universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"Prognostic and predictive biomarkers in non-small cell lung cancer with emphasis on epidermal growth factor receptor mutations"

verfasst von / submitted by

Christian Tomuta, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2016 / Vienna 2016

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 834
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Molekulare Biologie
Betreut von / Supervisor:	Assoc. Prof. Univ.-Doz. Mag. Dr. Martin Filipits

Table of Contents

Acknowledgements ..... - 1 -

Introduction ..... - 2 -

Incidence and mortality of lung cancer ..... - 2 -

Types, subtypes and conventional treatment of lung cancer ..... - 2 -

Biomarkers in non-small cell lung cancer ..... - 3 -

Definition ..... - 3 -

The epidermal growth factor receptor ..... - 3 -

Activating EGFR mutations ..... - 4 -

Tyrosine kinase inhibitors ..... - 7 -

Non-small cell lung cancer and the immune system ..... - 9 -

Other biomarkers ..... - 10 -

Materials & methods ..... - 11 -

Analysis of *EGFR* mutation status ..... - 11 -

Genomic DNA purification and isolation ..... - 11 -

Primer design and amplification of DNA regions of interest ..... - 12 -

Agarose gel electrophoresis ..... - 14 -

Pyrosequencing® ..... - 14 -

Method overview ..... - 14 -

Sequencing primers ..... - 16 -

Reagents and solutions for Pyrosequencing® ..... - 18 -

Platform and accessories for Pyrosequencing® ..... - 18 -

Preparation of samples, reagents & execution of Pyrosequencing® reaction ..... - 18 -

Analysis of immune markers and selected samples harboring *EGFR* mutations ..... - 20 -

Immunohistochemistry ..... - 20 -

Method overview ..... - 20 -

Reagents and solutions for immunohistochemistry ..... - 22 -

Study population ..... - 24 -

Data analysis & calculations ..... - 25 -

Results ..... - 25 -

DNA Mutations ..... - 25 -

Immune markers ..... - 32 -

Discussion ..... - 43 -

DNA mutations in lung cancer as molecular biomarkers ..... - 43 -

EGFR mutations ..... - 43 -
KRAS mutations ..... - 45 -
Immune markers in lung cancer ..... - 46 -
Mutational analysis and DNA sequencing ..... - 49 -
Immunohistochemistry ..... - 51 -
Concluding remarks ..... - 52 -
Reference list ..... - 54 -
List of figures ..... - 60 -
List of tables ..... - 61 -
Zusammenfassung / Abstract ..... - 63 -

Acknowledgements

First and foremost, I'd like to thank Assoc. Prof. Mag. Dr. Martin Filipits for accepting me into his lab, the interesting topic he assigned to me as well as his constant support during the practical, the writing of the thesis and the search for PhD positions.

Furthermore, I'd like to extent my gratitude to Anita Brandstetter, Mag. Jorun Buresch, Anna Buder, MSc., Stefan Traint, BSc., and Mag. Gerald Timelthaler for helping me to get accustomed to the institute, for teaching me the methods that I've applied and for the positive lab atmosphere they all created.

Most importantly, I want to thank my father, Dr. med. vet. Silviu Tomuta, for all his support – financially as well as private - throughout the years. I couldn't imagine a better father than him. The same amount of gratitude applies to my grandparents, Silvia and Adolf Müller. Furthermore, I thank the Austrian part of my family, Dr. Elena Tomuta and Dr. Bernhard Wall, for their support.

Last but not least, I'd like to thank my best friends, Varinia Neuhauser, Patrick Jambura and Sebastian Ruhnau for always being there for me. You are like family.

Introduction

Incidence and mortality of lung cancer

From all types of cancer, lung cancer has the highest incidence rate. In 2012, lung cancer made up 12.9% of all cancers diagnosed worldwide. Furthermore, due to the extent of tobacco abuse, the limited success of standard chemotherapy and the fact that in most cases, lung cancer is diagnosed in the late stages of the disease, it is the most common cancer death worldwide; the overall ratio of mortality to incidence is 0.87, making lung cancer a particularly fatal disease. Neoplasm formation in the lung is mainly attributable to tobacco smoking; consequentially, there are regional differences in incidence rates across different regions in the world and between men and women. While incidence rates are generally decreasing in industrialized nations and among males, they are increasing in developing countries and among females.[1]

Types, subtypes and conventional treatment of lung cancer

In clinical routine, lung cancer is divided into small cell lung cancer (SCLC), accounting for 10-15% of all lung cancers, and non-small cell lung cancer (NSCLC), accounting for 85-90% of all lung cancers. This classification arose due to the phenotype of the tumor cells under the microscope. Although somewhat outdated, medical professionals still use this classification up until today. SCLC is the most aggressive form of lung cancer, is strongly associated with cigarette smoking and tends to metastasize rapidly.[2] NSCLC can be further subdivided into adenocarcinoma, making up 40%, squamous cell (epidermoid) carcinoma, making up 25-30%, and large cell carcinoma, making up 10-15% of all lung cancers diagnosed.[3] Treatment options are generally limited by the stage in which lung cancer is diagnosed. In stages I & II, which describe the early and localized stages of the disease, surgery is the main therapy of choice. The tumor and / or part of the lung are removed and adjuvant chemotherapy is indicated in many cases in order to get rid of the complete tumor burden. Stage IIIA describes the locally advanced stage. Surgery along with mediastinal lymphadectomy and adjuvant chemotherapy is the therapy most commonly administered for patients in this stage. While treatment regimens for stages I through IIIA aim to cure the patient from the malignant disease, conventional treatment options in stages IIIB and IV, which are the late and metastatic stages of lung cancer, are limited to palliative care: chemotherapy mainly aims to prolong the patient's life and the palliation of symptoms.[4] Unfortunately, lung cancer is most often diagnosed in stage IV due to the absence of symptoms in the early stages, and the 5-year observed survival rate at this stage ranges around 1% even with chemotherapy. Furthermore, chemotherapy consists of cytotoxic drugs, thus they have an unfavorable side effect profile

with relatively low response rates. Also, drug resistance emerges after prolonged treatment, eventually exhausting all treatment options for patients. Due to these circumstances, molecular research in the field of biomarkers in non-small cell lung cancer became an important cornerstone. Rather than the application of cytotoxic agents which pose a considerable burden on the whole organism, molecular mechanisms that lead to malignant diseases are elucidated and monitored in order to characterize the disease and its course. Furthermore, novel drugs can be developed that tackle the exact mechanisms responsible for triggering the malignant disease. In this master thesis, the most important molecular biomarkers in non-small cell lung cancer will be characterized, as well as novel treatment options. Moreover, the results of the statistical analyses of certain biomarkers from 161 lung adenocarcinoma samples available on formalin-fixed, paraffin-embedded blocks will be discussed, followed by an interpretation of the results and a discussion of materials and methods.

Biomarkers in non-small cell lung cancer

Definition

Biomarkers are characteristics that are objectively measured and evaluated as indicators of normal biological processes, pathogenic processes, or pharmacologic responses to therapeutic intervention.[5] In clinical routine, prognostic and predictive biomarkers can be distinguished. A prognostic biomarker is a situation or condition, or a characteristic of a patient that can be used to estimate the chance of recovery from a disease or the chance of the disease recurring, while a predictive biomarker is a condition or finding that can be used to help predict whether a person's cancer will respond to a specific treatment. Predictive factor may also describe something that increases a person's risk of developing a condition or disease.[6]

The epidermal growth factor receptor

A large area of lung cancer biomarker research is devoted to the characterization and the treatment of aberrant epidermal growth factor receptor signaling. Of particular interest are somatic mutations in the gene encoding the epidermal growth factor receptor, which confer constitutive activity of EGFR signaling, thus delivering constant growth- and proliferation signals to the cell. The epidermal growth factor receptor belongs to the ErbB family of receptor tyrosine kinases and is also known as ErbB1 and HER1 in humans.[7] Like all receptor tyrosine kinases, each ErbB receptor comprises a large extracellular region, a single-spanning transmembrane (TM) domain, an intracellular juxtamembrane (JM) region, a tyrosine kinase domain and a C-terminal regulatory region.[8] The extracellular region of the epidermal growth factor receptor can be broken down into four domains (domains I, II, III, and IV). Of these,

domains I and III share 37% sequence homology, adopt β-helix folds and fulfill the function of ligand binding, while domains II and IV adopt laminin-like folds, are cysteine-rich and responsible for dimerization upon ligand binding. The tyrosine kinase domain can be subdivided into an N-terminal lobe and a C-terminal lobe.[9]

At least seven different EGF agonists are known: EGF itself, TGFα, amphiregulin, β-cellulin, epigen, and heparinbinding EGF-like growth factor.[8]

Upon binding of a ligand, dimerization of the extracellular region of EGFR takes place: a β-hairpin in domain II, referred to as the 'dimerization arm', makes extensive contacts with the domain II of its binding partner and reaches up to its opponent's domain I.[8] The tyrosine kinase domain consists of two subunits, called N-lobe and C-lobe. The N-lobe is composed of a five-stranded β sheet and one prominent α helix, called helix αC. The C lobe is larger and is predominantly helical. The substrate, ATP, is bound within a deep cleft between these two lobes. A structure within the N-lobe, called the P-loop (phosphate-binding loop[10] or nucleotide-binding loop[8]), comes into close proximity to the phosphates of ATP and coordinates them via backbone interactions. Since this loop is rich in glycines, it exhibits a certain flexibility, which facilitates the binding of small molecule inhibitors. An activation loop,[10] which is part of the C-lobe,[11] then provides a platform for the peptide substrate close to the γ-phosphate of ATP. In the active state of the kinase, the activation loop is phosphorylated, which leads to an open and extended conformation, thus making it permissive for substrate binding. The phosphorylated tyrosines of the peptide substrate serve as docking sites for proteins which initiate intracellular signaling via several pathways, such as the Ras/Raf/mitogen-activated protein kinase pathway, which is responsible for phosphorylation of transcription factors involved in cell survival and proliferation, the phosphatidylinositol 3-kinase/Akt pathway, involved in cell growth, apoptosis resistance, invasion and migration, phospholipase Cγ, responsible for activation of MAPK and c-jun activation. Furthermore, signal transducers and activators of transcription (STAT) proteins can drive the expression of specific target genes. Src kinase pathways, which are activated as well, can themselves activate a series of substrates, including STATs and focal adhesion kinases (FAKs), which regulate adhesion and migration.[12]

Activating EGFR mutations

The epidermal growth factor receptor plays a vital role in the development and maintenance of cancer diseases. *EGFR* overexpression through gene amplification and/or elevated levels of its cognate ligands causing autocrine stimulation were detected in head and neck, ovarian,

cervical, bladder and esophageal cancers, as well as gastric, breast, endometrial and colorectal cancers and can be used as prognostic indicators.[13] Another mechanism that leads to the rise of lung cancer is *EGFR* mutations. These activating mutations occur in 10-15% of all lung adenocarcinoma patients from European or African descent and in 35% of all East Asians with the disease.[14,15] Furthermore, it is most common among never-smokers and light smokers, younger and female patients and emerges almost exclusively in tumors with adenocarcinoma histology.[15] Germline mutations in the *EGFR* gene occur rarely; most of the mutations are of somatic nature.[16] All activating *EGFR* mutations occur within the exons encoding the components of the tyrosine kinase domain. The nature of the mutations is diverse: deletions, insertions and missense point mutations can occur. Of great significance is a point mutation in exon 21, which encodes the activation loop of the tyrosine kinase: the L858R mutation accounts for 40-45% of all tyrosine kinase activating mutations.[17] This single nucleotide substitution is located near the conserved aspartic acid-phenylalanine-glycine sequence that stabilizes the A-loop, thus facilitating downstream protein activation.[11] In-frame deletions (exon 19, ~40%) and in-frame duplications/insertions occur on either side of the C-helix. It has been hypothesized that these mutations narrow the ATP binding cleft, resulting in increased gene activation.[11,14,18] Furthermore, there is a mutation that arises after prolonged treatment with tyrosine kinase inhibitors as resistance mechanism. This gatekeeper mutation, so named due to its position close to the ATP binding cleft, is a point mutation in exon 20, leading to an exchange of threonine to methionine at position 790 and has been demonstrated to enhance the receptor's affinity to ATP.[19] Table 1 shows a summary of mutations known in the gene sequence of *EGFR*, their exon positions and their positions in the mature protein divided into sensitivity and resistance to the first-generation TKIs erlotinib and gefitinib.[17,20]

Exon	**Mutation associated with drug sensitivity**	**Mutation associated with drug resistance**
Exon 18	G719C	
	G719S	
	G719A	
	V689M	
	N700D	

	E709K/Q	
	S720P	
Exon 19	ΔE746-A750	
	ΔE746-T751	
	ΔE746-A750 (ins RP)	
	ΔE746-T751 (ins A/I)	
	ΔE746-T751 (insVA)	
	ΔE746-S752 (ins A/V)	
	ΔL747-E749 (A750P)	
	ΔL747-A750 (ins P)	
	ΔL747-T751	
	ΔL747-T751 (ins P/S)	
	ΔL747-S752	
	ΔL747-752 (E746V)	
	ΔL747-752 (P753S)	
	ΔL747-S752 (ins Q)	
	ΔL747-P753	
	ΔL747-P753 (ins S)	
	ΔS752-I759	
		D761Y
Exon 20	V765A	
	T783A	
		T790M
		D770_N771 (ins NPG)

		D770_N771 (ins SVQ)
		D770_N771 (ins G), N771T
		V769L
		S768I
		C797S
Exon 21	L858R	
	N826S	
	A839T	
	K846R	
	L861Q	
	G863D	

Table 1: EGFR mutations in non-small cell lung cancer according to sensitivity and resistance to gefitinib and erlotinib, two first-generation TKIs.[17]

Since the *EGFR* gene has the ability to maintain the malignant phenotype in tumor cells, it is termed an oncogene.[21] It has previously been shown that certain cancers that have developed through the normal process of multistage carcinogenesis are largely maintained because of a single mutation.[22] The aforementioned mutations in the *EGFR* gene are an example of "oncogene addiction"; meaning the cancer cells are physiologically dependent on the continued activity of specifically activated or overexpressed oncogenes for the maintenance of their malignant phenotype.[18] Since outstanding work has been conducted in determining the structure of the EGF receptor, it is now known where these mutations are located and how they contribute to the activating mutation phenotype. In the case of NSCLC, the existence of one of the aforementioned mutations in exons encoding the tyrosine kinase domain is sufficient for the maintenance of the malignant phenotype. Due to this fact, the epidermal growth factor receptor is a target for tyrosine kinase inhibitors (TKIs).

Tyrosine kinase inhibitors

In order to counteract the aforementioned epidermal growth factor receptor mutations, tyrosine kinase inhibitors (TKIs) have been developed and are administered to patients

harboring these mutations replacing conventional chemotherapy. Until today, three generations of TKIs are on the market and are frequently used in clinical routine. The first generation is the reversible inhibitors, such as gefitinib and erlotinib. They bind to the ATP binding pocket of the tyrosine kinase domain in a reversible fashion, thereby competing with its substrate ATP for the space.[23] While in some trials,[24] patients treated with reversible inhibitors show response rates and median progression-free survival rates superior to treatment with platinum-based chemotherapy, there was a desire to raise response rates even more and to also successfully treat patients with unkown activating *EGFR* mutations. Thus, irreversible TKIs were developed, which belong to the second generation of this drug group. Afatinib is a prominent example. Afatinib binds to the active center of the enzyme, thus blocking the location for its substrate ATP and shut it down permanently. However, given the fact that these drugs also permanently shut down wildtype EGF receptors, side effect rates using afatinib are higher than for first-generation TKIs.[24] Also, irreversible TKIs fail to inhibit EGFR in patients who developed the T790M mutation due to preceding treatment with reversible TKIs.[25] However, the recently published LUX Lung 7 study clearly shows improved clinical benefit of afatinib over gefitinib in the first-line treatment of patients with *EGFR* mutation-positive NSCLC: patients treated with afatinib have a 27% reduced risk of lung cancer progression and treatment failure than those treated with gefitinib.

Furthermore, even though not statistically significant, the median duration of response to afatinib outcompetes gefitinib by 1.7 months (10.1 months vs. 8.4 months, respectively).[26]

TKIs of the third generation, such as osimertinib and rociletinib, are mutant-selective, target mutated EGF receptors specifically and bind irreversibly.[20] They have been developed in response to emergence of the gatekeeper mutation T790M within exon 20, which is responsible for conferring resistance to first- and second generation TKIs. Several phase II and III clinical trials are currently ongoing to test osimertinib as first/second line treatment or maintenance treatment, including FLAURA, AURA 2 & 3, ADAURA and ASTRA. Furthermore, several phase I and II trials are ongoing that compare osimertinib and rociletinib with each other: response rates for osimertinib and rociletinib were 61% and 59%, respectively, and the median progression-free survival ranged between 9.6 months for osimertinib and 13.1% for rociletinib.[25,27,28] Treatment options involving TKIs are exhausted, however, after the patients have acquired the C797S mutation as a consequence of prolonged treatment with third generation TKIs. Another possibility of inactivating the receptor is monoclonal antibodies. Unlike TKIs, which bind to the tyrosine kinase domain of the receptor

and thus to an intracellular site, monoclonal antibodies bind extracellularly, thus competing with EGF for the ligand binding sites of the receptor. Several EGFR-directed monoclonal antibodies exist, including necitumumab, matuzumab, panitumumab, and, its most prominent representative, cetuximab.[29] Aside from its inhibitory role in EGFR signaling, cetuximab has recently been shown to inhibit T790M-mutated EGF receptors in lung adenorcarcinoma patient-derived xenograft mouse model,[30] thus prolonging the need for fourth-generation TKIs that are active against this specific mutation. Treatment with cetuximab further leads to enduring signal termination by antibody-mediated EGFR internalisation and degradation.[31,32] Moreover, it has been shown that cetuximab plays a role in antibody-dependent cellular cytotoxicity,[29,31] which further contributes to tumor elimination. Several clinical phase II and III trials have been conducted with cetuximab administered in a combination therapy with conventional chemotherapy; nearly all trials have demonstrated improved response and median survival rates compared to chemotherapy alone.[29]

It can be concluded that treatment for NSCLC becomes more and more individualized. Consequently, it is vital to assess the molecular pathology of the tumor. One method to do so is the sequencing of the DNA that encodes the epidermal growth factor receptor. With the mutational status of *EGFR* and the mutation frequency at hand, patients can receive the appropriate therapy, i.e. TKIs or monoclonal antibodies.

