Epidemiological and functional characterization of oncogenic drivers in lung cancer

Inaugural-Dissertation
zur Erlangung des Doktorgrades

Dr. nat. med.

der Medizinischen Fakultät und
der Mathematisch-Naturwissenschaftlichen
Fakultät der Universität zu Köln

vorgelegt von

Dipl. Biol. Danila Seidel

aus Cottbus
Oktober 2013

Die der vorliegenden Arbeit zugrunde liegenden Experimente wurden am Max-Planck-Institut für neurologische Forschung, sowie in der Abteilung Translationale Genomik der Universität zu Köln durchgeführt.

Prüfer

Prof. Dr. Mats Paulsson (Vorsitzender des IPMM-Ausschusses)
Prof. Dr. Peter Nürnberg (Gutachter)
Prof. Dr. Bernd Wollnik (Gutachter)
Prof. Dr. Jürgen Wolf
Prof. Dr. Günter Schwarz

Dr. Hildegard Büning (Beisitzerin)

Tag der mündlichen Prüfung: 26. Feb. 2014

Table of contents

Declaration of my contribution to this work:

Results presented here were generated as part of the **Clinical Lung Cancer Genome Project** Initiative that was launched in 2008 with the aim of creating a network of scientists and clinicians working together in a tight international collaboration to understand the genetic mechanisms underlying lung cancer and to bring new therapy opportunities to the lung cancer patients. All data presented here were analyzed and interpreted by me. I designed all experiments, organized, collected and prepared samples for pathological review including immunohistochemistry, mutation, copy number and gene expression analysis. I defined the gene set for genetic analysis, produced and validated all mutation data using mass spectrometry. I analyzed raw mutation and expression data using appropriate methods. Furthermore, I designed and performed all experiments and interpreted results of the functional characterization of the novel mutations *EGFR* K714N and *FGFR3* R248C that I identified. I received help from computational scientists that programmed algorithms for data analysis. Raw copy number data were processed to copy number calls by Dr. Martin Peifer (Figure 14). Scott Carter calculated the purity of tumor samples using segmented copy number data.[1] Finally, I led the final data analysis and interpretation of the entire study. For this part I received help from a board-certified oncologist, Priv.-Doz. Dr. Thomas Zander (Department I of Internal Medicine, University of Köln), who co-led the study with me. Detailed acknowledgment of the people involved in each step is given in the Material & Methods section.

Abstract

Tumors were traditionally classified based on morphological and immunohistochemical characteristics into adenocarcinoma, large cell carcinoma, squamous cell carcinoma and small cell carcinoma.[2] In extensive genomic analyses of lung tumors within the past decade several genetic alterations were identified in certain subtypes that have been proven to be useful targets for specific molecular therapies, such as mutant EGFR that predict response to EGFR inhibitors in adenocarcinoma. With the increasing number of genetic alterations identified in lung cancer that are or might be suitable for targeted therapies, traditional diagnostic approaches become more and more insufficient.

This thesis aimed at investigating genetic characteristics of all histological subtypes of lung tumors to define a genetically informed classification of lung cancers and to explore the eligibility of new molecular targets for targeted therapies. To identify genetic subgroups, mutations in selected genes, genome-wide copy number alterations and gene expression patterns were analyzed in 1,255 clinically annotated lung tumors. Most genomic alterations segregated with one of the major histological subtypes adeno-, squamous or small cell carcinoma but were not exclusive for one. In rare cases subtype-specific genetic alterations were also identified in other subtypes, emphasizing the need for reevaluating current genetic tests that are mainly assigned to specific histological subtypes. Large cell carcinomas that are morphologically and clinically heterogeneous did not reveal a distinct pattern of genetic alterations. Most of the tumors of this subtype could be reassigned to one of the other subtypes based on their genetic and expression profiles. Thus, immunohistochemical and genetic tests should be considered for sub-classifying tumors of this subtype into clinically relevant groups. Furthermore, in this study alterations in the tyrosine kinases FGFR1 and FGFR3 have been identified that – based on functional analyses - might be eligible for targeted therapies in lung cancer patients.

The described mutational catalog of primary lung tumors provides detailed information to further guide lung cancer diagnostics and treatment.

Zusammenfassung

Das Bronchialkarzinom ist die häufigste tödliche Krebserkrankung mit einer mittleren Fünfjahres-Überlebensrate von nur 15%. Morphologisch werden die maligne Tumore in die vier Haupttypen Kleinzelliges (SCLC), Adeno- (AD), Platten- (SQ) und Großzelliges (LCC) Karzinom eingeteilt, wobei die letzten drei als Nichtkleinzellige Karzinome zusammengefasst werden. Diese Einteilung erwies sich als notwendig bei der Wahl von Chemotherapeutika, hatte jedoch für die Nichtkleinzelligen Karzinome keine prognostische Relevanz. Als neue Technologien genomische Untersuchungen ermöglichten, wurde deutlich, dass Tumore durch eine Vielzahl von genetischen Veränderungen charakterisiert sind. Genetische Alterationen können zu einer veränderten Genexpression führen und Zellen durch den Erwerb neuer zellulärer Eigenschaften einen Wachstumsvorteil gegenüber normalen Zellen verschaffen. Der Einsatz von zielgerichteten Substanzen, die eine Inhibierung veränderter enzymatischer Prozesse in Tumorzellen bedingen, führte erstmals zu verbesserten Therapieergebnissen in Patientengruppen mit spezifischen genetischen Merkmalen. Zum Beispiel wurde 2004 gezeigt, dass die Hemmung des Epidermalen Wachstumsfaktor Rezeptors mit Erlotinib, einem niedermolekularen Inhibitor der die Tyrosinkinasedomäne des Rezeptors blockiert, zu einem verlängerten progressionsfreien Überleben in Patienten führt, deren Tumor eine genetische Aberration in diesem Rezeptor aufweisen.[3,4] Derartige neue Therapiemöglichkeiten führten zu drastischen Veränderungen in der Tumorklassifizierung von einer Histologie-basierten hin zu einer kombinierten morphologischen und molekularen Diagnostik zur Stratifizierung von Patienten für individualisierte, zielgerichtete Therapien.

