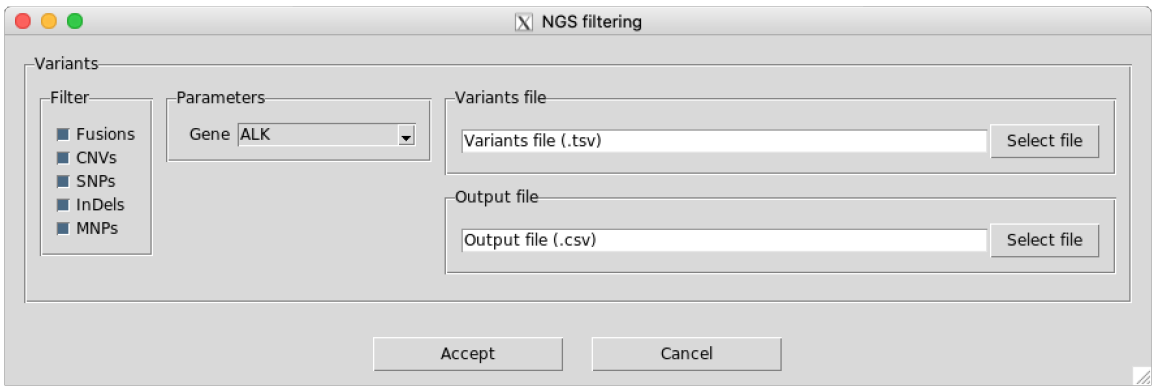
Based on data from confirmed ALK-positive samples, false positives are highly likely in MNPs. In this context, only variants that have been detected in at least 50 molecular counts with ≥ 200 reads and an AF ≥ 0.03 were considered for confirmation by dPCR.

On the other hand, abnormal Ion GeneStudio™ S5 Sequencer behavior at position chr2:29443611 of the ALK locus and involving the reading of 6 consecutive guanines (G) was observed. Thus, variants at that location were excluded from further analysis.

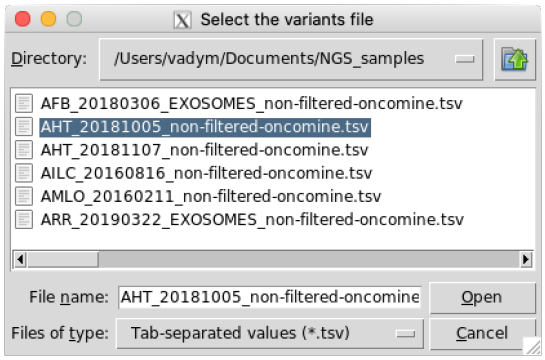
4.4.2 Graphical User Interface (GUI)

The implemented user interface was developed to facilitate data input and interpretation of results. The different layouts are shown in Figure 4.4.

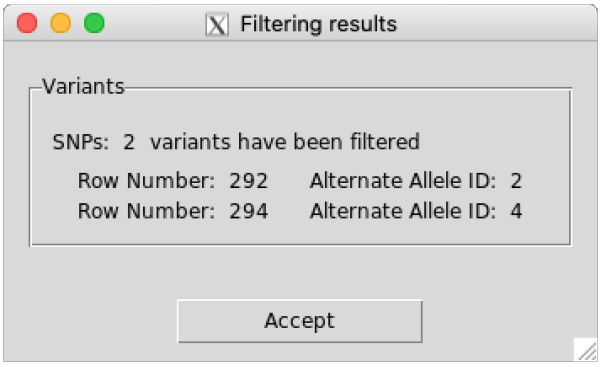
The start screen (Figure 4.4a) has several options that allow some variability when proceeding with the filtering process. On the one hand, it is possible to select the filter/s to apply to the *non-filtered-oncomine.tsv* file, as well as the gene to study. Currently, and as previously described, only the ALK gene has been addressed. On the other hand, the user can select both the source and output files through a pop-up screen, which is shown in Figure 4.4b. In case the output file does not exist, the program creates it automatically.



(a) Start screen. It allows selecting the type of variant/s to filter, the gene involved (ALK currently), as well as the source and output files.



(b) Pop-up window for selecting the *.tsv* source file containing raw variants.



(c) Results screen showing 2 filtered SNPs and their *.tsv* row number and alternate allele ID.

Figure 4.4. Implemented GUI showing the different display areas that appear throughout the filtering process.