Non-small cell lung cancer and the immune system

The role of the immune system in cancer becomes increasingly important in research. Destruction of tumor cells by the immune system has previously been described as an emerging hallmark of cancer by Hanahan and Weinberg.[33] The immune system's role is based on the finding that immune cells, especially cytotoxic T-cells, recognize tumor cells as foreign and destroy them, thus hindering tumor growth. Cancer cells, however, have developed mechanisms to evade immune destruction. Considering the interaction between host cells and tumor cells, three essential phases have been proposed: elimination, equilibrium, and escape, designated the "three E's".[34] During elimination, cells of the innate and adaptive immune system recognize malignant cells and subsequently destroy them. As tumor cells begin to acquire mechanisms to escape immune surveillance, elimination is followed by equilibrium. This phase is characterized by a balance between immune destruction and propagation of cells that developed a resistance to the immune system's mechanism. In the escape phase, tumor cells readily escape destruction by the immune system.[34]

In non-small cell lung cancer, one particular mechanism to evade tumor cell destruction by the immune system has been extensively studied. It involves the interaction between the programmed cell death (PD)-1 receptor, and the programmed cell death ligands (PD-L) 1 and 2.

PD-1 is encoded by the *Pdcd1* gene. It is a 50-55kDa type I transmembrane glycoprotein and is expressed on B and T cells.[35] PD-L1 and PD-L2 are encoded by the *CD274*[36] and the *PDCD1LG2*[37] genes, respectively, and are expressed on certain tumor cells, among other cell types.[14]

Interaction between PD-1, PD-L1 and/or PD-L2 leads to a process called "T-cell exhaustion", which describes the repression of T-cell maturation and the subsequent hindrance of the respective cell's destruction by the immune system. Under normal physiological processes, this system serves as a mechanism for the immune system to distinguish between "self" and "non-self" antigens. However, tumor cells have the ability to hijack this mechanism in order to avoid destruction by the immune system by expressing PD-L1 and/or PD-L2.[38]

In a study performed by Calles *et al.*[39], it has been elucidated that 24% of lung cancer patients with confirmed *KRAS*-mutation show PD-L1 expression, 47% show PD-L2 expression, and 68% show PD-1 expression. Furthermore, PD-L1 expression was demonstrated to correlate with a history of tobacco smoking in regard to prevalence and intensity.

Extensive research in the field of cancer immunity lead to the development of antibodies that disrupt the PD-1 / PD-L1/2 interaction by binding to either PD-1 (nivolumab, pembrolizumab) or PD-L1 (MPDL3280A, MEDI4736).[38] In a study performed by Herbst *et al.*[40], 1034 patients with a PD-L1 tumor proportion score of at least 1% were allocated to receive either pembrolizumab or docetaxel. It has been shown that treatment with pembrolizumab yields more favorable outcomes in respect to overall survival and progression-free survival, as well as a more acceptable side effect profile compared to treatment with docetaxel. Furthermore, a subgroup analysis of the patients revealed that all patient subgroups favor pembrolizumab over docetaxel in regard to overall survival.

Other biomarkers

As previously mentioned conventional chemotherapy as treatment for NSCLC is characterized by low response rates, severe side effect profiles and lacking personalization. The progress in the elucidation of the molecular mechanisms underlying the malignant disease, as well as the development of new drugs that can be implemented in novel treatment

plans that target these molecular mechanisms is continuous. Of particular interest are mutations in the *KRAS* gene, encoding the K-RAS protein, because these mutations occur in about 25%[24] of all adenocarcinomas and are thus the largest group of gene mutations in adenocarcinoma. Furthermore, these mutations are correlated with tobacco abuse, adding to their importance due to the fact that the vast majority of lung cancer cases are tobacco-induced. Until today, it is not possible to inhibit K-RAS signaling directly,[24] leaving the inhibition of downstream targets as the only option. In this context, a phase II study was conducted using the ERK inhibitor selumetinib in combination with the chemotherapeutic agent docetaxel for treatment of patients with advanced non-small cell lung cancer harboring *KRAS* mutations. Although this combination therapy has proven to be effective in terms of progression-free survival and overall survival compared to treatment with docetaxel alone, an increased number of adverse effects are occurring and further investigation in this field is still lacking.[41]

Another set of important biomarkers in NSCLC are *ALK* & *ROS1* rearrangements, as well as *MET* amplifications and mutations. Although they only make up 5, 1-2, 1-2, and 2-4% of all mutations in adenocarcinoma, respectively,[24] they are all targets of the drug crizotinib, an ALK inhibitor that has been investigated and is under current investigation in a number of clinical trials. It has been shown that patients harboring the aforementioned biomarkers treated with crizotinib show durable responses and, in some cases, more favorable outcomes than those treated with standard chemotherapy.[42]

It is important to note that all DNA mutations in a given tumor are of similar importance because malignant diseases are generally characterized by their extensive heterogeneity of mutations among clonal subpopulations.[43,44] Furthermore, it has been shown that the exerting mechanism of oncogene addiction can switch from one oncogene to another after exposure to treatment against the initial addiction. This model, termed “genetic streamlining”, describes the consequences of selective pressure on one pathway[44] and has been demonstrated in *EGFR*-mutated adenocarcinoma, in which tumors underwent amplification of *MET* in 15 to 20% of cases as a resistance mechanism towards EGFR TKIs.[45]

Materials & methods

Analysis of *EGFR* mutation status

Genomic DNA purification and isolation

For genomic DNA isolation, the QIAamp® DNA FFPE Tissue Kit (Qiagen, Hilden, Germany) was used according to the included protocol. Generally, DNA isolation from FFPE

blocks consists of six steps: paraffin removal with xylene, sample lysis under denaturing conditions with proteinase K, removal of formalin crosslinking by heating of the sample, DNA binding to membrane and flow through of contaminants, removal of residual contaminants by washing, and elution of the purified DNA.

From each sample from formalin-fixed, paraffin-embedded (FFPE) blocks, 4 sections, each 5µm thick, were cut using a microtome and placed in an Eppendorf reaction tube. DNA was eluted in 40µl ATE buffer.

After DNA isolation, we measured the DNA concentration in the samples using the Nanodrop 1000 spectrophotometer (Peqlab, Biotechnologie GmbH, Polling, Austria). Due to the variability of the DNA concentrations between the samples, all samples were diluted with ATE buffer to a final concentration of 2ng/µl DNA. In addition, some samples were later diluted to a concentration of 10ng/µl due to weak bands in the consecutive analytical gel electrophoresis following DNA amplification by polymerase chain reaction.

Primer design and amplification of DNA regions of interest

Polymerase chain reaction (PCR) was performed to amplify the DNA regions of interest. Amplicons for the analysis of mutations in codons 719, 768, 769, 790 and 858 through 861 were produced, as well as for deletions and complex mutations in exon 19:

Exon	Encoding TK subunit	Codon(s)
18	Nucleotide-binding loop[17]	719
19	C-helix[46]	several
20	C-helix[47]	768, 769, 790
21	Activation loop[17]	858 - 861

Table 2: Examined exons, codons and mutations as well as their role within the EGFR tyrosine kinase domain.

In total, four DNA amplicons were generated: one for codon 719, one for exon 19 deletions, one for codons 768, 769 & 790, and one for codons 858 - 861. The amplicon covering codons 768, 769 and 790 is divided into two sequencing reactions, whilst all other amplicons are sequenced in one reaction[48] (see later). Table 3 lists the forward and reverse primers that were used in the process of the PCR for each amplicon.

Amplicon	Forward Primer	Reverse Primer
Codon 719	AGGATCTTGAAGGAAACTGAATT	[Biotin]-TGCCAGGGACCTTACCTTATA
Exon 19 deletions	AGATCACTGGGCAGCATGT	[Biotin]-CAAAGCAGAAACTCACATCGA
Codons 768, 769, 790	CTCCCTCCAGGAAGCCTACG	[Biotin]-CTTTGTGTTCCCGGACATAGTC
Codons 858-861	AAACACCGCAGCATGTCAAGA	[Biotin]-TGCCTCCTTCTGCATGGTATTC

Table 3: EGFR amplicons and their PCR primers.

PCR was conducted using the PyroMark PCR Kit from QIAGEN. According to the handbook, the DNA amplified using this kit is especially suitable for Pyrosequencing® as a downstream application. Furthermore, due to the presence of CoralLoad Concentrate in the reaction solution which contains a gel loading reagent and two marker dyes, it is possible to load a fraction of the preparation directly on an agarose gel[49] for subsequent analysis of the PCR. The following reagents were used for PCR:

- PyroMark PCR Master Mix, 2x (contains HotStarTaq DNA Polymerase and optimized PyroMark Reaction Buffer containing 3 mM $MgCl_2$ and dNTPs)[49]
- CoralLoad® Concentrate, 10x
- RNase-Free Water

The preparation for one sample was pipetted as follows:

PyroMark PCR Master Mix, 2x	12.5	µl
CoralLoad Concentrate, 10x	2.5	µl
PCR Primers	1.0	µl
H_2O	4.0	µl
	20.0	µl
+ DNA/Control DNA/H_2O	5.0	µl
	25.0	µl

Depending on the number of samples per run, a master mix was prepared consisting of the first four reagents; each reagent multiplied with the number of samples and including one to four spare preparations to account for pipetting inaccuracies. Preparation of the reagents and the respective DNA included vortexing and subsequent downspinning. For each run, we also included a negative template control (NTC; DNA replaced with RNAse-Free Water from the kit), a wildtype control DNA sample and positive controls for certain amplicons. All reagents as well as the DNA were stored at -20°C for long-term, and 4°C for short-term storage.

As thermocycler, we used the C1000 Thermal Cycler, BioRad, California, which has a maximum capacity of 96 samples. The PCR program was the same for all amplicons generated:

Initial Activation Step:	15 minutes	95 °C
3-Step Cycling: (42 cycles)		
Denaturation	20 seconds	95 °C
Annealing	30 seconds	53 °C
Extension	20 seconds	72 °C
Final Extension:	5 minutes	72 °C

Agarose gel electrophoresis

For electrophoresis, 2% agarose gels with a size of 15x10cm (for 56 samples) were produced: 2g of agarose was mixed with 100ml of TBE (Tris/Borate/EDTA)-buffer in an Erlenmeyer flask. Agarose was dissolved by boiling the solution in the microwave until it became completely clear. After the solution cooled down to approximately 60°C, 10µl GelRed Nucleic Acid Stain was added and distributed by carefully slewing the flask. Solution was then transferred to the gel caster and the comb was carefully removed after solidification of the gel.

Pyrosequencing®

Method overview

Pyrosequencing® is a "sequencing-by-synthesis" method, meaning that a new DNA strand is synthesized in the process of sequencing. Key mechanism is the detection of pyrophosphate, which is released when a nucleotide is incorporated into a new strand, as first described by Hyman in 1988.[50] Ever since the discovery that pyrophosphate can be used to sequence a given DNA strand, the technique experienced substantial improvements. Nowadays, Pyrosequencing® is performed using four different enzymes that work in a reaction cascade in order to determine the sequence of a given DNA strand.

Pyrosequencing® offers a fast, easy and accurate way to sequence short DNA fragments that have been amplified beforehand by polymerase chain reaction: the DNA fragment to be sequenced was isolated from tissue and amplified using a biotinylated forward primer and a non-biotinylated backward primer (see table 3). Following strand separation, the biotinylated strands get immobilized on sepharose beads and the sequencing reaction begins: the

sequencing primer binds to the template to be sequenced and gets extended. Each time the correct dNTP gets incorporated into the new strand, pyrophosphate is released, which is converted to ATP by the APS present in the solution, catalyzed by the enzyme ATP sulfurylase. This ATP acts as energy substrate to convert luciferin to oxyluciferin by the presence of luciferase. Oxyluciferin generates visible light that is detected by a photosensitive detector. After incorporation of the correct dNTP, the other dNTPs are degraded by the enzyme apyrase and the polymerase that extents the sequencing primer moves to the next nucleotide. After sequencing, the correct sequence can be read out by evaluating the so-called pyrogram: each peak represents a signal that was generated by oxyluciferin. Since the signal strength is proportional to the amount of ATP present, two subsequent identical dNTPs generate a peak that is double as high.

In practice, Pyrosequencing® can be divided into the following workflow:

***In silico* steps**

- Sequencing primer design: the sequencing primer serves as template that is elongated by DNA polymerase during the sequencing reaction. In principle, the sequencing primer can start anywhere downstream of the sequence to analyze, however, due to the use of specialized plug-ins and the desire for short strands due to economic reasons, the sequencing primers for each sequencing reaction start exactly at the respective sequence to analyze.
- Assay design: for each sequencing run, an assay has to be designed *in silico* with the sequence to analyze, the variable positions and the dispensation order of the nucleotides. Furthermore, in order to receive an acceptable pyrogram, nucleotides are also dispensed at known positions. In this way, we're able to judge whether a particular run has worked in general. These peaks are called "reference peaks" in the pyrogram.
- Run setup: once the assay has been set up and all reagents and samples are available, a run is set up in which sample numbers, assay names and information on the cartridge used for substrate, enzyme and nucleotide dispensation are compiled. The system then generates the "pre run information", which contains information on the volumes of reagents that need to be applied to the cartridge. The run is then saved on a USB stick and connected to the Pyrosequencer.

***In vitro* steps**

- Immobilization of the PCR products on Streptavidin High Sepharose Performance Beads: this step serves to link the PCR product to beads by taking advantage of the strong binding forces between streptavidin and biotin. Since the reverse primers of all PCRs described in this thesis are biotinylated, binding occurs.
- Denaturation of the strands and annealing of sequencing primer: during this step, the double-stranded PCR product is denatured so that only one single-stranded product remains. This is accomplished by generating a vacuum between the beads carrying the amplicons and filter probes. These filter probes with the attached samples are subsequently lowered in 70% ethanol, denaturation solution, wash buffer and water so that only single-stranded product remains attached to the filter probes. At the end of this procedure, the occupied filter probes are lowered over wells containing annealing buffer and sequencing primer, the vacuum is switched off and the probes are dipped into the solution so that the single-stranded amplicons are mixed with the solution.
- Preparation of cartridge: enzyme mix (containing DNA polymerase, ATP sulfurylase, luciferase, apyrase and single-stranded binding protein), substrate mixture (containing adenosine 5' phosphosulfate [APS]) and nucleotides are pipetted into the corresponding wells of the cartridge according to the pre run information obtained during the *in silico* steps.
- Starting the Pyrosequencing® run: the plate with the wells containing the single-stranded PCR products with the annealed sequencing primer and the filled cartridge is inserted into the Pyrosequencer and the run is loaded from the USB stick.

Sequencing primers

Pyrosequencing® was conducted using the PyroMark® Q24 MDx (QIAGEN) platform. The aforementioned mutations in the exons encoding the tyrosine kinase domain of the epidermal growth factor receptor were assessed. In the Pyrosequencing® assays, the following sequencing primers were employed:

Assay (*EGFR*)	Sequencing primer	Sequence to analyze	Nucleotide disp. order	Analyzed mutations
719	AAAAAGATCAAAGTGCTG	DGCTCCGGTGC	ATGTCACTCGTG	G719A, G719C, G719S
Exon 19 Deletion	TTAAAATTCCCGTCGC	TATCAA[GGAATTAAGAGAAGC]AACATCTCCGAAAGCCAACAAGGA	CTATACTGTCAGCTCGATCGTCATCGTCACGC	20 deletions and complex mutations
768	GCCTACGTGATGGC	CAKCGTG	TCGAGTCGAT	S768I
790	TCCACCGTGCAGCTC	ATCAYGCAG	CATCGACTGCA	T790M
858-861	AGATCACAGATTTTGGG	CKGGCCAAACDGCTGGGT	ATCGTGCAAGCATGCTG	L858R, L861Q, L861R

Table 4: Assays and corresponding sequencing primers, sequences to analyze, nucleotide dispensation order and analyzed mutations for mutational analysis of EGFR.

Reagents and solutions for Pyrosequencing®

Reagent	Manufacturer
PyroMark® Binding Buffer	Qiagen (Hilden, Germany)
PyroMark® Denaturation Solution	Qiagen (Hilden, Germany)
PyroMark® Wash Buffer concentrate	Qiagen (Hilden, Germany)
PyroMark® Annealing Buffer	Qiagen (Hilden, Germany)
Streptavidin SepharoseTM High Performance	GE Healthcare (Uppsala, Sweden)
High-purity water ELGA Purelab Ultra(Milli-Q 18.2 MΩ)	ELGA VEOLIA (Paris, France)
Ethanol (70 %)	Chemicals VWR BDH Prolabo, (Radnor, Pennsylvania)
Enzyme Mixture	Qiagen (Hilden, Germany)
Substrate Mixture	Qiagen (Hilden, Germany)
dATαS (PyroMark® Q24 Gold Reagent)	Qiagen (Hilden, Germany)
dGTP (PyroMark® Q24 Gold Reagent)	Qiagen (Hilden, Germany)
dCTP (PyroMark® Q24 Gold Reagent)	Qiagen (Hilden, Germany)
dTTP (PyroMark® Q24 Gold Reagent)	Qiagen (Hilden, Germany)

Table 5: Reagents employed for Pyrosequencing® and their corresponding manufacturers. Reagents used for sample preparation are separated by a reinforced border from reagents used for the sequencing reaction proper.

Platform and accessories for Pyrosequencing®

- PyroMark® Q24 MDx Instrument
- PyroMark® Q24 Software
- PyroMark® Q24 Vacuum Workstation
- PyroMark® Q24 Plate Holder
- PyroMark® Q24 Instrument
- PyroMark® Q24 Troughs
- PyroMark® Q24 Cartridge

Preparation of samples, reagents & execution of Pyrosequencing® reaction

For Pyrosequencing®, PCR products were calibrated to room temperature and spinned down. Streptavidin Sepharose High Performance Beads were dissolved in their own solution by gentle slewing of the glass container. For DNA immobilization, the following preparation was pipetted:

PyroMark Binding Buffer	40 µl
Streptavidin Sepharose High Performance	2 µl
H_2O	18-28 µl
	60-70 µl

Depending on the number of samples per run, a master mix was prepared consisting of these three reagents; each reagent multiplied with the number of samples and including two spare preparations to account for pipetting inaccuracies. Mastermix was vortexed and distributed among PCR wells. Vortexing was repeated after every second pipetted well to avoid sedimentation of the beads. Either 10µl or 20µl of the PCR product was added, depending on the strength of the bands (and thus, the DNA concentration) in the analytical gel electrophoresis conducted beforehand. PCR well plate was applied to a microplate shaker and shaking was set to 3000rpm/minute to bind the PCR product to the beads.

The sequencing primer was prepared as follows:

PyroMark Annealing Buffer	22.5 µl
Sequencing Primer (10 µM)	2.5 µl
	25.0 µl

Likewise, a master mix was prepared depending on the number of samples and including spare preparations. Beforehand, the sequencing primer was calibrated to room temperature, vortexed and spinned down. After preparation of the master mix, it was distributed among the wells of a PyroMark® Q24 Plate.