In der hier präsentierten Studie wurden erstmals Lungentumore aller histologischen Subtypen genetisch analysiert, mit dem Ziel eine umfassende genetische Charakterisierung dieser Krebsart zu erlangen und klinisch relevante Subgruppen zu identifizieren, die als Grundlage einer neuen, molekularen Klassifikation dienen könnten. Dafür wurden mehr als 1,000 Tumore auf somatische chromosomale Aberrationen und Genmutationen hin untersucht. Die

Genexpression wurde von einem Teil der Tumore analysiert. Mutationsfrequenzen wurden in Genen untersucht, die bereits in vorherigen Studien in diversen Krebsarten als Onkogene oder Tumorsuppressorgene bestätigt wurden. In dieser Studie wurden mehrere neue potentielle Tumormarker in Lungentumoren identifiziert. Dazu gehören Aberrationen des Fibroblasten Wachstumsfaktor Rezeptors (FGFR), dessen Genkopien häufig in Plattenepithelkarzinomen erhöht waren (*FGFR1*) oder in seltenen Fällen dessen Genbereich, der für die extrazelluläre Region des Rezeptors kodiert, mutiert war (*FGFR3*). In beiden Fällen wurde in funktionellen Experimenten die Sensitivität gegenüber FGFR-Inhibitoren gezeigt.

In den histologische Subtypen AD, SQ und SCLC wurden jeweils charakteristische genetische Aberrationen beschrieben. Genalterationen, die typisch für einen bestimmten histologischen Subtyp waren, wurden in seltenen Fällen ebenfalls in Tumoren anderer Subtypen gefunden. So wurden zum Beispiel Mutationen in *EGFR* und *KRAS*, zwei Gene die häufig im Adenokarzinom alteriert sind, auch in Plattenepithel Tumoren identifiziert. Tumore des LCC Subtyps ließen kein spezifisches Markerprofil erkennen, stattdessen zeigten sie Mutationen, Genkopienzahl-Veränderungen und Genexpressionsprofile, die typisch für die andere Subtypen sind. Mit Hilfe immunohistologisch nachweisbarer linienspezifischer Antigene konnte die Mehrzahl der LCC in andere Lungenkrebs-Subtypen klassifiziert werden, die typische genetische Aberrationen aufwiesen.

Um effektive Strategien für die molekulare Diagnostik zu entwickeln ist es von hoher Bedeutung das Auftreten relevanter Marker in Tumoren zu kennen. Die hier präsentierten Ergebnisse lassen erahnen, dass Einzelgen-basierte molekulare Diagnostik, wie sie heute erfolgt, nicht mehr lange effizient sein wird. Da zahlreiche Therapeutika gegen molekulare Marker bereits in fortgeschrittenen Studienphasen getestet werden, wird eine effiziente Analyse von Tumoren zur molekularen Diagnose nur im Multiplexverfahren sinnvoll sein. Da in seltenen Fällen klinisch relevante Subtypen-charakteristische Aberrationen auch in Tumoren anderer Subtypen identifiziert wurden, sollten Tumore nicht auf Grund der Morphologie von molekularen Tests ausgeschlossen werden. Dies betrifft insbesondere die heterogene Gruppe der LCC, die Genaberrationen aufweisen, die typisch für alle

anderen Subtypen sind. Ob sich aus der Patientenstratifizierung auf Grund molekularer Marker eine Verbesserung der Prognose für Lungenkrebspatienten ergibt, wird sich in entsprechenden Studien zeigen.

Abbreviations

The international system of units (SI units) was used in this thesis.

AD	adenocarcinoma of the lung
ALK	anaplastic lymphoma receptor tyrosine kinase; cytogenetic band: 2p23
BCR	breakpoint cluster region; cytogenetic band: 22q11.23
BRAF	v-raf murine sarcoma viral oncogene homolog B1; cytogenetic band: 7q34
BRCA1/2	breast cancer 1 and 2, early onset; cytogenetic band: 17q21 (1), 13q12.3 (2)
CA	carcinoid of the lung
CD56	synonym: neural cell adhesion molecule (NCAM)
CDKN2A	cyclin-dependent kinase inhibitor 2A; cytogenetic band: 9p21
CK5/6	cytokeratin 5 and 6
CK7	cytokeratin 7
EGFR	epidermal growth factor receptor; cytogenetic band: 7p12
EML4	echinoderm microtubule associated protein like 4; cytogenetic band: 2p21
ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, synonym: HER2; cytogenetic band: 17q12
ERK	extracellular-signal-regulated kinase; cytogenetic band: 22q11
FCS	fetal calf serum
FFPE	Formaline fixed paraffin embedded
H&E	hematoxylin and eosin staining
HRAS	v-Ha-ras Harvey rat sarcoma viral oncogene homolog; cytogenetic band: 11p15.5
IHC	immunohistochemistry
KEAP1	Kelch-Like ECH-Associated Protein 1; cytoband: 19p13
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog; cytogenetic band:12p12.1
LCC	large cell carcinoma of the lung
LCNEC	large cell neuroendocrine carcinoma
Mb	megabase (1Mb = 1 mio. base pairs)
MAPK	Mitogen-activated protein kinase
MEK	mitogen-activated and extracellular-signal-regulated kinase kinase, synonym: MAPKK, MEK1; cytogenetic band: 15q22.1-q22.33