Finally, to fully characterize the variants that have passed a particular filter, the results are displayed specifying the mutation type, the row number of the variants in the *non-filtered-oncomine.tsv* file, and the identifier of the alternate allele (Figure 4.4c). Simultaneously, each of the identified variants is appended to the *.csv* output file along with its main characteristics.

It should also be noted that several additional screens have been implemented to offer the end-user information on the progress of the filtering process, on any errors that may have occurred during it, and on the non-identification of any variant of interest by the algorithm.

4.5 Performance of the Implemented Algorithm

The 39 initial sequenced samples were used to study the performance of the developed algorithm regarding the ALK mutations confirmed by dPCR, which was considered the gold standard. To assess this, sensitivity, specificity, and positive and negative predictive values, shown in Table 4.4, were calculated based on the 30 patients in this study. Furthermore, an ALK mutation prevalence of 20% was assumed for these calculations, since approximately 20–50% of patients treated with an ALK inhibitor end up developing some type of resistance [56, 60].

Statistic	Value	95% CI
Sensitivity	87.50%	47.35% to 99.68%
Specificity	81.82%	59.72% to 94.81%
Disease prevalence	20.00%	-
Positive Predictive Value	54.61%	32.31% to 75.20%
Negative Predictive Value	96.32%	80.55% to 99.40%
Accuracy	82.95%	64.83% to 94.13%

Table 4.4. Performance of the developed algorithm to filter ALK mutations. The results obtained by dPCR were taken as the gold standard.

In this context, the algorithm managed to successfully identify 7 of the 8 mutations confirmed by dPCR, reaching a specificity of 87.50%. Specifically, the *AHT* patient sample was discarded by the algorithm (Table 4.3). On the other hand, 4 patients were incorrectly classified as carriers of ALK mutations, thus a specificity of 81.82% was obtained. Finally, the algorithm reported an accuracy of 82.95%, with a remarkable performance in terms of discarding samples without any mutation since a negative predictive value of 96.32% was achieved.

Noteworthy, the Oncomine™ Variants v5.12 filter only detected ALK mutations in 3 patients.

Chapter 5

Discussion and Conclusions

5.1 Discussion

ALK is an established therapeutic target in NSCLC and several other hematologic and solid malignancies. Since its discovery as a fusion oncogene in 1994, numerous ALK inhibitors have been introduced into clinical practice. To date, drug-based treatments such as crizotinib, ceritinib, lorlatinib, and alectinib have become standard therapies for advanced ALK-positive NSCLC. However, despite the remarkable responses seen with ALK TKIs, patients invariably relapse due to acquired resistance, which is the main concern of targeted-therapies in oncogene-addicted NSCLC. Therefore, rebiopsies after progression on TKIs are essential in the treatment decision making. Overall, the logistics of obtaining repeat tumor biopsies are complicated and in most cases not clinically feasible since many patients are unable to undergo an invasive procedure. Additionally, it requires highly specialized personnel and equipment, which is expensive and leads to a delay in care. Furthermore, tissue biopsies may not provide a complete picture of tumor heterogeneity.

In this context, liquid biopsy emerges as an alternative to tissue biopsy techniques, overcoming many of the latter's drawbacks. Specifically, it provides spatial and temporal information on the heterogeneity of the tumor, is minimally invasive, and allows repeated sampling, offering evidence on treatment response and cancer progression. Moreover, it is less expensive and the sample preparation is faster. However, to as-

sess liquid biopsy samples and achieve reliable results, highly sensitive methods are required due to the scarcity of gene alterations.

Currently, ALK mutations can be detected via liquid biopsy using next-generation sequencing (NGS) methods and targeted PCR assays. Both methods are attractive for specific settings using ctDNA or exosome RNA. However, PCR and the approaches derived from it can only detect known genetic changes and the identification of several biomarkers simultaneously is limited. In contrast, NGS can identify multiple genetic alterations, including novel mutations. This advantage is due to the use of amplicon-based NGS assays, such as the one employed in this study (Oncomine™ Pan-Cancer Cell-Free Assay), which provides coverage of the major targetable gene alterations. In this study, with the increase in accessibility of both technologies, NGS and digital PCR had complementary roles in treatment monitoring in metastatic cancer. Specifically, the use of dPCR allowed the confirmation of low-frequency and tumor-specific mutations detected by NGS with a greater sensitivity, providing the high accuracy essential for clinical decision making.