Vacuum workstation was then prepared according to manufacturer's manual by filling the troughs with 70% ethanol, denaturation solution, wash buffer, and 2 x H_2O. Filter probes were washed according to the manufacturer's manual. PyroMark® Q24 Plate Holder was heated to 80°C and kept in the incubator. After 10 minutes, PCR-products bound to the beads and were immediately transferred to the PyroMark® Q24 Vacuum Workstation. Hand device with switched on vacuum was lowered into the PCR wells containing the beads with bound PCR products to bind the product to the filter tips. Filter tips were then sequentially lowered into 70% ethanol, denaturation solution, was buffer and H_2O according to the manual. Hand device was subsequently held over the wells of the PyroMark® Q24 Plate, vacuum was switched off and filter tips with the bound product were lowered into the sequencing primer / annealing buffer solution. For primer annealing, PyroMark® Q24 Plate was placed on the PyroMark® Q24 Plate Holder in the incubator for 2 minutes, then calibrated to room temperature for 10 minutes. In the meantime, enzyme and substrate mixtures, as well as the nucleotides, were applied to the PyroMark® Q24 cartridge according to the pre run information. Cartridge and PyroMark® Q24 Plate were then positioned in the PyroMark® Q24 MDx Instrument and the run was started from the respective file saved on the USB stick.

Cleaning of the appliances was performed according to the user's manual.

Analysis of immune markers and selected samples harboring *EGFR* mutations

Immunohistochemistry

Method overview

PD-L1 & PD-1 expression was assessed by immunohistochemistry. This method exploits the circumstance that specific antibodies are able to bind to their respective antigens. Thereby, the researcher is able to detect the presence of specific epitopes and also quantify their expression by evaluation of the staining intensity. Although immunohistochemistry is a rather old technique, it provides several advantages: it is easy to conduct, established protocols maximize specificity and sensitivity, and the variability (choice of buffers, dilution of the antibody solution, etc.) can be tested in a quick and simple manner, leading to the possibility of quick establishment of individual protocols for each antibody.

Immunohistochemistry protocols for a given antibody differ in the rinsing buffer, buffer for antigen, duration of primary antibody incubation, antibody dilution factor and choice of diluent, duration of DAB incubation and duration of exposure to haematoxylin. Aside from these variables, the workflow is always the same: tissue is cut using a microtome into 4µm thick sections and put on a glass slide. Specimen is then incubated for 10 minutes at 65°C in order to melt the surrounding paraffin. Sample is then subjected to consecutive incubation in xylene, 100% ethanol, 70% ethanol and deionized water. This step is necessary to remove the paraffin and to avoid drying of the sample. Since we're using a horseradish peroxidase-based detection system, incubation of the sample in H_2O_2 to block the endogenous peroxidase is necessary. In order to remove any H_2O_2 following this step, the sample is subsequently rinsed in the appropriate buffer. The next step involves the process of "antigen retrieval": our samples are generally fixed with formalin, which forms addition products (adducts) with uncharged reactive amino groups (-NH or NH_2). This leads to the eventual development of cross-links through the formation of a reactive hydroxymethyl compound and the evolution of methylene bridges between proteins.[51] To reverse these cross-links and to make the respective epitope available for antibody binding, the sample is subject to heat-induced epitope retrieval: it is boiled in an appropriate buffer for a fixed duration, including boiling under elevated pressure for a short amount of time. Although formalin fixation and the subsequent antigen retrieval denatures proteins, epitope binding by the antibody is still possible due to the fact that the primary amino acid sequence is sufficient for successful antibody binding.[52] After antigen retrieval, two consecutive rinse steps are performed in the appropriate buffer. Furthermore, a small amount of Tween® 20 is added to the rinse buffer. Tween® 20 is a nonionic detergent[53] that facilitates antigen-antibody binding by solubilizing membrane

proteins and by decreasing the surface tension of the liquid.[54] Following rinsing after antigen retrieval, the slide around the tissue sample is dried using a fine napkin and the tissue is encircled with a hydrophobic barrier to prevent the following solutions to pour away.

In general, immunohistochemistry protocols can be divided into “direct method” and “indirect method”. In the former case, a labeled antibody binds to its antigen and is detected directly. In the latter case, a primary, unlabelled antibody binds to its antigen and a secondary, labeled antibody, is raised against the primary antibody. Given the fact that the direct method omits at least one incubation step as well as several rinse steps, this method is easier to conduct and faster. However, secondary antibodies that are part of the indirect method bind to primary antibodies in a numerous way, thereby amplifying the signal and thus increasing sensitivity.[51] Furthermore, biotechnological companies produce commercially available kits with standardized protocols for antibody detection, hence bringing detection to near perfection. As detection system, we use the “Lab Vision™ UltraVision™ LP Detection System: HRP Polymer” from Thermo Fisher Scientific, consisting of a blocking solution, a primary antibody enhancer solution and an HRP polymer solution. The blocking solution serves as agent to mask nonspecific primary antibody sites, thereby reducing unspecific antibody binding. Other blocking agents might include normal serum and protein solutions, e.g. bovine serum albumin (BSA). Upon administering the blocking solution and incubation of the sample, the diluted antibody is applied.

Antibodies can be divided into monoclonal and polyclonal antibodies, and most antibodies are extracted from rabbit or mouse. Since the majority of rabbit monoclonal antibodies have a smaller K_D value[55] (and thus a higher affinity) than their mouse counterparts, we used rabbit mAbs wherever possible. Furthermore, we solely used monoclonal antibodies due to higher specificity of recognizing one single epitope. Antibody dilutions and the respective diluent generally depend on the solution the antibody was delivered in, as well as the source of the organism the antibody was extracted from. In some cases, the antibody comes in a ready-to-use solution and already diluted. In other cases, prefabricated antibody diluents have to be used. In most cases, however, dilution factor and diluent have to be tested. The standard protocol used in our lab demands to use the respective rinse buffer (PBS or TBS) with the addition of 0.1% Tween® 20 as diluent, as well as a 1:100 dilution of goat serum as blocking agent. Antibody dilution and incubation time depends on the affinity of the antibody; it is assessed by performing test stainings and by taking into account the manufacturer’s recommendation, where available.

For amplification of the primary antibody, the secondary antibody is applied after incubation of the former and two consecutive rinse steps. We solely use the antibody enhancer from the aforementioned detection system for this purpose. After incubation with the enhancer and two rinse steps, the HRP polymer solution of this detection kit is applied. The polymer in this solution is labeled with horseradish peroxidase and binds to the secondary antibody. After incubation, again two rinse steps are performed and 3,3'-diaminobenzidine (DAB), diluted in substrate buffer, is prepared. It is vital that the DAB dilution is prepared freshly (right before application to the sample) due to the unstable nature of the organic compound when dissolved in buffer. In the presence of horseradish peroxidase, DAB is oxidized and changes its color to dark-brown. In this way, the presence of antibody binding to its respective antigen can be indirectly detected. Duration of DAB incubation depends on the abundance of horseradish peroxidase and thus on the abundance of the antigen in the sample. Incubation time is assessed by monitoring the color change of the sample; too short incubation leads to low detection of the antigen, too long incubation leads to too strong background staining.

Following DAB incubation, the slides are rinsed with distilled water and the sample is counterstained with haematoxylin, which serves to color the chromatin of the cells. Duration of exposure to haematoxylin is generally dependent on the intensity of DAB staining and its location: the goal is to achieve an acceptable color discrimination (contrast) between the DAB staining and the counterstaining.[51] When the antigen and thus the bound primary and secondary antibodies are close to the nucleus, shorter haematoxylin incubation might be required as to not mask the DAB staining. When the antigen is at the surface of the cell, longer and thus stronger counterstaining is permitted since the target of haematoxylin is further away from the brownish-colored target.

Reagents and solutions for immunohistochemistry

TBS buffer (10x concentrated), pH 7.6

121g TRIS ultra pure (Biomol, ordering # 08003.1) and 400g sodium chloride for analysis (Merck, ordering # 1.06404.1000) are dissolved in 5000ml distilled water. The pH is set to 7.6 with 37% hydrochloric acid or sodium hydroxide, 2mol/l.

10mM citrate buffer for TBS, pH 6.0

5.88g tri-sodiumcitrate dihydrate (Merck, ordering # 1.06448.1000) is dissolved in 2000ml distilled water. The pH is set to 6.0 with 37% hydrochloric acid or sodium hydroxide, 2mol/l.

EDTA buffer

0.372g ethylenediamintetraacetic acid disodium salt (Sigma-Aldrich, ordering # E5134-50G) is dissolved in 1000ml distilled water. The pH is set to 8.0 with sodium hydroxide, 2mol/l.

Immunohistochemistry of PD-L1

For each sample, a 4µm thick section was cut from the respective formalin-fixed, paraffin-embedded (FFPE) block and mounted on a glass slide. All following steps were conducted at room temperature, unless otherwise noted. Glass slides with mounted samples were incubated for 10 minutes at 65°C to melt the surrounding paraffin and immediately subjected to 2 x 1 minute xylene, 2 x 1 minute 100% ethanol, 2 x 1 minute 70% ethanol and 2 x 1 minute distilled water and then incubated for 10 minutes in 0.3% H_2O_2 in TBS. Samples were then rinsed twice for three minutes each in TBS. For antigen retrieval, slides with samples were immersed in plastic cuvettes filled with 10mM citrate buffer for TBS, pH 6.0. Cuvettes were covered with aluminum foil to prevent vaporization during the following steps. Autoclave with basket was filled with 0.5l tap water and 0.5l distilled water, plastic cuvettes were placed in the basket, lid was positioned and locked and autoclave was heated to approximately 98°C. Upon reaching this temperature, the pressure valve was closed and samples were boiled at 1.5 bar for 2.5 minutes. After expiration of this time, autoclave was switched off and opened after pressure was back at 1 bar. Samples were then taken out, the aluminum foil was removed and left to cool for 15 minutes. Following, samples were transferred into TBS + 0.1% Tween® 20 for 3 minutes. In the meantime, humidity chambers were prepared with distilled water. Slides with samples were taken out, positioned in the humidifying chamber, dried using a napkin, and tissue was encircled with a hydrophobic barrier. "Lab Vision™ UltraVision™ LP Detection System: HRP Polymer" (Thermo Scientific) Ultra V Block was applied to each sample (2-3 drops per sample) and incubated for 5 minutes under the exclusion of light. In the meantime, the antibody dilution was prepared: antibody stock was vortexed, spinned and diluted 1:100 in SignalStain® Antibody Diluent #8112 (Cell Signaling). Antibody employed was PD-L1 (E1L3N®) XP® Rabbit mAb #13684 (Cell Signaling). Antibody was then applied to the samples with a volume of 100µl or 150µl per sample, depending on sample size and incubated in the humidifying chamber under the exclusion of light for 30 minutes. Subsequently, slides were rinsed in TBS + 0.1% Tween® 20 twice for 3 minutes each. Afterwards, the area around the hydrophobic barrier was dried in order to prevent liquids from passing over. Primary Antibody Enhancer from the detection kit was applied to each sample (2-3 drops per sample) and incubated for 10 minutes under the exclusion of light, followed by

two rinse steps for three minutes each. HRP Polymer from the detection kit was then applied for 15 minutes, again 2-3 drops per sample, under exclusion of light, and rinse steps were performed as before. In the meantime, the DAB substrate was prepared freshly: 20µl DAB + chromogen were mixed and vortexed per 1ml substrate buffer. 100 – 150µl of diluted DAB + chromogen were applied to each sample and incubated for 10 minutes under the exclusion of light. Slides were then rinsed in distilled water and counterstained with Gill's Haematoxylin No. 3 for 20 seconds, then rinsed twice in tab water and twice in distilled water. Tissue was dehydrated by immersing the slides in 70%, 80%, 96%, and 100% ethanol subsequently. As last step, slides were immersed in ethyl-n-*butyl* ether (EBE) and covered with appropriate coverslips using Entellan (Merck Millipore, Darmstadt, Germany catalogue# 1.07961.0100) as mounting agent.

Immunohistochemistry of PD-1

Immunohistochemistry of PD-1 was conducted with the following antibody: PD-1 (EH33) Mouse mAb #43248 (Cell Signaling) Staining protocol is the same as for PD-L1 with exception of the antigen retrieval buffer: instead of citrate buffer for TBS, EDTA buffer (pH 8.0) was used.

Immunohistochemistry of EGFR L858R & E746-A750del

Imunohistochemistry of EGFR L858R & E746-A750del was conducted with the following antibodies: EGF Receptor (L858R Mutant Specific) (43B2) Rabbit mAb #3197 (Cell Signaling) & EGF Receptor (E746-A750del Specific) (D6B6) XP® Rabbit mAb #2085 (Cell Signaling). Staining protocol is the same as for PD-1, with the following exceptions: after incubation for 10 minutes at 65°C, samples were incubated in xylol (3 x 5 minutes), 100% ethanol (2 x 10 minutes), 95% ethanol (2 x 10 minutes) and distilled water (2 x 5 minutes). Antigen retrieval was conducted subsequently. Boiling at 125°C (1.5 bar) was conducted for 30 seconds. Samples were left to cool for 10 minutes then rinsed in distilled water, following incubation in distilled water (3 x 5 minutes). Subsequently, samples were incubated in 3% H_2O_2, rinsed in distilled water (2 x 5 minutes) and incubated in TBS + 0.1% Tween® 20 for 5 minutes. Ultra V Block was then applied, and all following steps are the same as for PD-L1 immunohistochemistry, with the exception of the rinse steps: instead of 2 x 3 minutes, 3 x 5 minutes were used.

Study population

The study population consists of 161 Serbian patients with early stage operable lung adenocarcinoma. Samples were obtained during surgery, fixed in formalin and embedded in paraffin.

Data analysis & calculations

Data were analyzed with IBM SPSS Statistics Version 23 (SPSS Inc, Chicago, IL). To identify possible selection biases among patients in characteristics and to figure out the statistical significance between characteristics and certain parameters, Pearson's chi-square tests were performed where indicated in the results section. P values of less than 0.05 were considered statistically significant in differences between groups in respect to a certain parameter. All P values result from two-sided tests. *EGFR* and *KRAS* mutation status, as well as PD-L1 expression were compared to patient characteristics to analyze possible grouping of these parameters; furthermore, *EGFR*, *KRAS* and *PI3KCA* mutation status were analyzed in conjunction with PD-L1 expression and PD-1 expression and intensity of expression to analyze whether there is a connection between these variables. Moreover, *EGFR* and *KRAS* mutation status were compared to pack years by creation of box plots to find possible correlations between these parameters. To assess correlations between patient subgroups and PD-1 expression, a subgroup analysis was performed as previously described for PD-L1. Survival times were calculated as periods between date of diagnosis and relapse or death for relapse-free (RFS) and overall survival (OS), respectively. Relapse is defined as the return of a disease or the signs and symptoms of a disease after a period of improvement.[56] Relapse and survival probabilities as functions of PD-1 & PD-L1 expression were estimated with the product limit method according to Kaplan and Meier.[57] Differences in survival distributions resulting from Kaplan-Meier analysis were elucidated by the log-rank test. To describe the unadjusted effects of covariates on RFS and OS, univariate Cox proportional hazards regression models were used. Multivariate Cox proportional hazards regression models were used to assess the independent effects of PD-1 expression on RFS and OS.[58,59] Variables were coded as follows: age, <54, 55 – 64, >64; sex, male or female; stage, IA, IB, IIA, IIB, IIIA or IIIB; tumor size, 1a, 1b, 2a, 2b, 3 or 4; performance status, Eastern Cooperative Oncology Group (ECOG) 0, 1 or 2; type of surgery, pneumonectomy, bilobectomy, lobectomy or segmentectomy; adjuvant radiotherapy, yes or no; adjuvant chemotherapy, yes or no; smoking status, yes (>100 cigarettes / lifetime), no (<100 cigarettes / lifetime), former (>100 cigarettes / lifetime, >1 year smoke-free).

Results

DNA Mutations

DNA mutations were investigated for *EGFR*, *KRAS*, *NRAS*, *HRAS*, *PI3KCA*, and *BRAF* genes. Furthermore, *ALK* and *ROS1* translocations were analyzed. Of 161 tumor samples, 10 (6.2%) were *EGFR* mutation positive (of which one sample harbored two mutations), 68

(42.3%) were positive for *KRAS* mutations, 1 (0.6%) *BRAF* and 7 (4.4%) *PI3KCA* mutations were detected. Moreover, 7 (4.4%) and 3 (1.9%) *ALK* and *ROS1* rearrangements were found, respectively. Table 6 lists all mutations, including the affected amino acids, as well as their percentages in the total amount of samples. In total, 97 samples (60%) carried at least one mutation of any kind.

Types and prevalence of DNA mutations in the patient samples (N = 161)

Gene	Amino Acid Mutation	Number	Percent
EGFR	G719A / V769M	1	0.6
	G719C	1	0.6
	E746_A750del	3	1.9
	L747_P753>S	1	0.6
	L747_T751del	1	0.6
	L861Q	1	0.6
	L858R	2	1.2
Subtotal		10	6.2
KRAS	G12C	32	19.9
	G12V	15	9.3
	G12A	6	3.7
	G12D	4	2.5
	G13D	3	1.9
	G12S	1	0.6
	Q61L	4	2.5
	Q61H	3	1.9
Subtotal		68	42.3
BRAF	G469A	1	0.6
PI3KCA	E542K	4	2.5
	Q546P	1	0.6
	H1047L	1	0.6
	H1047R/ H1047L	1	0.6
Subtotal		7	4.43
***ROS1* (translocation)**	negative	154	95.7
	positive	7	4.3
***ALK* (translocation)**	negative	158	98.1
	positive	3	1.9
Total mutations		97	60

Table 6: Summary of types and prevalence of DNA mutations in the patient samples.

Furthermore, we performed a subgroup analysis for *EGFR* and *KRAS* mutation positive patients. Patient subgroups were divided into age, sex, tumor stage, tumor size, ECOG status, surgery history, radiotherapy history, chemotherapy history and smoking status. Results by patient characteristics, as well as P values for every subgroup are listed in tables 7 and 8.