MYC	v-myc myelocytomatosis viral oncogene homolog, transcription factor; cytogenetic band: 8q24.21
NFE2L2	nuclear factor (erythroid-derived 2)-like 2; cytogenetic band: 12q13
NKX2-1	NK2 homeobox 1, synonym: TTF-1; cytogenetic band: 14q13
NRAS	neuroblastoma RAS viral (v-ras) oncogene homolog; cytogenetic band: 1p13.2
P/S	Penicillin/Streptavidin
PI3K	phosphatidylinosito-4,5-bisphosphate-3-kinase; cytogenetic band: 3q26.3
PIK3CA	phosphatidylinositol-4,5-bisphosphate-3-kinase, catalytic subunit
PTEN	phosphatase and tensin homolog; cytogenetic band: 10q23.3
RAS	gene: Rat sarcoma - HRAS, NRAS, KRAS
RB1	retinoblastoma; cytogenetic band: 13q14.2
ROS1	c-ros oncogene 1, receptor tyrosine kinase; cytogenetic band: 6q22
SARC	sarcomatoid of the lung
SCLC	small cell carcinoma of the lung
SCNA	somatic copy number alteration (gains and losses)
SOX2	SRY (Sex Determining Region Y)-Box 2, transcription factor; cytogenetic band: 3q26.3-q27
SQ	squamous cell carcinoma of the lung
TP53	tumor protein p53; cytogenetic band: 17p13.1
TTF-1	thyroid transcription factor-1, see NKX2-1
VEGFR	vascular endothelial growth factor receptor 1, 2 and 3
WHO	World Health Organization

Descriptions

Driver mutation: is a genetic alteration that is sufficient to confer growth and survival advantage to a cell and whose inhibition leads to cell death of cells harboring this alteration.

Five-year survival rate: is the percentage of patients who live at least five years after being diagnosed with cancer.

Kaplan-Meier survival curve: shows the probability of survival in a given time frame taking patient information into account where definite endpoint observation (date of death) is missing. For these *censored* cases date of last follow-up determines the endpoint of observation.

Median survival time: is the time since diagnosis when relative survival is at 50% (time when half of the patients have survived).

Targeted Therapy: is a type of treatment that uses small molecules or monoclonal antibodies that were designed to target and thereby inhibiting the action of certain proteins or other molecules in cancer cells.

Tumorigenesis: is the formation and development of tumors.

Overall survival: is a term used to denote the percentage of patients who are alive after a specified duration of time after being diagnosed with cancer (often reported as 5-year survival rate; for example a 5-year survival rate of 5% means that 5% of the patients are alive after five years).

Progression-free survival: is the time from start of treatment until progression of the disease.

Figures and Tables

Introduction

1. Cancer

One out of three people will get cancer in the course of their lives and one out of four of them will die of it. Every day about 20.000 people worldwide die of cancer, which is therefore the second most common cause of death after cardiovascular disease (7.6 vs. 17 million deaths in 2008, respectively). Due to population growth and increased longevity, statistical trends suggest an increase of the total numbers of annual cancer deaths to approximately 13 million by the year 2030.[5]

1.1. Cancer is a disease of the genome

Cancer initiation and tumorigenesis is a multifactorial, multistage process. The mechanisms inducing cancer development are at this point not fully understood for every cancer type. Several toxins have been linked to certain cancer types such as aflatoxin and alcohol abuse to liver cancer, asbestos to mesothelioma, and cigarette smoking to lung cancer.[6] Furthermore, epidemiological studies and functional analyses revealed that a variety of pathogens are associated with many cancer types to either induce tumor formation or to contribute to tumorigenesis. Most prominently, cervical cancer was found to be induced by human papilloma viruses and liver cancer by hepatitis V viruses. Regardless of the initiating factors, tumors are characterized to accumulate several alterations in the genome of somatic cells inevitably leading to cellular transformation. Tumors consequently arise from somatic evolution. Genetic changes that lead to enhanced or disrupted function of gene products in the affected cell, can eventually confer a selective proliferation and thus lead to the malignant phenotype.[7] Classically it is believed that a tumor emerges from a single mutated cell by clonal expansion and acquisition of additional alterations during proliferation.[7] Today it is assumed that in certain cancer types tumor mass formation is driven only by the proliferation of a distinct subpopulation of cells, which have enhanced self-renewal potential and are also responsible for tumor heterogeneity.[8,9]

The idea that cancer is caused by genomic alterations arose in the early 20th century when Boveri postulated that aberrations affecting chromosomes are associated with malignant growth of tissue.[10] In 1959, the Philadelphia

Chromosome, later identified to be a reciprocal translocation of chromosomes 9 and 22 causing a fusion of the *BCR* and *ABL* genes, was discovered in patients with Chronic Myeloid Leukemia (CML).[11] Since then and with technological advances various cancer types have been studied by global sequencing and SNP (single nucleotide polymorphism) array analyses with the aim to identify genetic alterations to understand molecular changes underlying cancer.[12-18] From the unexpected high amount of genetic alterations identified in those studies the majority are found to be specific for certain cancer types. Only a limited number of genes are repeatedly mutated across many cancer types (for example *BRAF, CDKN2A, KRAS, PIK3CA, PTEN,* and *TP53*) suggesting their impact on the tumorigenic phenotype.[19,20]