Regarding NGS, it generates a large amount of data that must be correctly interpreted and contrasted to achieve rigorous and reliable outcomes. Because of the lack of published guidance, the way of approaching this analysis changes according to each laboratory, whose members must carry out a thorough validation by themselves to have high confidence in the performance of the NGS results. However, this process of interpreting variants to determine whether they should be considered of clinical significance is complex and often prone to subjectivity and human error. Additionally, processing raw sequence data to detect genomic abnormalities has a direct impact on disease management and patient care. Therefore, appropriate automation of this procedure will improve productivity, contribute to optimized test turnaround time, and improve the accuracy of the outcomes, decreasing the number of unnecessary dPCR confirmation runs and saving resources.

In this context, an algorithm has been developed to fully automate the process of interpreting NGS results from a study population (N=30) diagnosed with ALK-positive NSCLC. This maintainable infrastructure, apart from showing good perfor-

mance in filtering mutations, has considerably reduced the time invested in it. In the first part of this study, an average of 15–20 minutes per sample was invested in manually analyzing the sequence results (39 samples). With the implemented pipeline, this analysis is performed in seconds. Overall, it has also been possible to improve the filtering results on this study population according to the internal filters of the Ion Reporter™ platform (Oncomine™ Variants v5.12), which managed to identify only mutations in 3 (37.5%) out of 8 ALK-positive confirmed patients, compared to 7 (87.5 %) using the designed algorithm. Specifically, for the patient sample (AHT) that did not pass the implemented filter, only 1 molecular family with 12 reads was assessed.

It should be noted that to test the real effectiveness of the algorithm it would be necessary to analyze a more extensive and external data set. The reason is that the developed pipeline may correspond too closely or exactly to the particular mutations in this study, and may, therefore, fail to fit additional data or predict future observations reliably. Therefore, the main limitation of this study is the number of patients included. In this context, larger studies will be needed to validate the findings from this thesis. Nonetheless, this study adds to the evidence that NGS, the developed algorithm, and dPCR have high concordance and accuracy in detecting ALK resistance mutations in lung cancer patients.

Regarding the clinical findings, the average number of mutations per patient was 1.7. Within the spectrum of ALK variants, it should be pointed out that most of them (87.5 %) were identified jointly with TP53 concurrent alterations. The role of TP53 co-mutations in ALK-positive NSCLC is usually associated with increased degrees of genomic copy number instability and a higher somatic mutation burden [26]. Therefore, co-occurring TP53 mutations predict an unfavorable outcome of therapies with TKIs, and in general of systemic therapies. On the other hand, the most prevalent mutation, G1202R, was identified in 4 different patients who had progressed on alectinib (N=3) and ceritinib (N=1). These resistances to second-generation ALK TKIs were reported to oncologists in order to switch the current treatment to another ALK inhibitor such as brigatinib or lorlatinib, which targets G1202R mutations (Table 1.6). Additionally, as shown in Table 4.3, the patient for whom ceritinib

treatment had failed (*ARS*), previously had also developed resistance to crizotinib (G1269A mutation), which is in line with previous studies that demonstrated a progressive appearance of specific mutations for each drug [60, 64]. Finally, it is worth noting the presence of a KRAS mutation in the *JRR* patient, which with a more detailed study could lead to future ALK and EGFR mutations being discarded, since KRAS alterations seem to exclude them [22, 41]. Additionally, since a drug for the G12C mutation in the KRAS gene is currently in the approval process, patients with these characteristics may benefit from this targeted therapy in the near future [89].

As it was observed through this study, treatment decision making is becoming more individualized owing to molecular testing and liquid biopsy. Furthermore, it has been proven that the detection of ALK variants using NGS, the filtering algorithm, and dPCR is useful in selecting the most optimal TKI for therapy. In this regard, the excellent agreement found between these technologies is ideal for monitoring patients. While NGS is an adequate technique for the identification of infrequent or novel molecular alterations, dPCR is responsible for making these confirmations and perform periodic tests on already diagnosed patients with a specific mutation. Additionally, the prescription of second-generation ALK inhibitors is currently performed empirically, without knowing the tumor's molecular profile at progression, which underscores, even more, the importance of introducing these techniques into clinical practice.