Characteristics of patients with *EGFR*-mutated tumors

Characteristic	All Patients (N = 161)	Wildtype (N = 151)	*EGFR* mutations (N = 10)	P Value
Age				0.017
<54	32 (19.9)	30 (19.9)	2 (20.0)	
55-64	78 (48.4)	77 (51.0)	1 (10.0)	
>64	51 (31.7)	44 (29.1)	7 (70.0)	
Sex				0.09
Male	90 (55.9)	87 (57.6)	3 (30.0)	
Female	71 (44.1)	64 (42.4)	7 (70.0)	
Stage				0.69
IA	25 (15.5)	23 (15.2)	2 (20.0)	
IB	30 (18.6)	29 (19.2)	1 (10.0)	
IIA	35 (21.7)	31 (20.5)	4 (40.0)	
IIB	33 (20.5)	32 (21.2)	1 (10.0)	
IIIA	34 (21.1)	32 (21.2)	2 (20.0)	
IIIB	4 (2.5)	4 (2.6)	0 (0.0)	
T				0.13
1a	15 (9.3)	14 (9.3)	1 (10.0)	
1b	18 (11.2)	15 (9.9)	3 (30.0)	
2a	48 (29.8)	43 (28.5)	5 (50.0)	
2b	38 (23.6)	37 (24.5)	1 (10.0)	
3	38 (23.6)	38 (25.2)	0 (0.0)	
4	4 (2.5)	4 (2.6)	0 (0.0)	
ECOG				0.65
0	45 (28.0)	41 (27.2)	4 (40.0)	
1	114 (70.8)	108 (71.5)	6 (60.0)	
2	2 (1.2)	2 (1.3)	0 (0.0)	
Surgery				0.35
Pneumonectomy	26 (16.1)	25 (16.6)	1 (10.0)	
Bilobectomy	11 (6.8)	10 (6.6)	1 (10.0)	
Lobectomy	113 (70.2)	107 (70.9)	6 (60.0)	
Segmentectomy	11 (6.8)	9 (6.0)	2 (20.0)	
Adjuvant radiotherapy				0.68
No	150 (93.2)	141 (93.4)	9 (90.0)	
Yes	11 (6.8)	10 (6.6)	1 (10.0)	
Adjuvant chemotherapy				0.36
No	87 (54.0)	83 (55.0)	4 (40.0)	
Yes	74 (46.0)	68 (45.0)	6 (60.0)	
Smoking				<0.001
No	20 (12.4)	15 (9.9)	5 (50.0)	
Former	41 (25.5)	37 (24.5)	4 (40.0)	

Yes	100 (62.1)	99 (65.6)	1 (10.0)	

Table 7: Characteristics of patients with EGFRmutated tumors. P values were calculated by performing Pearson's chi-square test.

Characteristics of patients with *KRAS*-mutated tumors

Characteristic	All Patients (N = 161)	Wildtype (N = 93)	*KRAS* mutations (N = 68)	P Value
Age				0.003
<54	32 (19.9)	20 (21.5)	12 (17.6)	
55-64	78 (48.4)	35 (37.6)	43 (63.2)	
>64	51 (31.7)	38 (40.9)	13 (19.1)	
Sex				0.003
Male	90 (55.9)	61 (65.6)	29 (42.6)	
Female	71 (44.1)	32 (34.4)	39 (57.4)	
Stage				0.41
IA	25 (15.5)	16 (17.2)	9 (13.2)	
IB	30 (18.6)	18 (19.4)	12 (17.6)	
IIA	35 (21.7)	16 (17.2)	19 (27.9)	
IIB	33 (20.5)	20 (21.5)	13 (19.1)	
IIIA	34 (21.1)	22 (23.7)	12 (17.6)	
IIIB	4 (2.5)	1 (1.1)	3 (4.4)	
T				0.58
1a	15 (9.3)	11 (11.8)	4 (5.9)	
1b	18 (11.2)	10 (10.8)	8 (11.8)	
2a	48 (29.8)	28 (30.1)	20 (29.4)	
2b	38 (23.6)	18 (19.4)	20 (29.4)	
3	38 (23.6)	24 (25.8)	14 (20.6)	
4	4 (2.5)	2 (2.2)	2 (2.9)	
ECOG				0.43
0	45 (28.0)	27 (29.0)	18 (26.5)	
1	114 (70.8)	64 (68.8)	50 (73.5)	
2	2 (1.2)	2 (2.2)	0 (0.0)	
Surgery				0.41
Pneumonectomy	26 (16.1)	14 (15.1)	12 (17.6)	
Bilobectomy	11 (6.8)	9 (9.7)	2 (2.9)	
Lobectomy	113 (70.2)	64 (68.8)	49 (72.1)	
Segmentectomy	11 (6.8)	6 (6.5)	5 (7.4)	
Adjuvant radiotherapy				0.82
No	150 (93.2)	87 (93.5)	63 (92.6)	
Yes	11 (6.8)	6 (6.5)	5 (7.4)	
Adjuvant chemotherapy				0.94
No	87 (54.0)	50 (53.8)	37 (54.4)	
Yes	74 (46.0)	43 (46.2)	31 (45.6)	

Smoking				0.006
No	20 (12.4)	17 (18.3)	3 (4.4)	
Former	41 (25.5)	27 (29.0)	14 (20.6)	
Yes	100 (62.1)	49 (52.7)	51 (75.0)	

Table 8: Characteristics of patients with KRASmutated tumors. P values were calculated by performing Pearson's chi-square test.

In order to visualize the correlation of intensity of tobacco abuse (described as pack-years) and the type of gene mutation (*KRAS* and *EGFR*), we created box plots. These can be seen in figures 1 and 2.

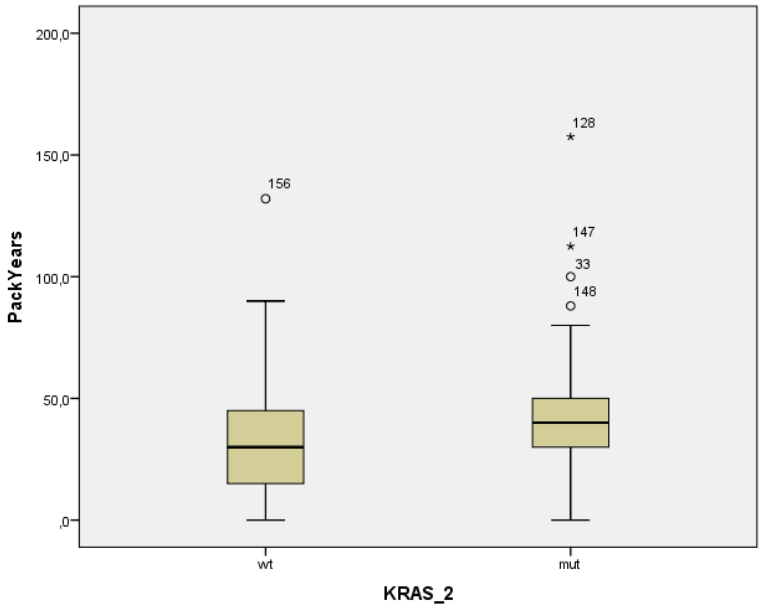

Figure 1: Box plot showing the correlation between pack-years and KRAS mutation status. Lower and upper part of the boxes describes the first and third quartiles, respectively, while the bar inside the box shows the position of the second quartile (the median). The whiskers mark the upper and lower 1.5 interquartile ranges, while the dots and asterisks represent weak and strong outliers, respectively.

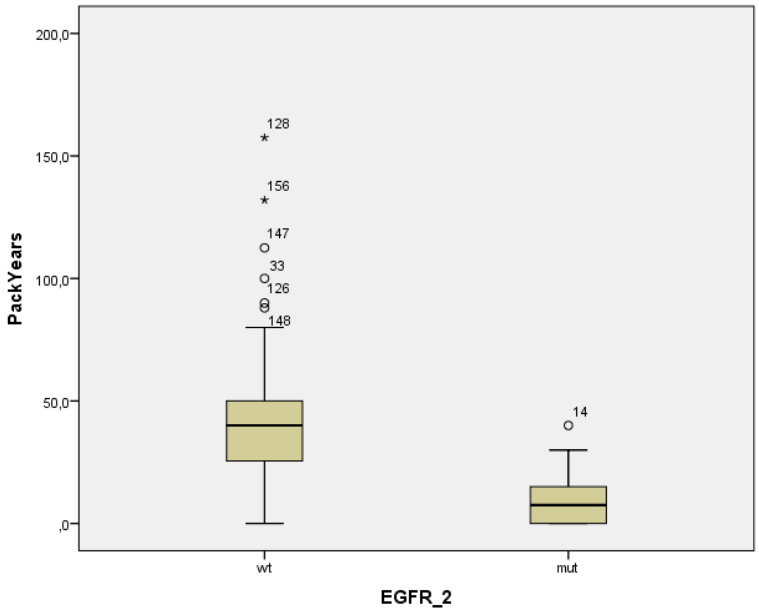

Figure 2: Box plot showing the correlation between pack-years and EGFR mutation status. Lower and upper part of the boxes describes the first and third quartiles, respectively, while the bar inside the box shows the position of the second quartile (the median). The whiskers mark the upper and lower 1.5 interquartile ranges, while the dots and asterisks represent weak and strong outliers, respectively.

Moreover, we were interested in the extent and the intensity of mutant receptor expression. We thus performed immunohistochemistry using antibodies as described in the materials & methods section on the two tissue samples harboring L858R- and the three tissue samples harboring E746_A750del-mutant receptors. Two representative pictures can be seen in figures 3 and 4.

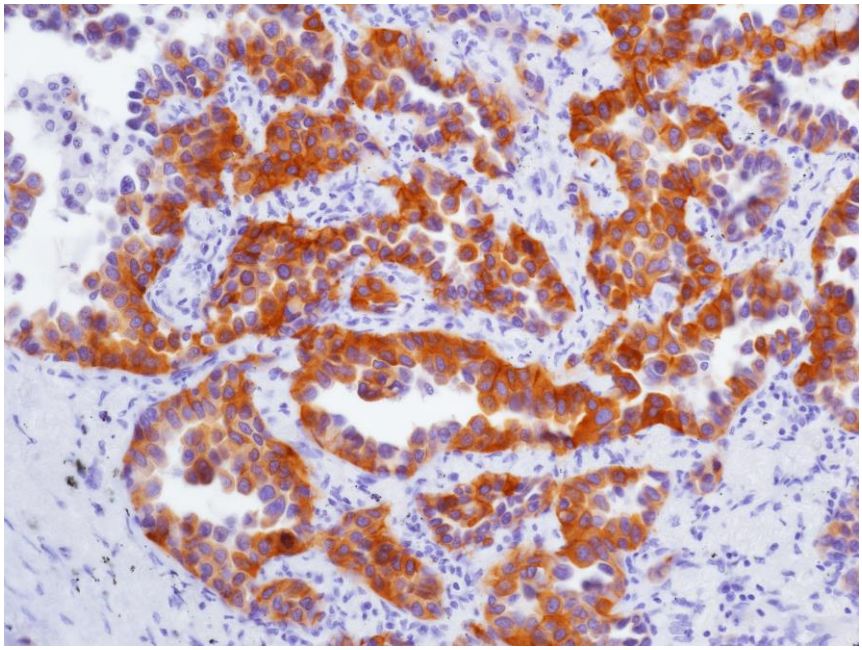

Figure 3: Immunohistochemistry of tissue sample with tumor cells harboring an EGFR L858R mutation. Blue staining stems from the haematoxylin, while the brownish staining arose from the DAB incubation. Magnification 100x.

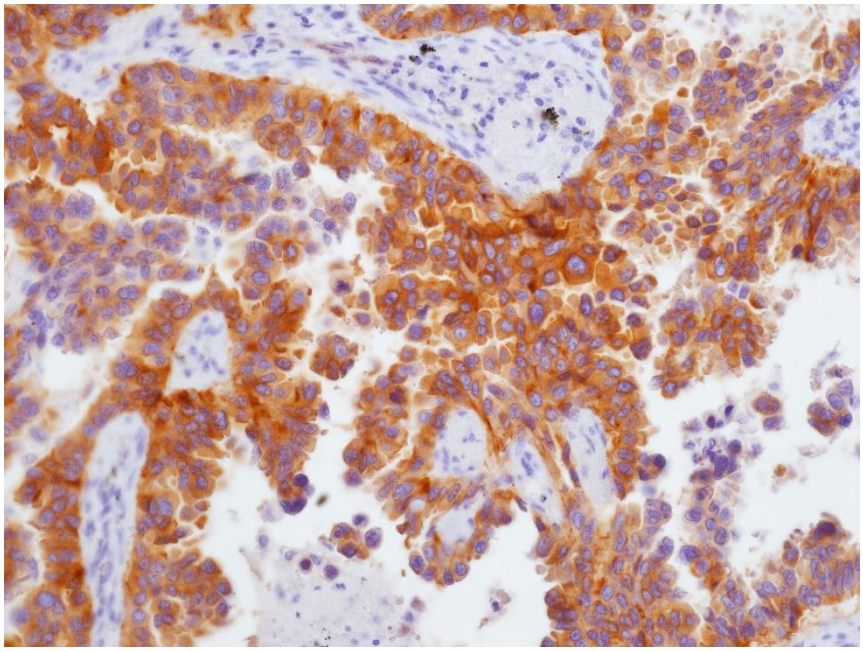

Figure 4: Immunohistochemistry of tissue sample with tumor cells harboring an EGFR E746_A750 deletion mutation. Blue staining stems from the haematoxylin, while the brownish staining arose from the DAB incubation. Magnification 100x.

As expected, all samples stained positive for their relevant mutation. We found that staining intensity was distributed across the whole respective tumor sample.

Immune markers

We further investigated the expression of the transmembrane protein PD-L1 in tumor cells by employing an anti-PD-L1 antibody for immunohistochemistry. Only specimens with ≥100 tumor cells were considered. Only tumor cell partial or complete membrane staining was evaluated. In total, 69 samples (62.9%) were negative, 31 samples (19.2%) showed a membrane staining on 1 – 49% of tumor cells, and 28 samples (17.4%) exhibited a membrane staining of more than or equal to 50% of tumor cells. Table 9 shows all results as well as P values for each patient subgroup.

Characteristic	**All Patients (N = 161)**	**Percentage of tumor cells positive for PD-L1 (%)**			**P Value**
		0	**1 - 49**	**≥50**	
Age					0.77
<54	32 (19.9)	18 (17.6)	7 (22.6)	7 (25.0)	
55-64	78 (46.6)	51 (50.0)	16 (51.6)	11 (39.3)	
>64	51 (31.7)	33 (32.4)	8 (25.8)	10 (35.7)	
Sex					0.56
Male	90 (55.9)	55 (53.9)	20 (64.5)	15 (55.9)	

Female	71 (44.1)	47 (46.1)	11 (35.5)	13 (46.4)	
Stage					0.41
IA	25 (15.5)	19 (18.6)	5 (16.1)	1 (3.6)	
IB	30 (18.6)	17 (16.7)	6 (19.4)	7 (25.0)	
IIA	35 (22.7)	24 (23.5)	8 (25.8)	3 (10.7)	
IIB	33 (20.5)	21 (20.6)	4 (12.9)	8 (28.6)	
IIIA	34 (21.1)	18 (17.6)	8 (25.8)	8 (28.6)	
IIIB	4 (2.5)	3 (2.9)	0 (0.0)	1 (0.6)	
T					0.52
1a	15 (9.3)	10 (9.8)	4 (12.9)	1 (3.6)	
1b	18 (11.1)	15 (14.7)	2 (6.5)	1 (3.6)	
2a	48 (29.8)	28 (27.5)	10 (32.3)	10 (35.7)	
2b	38 (23.6)	25 (24.5)	7 (22.6)	6 (21.4)	
3	38 (23.6)	23 (22.5)	7 (22.6)	8 (28.6)	
4	4 (2.5)	1 (1.0)	1 (3.2)	2 (7.1)	
ECOG					0.83
0	45 (28.0)	27 (26.5)	9 (29.0)	9 (32.1)	
1	114 (70.8)	73 (71.6)	22 (71.0)	19 (67.9)	
2	2 (1.2)	2 (2.0)	0 (0.0)	0 (0.0)	
Surgery					0.026
Pneumonectomy	26 (16.1)	12 (11.8)	5 (16.1)	9 (32.1)	
Bilobectomy	11 (6.8)	11 (10.8)	0 (0.0)	0 (0.0)	
Lobectomy	113 (70.2)	74 (72.5)	23 (74.2)	16 (57.1)	
Segmentectomy	11 (6.8)	5 (4.9)	3 (9.7)	3 (10.7)	
Adjuvant radiotherapy					0.73
No	150 (93.2)	94 (92.2)	29 (93.5)	27 (96.4)	
Yes	11 (6.8)	8 (7.8)	2 (6.5)	1 (3.6)	
Adjuvant chemotherapy					0.029

No	87 (54.0)	58 (56.9)	20 (64.5)	9 (32.1)	
Yes	74 (46.0)	44 (43.1)	11 (35.5)	19 (67.9)	
Smoking					0.32
No	20 (12.4)	15 (14.7)	3 (9.7)	2 (7.1)	
Former	41 (25.5)	30 (29.4)	6 (19.4)	5 (17.9)	
Yes	100 (62.1)	57 (55.9)	22 (71.0)	21 (75.0)	

Table 9: Patient characteristics and PD-L1 expression status results. P values were calculated by performing Pearson's chi-square test.

Furthermore, we compared PD-L1 staining to *EGFR*, *KRAS* and *PI3KCA* mutation status to elucidate possible co-occurrences between these biomarkers. The results and the P values for the DNA mutation subgroups are summarized in table 10:

Gene	**All Patients (N = 161)**	**Percentage of tumor cells positive for PD-L1 (%)**			**P Value**
		0	**1 - 49**	**≥50**	
EGFR					0.53
wildtype	151 (93.8)	94 (92.2)	30 (96.8)	27 (96.4)	
mutated	10 (6.2)	8 (7.8)	1 (3.2)	1 (3.6)	
KRAS					0.021
wildtype	93 (57.8)	66 (64.7)	17 (54.8)	10 (35.7)	
mutated	68 (42.2)	36 (35.3)	14 (45.2)	18 (64.3)	
PI3KCA					0.11
wildtype	154 (95.7)	100 (98.0)	29 (93.5)	25 (89.3)	
mutated	7 (4.3)	2 (2.0)	2 (6.5)	3 (10.7)	

Table 10: PD-L1 expression status and DNA mutations. P values were calculated by performing Pearson's chi-square test.

In order to determine survival probabilities according to PD-L1 expression frequency and intensity for RFS and OS, Kaplan-Meier analyses were performed. We focused on PD-L1 expression frequencies and intensities on tumor cells (figures 5, 6, 7 and 8), as well as on lymphocytes (figures 9 and 10). For simplification, expression variables were divided into "negative" or "positive" for Kaplan-Meier analyses, as opposed to the aforementioned arrangement of expression in three categories.