It became apparent that tumor formation is in general not caused by a single genetic event but requires several alterations affecting proto-oncogenes and tumor suppressor genes that in combination stimulate uncontrolled growth, division and survival of the cell.[19,21,22] The flood of sequencing data revealed that cancer types are distinguishable by different landscapes of somatic alterations and that tumors of the same type can present different genetic alterations.[16,18,23] Despite the heterogeneity of tumors, all cancer types share the same characteristics that distinguish them from normal cells, the so-called *Hallmarks of Cancer*, summarized by Hanahan and Weinberg in 2011.[22] Cancer cells

a. sustain proliferative signaling by producing growth signals (for instance PDGF production in glioblastoma), structural alterations or increase of receptors to facilitate ligand-independent signaling (for example *ERBB2* overexpression in breast and stomach cancer),
b. overcome negative regulation of proliferation, which mainly depends on tumor suppressor genes such as *RB* and *TP53*,
c. evade programmed cell death by up-regulation of anti-apoptotic and survival signals and down regulation of pro-apoptotic factors,
d. are replicative immortal through expression of telomerase, that adds telomere repeats to chromosome ends whose shortening with every replication step would eventually lead to senescence,

e. induce angiogenesis, the formation of blood vessels, to provide the growing tumor with nutrients and oxygen and evacuate metabolic waste and carbon dioxide by inductive signals (for example VEGF) or expression of angiogenic factors caused by oncogenes (for example *KRAS* or *MYC*),
f. invade surrounding tissue and spread to distant sites of the organism (metastasis),
g. accumulate genetic alterations that can affect single nucleotides within specific genes up to complete chromosomes due to replication errors or deficiencies in repair mechanisms,
h. are supplied with growth and survival factors, signals inducing angiogenesis, invasion and metastasis through inflammatory cells infiltrating the tumor,
i. metabolize glucose through glycolysis which allows metabolic intermediates to be used for biosynthesis of macromolecules required for cell assembly, and
j. evade the immune system that has been associated with suppression of tumor formation in mice.

1.2. Pathways that are altered in human cancer

Heterogeneous cell-signaling pathways are disrupted promoting cells to generate their own mitogenic signals, resist growth-inhibitory signals, evade apoptosis, proliferate inordinate, and eventually invade and metastasize to other sites of the body.[7] Signaling pathways generally proceed from the cell surface, to cytoplasmic intermediates to nuclear transcription factors leading to transcription of effector genes involved in specific cellular processes. If alterations occur in any of the proteins involved in cellular signaling pathways, accurate transduction might no longer be assured, which can result in cell transformation. Most genes that are frequently altered in cancer encode signaling molecules, such as protein kinases that transfer phosphate groups from adenosine triphosphate to specific amino acid residues (serine, threonine or tyrosine) and thereby modifying the functional properties of their substrates, phosphatases that counteract protein kinases by removing phosphate residues on target proteins, guanosine triphosphate (GTP)-binding proteins, and transcription factors.

The Ras-Raf-MEK and PI3K-Akt pathways are the most frequently disrupted signaling pathways in cancer (Figure 1).[7] Despite high complexity of cell signaling networks in each pathway, specific key regulators are involved that propagate in a linear fashion. Activation of these pathways is mediated by binding of a ligand to a tyrosine kinase receptor (for example, binding of the epidermal growth factor (EGF) to the EGF receptor (EGFR) or the fibroblast growth factor (FGF) to the FGF receptor (FGFR)). In the Ras-Raf-MEK pathway a guanosine exchange factor (GEF) binds to the dimerized and autophosphorylated receptor and catalyzes the exchange of the Ras-bound guanosine diphosphate (GDP) against guanosine triphosphate (GTP). This exchange activates Ras, an intracellular membrane-associated GTPase, that promotes phosphorylation and thereby activation of several downstream effectors. One of the most important effectors is B-Raf. Phosphorylated B-Raf activates the kinases MEK1/2, which in turn activate the kinases ERK1/2. ERK is translocated into the nucleus where it activates several transcription factors that initiate expression of cyclines and cytokines (that play a major role in the cell cycle) and growth factors amongst others. The immediate result is the activation of cell proliferation.

In the PI3K-AKT pathway first PI3K is activated either through transfer of a phosphate from the tyrosine kinase receptor dimers or from membrane-associated active GTP-bound Ras proteins (i.e., H-Ras, K-Ras and N-Ras). PI3K phosphorylates the membrane component phosphatidylinositol-4,5-bisphosphat (PIP2) to PIP3. The function of PI3K can be antagonized by the phosphatase and tumor suppressor PTEN (phosphatase and tensin homolog) that converts PIP3 back to PIP2. PIP3 activates AKT that can then inhibit apoptosis and activate translation (through activation of mTOR and the S6K transcription factor). The PI3K-AKT pathway can also be activated by dimerization of other receptors such as ErbB2, which does not need a ligand for activation but forms heterodimers with other receptor monomers of the same family (i.e., EGFR, ErbB3 or ErbB4) when their ligand is bound.[7,24]