In summary, with this study, it has been possible to verify the utility of massive sequencing together with the algorithm implemented for the detection and quantification of biomarkers from liquid biopsy samples in ALK-positive NSCLC patients. In this sense, these techniques have been applied to detect molecular markers, to identify and monitor resistance mutations, and to predict clinical outcomes.

5.2 Conclusions

1. The methodology proposed in this study (liquid biopsy analysis using NGS, filtering of the sequenced data by the implemented algorithm, and confirmation

of the identified alterations by dPCR) allows the detection of mutations in the ALK gene in patients with NSCLC with a detection limit of 0.1%.

2. ALK mutations can be assessed successfully from different types of liquid biopsies: circulating tumor DNA, exosome RNA, and cerebrospinal fluid (to study the central nervous system).

3. The study of ALK mutations through liquid biopsies is feasible and of great clinical utility to identify resistance alterations and monitor the response to TKI treatments. Furthermore, it allows predicting clinical outcomes and detecting oncogenic drivers earlier.

4. Using custom algorithms to filter sequencing results improves the performance of data analysis services offered by NGS platform manufacturers such as Thermo Fisher and improves the efficiency and effectiveness of the entire process.

5. NGS and dPCR are useful technologies for the study of biomarkers and a great complement to each other.

5.3 Future Directions

The implemented platform has been developed to be maintainable and malleable over time as new data is collected. Therefore, the modification of the filter parameters for the ALK gene, as well as the inclusion of new oncogenes (EGFR, ROS1, etc.) should be immediate.

The design of a second filter for the EGFR gene, whose mutation prevalence is even higher than that of ALK, is currently being carried out in the laboratory. Specifically, data collection has started and the first parameters have begun to be established in a procedure similar to that followed for the design of the ALK mutation filtering algorithm.

On the other hand, and in an environment with a considerable amount of data, it would be very convenient to implement a machine learning model. In this sense,

the training process and estimation of parameters could be automated as new data is incorporated.

Finally, it should be noted that part of the results captured in this study will be incorporated into a scientific publication within the context of the project in which this thesis is included.

References

[1] B. Stewart and P. Kleihues, *World Cancer Report*. IARC Press, 2003.

[2] I. Siddiqui *et al.*, "Resveratrol nanoformulation for cancer prevention and therapy," *Annals of the New York Academy of Sciences*, 2015, doi: 10.1111/nyas.12811.

[3] C. L. Chaffer and R. A. Weinberg, "A perspective on cancer cell metastasis," *Science*, 2011, doi: 10.1126/science.1203543.

[4] D. Hanahan and R. Weinberg, "Hallmarks of cancer: The next generation," *Cell*, 2011, doi: 10.1016/j.cell.2011.02.013.

[5] B. F. *et al.*, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, 2018, doi: 10.3322/caac.21492.

[6] Global Cancer Observatory (GCO), "Cancer today," 2018, http://gco.iarc.fr/today/home, Last accessed on 2020-06-16.

[7] National Cancer Institute, "Surveillance, epidemiology, and end results program," 2010-2016, https://seer.cancer.gov/statfacts/html/lungb.html, Last accessed on 2020-06-16.

[8] S. Walters *et al.*, "Lung cancer survival and stage at diagnosis in australia, canada, denmark, norway, sweden and the uk: a population-based study, 2004–2007," *Thorax*, 2013, doi: 10.1136/thoraxjnl-2012-202297.

[9] W. D. Travis *et al.*, "The 2015 world health organization classification of lung tumors: Impact of genetic, clinical and radiologic advances since the 2004 classification," *Journal of Thoracic Oncology*, 2015, doi: 10.1097/JTO.0000000000000630.

[10] G. P. Kalemkerian *et al.*, "Small cell lung cancer," *Journal of the National Comprehensive Cancer Network : JNCCN*, 2013, doi: 10.6004/jnccn.2013.0011.

[11] J. Calbó *et al.*, "Genotype-phenotype relationships in a mouse model for human small-cell lung cancer," *Cold Spring Harbor symposia on quantitative biology*, 2005, doi: 10.1101/sqb.2005.70.026.

[12] C. Zappa and S. A. Mousa, "Non-small cell lung cancer: current treatment and future advances," *Translational lung cancer research*, 2016, doi: 10.21037/tlcr.2016.06.07.

[13] R. N. Proctor, "Tobacco and the global lung cancer epidemic," *Nature Reviews Cancer*, 2001, doi: 10.1038/35094091.