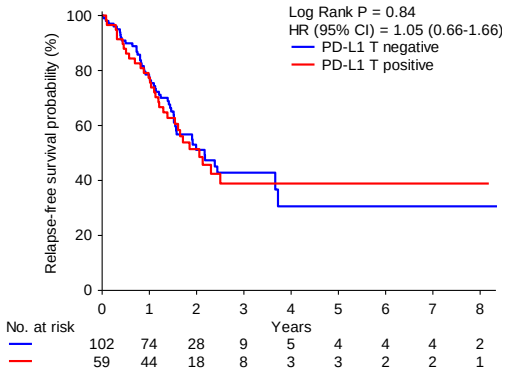

Figure 5: Kaplan-Meier curves for the effect of PD-L1 expression in tumor cells on relapse-free survival.

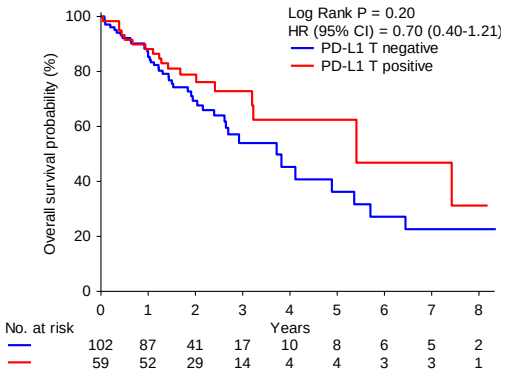

Figure 6: Kaplan-Meier curves for the effect of PD- L1 expression in tumor cells on OS.

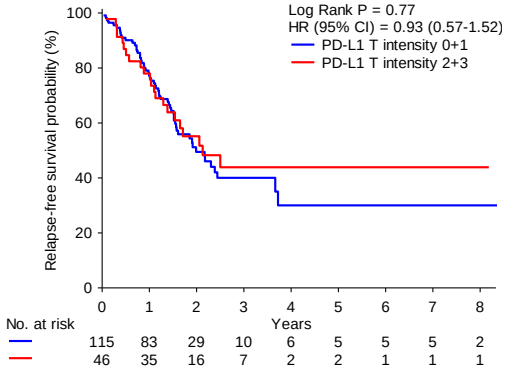

Figure 7: Kaplan-Meier curves for the effect of PD-L1 expression intensity in tumor cells on relapse-free survival.

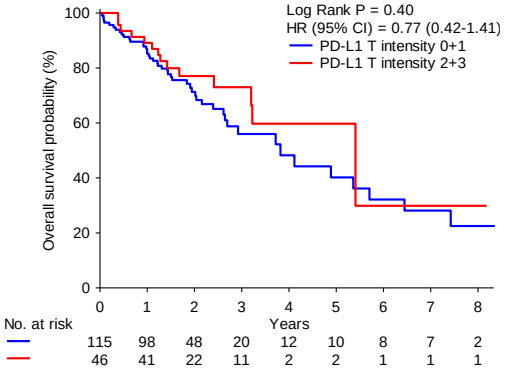

Figure 8: Kaplan-Meier curves for the effect of PD-L1 expression intensity in tumor cells on OS

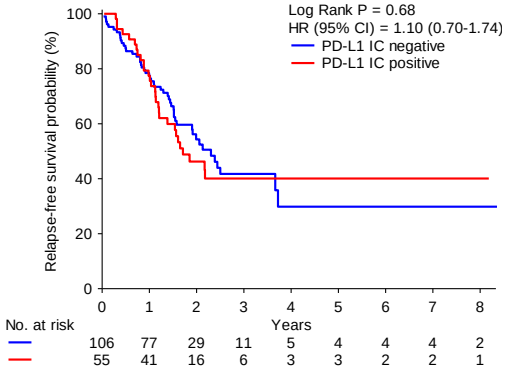

Figure 9: Kaplan-Meier curves for the effect of PD-L1 expression in lymphocytes on relapse-free survival.

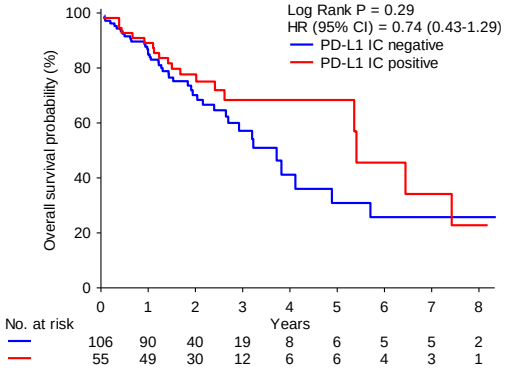

Figure 10: Kaplan-Meier curves for the effect of PD-L1 expression in lymphocytes on overall survival.

Moreover, we investigated PD-1 expression on tumor-infiltrating lymphocytes and its role as prognostic marker. PD-1 expression frequency and intensity was assessed byimmunohistochemistry. In total, 159 out of 161 samples were evaluable. Expression frequency was divided into negative (no staining) and positive (≥ 1% staining). Expression intensity was divided into weak and strong staining. Expression frequencies and intensities according to patient characteristics are listed in table 11.

Characteristics of patients with PD-1 lymphocyte staining

Characteristic	All Patients (N = 159)	Negative (N = 88)	Positive (N = 71)	P Value
Age				0.18
<54	32 (20.1)	17 (19.3)	15 (21.1)	
55-64	77 (48.4)	38 (43.2)	39 (54.9)	
>64	50 (31.4)	33 (37.5)	17 (23.9)	
Sex				0.09
Male	88 (55.3)	54 (61.4)	34 (47.9)	
Female	71 (44.7)	34 (38.6)	37 (52.1)	
Stage				0.24
IA	25 (15.7)	10 (11.4)	15 (21.1)	
IB	30 (18.9)	16 (18.2)	14 (19.7)	
IIA	35 (22.0)	22 (25.0)	13 (18.3)	
IIB	31 (19.5)	16 (18.2)	15 (21.1)	
IIIA	34 (21.4)	20 (22.7)	14 (19.7)	
IIIB	4 (2.5)	4 (4.5)	0 (0.0)	
T				0.33
1a	15 (9.4)	4 (4.5)	11 (15.5)	
1b	18 (11.3)	11 (12.5)	7 (9.9)	
2a	48 (30.2)	28 (31.8)	20 (28.2)	
2b	37 (23.3)	21 (23.9)	16 (22.5)	
3	37 (23.3)	22 (25.0)	15 (21.1)	
4	4 (2.5)	2 (2.3)	2 (2.8)	
ECOG				0.31
0	45 (28.3)	27 (30.7)	18 (25.4)	
1	112 (70.4)	59 (67.0)	53 (74.6)	
2	2 (1.3)	2 (2.3)	0 (0.0)	
Surgery				0.90
Pneumonectomy	24 (15.1)	15 (17.0)	9 (12.7)	
Bilobectomy	11 (6.9)	6 (6.8)	5 (7.0)	

Lobectomy	113 (71.1)	61 (69.3)	52 (73.2)	
Segmentectomy	11 (6.9)	6 (6.8)	5 (7.0)	
Adjuvant radiotherapy				0.23
No	148 (93.1)	80 (90.9)	68 (95.8)	
Yes	11 (6.9)	8 (9.1)	3 (4.2)	
Adjuvant chemotherapy				0.49
No	87 (54.7)	46 (52.3)	41 (57.7)	
Yes	72 (45.3)	42 (47.7)	30 (42.3)	
Smoking				0.41
No	20 (12.6)	13 (14.8)	7 (9.9)	
Former	40 (25.2)	19 (21.6)	21 (29.6)	
Yes	99 (62.3)	56 (63.6)	43 (60.6)	

Table 11: Patient characteristics and PD-1 expression status results. P values were calculated by performing Pearson's chi-square test.

Furthermore, a patient subgroup analysis for PD-1 staining intensity was conducted. The results can be seen in table 12:

Characteristics of patients and intensity of PD-1 lymphocyte staining

Characteristic	All Patients (N = 159)	Weak staining (N = 45)	Strong staining (N = 114)	P Value
Age				0.40
<54	32 (20.1)	11 (24.4)	21 (18.4)	
55-64	77 (48.4)	18 (40.0)	59 (51.8)	
>64	50 (31.4)	16 (35.6)	34 (29.8)	
Sex				0.27
Male	88 (55.3)	28 (62.2)	60 (52.6)	
Female	71 (44.7)	17 (37.8)	54 (47.4)	
Stage				0.76
IA	25 (15.7)	7 (15.6)	18 (15.8)	
IB	30 (18.9)	8 (17.8)	22 (19.3)	
IIA	35 (22.0)	11 (24.4)	24 (21.1)	
IIB	31 (19.5)	6 (13.3)	25 (21.9)	
IIIA	34 (21.4)	11 (24.4)	23 (20.2)	
IIIB	4 (2.5)	2 (4.4)	2 (1.8)	
T				0.88
1a	15 (9.4)	3 (6.7)	12 (10.5)	
1b	18 (11.3)	6 (13.3)	12 (10.5)	

2a	48 (30.2)	14 (31.1)	34 (29.8)	
2b	37 (23.3)	10 (22.2)	27 (23.7)	
3	37 (23.3)	10 (22.2)	27 (23.7)	
4	4 (2.5)	2 (4.4)	2 (1.8)	
ECOG				0.01
0	45 (28.3)	20 (44.4)	25 (21.9)	
1	112 (70.4)	24 (53.3)	88 (77.2)	
2	2 (1.3)	1 (2.2)	1 (0.9)	
Surgery				0.82
Pneumonectomy	24 (15.1)	7 (15.6)	17 (14.9)	
Bilobectomy	11 (6.9)	4 (8.9)	7 (6.1)	
Lobectomy	113 (71.1)	32 (71.1)	81 (71.1)	
Segmentectomy	11 (6.9)	2 (4.4)	9 (7.9)	
Adjuvant radiotherapy				0.44
No	148 (93.1)	43 (95.6)	105 (92.1)	
Yes	11 (6.9)	2 (4.4)	9 (7.9)	
Adjuvant chemotherapy				0.40
No	87 (54.7)	27 (60.0)	60 (52.6)	
Yes	72 (45.3)	18 (40.0)	54 (47.4)	
Smoking				0.56
No	20 (12.6)	7 (15.6)	13 (11.4)	
Former	40 (25.2)	9 (20.0)	31 (27.2)	
Yes	99 (62.3)	29 (64.4)	70 (61.4)	

Table 12: Patient characteristics and PD-1 expression intensity results. P values were calculated by performing Pearson's chi-square test.

Univariate analysis was performed to investigate correlations between clinical parameters and RFS and OS. Results can be found in table 13. We were able to show that lower stage at diagnosis, certain treatment regimes (surgery and chemotherapy) and PD-1 expression frequency and intensity correlated positively with RFS and OS.

Variable	Univariate		
	HR	**95% CI**	**P**
RFS			
Age	0.90	0.66 to 1.23	0.51
Sex	0.87	0.56 to 1.36	0.54
Stage	1.39	1.19 to 1.64	<0.001
ECOG	0.78	0.49 to 1.25	0.30
Type of surgery	0.65	0.50 to 0.84	0.001
Adjuvant radiotherapy	1.45	0.63 to 3.36	0.38
Adjuvant chemotherapy	2.05	1.30 to 3.24	0.002
Smoking status	0.73	0.51 to 1.05	0.09
PD-1 frequency	0.53	0.33 to 0.84	0.008
PD-1 intensity	0.51	0.32 to 0.80	0.004
OS			
Age	1.21	0.85 to 1.72	0.29
Sex	0.94	0.56 to 1.56	0.81
Stage	1.41	1.16 to 1.72	0.001
ECOG	0.88	0.52 to 1.49	0.63
Type of surgery	0.65	0.48 to 0.87	0.004
Adjuvant radiotherapy	1.24	0.49 to 3.01	0.65
Adjuvant chemotherapy	2.57	1.50 to 4.38	0.001
Smoking status	0.75	0.51 to 1.12	0.16
PD-1 frequency	0.47	0.27 to 0.80	0.006
PD-1 intensity	0.39	0.23 to 0.65	<0.001

Table 13: Univariate Cox proportional hazards regression models for relapse-free and overall survival for PD-1 expression

In order to determine survival probabilities according to PD-1 expression frequency and intensity for RFS and OS, Kaplan-Meier analyses were performed. For simplification of the curves, PD-1 expression was divided into ”0“ (no expression) and “1“ (expression) and “0“ (weak intensity) and “1“ (strong intensity). We were able to show an improved outcome in RFS (P = 0.007) and OS (P = 0.005) in patients with PD-1 expression frequency and an improved outcome in RFS (P = 0.003) and OS (P < 0.001) in patients with stronger PD-1 expression intensity.

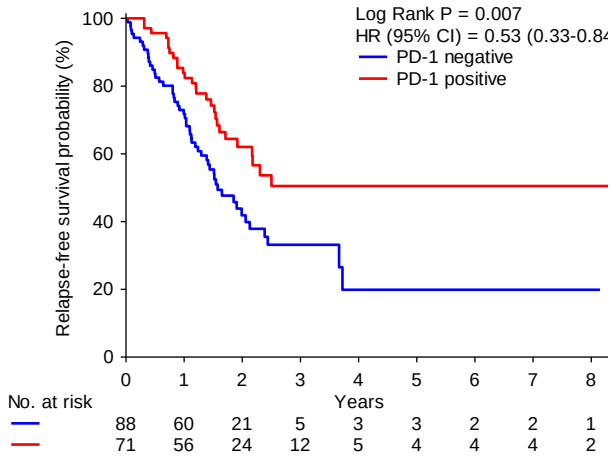

Figure 11: Kaplan-Meier curves for the effect of PD-1 expression in lymphocytes on relapse-free survival.

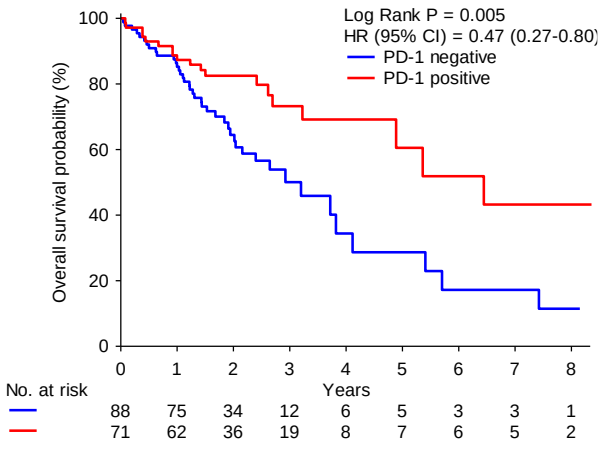

Figure 12: Kaplan-Meier curves for the effect of PD-1 expression in lymphocytes on overall survival.

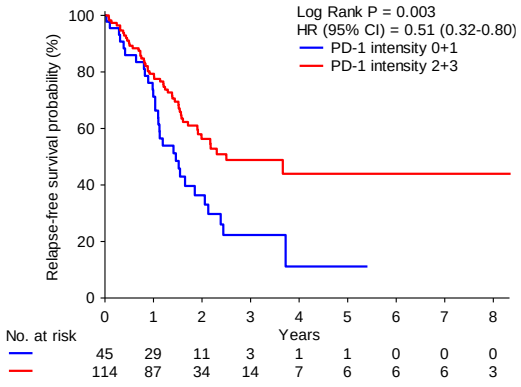

Figure 13: Kaplan-Meier curves for the effect of PD-1 expression intensity in lymphocytes on relapse-free survival.

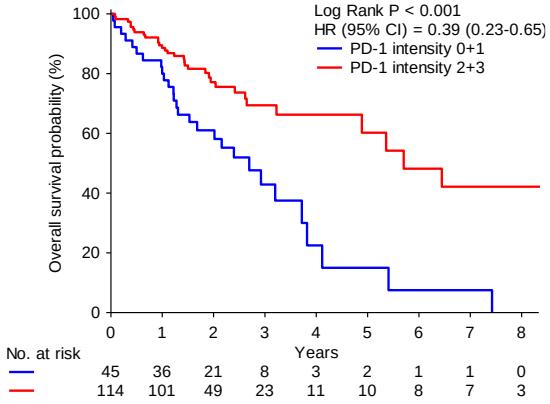

Figure 14: Kaplan-Meier curves for the effect of PD-1 expression intensity in lymphocytes on overall survival.

To identify independent variables for the effect of PD-1 expression frequency and expression intensity on RFS and OS, we performed a multiple Cox proportional hazards regression analysis with the aforementioned clinical parameters. Its results are shown in tables 14 and 15. As can be seen, both PD-1 expression frequency and intensity are independent prognostic biomarkers (0.026, 0.008, 0.005, and <0.001) for PD-1 expression frequency in RFS and OS and for expression intensity in RFS and OS, respectively). Furthermore, we were able to show that stage has independent prognostic relevance for RFS and OS, with early stages posing a lower risk in these two categories.

Variable	Multivariate		
	HR	**95% CI**	**P**
RFS			
Age	0.99	0.73 to 1.36	0.96
Sex	0.94	0.59 to 1.50	0.81
Stage	1.28	1.07 to 1.54	0.007
ECOG	0.77	0.47 to 1.27	0.31
Adjuvant radiotherapy	0.90	0.38 to 2.13	0.81
Adjuvant chemotherapy	1.52	0.90 to 2.59	0.12
PD-1 frequency	0.58	0.36 to 0.94	0.026
OS			
Age	1.31	0.91 to 1.88	0.14
Sex	1.13	0.66 to 1.93	0.65
Stage	1.26	1.00 to 1.59	0.047
ECOG	0.72	0.41 to 1.27	0.26
Adjuvant radiotherapy	0.63	0.24 to 1.65	0.34
Adjuvant chemotherapy	2.17	1.15 to 4.08	0.017
PD-1 frequency	0.46	0.26 to 0.82	0.008

Table 14: Multivariate Cox proportional hazards regression models for relapse-free and overall survival for PD-1 expression.

Variable	Multivariate		
	HR	**95% CI**	**P**
RFS			
Age	1.00	0.74 to 1.37	0.98
Sex	1.00	0.62 to 1.59	0.99
Stage	1.29	1.07 to 1.55	0.007
ECOG	0.88	0.53 to 1.45	0.60
Adjuvant radiotherapy	1.15	0.49 to 2.70	0.76
Adjuvant chemotherapy	1.60	0.94 to 2.72	0.09
PD-1 intensity	0.50	0.31 to 0.82	0.005
OS			
Age	1.28	0.90 to 1.82	0.18
Sex	1.08	0.64 to 1.84	0.77
Stage	1.27	1.01 to 1.60	0.04
ECOG	0.88	0.50 to 1.55	0.65
Adjuvant radiotherapy	0.85	0.33 to 2.22	0.74
Adjuvant chemotherapy	2.17	1.16 to 4.05	0.015
PD-1 intensity	0.37	0.22 to 0.62	<0.001

Table 15: Multivariate Cox proportional hazards regression models for relapse-free and overall survival for PD-1 expression intensity.