Several other pathways regulate cell proliferation, apoptosis and DNA repair. Cell divisions are controlled during cell cycle by several proteins that suppress or activate transition from one state to the next involving DNA synthesis (S phase), mitosis (M phase) and gaps (G phases) where integrity of the cell is assured at

certain checkpoints. Two proteins are mainly involved in the cell cycle control: Rb controls the progression into S phase and p53 regulates expression of proliferation-inhibiting and apoptosis-promoting proteins in response to DNA damage. If Rb or p53 are disrupted cells can divide irrespective of alterations in the DNA sequence.[7]

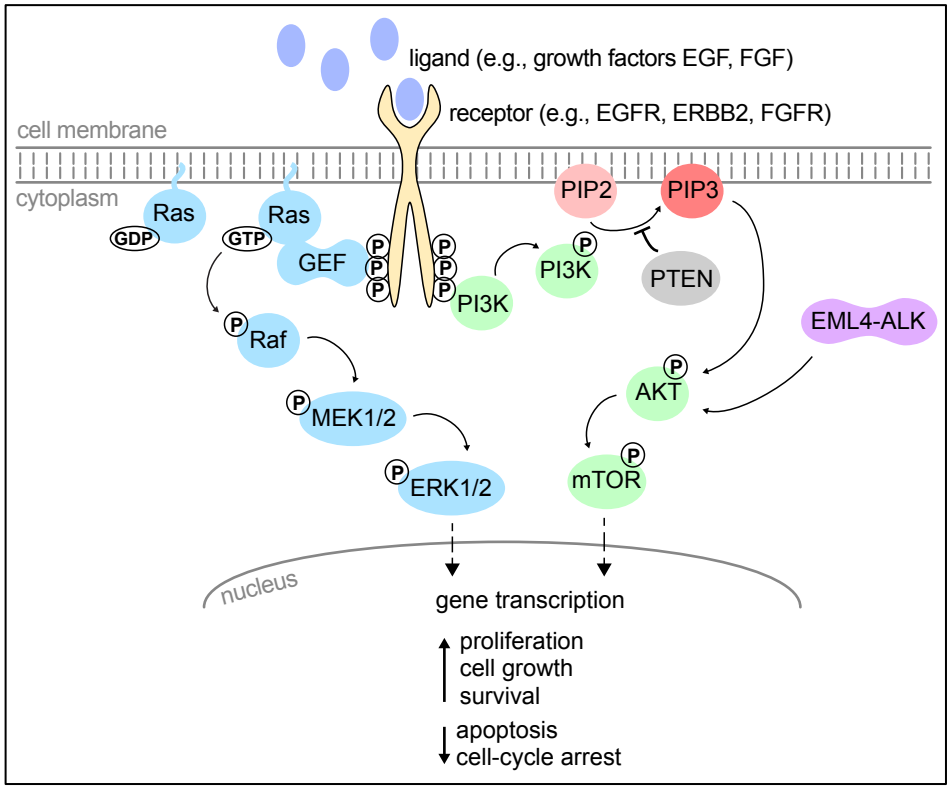

Figure 1 Ras-Raf-MEK (blue) and PI3K-AKT (green) signaling pathways. The schematic illustrates the signal transduction from extracellular stimuli (ligand) through receptors (yellow) to the cell nucleus where gene expression is activated. The direction of information flow between the main signaling molecules is indicated by arrows.

Signaling pathways form a complex network with parallel and vertical pathways. Every component can be disrupted by various mechanisms through different causes that eventually lead to either apoptosis or to the transformation of the cell through activation of cell signaling or inactivation of inhibitory mechanisms.[7] Transformation of normal cells into malignant tumors is a multistep process involving step-wise acquisition of chromosomal and cellular alterations[7] that suggests simultaneous occurrence of alterations. Analyses of retroviral

insertional mutagenesis screens in mice revealed preferential co-mutations of specific combinations of genes that causes uncontrolled cell proliferation and eventually leading to tumor formation.[25] It has been widely accepted that alterations in proteins within the same signaling pathway tend to be mutually exclusive.[7]

1.3. Oncogenes and Tumor Suppressor Genes

Genes that control cell growth and proliferation in normal cells can be grouped into two classes: proto-oncogenes and tumor suppressor genes (TSG). Whereas specific alterations in proto-oncogenes lead to activation of their function, alterations in TSGs lead to their inactivation (Figure 2).[7,26]

Proto-oncogenes often encode proteins that stimulate cell division and decrease or inhibit cell differentiation and cell death by receiving and processing growth-stimulatory signals that originate in the extracellular matrix.[7,26] Proto-oncogenes have been identified at all levels of all signal transduction cascades and include receptor growth factors (e.g., EGF, VEGF), receptor and non-receptor tyrosine kinases (e.g., ALK, EGFR and VEGFR; and ABL1, respectively), serine-threonine kinases (e.g., Akt and B-Raf), G-proteins (e.g., Ras), and transcription factors (e.g., Myc). Activating alterations in proto-oncogenes lead to continuous activation or increased production of the gene products and thus to an enhancement of the malignant phenotype.[7] Mutated proto-oncogenes are called oncogenes. Alterations involve substitutions of single or multiple nucleotides (e.g., mutations in codon 12 and 13 in the *RAS* genes in different cancers as well as mutations in the ELREA motif in *EGFR* in lung tumors[18]), gene amplifications (e.g., *ERBB2* amplification in breast cancer[7,27]) or chromosomal rearrangements (e.g., *ALK* rearrangements in lung cancer[28]). Those alterations are typically dominant in nature, meaning an alteration in only one allele is sufficient in order to observe the mutant phenotype (Figure 2).[7] Thus, different mechanisms can lead to an activation of proto-oncogenes. For example, alterations in receptor proteins may lead to ligand-independent dimerization and thereby constitutive activation of downstream signaling. Modifications in the extracellular domain can lead to stable