[14] T. Walser *et al.*, "Smoking and lung cancer," *Proceedings of the American Thoracic Society*, 2008, doi: 10.1513/pats.200809-100TH.

[15] L. E. Quint *et al.*, "Distribution of distant metastases from newly diagnosed non-small cell lung cancer," *The Annals of Thoracic Surgery*, 1996, doi: 10.1016/0003-4975(96)00220-2.

[16] American Lung Association, "Lung cancer basics," 2020, https://www.lung.org/lung-health-diseases/lung-disease-lookup/lung-cancer/learn-about-lung-cancer/what-is-lung-cancer/lung-cancer-basics, Last accessed on 2020-06-16.

[17] Z. Zhu *et al.*, "Survival analysis in caucasian pulmonary adenocarcinoma patients based on differential targets between caucasian and asian population," *Saudi Journal of Biological Sciences*, 2018, doi: 10.1016/j.sjbs.2018.05.023.

[18] R. Berardi *et al.*, "Squamous cell carcinoma of the lung: clinical criteria for treatment strategy," *Journal of Cancer Metastasis and Treatment*, 2015, doi: 10.4103/2394-4722.157974.

[19] K. E. Rosenzweig *et al.*, "35 - tumors of the lung, pleura, and mediastinum," in *Leibel and Phillips Textbook of Radiation Oncology (Third Edition)*. W.B. Saunders, 2010, doi: 10.1016/B978-1-4160-5897-7.00036-6.

[20] D. C. Ihde and J. D. Minna, "Non-small cell lung cancer part: I biology, diagnosis, and staging," *Current Problems in Cancer*, 1991, doi: 10.1016/0147-0272(91)90014-2.

[21] J. E. Muscat *et al.*, "Cigarette smoking and large cell carcinoma of the lung," *Cancer Epidemiology and Prevention Biomarkers*, 1997.

[22] O. Calvayrac *et al.*, "Molecular biomarkers for lung adenocarcinoma," *European Respiratory Journal*, 2017, doi: 10.1183/13993003.01734-2016.

[23] D. Sidransky, "Emerging molecular markers of cancer," *Nature Reviews Cancer*, 2002, doi: 10.1038/nrc755.

[24] T. Santarius *et al.*, "A census of amplified and overexpressed human cancer genes," *Nature Reviews Cancer*, 2010, doi: 10.1038/nrc2771.

[25] E. A. Kastelijn, A. J. de Langen, and B. J. M. Peters, "Treatment of oncogene-driven non-small cell lung cancer," *Current opinion in pulmonary medicine*, 2019, doi: 10.1097/mcp.0000000000000572.

[26] F. Skoulidis and J. V. Heymach, "Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy," *Nature Reviews Cancer*, 2019, doi: 10.1038/s41568-019-0179-8.

[27] F. Barlesi *et al.*, "Routine molecular profiling of patients with advanced non-small-cell lung cancer: results of a 1-year nationwide programme of the french cooperative thoracic intergroup (IFCT)," *The Lancet*, 2016, doi: 10.1016/s0140-6736(16)00004-0.

[28] N. I. Lindeman *et al.*, "Updated molecular testing guideline for the selection of lung cancer patients for treatment with targeted tyrosine kinase inhibitors," *The Journal of Molecular Diagnostics*, 2018, doi: 10.1016/j.jmoldx.2017.11.004.

[29] N. B. Leighl *et al.*, "Molecular testing for selection of patients with lung cancer for epidermal growth factor receptor and anaplastic lymphoma kinase tyrosine kinase inhibitors: American society of clinical oncology endorsement of the college of american pathologists/international association for the study of lung cancer/association for molecular pathology guideline," *Journal of Clinical Oncology*, 2014, doi: 10.1200/jco.2014.57.3055.

[30] P. A. Jänne *et al.*, "Selumetinib plus docetaxel for KRAS-mutant advanced non-small-cell lung cancer: a randomised, multicentre, placebo-controlled, phase 2 study," *The Lancet Oncology*, 2013, doi: 10.1016/s1470-2045(12)70489-8.

[31] L. M. Sholl *et al.*, "Multi-institutional oncogenic driver mutation analysis in lung adenocarcinoma: The lung cancer mutation consortium experience," *Journal of Thoracic Oncology*, 2015, doi: 10.1097/jto.0000000000000516.