Finally, we were interested in the distribution of PD-1 expression frequency (table 16) and intensity (table 17) among the DNA mutation groups to elucidate whether there are

correlations between PD-1 expression and the emergence of certain DNA mutations. No statistically significant correlations could be found.

Gene	All Patients (N = 159)	PD-1 expression status		P Value
		negative	positive	
EGFR				0.73
wildtype	149 (93.7)	83 (94.3)	66 (93.0)	
mutated	10 (6.3)	5 (5.7)	5 (7.0)	
KRAS				0.88
wildtype	93 (58.5)	51 (58.0)	42 (59.2)	
mutated	66 (41.5)	37 (42.0)	29 (40.8)	
PI3KCA				0.92
wildtype	152 (95.6)	84 (95.5)	68 (95.8)	
mutated	7 (4.4)	4 (4.5)	3 (4.2)	

Table 16: PD-1 expression status and DNA mutations. P values were calculated by performing Pearson's chi-square test.

Gene	All Patients (N = 159)	PD-1 expression intensity		P Value
		negative	positive	
EGFR				0.90
wildtype	149 (93.7)	42 (93.3)	107 (93.9)	
mutated	10 (6.3)	3 (6.7)	7 (6.1)	
KRAS				0.81
wildtype	93 (58.5)	27 (60.0)	66 (57.9)	
mutated	66 (41.5)	18 (40.0)	48 (42.1)	
PI3KCA				0.40
wildtype	152 (95.6)	44 (97.8)	108 (94.7)	
mutated	7 (4.4)	1 (2.2)	6 (5.3)	

Table 17: PD-1 expression intensity and DNA mutations. P values were calculated by performing Pearson's chi-square test.

Discussion

DNA mutations in lung cancer as molecular biomarkers

EGFR mutations

Molecular biomarkers in lung cancer, particularly in adenocarcinoma, play a key role in modern diagnostic testing and in the field of prognosis and prediction concerning the disease. Since most patients from the study group investigated in this thesis received standard chemotherapy and / or radiotherapy disregarding the mutation status of the aforementioned genes, only the prognostic value of these biomarkers can be discussed here.

Since the study population was not preselected for any mutations, we can assume a random distribution regarding the aforementioned biomarkers. As mentioned in the results section, 6.2% of the investigated samples contained *EGFR* mutations. This value is somewhat lower compared to the incidence of *EGFR* mutations in lung cancer in Caucasians described in the literature, which is around 10-15%.[15] Furthermore, 70% of *EGFR* mutations were detected in patients old than 64 years; contrasting the finding that these mutations mostly occur in younger patients.[60] However, although this result is statistically significant ($P = 0.017$), the small sample size (10 *EGFR* mutated patients) has to be taken into account. Considering the distribution of *EGFR* mutations in the different sexes, we found a non-significant ($P = 0.09$) accumulation in the female sex. This is in accordance with observations described in the literature.[14] Probably the strongest correlation ($P < 0.001$) between *EGFR* mutation status and a particular patient characteristic can be found in the smoking status: 50% of all *EGFR* mutations occurred in non-smokers. This is in strong conformity with published data[14] that indicate that the vast majority of *EGFR*-mutated patients are never-smokers and that *EGFR* mutations surmount *KRAS* mutations in patients with a smoking history of ≤10 pack years.[61] Furthermore, although not statistically significant, we can deduct that patients with *EGFR*-mutated tumors have less median pack-years of smoking history than patients with *EGFR*-wildtype tumors (figure 2), a finding that is inverted for *KRAS* (figure 1). We can thus conclude that tobacco smoke is not a carcinogen that drives *EGFR* mutations,[11] which signifies that other factors predispose for these mutations. Given the strong correlation with the ethnicity of the patients, one explanation for inclination towards development of these mutations might be the presence of certain polymorphisms.[11] While these polymorphisms explain ethnic variations in *EGFR* expression levels, they do not elucidate the reason for the emergence of mutations. A study[62] has shown that genetics seem to be more pivotal than environmental factors; however, the study group was small and thus, this finding merits further investigation. Furthermore, it is yet unknown why *EGFR* mutations are predominantly

prevalent in adenocarcinoma. In order to prevent new lung cancer cases with tumors harboring these mutations, extensive research is necessary.

We furthermore examined the types of *EGFR* mutations. As can be seen in table 6, 2 mutations (20%) occurred in exon 18, 5 mutations (50%) were exon 19 deletions or insertions, 2 mutations (20%, of which one co-occurred with a mutation in exon 18) were found in exon 20, and 2 mutations (20%) in exon 21. These results are slightly different compared to the prevalence of EGFR mutations described in the literature, with exon 18 mutations making up 5%, exon 19 indels making up 45%, exon 20 mutations making up <1%, and exon 21 making up 40 – 45% of all mutant cases.[17] As previously mentioned, however, due to the small sample size of *EGFR*-mutated tumor samples, no statistically significant statements can be made in respect to prevalence and distribution of *EGFR* mutation types.

Nine out of 10 detected *EGFR* mutations are definitely sensitive to tyrosine kinase inhibitors. The mutation in exon 20, V769M, is of particular interest: according to the COSMIC Database,[63] this mutation was previously detected in only one other sample, along with an in-frame deletion in exon 19, and not further investigated.[64] In our case, the V769M mutation co-occurs with a point mutation in exon 18. Thus, nothing is known about sensitivity to tyrosine kinase inhibitors of tumors harboring this mutation and it is unclear whether these co-occurrences are random or systematic.

To further confirm our results and to further investigate the distribution of the expression of mutant receptors within the tumor area, we conducted immunohistochemistry for the L858R and the E746-A750 deletion mutations. For all five samples, we were able to show expression of the mutant receptor. For the three deletion mutation samples, we were able to find that staining (and thus, expression) intensity is linked to the abundance of the mutant receptor, with stronger staining in the case of heavily mutated samples. This might also be valid for L858R mutations; however, due to the small sample size and the similar extent of the mutation in these two samples (both around 14% as assessed by Pyrosequencing®), no assumptions can be made.

The two antibodies had been extensively tested in other experiments[65] and had shown remarkable results in sensitivity and specificity for the detection of the L858R and the E746-A750 deletion mutations. Moreover, since immunohistochemistry proves to be a cost-effective, quick and established method and can be automatized, clinical implementation of this method for *EGFR* mutation status testing might be an option, especially so because

immunohistochemistry produces a visual image of the tumor cells harboring these mutations. With a standardized interpretation manual, results can be quantified, interpretation biases can be minimized and decisions can be made whether a certain result is predictive for TKI response. However, as is always the case with immunohistochemistry, interpretation might in some cases be difficult due to the nature of the sample and similar phenotypes of different cells. Thus, more experience in the interpretation of the staining results is necessary than when handling data from Pyrosequencing®. Furthermore, antibodies employed in IHC only bind well to the structures that they have been raised against. In the study cited above,[65] it was shown that the specific antibody binds well to 15bp exon 19 deletions, but that sensitivity significantly decreases when used to detect deletions of other lengths. This obstacle is similar to the limitation of PCR-based mutation detection assays, where each mutation requires its own primer. In Pyrosequencing®, on the other hand, different mutations can be detected easily without the need of multiple antibodies. In conclusion, both methods have their advantages and their pitfalls. Their use depends on the type of questions asked. However, in clinical routine DNA sequencing represents currently the gold-standard for the detection of DNA mutations.

Due to the fact that these two kinds of mutations belong to the group of activating mutations in adenocarcinoma, mutant receptor expression was solely found on tumor cells as expected. Furthermore, we were able to see that staining frequency and intensity varies between tumor cells. This finding is in concordance with the concept of multistep cancerogenesis, which states that a variety of mutations are necessary for the neoplasic phenotype to arise. Clonal expansion of tumor cells, each with differences in their mutation makeup, leads to the possibility of the expression of driver mutations other than those found in the *EGFR* gene. As can be deducted from these results, not only the emergence of resistance mutations leads to eventual failure of TKI therapy, but also the fact that some tumor cells do not contain the respective EGFR mutation and, therefore, are intrinsically resistant to TKI therapy.

KRAS mutations

KRAS mutations are of particular importance in adenocarcinoma biomarker research because of their abundance: they are detected in about 26% of all Western patients with lung cancer of non-squamous histology,[66,67] making them one of the most commonly detected mutation in lung cancer. However, there are currently no drugs available for blocking constitutive K-RAS downstream signaling which is conferred by these activating mutations.

From our samples, a total of 68, or 42.3%, were *KRAS*-mutation positive, yielding a strong overrepresentation. As seen in table 8, we were able to provide statistically significant evidence that the mutations occur predominantly in the age group of 55 – 64 and in females (P = 0.003 for both characteristics). Furthermore, *KRAS* mutation prevalence correlates with smoking history (P = 0.006); a finding that is fortified to a certain extent by the fact that median pack-years are higher in patients with *KRAS*-mutated tumors than in patients with *KRAS*-wildtype tumors (figure 1, not significant). As described in table 6, point mutations in codon 12 were found to be most common with over 80% of all *KRAS* mutations, which coincides with observations described in the literature.[24] It has now been widely established that *KRAS* mutation status strongly correlates with exposure to tobacco. In a study conducted in 2012, it has been found that in patients with a smoking history of >10 pack years, *KRAS* mutations are more common than *EGFR* mutations. The same study also explains the role of sex differences and tobacco exposure; with women more prone to lung cancer than men.[61,68] This finding serves to explain the statistically significant difference in *KRAS*-mutated tumors examined during this project.

The use of routine *KRAS* testing in clinical practice is currently debated due to lack of drugs that target K-RAS specifically. This circumstance is not expected to change in the near future, leaving a combination of docetaxel with the MEK inhibitor selumetinib as most promising option.[24]

Immune markers in lung cancer

Given the fact that many lung cancer patients acquired the disease due to prolonged tobacco exposure, the probability of harboring *EGFR* driver mutations is low and thus tyrosine kinase inhibitors are not effective. With no direct treatment against K-RAS mutations, patients are left with no targeted treatment approaches and thus have to surrender to conventional chemotherapy. Because of this *status quo*, PD-L1 and PD-L2 expression status in tumor cells and PD-1 expression in tumor-infiltrating lymphocytes play an increasingly important role in lung cancer biomarker research, especially so after the discovery of the immune system's role in cancer as emerging hallmark by Hanahan and Weinberg.[33] In a study conducted by Calles *et al*.[39], it was shown that smokers demonstrate increased incidence of PD-L1 expression and increased expression levels, which correlate with intensity of the smoking history as measured in pack-years. Our results, however, demonstrated no statistical significant correlation between PD-L1 expression, its intensity and the smoking status; from 100 current smokers, 22 patients had a PD-L1 expression of 1 – 49% on their tumor cells, and 21 patients had a PD-L1

expression of ≥50% (P = 0.32). The reason for a potential connection between smoking status and / or smoking history and PD-L1 expression is not yet fully understood. In some publications, researchers draw the conclusion that smoking-induced inflammatory response involves T-cell proinflammatory cytokines, including interferon-γ, that is known to induce PD-L1 expression.[39,69,70] Furthermore, since smoking-associated lung cancers have a higher mutational load, more tumor antigens are created, which in turn results in increased immunogenicity.[71,72,73] By analysis of the PD-L1 expression status and the DNA mutations, we were able to bring in context the incidence of *KRAS* mutations and PD-L1 expression: as can be seen in table 10, PD-L1 expression significantly correlates with *KRAS* mutations (P = 0.021), with an increase in *KRAS* mutations accompanying tumor cell PD-L1 expression. This result might be foreseeable due to high incidence of *KRAS* mutations along with high expression of PD-L1 among current and former smokers; however, other published studies[74,75] could not find a statistically significant correlation between *KRAS* muation status and PD-L1 expression status. The reasons for the difference between our results and the results of other studies can be manifold, with the two main considerations being a too small study population and differing scoring systems of PD-L1 immunohistochemistry. More research has to be conducted in this field to investigate the relationship between immune markers such as PD-L1 and PD-1 and DNA mutations. Anyhow, given the availability of targeted treatment options, immune therapy offers a good alternative for patients who have unknown or untargetable oncogene-driven lung cancer.

An important question to be addressed is the prognostic and predictive values that might be attributed to these biomarkers. In a multivariate analysis, we were able to show that PD-1 expression frequency and intensity correlates with RFS and OS (P = 0.026, 0.008 and 0.005, <0.001, respectively), making PD-1 expression in TILs a favorable, independent prognostic biomarker. This result is foreseeable due to the fact that PD-1 expression was assessed in T-lymphocytes, their primary location. Tumor-infiltrating lymphocytes play a major role in the tumor microenvironment: they possess a cytotoxic role and are able to recognize tumor cells and destroy them, with the aforementioned PD-1 / PD-L1/2 mechanism counteracting this mechanism. However, this interaction, which suppresses immune destruction, is only apparent in a certain subset of lymphocytes, which are in most cases part of a larger population of T-lymphocytes invading the tumor. It has previously been shown[76] that certain TILs have prognostic value and that high levels of certain TILs ($CD3^+$, $CD4^+$, $CD8^+$) can improve outcome, while high levels of others ($FoxP3^+$) are associated with a poor prognosis.[77] Although playing a key role in immune evasion, PD-1 is in most cases only expressed in a

subgroup of lymphocytes; thus, total extent of TILs is a favorable prognostic biomarker in NSCLC. It can be speculated that the fraction of lymphocytes expressing PD-1 aid in tumor immune evasion within the PD-1 / PD-L1/2 framework, and that disruption of this interaction could further improve outcome in patients with huge areas of TILs due to the fact that these lymphocytes then also become available for tumor cell destruction by the immune system. However, even in cases with huge areas of lymphocyte PD-1 expression, these lymphocytes would need to be in physical contact with tumor cells expressing PD-L1 or PD-L2 and the system of T-cell exhaustion would have to be readily activated in order for PD-1 expression to be considered a poor prognostic factor. Taken all these points into consideration, further investigation, especially with focus on the context in which PD-1 is expressed, needs to be conducted in order to establish the value of PD-1 expression as biomarker.

Assessment of PD-L1 expression by immunohistochemistry poses certain difficulties: PD-L1 can be expressed in tumor cells, lymphocytes and macrophages. Only expression in viable tumor cells can be used as predictable biomarker in this context. Given the circumstance that macrophages and adenocarcinoma tumor cells have a similar appearance after IHC staining, well-trained personnel is required to evaluate these patient samples. While also true for PD-1, as protein marker, PD-L1 expression is subject to change depending on treatment and time, and success of immunohistochemistry is dependent on the antibody and the protocol used. In order to obtain and record reproducible results, scoring systems and cut-off points should be clearly defined, which is not always the case until now.[78] Notwithstanding these considerations, we cannot confirm that PD-L1 expression frequency and intensity are in any way prognostic for RFS and OS: no significant differences could be found in the results of Kaplan-Meier analysis; neither for PD-L1 expression on tumor cells (figures 5, 6, 7 and 8), nor for its expression on lymphocytes (figures 9 and 10). In some studies, high PD-L1 expression was associated with a poor prognosis,[79] while in other studies, PD-L1 expression correlates with higher overall survival.[80]

Although not investigated in this thesis, the identification of predictive biomarkers for assessment of response to antibody therapy that disrupts the PD-1 / PD-L1/2 interaction is an emerging field. PD-1, PD-L1 and PD-L2 expression levels themselves are under current investigation. Several studies suggest that PD-L1 expression is a suitable predictive biomarker in respect to response rates, RFS and OS. Since PD-L1 is induced by interferon-γ, its expression was also tested as potential predictive biomarker for response to anti-PD-L1

antibodies in several studies. It was found that IFNγ expression correlates with response rate and overall survival, however stronger in melanoma patients than in patients with NSCLC.[77]

Mutational analysis and DNA sequencing

Due to the ever increasing importance of the analysis of molecular biomarkers in malignant diseases, DNA mutational analyses play a key role in research. Nowadays, numerous analysis techniques from many biotech companies exist, which poses certain challenges to the researcher in regards to the choice of technique. In our case, the demands are relatively clear: rather than a platform that was designed for high throughput sequencing, we were interested in a system that detects point mutations and small insertions and deletions with high accuracy. Furthermore, we were interested in a method that detects mutations in a given sample in a relatively low frequency due to the fact that some of our samples might contain a relatively low amount of tumor cells and not all of those might contain the mutation in question. For this purpose, two approaches exist: sequencing-based techniques and PCR-based techniques. Sequencing-based techniques exploit the possibility of nucleotide detection, while in PCR-based techniques, allele-specific primers are used. Depending on the individual experiment and the questions asked by the experimenter, each technique has its own set of advantages and drawbacks. Above all, the ability to detect mutations should be sensitive (correct detection of true positives) and specific (correct detection of true negatives). Furthermore, the results should be reproducible to ensure their scientific significance. Also, it is highly desirable that test is robust, meaning that it should not be influenced by varying parameters such as tumor tissue ratio across different samples, and there should be the possibility of automated reporting to compensate for judgment deviations across individual evaluators.[81] Finally, it may be advantageous to use a system that only requires a small amount of DNA.

For our purpose, Pyrosequencing® is the most suitable method. Pyrosequencing® enables us to qualitatively and quantitatively identify mutations, it has a relatively low limit of detection, compensating for samples with low amounts of tumor cells and the heterogeneity within the tumor cells or with relatively few mutant alleles, protocols for *EGFR* mutational analysis are standardized to a great extent due to the existence of commercially available kits and the established methods used for sample preparation, and it identifies all nucleotides (and thereby all possible mutations) in a given amplicon, abolishing the need for numerous primers in order to detect all mutations of a given sample in allele-specific PCR mutational analyses. Furthermore, the provided software of the platform allowed easy assay design and evaluation of the results, minimizing subjective errors.