dimerization of two transmembrane receptors such as FGFR harboring a corresponding alteration. The fibroblast growth factor (FGF) receptor family consists of four transmembrane tyrosine kinase receptors (FGFR1, FGFR2, FGFR3, and FGFR4) that are activated by ligand dependent dimerization leading to a conformational shift and thereby intracellular transphosphorylation of the tyrosine kinase domains.[29] Aberrant FGF signaling has been described in several cancers, but alterations in the receptor are most common in bladder cancer, where approximately 50% of the tumors harbor a mutation in *FGFR3*.[29] The majority of those alterations occur in the extracellular domain of the receptor with the substitution of serine with cysteine at amino acid (AA) position 249 being the most prevalent. This alteration leads to ligand independent constitutive dimerization through formation of an intermolecular cysteine disulfide bridge and thereby activation of the receptor.[29] Also alterations in the intracellular domain of receptors, such as the kinase domain of EGF receptor, can constitutively activate the protein. Such activating mutations mainly occur in the adenosine triphosphate (ATP)-binding pocket of the protein.[30] About 90% of alterations found in EGFR affected the amino acid sequence ELREA (746 to 750) and a leucine at position 858.[31] Mutant EGFR has multiple tyrosine residues phosphorylated at the C-terminus in the absence of the ligand,[32] thereby leading to the constitutive activation of downstream signaling without extracellular stimulation. *ERBB2* is closely related to *EGFR* and frequently overexpressed, for example in breast cancer, leading to expression levels that may be 100 times higher than in normal cells which promotes spontaneous receptor dimerization in the absence of a ligand.[33,34] *KRAS* activating mutations occur in codon 12 or 13 that encode for a protein fragment adjacent to the GDP/GTP binding pocket and thereby negatively regulate the function of the protein; mutations in this region consequently lock the K-Ras protein in the active state. In normal cells transmission of growth-promoting signals from cell surface receptors through Ras is balanced through binding of GTP (active state) and hydrolysis of GTP to GDP (inactive state). If mutated, its ability to hydrolyze GTP is inactivated and downstream signaling thus remains activated independent of an extracellular stimulus, which usually initiates the signaling cascade. Proto-oncogenes can also be activated through fusion of DNA-sequences of

two different genes that can lead to a consecutive activation of the gene product. One example is the fusion between DNA-sequence encoding for the N-terminus of EML4 with the sequence encoding for the receptor tyrosine kinase domain of ALK. EML4 mediates the ligand-independent homodimerization and thus activation of the gene product even in cells where ALK is normally not expressed or active.[35,36]

Alterations in TSGs on the other hand mainly lead to loss of function and (following the Knudson 2 hit hypothesis) most alterations occurring in these genes are recessive; both alleles need to be mutated in order to completely inactivate the function of the gene.[7] TSGs control the cell cycle in normal cells through stimulation of differentiation or inhibition of proliferation of the cell (e.g., pRb, CDKN2A and p53), stimulation of apoptosis (e.g., p53), and they are also involved in DNA damage recognition and DNA repair (e.g., BRCA1 and 2, and p53). Even though TSGs involved in DNA repair do not directly control cell proliferation, cells that compromise their ability to repair DNA damages can acquire additional mutations in other TSGs or proto-oncogenes.

To inactivate a TSG either the whole chromosomal region needs to be deleted in both chromatids during mitosis or two independent somatic alterations are required. Often a mutation in one copy of the gene occurs sporadically. The other gene copy is then either lost through mitotic recombination (i.e. recombination of the chromosome arm carrying the mutation and the one carrying the wild type allele during mitosis eventually leading to a daughter cell lacking both wild type alleles), gene conversion (i.e. chromosome with the wild type allele uses the other chromosome with mutant allele as a template during DNA replication) or loss of the entire chromosome carrying the wild type allele during mitosis. Mitotic recombination or chromosomal losses occur between one and two magnitudes more frequently per cell generation than mutations and are therefore more likely to contribute to alteration of the second allele than another mutation.[7] The two-hit hypothesis was postulated in 1971 to explain symptoms of inherited and sporadic retinoblastoma, where children develop tumors in one eye only or in both eyes in association to family history of the disease.[37] Germline mutations in the TSG *RB1* predispose individuals to the development of tumors in both eyes early in life and they have a high risk of developing other tumors later in life. Sporadic *RB1*

mutations on the other hand may cause one tumor in one eye only and surgery can cure the patient.[7] In individuals who inherited one mutant allele, the gene can be inactivated by mutation of the normal allele whereas without an inherited mutation two mutations need to occur in the same cell in order to inactivate the gene. Since mutation rate is equal for both alleles, inactivation of the gene with an inherited mutant allele occurs more frequent than inactivation by mutation of both alleles in the normal gene.