[32] Jie-Liu *et al.*, "Genotype–phenotype correlation in chinese patients with pulmonary mixed type adenocarcinoma: Relationship between histologic subtypes, TITF-1/SP-a expressions and EGFR mutations," *Pathology - Research and Practice*, 2014, doi: 10.1016/j.prp.2013.11.013.

[33] S. V. Sharma *et al.*, "Epidermal growth factor receptor mutations in lung cancer," *Nature Reviews Cancer*, 2007, doi: 10.1038/nrc2088.

[34] S. P. Ducray *et al.*, "The transcriptional roles of ALK fusion proteins in tumorigenesis," *Cancers*, 2019, doi: 10.3390/cancers11081074.

[35] S. R. Sabir *et al.*, "EML4-ALK variants: Biological and molecular properties, and the implications for patients," *Cancers*, 2017, doi: 10.3390/cancers9090118.

[36] M. Soda *et al.*, "Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer," *Nature*, 2007, doi: 10.1038/nature05945.

[37] M. Patel, J. Malhotra, and S. K. Jabbour, "Examining EML4-ALK variants in the clinical setting: the next frontier?" *Journal of Thoracic Disease*, 2018, doi: 10.21037/jtd.2018.11.07.

[38] T. Sasaki *et al.*, "The biology and treatment of EML4-ALK non-small cell lung cancer," *European Journal of Cancer*, 2010, doi: 10.1016/j.ejca.2010.04.002.

[39] W. Lim *et al.*, "The 8th lung cancer TNM classification and clinical staging system: review of the changes and clinical implications," *Quantitative Imaging in Medicine and Surgery*, 2018, doi: 10.21037/qims.2018.08.02.

[40] P. Goldstraw *et al.*, "The IASLC lung cancer staging project: Proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer," *Journal of Thoracic Oncology*, 2016, doi: 10.1016/j.jtho.2015.09.009.

[41] F. R. Hirsch *et al.*, "Lung cancer: current therapies and new targeted treatments," *The Lancet*, 2017, doi: 10.1016/s0140-6736(16)30958-8.

[42] D. Divisi *et al.*, "National adoption of video-assisted thoracoscopic surgery (VATS) lobectomy: the italian VATS register evaluation," *Journal of Thoracic Disease*, 2018, doi: 10.21037/jtd.2017.11.133.

[43] M. G. Kris *et al.*, "Adjuvant systemic therapy and adjuvant radiation therapy for stage i to IIIA completely resected non–small-cell lung cancers: American society of clinical oncology/cancer care ontario clinical practice guideline update," *Journal of Clinical Oncology*, 2017, doi: 10.1200/jco.2017.72.4401.

[44] H. Lemjabbar-Alaoui *et al.*, "Lung cancer: Biology and treatment options," *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 2015, doi: 10.1016/j.bbcan.2015.08.002.

[45] G. da Cunha Santos, F. A. Shepherd, and M. S. Tsao, "EGFR mutations and lung cancer," *Annual Review of Pathology: Mechanisms of Disease*, 2011, doi: 10.1146/annurev-pathol-011110-130206.

[46] C. S. Kim and M. D. Jeter, *Radiation Therapy, Early Stage Non-Small Cell Lung Cancer*. StatPearls Publishing, Treasure Island (FL), 2019.

[47] D. Owen and J. E. Chaft, "Immunotherapy in surgically resectable non-small cell lung cancer," *Journal of Thoracic Disease*, 2018, doi: 10.21037/jtd.2017.12.93.

[48] M. G. Kris *et al.*, "Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs," *JAMA*, 2014, doi: 10.1001/jama.2014.3741.

[49] E. A. Collisson *et al.*, "Comprehensive molecular profiling of lung adenocarcinoma," *Nature*, 2014, doi: 10.1038/nature13385.

[50] M. Román *et al.*, "KRAS oncogene in non-small cell lung cancer: clinical perspectives on the treatment of an old target," *Molecular Cancer*, 2018, doi: 10.1186/s12943-018-0789-x.

[51] J. G. Bustamante and G. A. Otterson, "Agents to treat BRAF-mutant lung cancer," *Drugs in Context*, 2019, doi: 10.7573/dic.212566.