There are several studies that compare different platforms for mutational analyses. One disadvantage of Pyrosequencing® is that it requires a time-consuming PCR step as sample preparation, making it prone to possible DNA mutations that arise during the amplification process. These mutations might masquerade as sequence variants during the analysis process.[82] Moreover, the Pyrosequencing® technology requires relatively short sequences to analyze, as longer DNA strands might form secondary structures. This would render the strand inaccessible for the DNA polymerase and thus, sequencing could not be performed. Furthermore, the light signal generated during the sequencing reaction is only proportional to up to six incorporated dNTPs, making the identification of longer homopolymers problematic.[83] In addition, it has been shown that Pyrosequencing® frequently introduces errors during the sequencing of insertions and/or deletions (indels).[84] Due to these limitations, the system dictates a limit of detection that is characteristic for each mutation. All detected mutations below this limit are low-level mutations and were judged as wildtype when unconfirmed in duplicate. Since in our case mutational analysis of tumor tissue is performed in order to judge whether a patient will benefit from a therapy using tyrosine kinase inhibitors, only short amplicons were analyzed and no extensive homopolymers were sequenced, these limitations only play a minor role. However, in our experiments, the mutation frequency for some samples differed significantly when tested multiple times, giving us a rather randomized result. Moreover, some samples yielded mutation frequencies of far over 50%, which is a result that is highly unrealistic: germline *EGFR* mutations in patients with adenocarcinoma are extremely rare,[16] so the probability of two mutant alleles and thus a mutation frequency of over 50% is highly unrealistic. The fact that most of our samples do not solely contain tumor tissue and the presence molecular heterogeneity of tumors further reduces mutation frequencies. One explanation for the presence of highly mutated *EGFR* could be aneuploidy and amplification of chromosomal segments which is frequently found in cancer;[85] however, it is highly doubtful that these mechanisms lead to results of over 50% in our samples. Samples exhibiting discrepancies in their results should therefore be tested in duplicate, and if uncertainties still persist, the results should be confirmed with a second, PCR-based method, such as the cobas® EGFR Mutation Test.

In summary, we can conclude that Pyrosequencing® offers a fast, sensitive and specific technique for the qualitative detection of point mutations and indels. Although a huge financial investment is initially necessary to acquire a Pyrosequencer and the accompanying software, costs can be reduced by the development of protocols for the desired sequencing reactions, abolishing the need to purchase costly kits. Furthermore, Pyrosequencing® is easy

to conduct and automated reporting is available for many mutations. However, certain limitations persist in regard to allowed amplicon length, the presence of homopolymers and indels. Also, for some mutations, the limit of detection in Pyrosequencing® might give inaccurate results when detecting low mutation frequencies. As previously mentioned, the extended characterization of certain samples by a PCR-based method is recommended, especially in the case of diagnostic purposes.

Immunohistochemistry

Immunohistochemistry is a relatively simple method to detect antigens. Although somewhat arduous in the light of a long-lasting hands-on time even when using an automated stainer, protocol development is relatively easy due to the limited number of variable incubation times and reagents. Once a protocol for the detection of a certain antigen using its specific antibody has been developed, reproducibility is greatly ensured. Due to the fact that the method of immunohistochemistry itself exists since over 40 years (first described by Taylor und Burns in 1974[86]), it is a well-established, relatively inexpensive and renowned system. Its particular strength lies in the possibility of not only the sheer detection of certain antigens, but by their visualization within the framework of a tissue section. This way, it is possible to determine which type of cell expresses the antigen of interest, we can draw conclusions about the degree of expression of the antigen and by identification of the cell types in the vicinity, it is possible to draw conclusions about the course of a malignant disease; e.g. by quantification of tumor-infiltrating lymphocytes. Furthermore, double staining allows for the detection of interactions between antigens and stained samples on glass slides can be stored over long periods of time and thus be re-evaluated if desired. However, there are certain drawbacks to immunohistochemistry. Above all, it is imperative that the protocols for the visualization of a certain antibody-antigen interactions are sophistically elaborated, because quite frequently, background staining arises, which may lead to an incorrect staining interpretation and thus to false-positive results. Additionally, when one wishes to detect antigens that are in close proximity to the cell nucleus, haematoxylin counterstaining might mask the brown color arising from the reaction with the DAB. Probably the biggest disadvantage of immunohistochemistry lies in the nature of the technique itself: in order to evaluate the result of a particular experiment, it is almost always vital to identify the cell types which express the antigen of interest. Given the fact that this is demanding in many cases, this judgment requires an experienced pathologist for the interpretation of a particular result. However, in many, if not most cases, these judgments even differ among experienced pathologists. In the end, it is to mention that to detect a certain antigen, the corresponding antibody must exist and must

have been tested rigorously. Unfortunately, we have experienced that in certain cases, a commercially available antibody does not function. In that case, immunohistochemistry is not possible.

Concluding remarks

Lung cancer is one of the most fatal malignant diseases, accounting for more deaths than colon, breast and prostate cancer combined.[87,88] Although the role of tobacco abuse in respect to the incidence of lung cancer has been studied extensively and there is no doubt that most lung cancer cases can be prevented, incidence rates are still increasing, especially in the world's developing regions and among women. Numerous traditional treatment options exist, such as surgical removal of the tumor, chemotherapy and radiation therapy. These options have been extensively tested in large clinical trials and treatment regimens have been constantly improved over the past decades; with the emergence of novel chemotherapeutic agents and advancements in modern radiology. However, as previously mentioned, treatment options are significantly limited in stage IV lung cancers, the stage in which the disease is most often diagnosed. Due to the fact that even modern chemotherapeutic agents have severe side effect profiles because they also target normal cells, relatively low response rates due to inherent or acquired drug resistance, and surgical removal of the tumor is only recommended in early stages, new treatment options had to be developed in order to increase survival rates and decrease side effects. In the light of modern oncology, biomarker research has contributed significantly to the emergence of personalized therapy, standing at the crossroads of basic research, translational research, drug development and clinical practice. It is thus evident that biomarker research, especially in the field of molecular biology, offers an abundance of possibilities to even revolutionize modern oncology and personalized medicine. However, due to the fact that even in the case of individual objects, professionals from many different disciplines are involved, making close collaborations between basic and translational researchers, as well as clinicians, imperative. Another obstacle for the implementation of biomarkers into clinical practice is the lack of sensitivity and specificity of most candidate biomarkers. In this consideration, the prevalence of the corresponding cancer disease plays a vital role. For example, if a certain malignant disease has a prevalence of 6 cases in every 10,000 people, the biomarker must have a sensitivity (the correct identification of true positives) and specificity (the correct identification of true negatives) of 100% and 99.4%, respectively, to identify one true positive in 10 false negatives.[89] Needless to say, the identification of powerful biomarkers with acceptable positive and negative predictive values

remains an arduous task. Once a biomarker has passed these initial tests, implementation into clinical routine is another hurdle. The biomarker should be easily accessible and ideally, testing should be fast, simple and economically reasonable to conduct. Furthermore, the biomarker should have both strong prognostic and predictive value, and professionals should be able to utilize the biomarker in the study of therapy success and disease progression.

In the light of these considerations, biomarker research in the field of lung cancer is particularly interesting. Several demands and characteristics are important for a biomarker to be suitable for clinical implementation in this area. Naturally, an optimal biomarker offers high sensitivity and specificity. Furthermore, it allows for swift and cost-effective screening of high-risk groups (e.g. heavy smokers) as a preventive measure. Due to the fact that tumor biopsies are in many cases laborious to obtain due to the location of the tumor,it is of interest that the biomarker be detectable in other body parts as well. This is a particularly fastidious demand when monitoring the success of a therapy or disease progression, because samples need to be obtained at multiple points in time. Moreover, as previously mentioned, lung cancer is mainly detected in stage IV, in which treatment options are severely limited and no curative therapies exist. Also, even at this stage, many tumors remain undetected for a prolonged period of time because of the shortcomings of radiographic methods. It is thus desirable that the biomarker be detectable in the early stages of the disease to limit metastasis and to be able to exploit more treatment options. Additionally, the biomarker should be able to deliver significant prognostic and predictive values.

There have been significant advancements in the discovery of molecular biomarkers in non-small cell lung cancer, particularly in the identification of driver mutations in adenocarcinoma. An established range of screening methods have become available and detection of certain biomarkers, such as the *EGFR* mutation status, as well as *ALK* rearrangements, has nowadays become a standard procedure. Furthermore, as previously mentioned, other biomarkers have been identified and extensively studied as well, including but not limited to *KRAS*, *BRAF*, *HER2*, *PIK3CA*, *AKT1*, *MEK1*, *NRAS* and *ROS1*.[67]

An obstacle in the identification of driver mutations in lung cancer remains the issue of specimen acquisition. Oftentimes, the tumor is located deep within the bronchia and thus difficult to access, making biopsies an arduous process. Liquid biopsy is a novel method of sample acquisition; it opens the doors to uncomplicated biomarker screening and therapy monitoring due to the analysis of the blood plasma. Promising trials are currently underway;

however, extensive further research is necessary to unleash the full potential of liquid biopsies.[90]

Reference list

1. Lung Cancer Estimated Incidence, Mortality and Prevalence Worldwide in 2012. http://globocan.iarc.fr/old/FactSheets/cancers/lung-new.asp. Published 2012. Accessed March 24, 2016.
2. Jackman DM, Johnson BE. Small-cell lung cancer. *Lancet*. 2005;366(9494):1385-1396. doi:10.1016/S0140-6736(05)67569-1.
3. Detailed guide to non-small cell lung cancer. http://www.cancer.org/cancer/lungcancer-non-smallcell/detailedguide/non-small-cell-lung-cancer-what-is-non-small-cell-lung-cancer. Accessed March 24, 2016.
4. Zarogoulidis K, Zarogoulidis P, Darwiche K, et al. Treatment of non-small cell lung cancer (NSCLC). *J Thorac Dis*. 2013;5(SUPPL.4):2-9. doi:10.3978/j.issn.2072-1439.2013.07.10.
5. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin Pharmacol Ther*. 2001;69(3):89-95. doi:10.1067/mcp.2001.113989.
6. NCI Dictionary of Cancer Terms. http://www.cancer.gov/publications/dictionaries/cancer-terms?expand=P. Accessed March 24, 2016.
7. Ren X, Ma W, Lu H, et al. Modification of cytokine-induced killer cells with chimeric antigen receptors (CARs) enhances antitumor immunity to epidermal growth factor receptor (EGFR)-positive malignancies. *Cancer Immunol Immunother*. 2015;64(12):1517-1529. doi:10.1007/s00262-015-1757-6.
8. Ferguson KM. A structure-based view of Epidermal Growth Factor Receptor regulation. *Annu Rev Biophys*. 2008;37(215):353-373. doi:10.1146/annurev.biophys.37.032807.125829.A.
9. Burgess AW, Cho HS, Eigenbrot C, et al. An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Mol Cell*. 2003;12(3):541-552. doi:10.1016/S1097-2765(03)00350-2.
10. Huse M, Kuriyan J. The Conformational Plasticity\tReview of Protein Kinases. *Cell*. 2011;109:1-8. doi:10.1016/S0092-8674(02)00741-9.
11. Shigematsu H, Gazdar AF. Somatic mutations of epidermal growth factor receptor signaling pathway in lung cancers. *Int J Cancer*. 2006;118(2):257-262. doi:10.1002/ijc.21496.
12. Scaltriti M, Baselga J. The Epidermal Growth Factor Receptor Pathway : A Model for Targeted Therapy The Epidermal Growth Factor Receptor Pathway : A Model for T argeted Therapy. *Clin Cancer Res*. 2006;12(18):5268-5272. doi:10.1158/1078-0432.CCR-06-1554.
13. Nicholson RI, Gee JM, Harper ME. EGFR and cancer prognosis. *Eur J Cancer*. 2001;37 Suppl 4:S9-S15. doi:10.1016/S0959-8049(01)00231-3.

14. Shigematsu H, Lin L, Takahashi T, et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst*. 2005;97(5):339-346. doi:10.1093/jnci/dji055.

15. Jorge SEDC, Kobayashi SS, Costa DB. Epidermal growth factor receptor (EGFR) mutations in lung cancer: preclinical and clinical data. *Braz J Med Biol Res*. 2014;47(11):929-939. doi:10.1590/1414-431X20144099.

16. Centeno I, Blay P, Santamaría I, et al. Germ-line mutations in epidermal growth factor receptor (EGFR) are rare but may contribute to oncogenesis: A novel germ-line mutation in EGFR detected in a patient with lung adenocarcinoma. *BMC Cancer*. 2011;11:172. doi:10.1186/1471-2407-11-172.

17. Sharma S V, Bell DW, Settleman J, Haber DA. Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer*. 2007;7(3):169-181. doi:nrc2088 [pii]\r10.1038/nrc2088.

18. Gazdar AF, Shigematsu H, Herz J, Minna JD. Mutations and addiction to EGFR: The Achilles "heal" of lung cancers? *Trends Mol Med*. 2004;10(10):481-486. doi:10.1016/j.molmed.2004.08.008.

19. Yun C-H, Mengwasser KE, Toms A V, et al. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc Natl Acad Sci U S A*. 2008;105(6):2070-2075. doi:10.1073/pnas.0709662105.

20. Thress KS, Paweletz CP, Felip E, et al. Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. *Nat Med*. 2015;21(6):560-562. doi:10.1038/nm.3854.

21. Croce CM. Oncogenes and cancer. *Science*. 1995;267(5203):1408-1409. doi:10.1126/science.7878455.

22. Weinstein IB, Joe AK. Mechanisms of disease: Oncogene addiction--a rationale for molecular targeting in cancer therapy. *Nat Clin Pract Oncol*. 2006;3(8):448-457. doi:10.1038/ncponc0558.

23. Tebbutt N, Pedersen MW, Johns TG. Targeting the ERBB family in cancer: couples therapy. *Nat Rev Cancer*. 2013;13(9):663-673. doi:10.1038/nrc3559.

24. Shea M, Costa DB, Rangachari D. Management of advanced non-small cell lung cancers with known mutations or rearrangements: latest evidence and treatment approaches. *Ther Adv Respir Dis*. 2015:1-17. doi:10.1177/1753465815617871.

25. Jänne P a, Yang JC-H, Kim D-W, et al. AZD9291 in EGFR Inhibitor-Resistant Non-Small-Cell Lung Cancer. *N Engl J Med*. 2015;372(18):1689-1699. doi:10.1056/NEJMoa1411817.

26. LUX-Lung 7 Results Infographic. http://www.giotrif.com/content/dam/internet/pm/giotrif1/com_EN/documents/2016_01_19_-_LUX-Lung_7_results_infographic_4_0.pdf. Accessed March 24, 2016.

27. Pirker R. Third-generation epidermal growth factor receptor tyrosine kinase inhibitors in advanced nonsmall cell lung cancer. *Curr Opin Oncol*. 2015:115-121. doi:10.1097/CCO.0000000000000260.

28. Mitsudomi T, Tsai C, Shepherd F et al. *AZD9291 in Pretreated T790M Positive Advanced NSCLC: AURA2 Phase II Study.* Denver, CO

29. Pirker R. EGFR-directed monoclonal antibodies in non-small cell lung cancer. *Target Oncol*. 2013;8(1):47-53. doi:10.1007/s11523-012-0244-7.

30. Martin P, Stewart E, Pham N-A, et al. Cetuximab Inhibits T790M-Mediated Resistance to Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor in a Lung Adenocarcinoma Patient-Derived Xenograft Mouse Model. *Clin Lung Cancer*. March 2016. doi:10.1016/j.cllc.2016.01.002.

31. Kurai J, Chikumi H, Hashimoto K, et al. Antibody-dependent cellular cytotoxicity mediated by cetuximab against lung cancer cell lines. *Clin Cancer Res*. 2007;13(5):1552-1561. doi:10.1158/1078-0432.CCR-06-1726.

32. Kawamoto T, Sato JD, Le a, Polikoff J, Sato GH, Mendelsohn J. Growth stimulation of A431 cells by epidermal growth factor: identification of high-affinity receptors for epidermal growth factor by an anti-receptor monoclonal antibody. *Proc Natl Acad Sci U S A*. 1983;80(5):1337-1341. doi:10.1073/pnas.80.5.1337.

33. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell*. 2011;144(5):646-674. doi:10.1016/j.cell.2011.02.013.

34. Kim R, Emi M, Tanabe K. Cancer immunoediting from immune surveillance to immune escape. *Immunology*. 2007;121:1-14. doi:10.1111/j.1365-2567.2007.02587.x.

35. Okazaki T, Honjo T. PD-1 and PD-1 ligands: From discovery to clinical application. *Int Immunol*. 2007;19(7):813-824. doi:10.1093/intimm/dxm057.

36. Marzec M, Zhang Q, Goradia A, et al. Oncogenic kinase NPM/ALK induces through STAT3 expression of immunosuppressive protein CD274 (PD-L1, B7-H1). *Proc Natl Acad Sci U S A*. 2008;105(52):20852-20857. doi:10.1073/pnas.0810958105.

37. Latchman Y, Wood CR, Chernova T, et al. PD-L2 is a second ligand for PD-1 and inhibits T cell activation. *Nat Immunol*. 2001;2(3):261-268. doi:10.1038/85330.

38. Momtaz P, Postow MA. Immunologic checkpoints in cancer therapy: Focus on the programmed death-1 (PD-1) receptor pathway. *Pharmgenomics Pers Med*. 2014;7:357-365. doi:10.2147/PGPM.S53163.

39. Calles A, Liao X, Sholl LM, et al. Expression of PD-1 and its ligands, PD-L1 and PD-L2, in smokers and never smokers with KRAS mutant lung cancer. *J Thorac Oncol*. 2015;10(12):1726-1735. doi:10.1097/JTO.0000000000000687.

40. Herbst RS, Baas P, Kim DW, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): A randomised controlled trial. *Lancet*. 2015;6736(15). doi:10.1016/S0140-6736(15)01281-7.

41. Jänne PA, Shaw AT, Pereira JR, et al. Selumetinib plus docetaxel for KRAS-mutant advanced non-small-cell lung cancer: A randomised, multicentre, placebo-controlled, phase 2 study. *Lancet Oncol*. 2013;14(1):38-47. doi:10.1016/S1470-2045(12)70489-8.

42. Roberts PJ. Clinical use of crizotinib for the treatment of non-small cell lung cancer. *Biologics*. 2013;7:91-101. doi:10.2147/BTT.S29026.

43. Michael W. Schmitt, Marc J. Prindle and LAL. Implications of genetic heterogeneity in cancer. *Ann N Y Acad Sci*. 2012;September(1267):110-116. doi:10.1111/j.1749-6632.2012.06590.x.Implications.

44. Torti D, Trusolino L. Oncogene addiction as a foundational rationale for targeted anti-cancer therapy: Promises and perils. *EMBO Mol Med*. 2011;3(11):623-636. doi:10.1002/emmm.201100176.

45. Sequist L V, Waltman B a, Dias-Santagata D, et al. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci Transl Med*. 2011;3(75):75ra26. doi:10.1126/scitranslmed.3002003.

46. Yun CH, Boggon TJ, Li Y, et al. Structures of Lung Cancer-Derived EGFR Mutants and Inhibitor Complexes: Mechanism of Activation and Insights into Differential Inhibitor Sensitivity. *Cancer Cell*. 2007;11(3):217-227. doi:10.1016/j.ccr.2006.12.017.