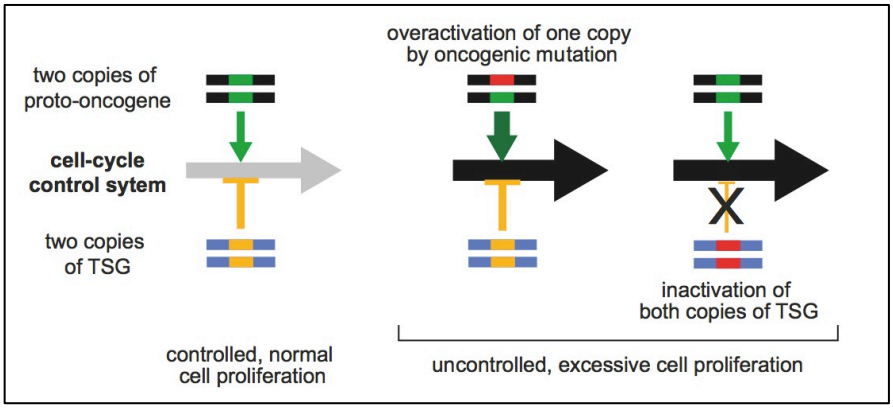

Figure 2 Proto-oncogenes and tumor suppressor genes regulate the cell proliferation. An alteration in one copy of the proto-oncogene or inactivation of both copies of a tumor suppressor gene (TSG) can cause excessive cell proliferation and can therefore contribute to cancer. (Picture adapted from Alberts. Molecular Biology of the Cell. 5th edition)

Whereas proto-oncogenes can be activated by alterations only at limited positions within the gene, TSGs can be inactivated through alterations of any kind at almost every site in the gene.[7] As aforementioned, there are several types of alterations that can affect single nucleotides, whole genes, long chromosomal regions, chromosome arms or complete chromosomes.[7]

2. Somatic genomic alterations

Somatic mutations are changes in the sequence of the genome that result from unrepaired DNA damage and occur in only those cells, which do not belong to the germline reproductive cell system; somatic mutations will therefore not be inherited to the next generation. Few germ-line mutations had been described that

are associated with specific patterns of tumor formation. Often these alterations affect TSGs or weak oncogenes that allow normal development but predispose to tumor formation often in early years as for instance in retinoblastoma (see above).

Somatic alterations that confer selective growth advantage on cells underlie all cancers.[7] In normal cells genetic alterations will be corrected or cells undergo apoptosis when alterations are severe, thus only specific alterations that will be tolerated and alter the function of gene products contribute to transformation of the cell. Recurrence of alterations across and within certain cancer subtypes suggests a selective advantage for such events. During transformation cells lose cell-cycle control and precision of DNA replication that might change the chromosomal structure and copy number and cause mutations in the genome. The most common alterations include nucleotide substitution mutations and small insertion and deletions, copy number gains and losses, and chromosomal rearrangements.[38]

Somatic mutations can occur spontaneously through replication and recombination mistakes during mitosis or can be induced by radiation or chemical mutagens that either directly modify nucleotides or cause errors during replication. Changes of nucleotides (i.e., the exchange, deletion or insertion of certain nucleotides) in the coding sequence of a gene can cause changes in the amino acid (AA) sequence that - if the alteration does not result in failure of transcription - might alter the structure and consequently the function of the gene product. Mutations can affect single or multiple nucleotides. Single nucleotide changes that alter the AA are called non-synonymous mutations (further divided into missense mutations, if the nucleotide change causes a change of the AA and nonsense mutations, if resulting AA encodes for a stop codon that causes premature termination of the protein during translation). Mutations in the coding region that do not cause changes in the amino acid sequence are called synonymous or silent. It is believed that synonymous mutations do not contribute to any structural changes of the gene product but can influence gene expression through altered folding of the mRNA during transcription.[39] Deletions and insertions in coding regions that do not comprise a multitude of three nucleotides and thereby shift the coding sequence

after the affected nucleotide(s) are known as frameshift alterations. These alterations change the complete AA sequence and increase the probability of introducing a stop codon that again will lead to a premature transcription stop and abnormal gene products.[40] TSG often harbor frameshift or nonsense mutations that disrupt the function of the gene product whereas these types of alterations are rarely found in oncogenes. Most oncogenes harbor missense mutations affecting only specific AAs (so-called hotspots) mostly in regulatory regions of the protein that cause continuous activity of the mutated protein.[7,41] In KRAS, for example, oncogenic missense mutations occur mainly in two consecutive glycine residues at AA position 12 and 13. *EGFR* is frequently altered through small in-frame deletions encompassing the five AA from codons 746 to 750 (exon 19 ELREA motif) or the missense mutation resulting in a substitution of leucine with arginine at codon 858 in the kinase domain leading to a constitutively active mutant protein.[42]

Genome copy number changes through amplifications, deletions, chromosome loss or duplication as well as changes in gene and chromosome structure through chromosomal translocation, inversion or other rearrangements, mainly occur during recombination due to unequal crossing-over, exon-shuffling (one exon from gene 1 is inserted into gene 2) or transposition (transfer of a DNA segment from one site to another on the same or different chromosome).[7] Gene amplification may lead to increased expression of the gene product and by this to a selective advantage during cell growth (e.g., amplification in *EGFR* in lung cancer or *ERBB2* in breast cancer are associated with increased expression of the respective protein).[34,43,44] Copy number amplification is a frequent event which causes increased gene expression of oncogenes, whereas deletions rather inactivate tumor suppressor genes.[38,44]

Structural rearrangements in the genome are caused by breakage of the DNA double helices and rejoining of chromosomal ends at different locations within the same or on a different chromosome. Translocations can generate fusion genes that encode proteins with novel properties. For example, an inversion on chromosome 2 brings the coding region of the regulatory domain of the *EML4* gene in close

proximity to the gene region of the proto-oncogene *ALK* encoding for the catalytic domain. The resulting chimeric fusion protein is constitutively active. Some proto-oncogenes are involved in different translocations, for example, several fusion partner for the kinases ALK or ROS1 have been identified in different cancer types resulting in an increased kinase activity.[45,46]