[52] G. Rossi *et al.*, "Detection of ROS1 rearrangement in non-small cell lung cancer: current and future perspectives," *Lung Cancer: Targets and Therapy*, 2017, doi: 10.2147/lctt.s120172.

[53] G. Bronte *et al.*, "Targeting RET-rearranged non-small-cell lung cancer: future prospects," *Lung Cancer: Targets and Therapy*, 2019, doi: 10.2147/lctt.s192830.

[54] J. Lung *et al.*, "MET exon 14 skipping mutations and gene amplification in a taiwanese lung cancer population," *PLOS ONE*, 2019, doi: 10.1371/journal.pone.0220670.

[55] C. M. Connell and G. J. Doherty, "Activating HER2 mutations as emerging targets in multiple solid cancers," *ESMO Open*, 2017, doi: 10.1136/esmoopen-2017-000279.

[56] D. R. Camidge, W. Pao, and L. V. Sequist, "Acquired resistance to TKIs in solid tumours: learning from lung cancer," *Nature Reviews Clinical Oncology*, 2014, doi: 10.1038/nrclinonc.2014.104.

[57] J. J. Lin, G. J. Riely, and A. T. Shaw, "Targeting ALK: Precision medicine takes on drug resistance," *Cancer Discovery*, 2017, doi: 10.1158/2159-8290.cd-16-1123.

[58] S.-H. I. Ou *et al.*, "Crizotinib for the treatment of ALK-rearranged non-small cell lung cancer: A success story to usher in the second decade of molecular targeted therapy in oncology," *The Oncologist*, 2012, doi: 10.1634/theoncologist.2012-0311.

[59] A. T. Shaw *et al.*, "Crizotinib versus chemotherapy in advanced ALK-positive lung cancer," *New England Journal of Medicine*, 2013, doi: 10.1056/nejmoa1214886.

[60] J. F. Gainor *et al.*, "Molecular mechanisms of resistance to first- and second-generation ALK inhibitors in ALK-rearranged lung cancer," *Cancer Discovery*, 2016, doi: 10.1158/2159-8290.cd-16-0596.

[61] K. Reckamp *et al.*, "Comparative efficacy of brigatinib versus ceritinib and alectinib in patients with crizotinib-refractory anaplastic lymphoma kinase-positive non-small cell lung cancer," *Current Medical Research and Opinion*, 2018, doi: 10.1080/03007995.2018.1520696.

[62] L. Friboulet *et al.*, "The ALK inhibitor ceritinib overcomes crizotinib resistance in non-small cell lung cancer," *Cancer Discovery*, 2014, doi: 10.1158/2159-8290.cd-13-0846.

[63] S. Peters *et al.*, "Alectinib versus crizotinib in untreated ALK-positive non–small-cell lung cancer," *New England Journal of Medicine*, 2017, doi: 10.1056/nejmoa1704795.

[64] V. Mazza and C. Dazzi, "Targeting ALK-positive non-small-cell lung cancer–novel inhibitors beyond crizotinib," *Cancer Reports and Reviews*, 2018, doi: 10.15761/crr.1000168.

[65] J. J. Lin, G. J. Riely, and A. T. Shaw, "Targeting ALK: Precision medicine takes on drug resistance," *Cancer Discovery*, 2017, doi: 10.1158/2159-8290.cd-16-1123.

[66] J. F. Gainor *et al.*, "Molecular mechanisms of resistance to first- and second-generation alk inhibitors in alk-rearranged lung cancer," *Cancer Discovery*, 2016, doi: 10.1158/2159-8290.CD-16-0596.

[67] B. Solomon, T. Bauer, and E. Felip, "Safety and efficacy of lorlatinib (PF-06463922) from the dose-escalation component of a study in patients with advanced ALK+ or ROS1+ non-small cell lung cancer (nsclc)," *ASCO Meeting Abstracts*, 2016, doi: 10.1200/JCO.2016.34.15_suppl.9009.

[68] G. Recondo *et al.*, "Diverse resistance mechanisms to the third-generation ALK inhibitor lorlatinib in ALK-rearranged lung cancer," *Clinical Cancer Research*, 2019, doi: 10.1158/1078-0432.ccr-19-1104.

[69] C. A. Perez *et al.*, "Overcoming the resistance to crizotinib in patients with non-small cell lung cancer harboring EML4/ALK translocation," *Lung Cancer*, 2014, doi: 10.1016/j.lungcan.2014.02.001.