47. Yasuda H, Park E, Yun C-H, et al. Structural, biochemical and clinical characterization of epidermal growth factor receptor (EGFR) exon 20 insertion mutations in lung cancer. *Sci Transl Med December*. 2013;18(5216):216-177. doi:10.1126/scitranslmed.3007205.

48. *Therascreen EGFR Pyro Kit Handbook, Version 1, Reference #971480*.

49. *PyroMark PCR Handbook, Ref. Date: May 2009*.

50. Hyman ED. A new method of sequencing DNA. *Anal Biochem*. 1988;174(2):423-436.

51. Ramos-Vara JA, Miller MA. When tissue antigens and antibodies get along: revisiting the technical aspects of immunohistochemistry--the red, brown, and blue technique. *Vet Pathol*. 2014;51(1):42-87. doi:10.1177/0300985813505879.

52. Sompuram SR, Vani K, Hafer LJ, Bogen SA. Antibodies Immunoreactive With Formalin-Fixed Tissue Antigens Recognize Linear Protein Epitopes. *Am J Clin Pathol*. 2006;125(1):82-90. http://ajcp.oxfordjournals.org/content/125/1/82.abstract.

53. Tween 20 product description. http://www.sigmaaldrich.com/catalog/product/sigma/p9416?lang=de&region=AT. Accessed March 24, 2016.

54. Ramos-Vara JA et al. Making and Using Antibodies: A Practical Handbook. In: Boca Raton, FL; 2007:273-314.

55. KD value: A quantitative measurement of antibody affinity. http://www.abcam.com/index.html?pageconfig=resource&rid=15749&source=pagetrap&viapagetrap=kd. Accessed March 25, 2016.

56. NCI Dictionary of Cancer Terms. http://www.cancer.gov/publications/dictionaries/cancer-terms?search=relapse. Accessed March 25, 2016.

57. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. *J Am Stat Assoc*. 1958;53(282):457-481. doi:10.2307/2281868.

58. Filipits M, Pohl G, Rudas M, et al. Clinical role of multidrug resistance protein 1 expression in chemotherapy resistance in early-stage breast cancer: The Austrian Breast and Colorectal Cancer Study Group. *J Clin Oncol*. 2005;23(6):1161-1168. doi:10.1200/JCO.2005.03.033.

59. Cox DR. Regression Models and Life-Tables. *J R Stat Soc Ser B*. 1972;34(2):187-220. doi:10.2307/2985181.

60. Prabhakar CN. Epidermal growth factor receptor in non-small cell lung cancer. *Transl lung cancer Res*. 2015;4(2):110-118. doi:10.3978/j.issn.2218-6751.2015.01.01.

61. Dogan S, Shen R, Ang DC, et al. Molecular epidemiology of EGFR and KRAS mutations in 3,026 lung adenocarcinomas: Higher susceptibility of women to smoking-related KRAS-mutant cancers. *Clin Cancer Res*. 2012;18(22):6169-6177. doi:10.1158/1078-0432.CCR-11-3265.

62. Tsao AS, Tang XM, Sabloff B, et al. Clinicopathologic characteristics of the EGFR gene mutation in non-small cell lung cancer. *J Thorac Oncol*. 2006;1(3):231-239. doi:01243894-200603000-00008 [pii].

63. Catalogue of somatic mutations in cancer.

64. Huang S-F, Liu H-P, Li L-H, et al. High frequency of epidermal growth factor receptor mutations with complex patterns in non-small cell lung cancers related to gefitinib responsiveness in Taiwan. *Clin Cancer Res*. 2004;10(24):8195-8203. doi:10.1158/1078-0432.CCR-04-1245.

65. Brevet M, Arcila M, Ladanyi M. Assessment of EGFR Mutation Status in Lung Adenocarcinoma by Immunohistochemistry Using Antibodies Specific to the Two Major Forms of Mutant EGFR. *J Mol Diagn*. 2010;12(2):169-176. doi:10.2353/jmoldx.2010.090140.

66. Dearden S, Stevens J, Wu Y-L, Blowers D. Mutation incidence and coincidence in non small-cell lung cancer: meta-analyses by ethnicity and histology (mutMap). *Ann Oncol*. 2013;24(9):2371-2376. doi:10.1093/annonc/mdt205.

67. Califano R, Abidin A, Tariq N-A, Economopoulou P, Metro G, Mountzios G. Beyond EGFR and ALK inhibition: Unravelling and exploiting novel genetic alterations in advanced non small-cell lung cancer. *Cancer Treat Rev*. 2016;41(5):401-411. doi:10.1016/j.ctrv.2015.03.009.

68. Zang EA, Wynder EL. Differences in Lung Cancer Risk Between Men and Women: Examination of the Evidence. *J Natl Cancer Inst* . 1996;88 (3-4):183-192. doi:10.1093/jnci/88.3-4.183.

69. Hodge G, Nairn J, Holmes M, Reynolds PN, Hodge S. Increased intracellular T helper 1 proinflammatory cytokine production in peripheral blood, bronchoalveolar lavage and intraepithelial T cells of COPD subjects. *Clin Exp Immunol*. 2007;150(1):22-29. doi:10.1111/j.1365-2249.2007.03451.x.

70. Spranger S, Spaapen RM, Zha Y, et al. Up-Regulation of PD-L1, IDO, and T(regs) in the Melanoma Tumor Microenvironment Is Driven by CD8(+) T Cells. *Sci Transl Med*. 2013;5(200):200ra116-ra200ra116. doi:10.1126/scitranslmed.3006504.

71. Soo RA. Shedding light on the molecular determinants of response to anti-PD-1 therapy. *Transl Lung Cancer Res*. 2015;4(6):816-819. doi:10.3978/j.issn.2218-6751.2015.05.06.

72. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer genes. *Nature*. 2013;499(7457):214-218. doi:10.1038/nature12213.

73. Snyder A, Makarov V, Merghoub T, et al. Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *N Engl J Med*. 2014;371(23):2189-2199.

doi:10.1056/NEJMoa1406498.

74. D'Incecco A, Andreozzi M, Ludovini V, et al. PD-1 and PD-L1 expression in molecularly selected non-small-cell lung cancer patients. *Br J Cancer*. 2015;112(1):95-102. doi:10.1038/bjc.2014.555.

75. Yang CY, Lin MW, Chang YL, Wu CT, Yang PC. Programmed cell death-ligand 1 expression in surgically resected stage i pulmonary adenocarcinoma and its correlation with driver mutations and clinical outcomes. *Eur J Cancer*. 2014;50(7):1361-1369. doi:10.1016/j.ejca.2014.01.018.

76. Djenidi F, Adam J, Goubar A, et al. CD8+CD103+ Tumor–Infiltrating Lymphocytes Are Tumor-Specific Tissue-Resident Memory T Cells and a Prognostic Factor for Survival in Lung Cancer Patients. *J Immunol* . 2015;194 (7):3475-3486. doi:10.4049/jimmunol.1402711.

77. Zeng D-Q, Yu Y-F, Ou Q-Y, et al. Prognostic and predictive value of tumor-infiltrating lymphocytes for clinical therapeutic research in patients with non-small cell lung cancer. *Oncotarget; Vol 7, No 12*. 2016. http://www.impactjournals.com/oncotarget/index.php?journal=oncotarget&amp.

78. Shukuya T, Carbone DP. Predictive markers for the efficacy of anti-PD-1/PD-L1 antibodies in lung cancer. *J Thorac Oncol*. doi:http://dx.doi.org/10.1016/j.jtho.2016.02.015.

79. Wang A, Wang HY, Liu Y, et al. The prognostic value of PD-L1 expression for non-small cell lung cancer patients: A meta-analysis. *Eur J Surg Oncol*. 2015;41(4):450-456. doi:http://dx.doi.org/10.1016/j.ejso.2015.01.020.

80. Cooper WA, Tran T, Vilain RE, et al. PD-L1 expression is a favorable prognostic factor in early stage non-small cell carcinoma. *Lung Cancer*. 2015;89(2):181-188. doi:http://dx.doi.org/10.1016/j.lungcan.2015.05.007.

81. Lopez-Rios F, Angulo B, Gomez B, et al. Comparison of molecular testing methods for the detection of EGFR mutations in formalin-fixed paraffin-embedded tissue specimens of non-small cell lung cancer. *J Clin Pathol*. 2013;66(5):381-385. doi:10.1136/jclinpath-2012-201240.

82. Metzker ML. Sequencing technologies [mdash] the next generation. *Nat Rev Genet*. 2010;11(1):31-46. http://dx.doi.org/10.1038/nrg2626.

83. Ronaghi M. Pyrosequencing Sheds Light on DNA Sequencing Pyrosequencing Sheds Light on DNA Sequencing. *Genome*. 2001;11(650):3-11. doi:10.1101/gr.150601.

84. Margulies M, Egholm M, Altman WE, et al. Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. *Nature*. 2005;437(7057):376-380. doi:10.1038/nature03959.

85. Giam M, Rancati G. Aneuploidy and chromosomal instability in cancer: a jackpot to chaos. *Cell Div*. 2015;10:3. doi:10.1186/s13008-015-0009-7.

86. Taylor CR, Burns J. The demonstration of plasma cells and other immunoglobulin-containing cells in formalin-fixed, paraffin-embedded tissues using peroxidase-labelled antibody. *J Clin Pathol*. 1974;27(1):14-20. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC477979/.

87. Schiller JH, Harrington D, Belani CP, et al. Comparison of Four Chemotherapy

Regimens for Advanced Non–Small-Cell Lung Cancer. *N Engl J Med*. 2002;346(2):92-98. doi:10.1056/NEJMoa011954.

88. Rapp E, Pater JL, Willan A, et al. Chemotherapy can prolong survival in patients with advanced non-small-cell lung cancer--report of a Canadian multicenter randomized trial. *J Clin Oncol* . 1988;6 (4):633-641. http://jco.ascopubs.org/content/6/4/633.abstract.

89. Wagner PD, Srivastava S. New Paradigms in Translational Science Research in Cancer Biomarkers. *Transl Res*. 2012;159(4):343-353. doi:10.1016/j.trsl.2012.01.015.

90. Buder A, Tomuta C, Filipits M. The potential of liquid biopsies. *Curr Opin Oncol*. 2016;28(2). http://journals.lww.com/co-oncology/Fulltext/2016/03000/The_potential_of_liquid_biopsies.7.aspx.

List of figures

Figure 1: Box plot depicting the correlation between pack-years and KRAS mutation status. Lower and upper parts of the boxes describe the first and third quartiles, respectively, while the bar inside the box shows the position of the second quartile (the median). The whiskers mark the upper and lower 1.5 interquartile ranges, while the dots and asterisks represent weak and strong outliers, respectively.

Figure 2: Box plot depicting the correlation between pack-years and EGFR mutation status. Lower and upper parts of the boxes describe the first and third quartiles, respectively, while the bar inside the box shows the position of the second quartile (the median). The whiskers mark the upper and lower 1.5 interquartile ranges, while the dots and asterisks represent weak and strong outliers, respectively.

Figure 3: Immunohistochemistry of tissue sample with tumor cells harboring an EGFR L858R mutation. Blue staining stems from the haematoxylin, while the brownish staining arose from the DAB incubation. Magnification 200x.

Figure 4: Immunohistochemistry of tissue sample with tumor cells harboring an EGFR E746_A750 deletion mutation. Blue staining stems from the haematoxylin, while the brownish staining arose from the DAB incubation. Magnification 200x.

Figure 5: Kaplan-Meier curves for the effect of PD-L1 expression in tumor cells on relapse-free survival.

Figure 6: Kaplan-Meier curves for the effect of PD-L1 expression in tumor cells on overall survival.

Figure 7: Kaplan-Meier curves for the effect of PD-L1 expression intensity in tumor cells on relapse-free survival.

Figure 8: Kaplan-Meier curves for the effect of PD-L1 expression intensity in tumor cells on overall survival.

Figure 9: Kaplan-Meier curves for the effect of PD-L1 expression in lymphocytes on relapse-free survival.

Figure 10: Kaplan-Meier curves for the effect of PD-L1 expression in lymphocytes on overall survival.

Figure 11: Kaplan-Meier curves for the effect of PD-1 expression in lymphocytes on relapse-free survival.

Figure 12: Kaplan-Meier curves for the effect of PD-1 expression in lymphocytes on overall survival.

Figure 13: Kaplan-Meier curves for the effect of PD-1 expression intensity in lymphocytes on relapse-free survival.

Figure 14: Kaplan-Meier curves for the effect of PD-1 expression intensity in lymphocytes on overall survival.

List of tables

Table 1: EGFR mutations in non-small cell lung cancer according to sensitivity and resistance to gefitinib and erlotinib, two first-generation TKIs.

Table 2: Examined exons, codons and mutations as well as their role within the EGFR tyrosine kinase domain.

Table 3: EGFR amplicons and their PCR primers.

Table 4: Assays and corresponding sequencing primers, sequences to analyze, nucleotide dispension order and analyzed mutations for mutational analysis of EGFR.

Table 5: Reagents employed for Pyrosequencing® and their corresponding manufacturers. Reagents used for sample preparation are separated by a reinforced border from reagents used for the sequencing reaction proper.

Table 6: Summary of types and prevalence of DNA mutations in the patient samples.

Table 7: Characteristics of patients with EGFR-mutated tumors. P values were calculated by performing Pearson's chi-square test.

Table 8: Characteristics of patients with KRAS-mutated tumors. P values were calculated by performing Pearson's chi-square test.

Table 9: Patient characteristics and PD-L1 expression status results. P values were calculated by performing Pearson's chi-square test.

Table 10: PD-L1 expression status and DNA mutations. P values were calculated by performing Pearson's chi-square test.

Table 11: Patient characteristics and PD-1 expression status results. P values were calculated by performing Pearson's chi-square test.

Table 12: Patient characteristics and PD-1 expression intensity results. P values were calculated by performing Pearson's chi-square test.

Table 13: Univariate Cox proportional hazards regression models for relapse-free and overall survival for PD-1 expression

Table 14: Multivariate Cox proportional hazards regression models for relapse-free and overall survival for PD-1 expression.

Table 15: Multivariate Cox proportional hazards regression models for relapse-free and overall survival for PD-1 expression intensity.

Table 16: PD-1 expression status and DNA mutations. P values were calculated by performing Pearson's chi-square test.

Table 17: PD-1 expression intensity and DNA mutations. P values were calculated by performing Pearson's chi-square test.

Zusammenfassung / Abstract

In der vorliegenden Masterarbeit wurde der Mutationsstatus von *EGFR*, *KRAS*, *NRAS*, *HRAS*, *PI3KCA* und *BRAF* von 161 Patienten mit Lungen – Adenokarzinom untersucht. Die Mutationen wurden auf formalinfixierten, paraffineingebetteten Tumorblöcken durch die Isolation der DNA, Amplifikation der kodierenden Region und durch darauffolgendes Pyrosequencing® nachgewiesen. Des Weiteren wurde die Proteinexpression der Immunmarker PD-1 und PD-L1 in diesen Proben mittels Immunhistochemie untersucht. Durch die Berechnung der P-Werte mittels chi-Quadrat Test nach Pearson sowie durch Regressionsanalysen nach Cox wurde die Rolle der Immunmarker als potentielle prognostische Biomarker getestet; außerdem wurde die Verteilung der *EGFR*-Mutationen sowie weiterer DNA-Mutationen, welche als Teil eines anderen Projektes erhoben wurden, unter Patienten-Untergruppen sowie Mutationsuntergruppen ausgewertet.

Von 161 Tumorproben waren 10 (6,2%) positiv für *EGFR*-Mutationen (von welchen eine Probe zwei Mutationen aufwies), 68 (42,3%) waren positiv für *KRAS*-Mutationen, 1 (0,6%) *BRAF* und 7 (4,4%) *PI3KCA*-Mutationen wurden detektiert. Ferner wurden jeweils 7 (4,4%) und 3 (1,9%) ALK bzw. ROS1-Mutationen nachgewiesen. Im Hinblick auf die PD-L1-Immunhistochemie waren 69 (62,9%) Proben negativ, 31 (19,2%) Proben wiesen eine Membranfärbung zwischen 1% und 49% aller Tumorzellen auf. Weitere 28 (17,4%) Proben wiesen eine Membranfärbung von ≥50% aller Tumorzellen auf. PD-1 Expressionsfrequenz und Expressionsintensität auf tumorinfiltrierenden Lymphozyten wurde mittels Immunhistochemie bestimmt. Insgesamt waren 159 von 161 Proben auswertbar. Die Expressionfrequenz wurde unterteilt in *negativ* (keine Färbung; n=88) und *positiv* (≥1% Färbung; n=71). Expressionsintensität wurde in schwache (n=45) und starke (n=114) Färbung unterteilt. Die multivariate Analyse hat ergeben, dass die Expressionsfrequenz- und Intensität von PD-1, jedoch nicht von PD-L1, unabhängige, prognostische Biomarker fur rezidivfreies Überleben und die Gesamtüberlebenszeit sind.

In the present master thesis, we investigated the *EGFR, KRAS, NRAS, HRAS, PI3KCA*, and *BRAF* mutation status of 161 lung adenocarcinoma patients. Mutations were detected on formalin-fixed, paraffin-embedded tumor blocks by isolation of the DNA, amplification of the encoding region and subsequent Pyrosequencing®. Moreover, we investigated the protein expression of the immune markers PD-1 and PD-L1 in these samples by immunohistochemistry. By calculation of the P-values using Pearson's chi-square test, as well as by setup of Cox proportional hazards regression models, we tested the role of the immune

markers as potential prognostic biomarkers and evaluated the distribution of *EGFR* mutations, as well as other DNA mutations that have been assessed as part of a different project, among patient and mutational subgroups.

Of 161 tumor samples, 10 (6.2%) were *EGFR* mutation positive (of which one sample harbored two mutations), 68 (42.3%) were positive for *KRAS* mutations, 1 (0.6%) *BRAF* and 7 (4.4%) *PI3KCA* mutations were detected. Moreover, 7 (4.4%) and 3 (1.9%) *ALK* and *ROS1* rearrangements were found, respectively. Regarding PD-L1 immunohistochemistry, 69 samples (62.9%) were negative, 31 samples (19.2%) showed a membrane staining on 1 – 49% of tumor cells, and 28 samples (17.4%) exhibited a membrane staining of more than or equal to 50% of tumor cells. PD-1 expression frequency and intensity was assessed on tumor-infiltrating lymphocytes by immunohistochemistry. In total, 159 out of 161 samples were evaluable. Expression frequency was divided into negative (no staining; n=88) and positive (≥ 1% staining; n=71). Expression intensity was divided into weak (n=45) and strong staining (n=114). Multivariate analyses revealed that PD-1 but not PD-L1 expression frequency and intensity are independent prognostic biomarkers for relapse-free survival and overall survival of the patients.