[70] D. R. Camidge, W. Pao, and L. V. Sequist, "Acquired resistance to TKIs in solid tumours: learning from lung cancer," *Nature Reviews Clinical Oncology*, 2014, doi: 10.1038/nrclinonc.2014.104.

[71] P. Maione *et al.*, "Overcoming resistance to targeted therapies in NSCLC: current approaches and clinical application," *Therapeutic Advances in Medical Oncology*, 2015, doi: 10.1177/1758834015595048.

[72] F. Cabillic *et al.*, "Parallel FISH and immunohistochemical studies of ALK status in 3244 non–small-cell lung cancers reveal major discordances," *Journal of Thoracic Oncology*, 2014, doi: 10.1097/jto.0000000000000072.

[73] D. Trombetta *et al.*, "Liquid biopsy and NSCLC," *Lung Cancer Management*, 2016, doi: 10.2217/lmt-2016-0006.

[74] M. Santarpia *et al.*, "Liquid biopsy for lung cancer early detection," *Journal of Thoracic Disease*, 2018, doi: 10.21037/jtd.2018.03.81.

[75] A. Tognela *et al.*, "Predictive and prognostic value of circulating tumor cell detection in lung cancer: A clinician's perspective," *Critical Reviews in Oncology/Hematology*, 2015, doi: 10.1016/j.critrevonc.2014.10.001.

[76] W. J. Allard, "Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases," *Clinical Cancer Research*, 2004, doi: 10.1158/1078-0432.ccr-04-0378.

[77] C. Bettegowda *et al.*, "Detection of circulating tumor DNA in early- and late-stage human malignancies," *Science Translational Medicine*, 2014, doi: 10.1126/scitranslmed.3007094.

[78] V. Gedvilaitė, D. Schveigert, and S. Cicėnas, "Cell-free DNA in non-small cell lung cancer," *Acta medica Lituanica*, 2017, doi: 10.6001/actamedica.v24i2.3495.

[79] L. A. Diaz and A. Bardelli, "Liquid biopsies: Genotyping circulating tumor DNA," *Journal of Clinical Oncology*, 2014, doi: 10.1200/jco.2012.45.2011.

[80] C. Rolfo *et al.*, "Liquid biopsy for advanced non-small cell lung cancer (NSCLC): A statement paper from the IASLC," *Journal of Thoracic Oncology*, 2018, doi: 10.1016/j.jtho.2018.05.030.

[81] I. Vanni *et al.*, "Exosomes: a new horizon in lung cancer," *Drug Discovery Today*, 2017, doi: 10.1016/j.drudis.2017.03.004.

[82] D. Pérez-Callejo *et al.*, "Liquid biopsy based biomarkers in non-small cell lung cancer for diagnosis and treatment monitoring," *Translational Lung Cancer Research*, 2016, doi: 10.21037/tlcr.2016.10.07.

[83] L. Crino and G. Metro, "Therapeutic options targeting angiogenesis in nonsmall cell lung cancer," *European Respiratory Review*, 2014, doi: 10.1183/09059180.00008913.

[84] M. G. Best *et al.*, "RNA-seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics," *Cancer Cell*, 2015, doi: 10.1016/j.ccell.2015.09.018.

[85] S. Goodwin, J. D. McPherson, and W. R. McCombie, "Coming of age: ten years of next-generation sequencing technologies," *Nature Reviews Genetics*, 2016, doi: 10.1038/nrg.2016.49.

[86] Thermo Fisher Scientific, "Oncomine Pan-Cancer Cell-Free Assay," 2018, https://www.thermofisher.com/order/catalog/product/A37664#/A37664, Last accessed on 2020-06-16.

[87] M. M. Li *et al.*, "Standards and guidelines for the interpretation and reporting of sequence variants in cancer," *The Journal of Molecular Diagnostics*, 2017, doi: 10.1016/j.jmoldx.2016.10.002.

[88] Thermo Fisher Scientific, "QuantStudio™ 3D Digital PCR System USER GUIDE," 2015, https://tools.thermofisher.com/content/sfs/manuals/MAN0007720.pdf, Last accessed on 2020-06-16.

[89] J. Canon *et al.*, "The clinical KRAS(g12c) inhibitor AMG 510 drives anti-tumour immunity," *Nature*, 2019, doi: 10.1038/s41586-019-1694-1.