3. Lung Cancer

Lung cancer is worldwide the most commonly diagnosed cancer type in men and the fourth in women, with 1.1 and 0.52 million new cases registered in 2008.[47] Lung cancer rates have reached a plateau for men in many regions of the world but continue to rise in women. Differences in lung cancer incidences are mainly a consequence of gender disparity of cigarette smoking, which is the number one risk factor for lung cancer[5] and may also explain the higher percentage of never-smokers in female lung cancer patients. Furthermore, epidemiologic associations explain the prevalence of lung cancer but these are not well defined yet.[6]

Overall lung cancer is the number one cause of cancer-related deaths accounting for 1.38 million deaths annually.[5] Only half of all lung cancer patients will survive for one year after diagnosis and only 15% are alive after five years.[6] Survival time depends on growth of the tumor and whether it is local or has spread from its original location. In general, survival is better the earlier the tumor is detected. Only 15% of the lung cancer cases are diagnosed at an early stage with a localized tumor when five-year survival rate is greater than 50%.[6] Whereas in the past 40 years for many other cancer types median survival time improved due to a combination of preventive care, early detection and better treatments (e.g., colon and breast cancer), survival rates in lung cancer patients improved barely (from 11 to 20 weeks with standard therapies).[48]

3.1. Histology and Classification

Tumors of the lung can arise from the epithelium, lymphs, mesothelium or soft tissue. Epithelial tumors (carcinoma) are most common and exhibit the greatest diversity. They can be divided into four major types: adenocarcinoma, large cell

carcinoma, squamous cell carcinoma, and small cell carcinoma. Mixed tumors exist and are mainly classified according to the most predominant cell type found. Small cell cancers (SCLC) account for about 15% of lung cancer cases and are clinically distinguished from others, because they respond well to chemotherapy. Tumors of the other three types, combined into non-small cell carcinomas, are removed surgically if possible. Carcinoids represent a fifth class that comprises benign tumors. Resected lung tumors are classified based on morphology according to the WHO criteria, supported by immunohistochemistry in case of poorly differentiated tumor samples.[2]

Non-small cell carcinoma

Non-small cell lung cancers (NSCLC) are staged (I - IV) according to established international criteria and grouped based on the TNM classification system combining tumor size and location (T), lymph node involvement (N) and presence of distant metastasis (M). In general, higher stage (overall and within each category) correlates with a poorer prognosis with a five-year survival of about 60% for stage I (T1-2, N0, M0; primary tumor has not spread to other sites) and less than 5% for stage IV (TX, NX, M1) disease.[5,6] The major aim of staging is to select cases suitable for surgery, which offers the highest possibility for a cure. Furthermore, staging helps to identify those candidates eligible for chemotherapy and radiotherapy and specific treatment options that are approved for advanced stage non-small cell lung cancer. The term non-small cell lung cancer does not reveal taxonomic precision in diagnosing these tumors and should therefor be avoided.

Adenocarcinoma

The proportion of adenocarcinoma (AD) in lung cancer has increased from approximately 5% in the 1950th to about 40% and is now the most common type of NSCLC. This trend can be explained by the fact that filtered cigarettes became available and people inhaled deeper into their lungs.[49] ADs occur mainly peripheral in the lung and in order to further distinguish them from the morphologically

similar metastatic AD the transcription factor TTF-1, which is expressed in the alveolar epithelium in the periphery of the lung, serves as a useful marker. ADs are characterized by either mucus formation, which may be discrete or intracellular (solid adenocarcinoma) or by distinct growth patterns such as acinar growth, papillar differentiation or a single-layer spread along bronchioles (characteristic for bronchioloalveolar carcinoma (BAC)) and the alveolar septum. They show frequent histologic heterogeneity that are more common than tumors consisting purely of a single pattern of solid adenocarcinoma with mucin formation, acinar, papillary or bronchioloalveolar adenocarcinoma.[2] BACs are characterized by proliferation of type II pneumocytes and/or clara cells in a lepidic fashion along the alveolar walls without destruction of lung architecture and are, in cases with predominant lepidic growth pattern, associated with 100% or near 100% disease-free survival after complete resection.[50-52] Patients with predominantly acinar and papillary or solid adenocarcinoma have a five-year survival of about 70% and 40%, respectively.[52]

The 2004 WHO Classification of adenocarcinoma has been revised in 2011 by a multidisciplinary expert panel representing the IASLC, the American Thoracic Society and the European Respiratory Society. The major adjustment involved the diagnosis of BAC, because the criteria for this classification were not clear and could range from pure BAC to mixed adenocarcinoma with minor components of BAC pattern. Prognosis therefore widely ranged from a 100% to a 10% five-year survival rate.[53] BAC is now subdivided into categories that correlate with clinical outcome ((a) adenocarcinoma in situ (noninvasive, lepidic pattern), (b) minimally invasive adenocarcinoma (predominantly lepidic pattern and invasion of less than 5 mm into the surrounding tissue), (3) lepidic predominant adenocarcinoma, (4) invasive (mucinous or nonmucinous, invasion of more than 5 mm)). Further micropapillary adenocarcinoma, which was not part of the 2004 WHO Classification, emerged as an important variant of papillary adenocarcinoma.[51,54] Pictures of hematoxylin-eosin stained tumor sections in Figure 3 demonstrate characteristic features of the five major subtypes of adenocarcinoma that were first presented at the 13th World Lung Conference in San Francisco, USA in 2009[55] and published by Travis *et al.* in 2011.

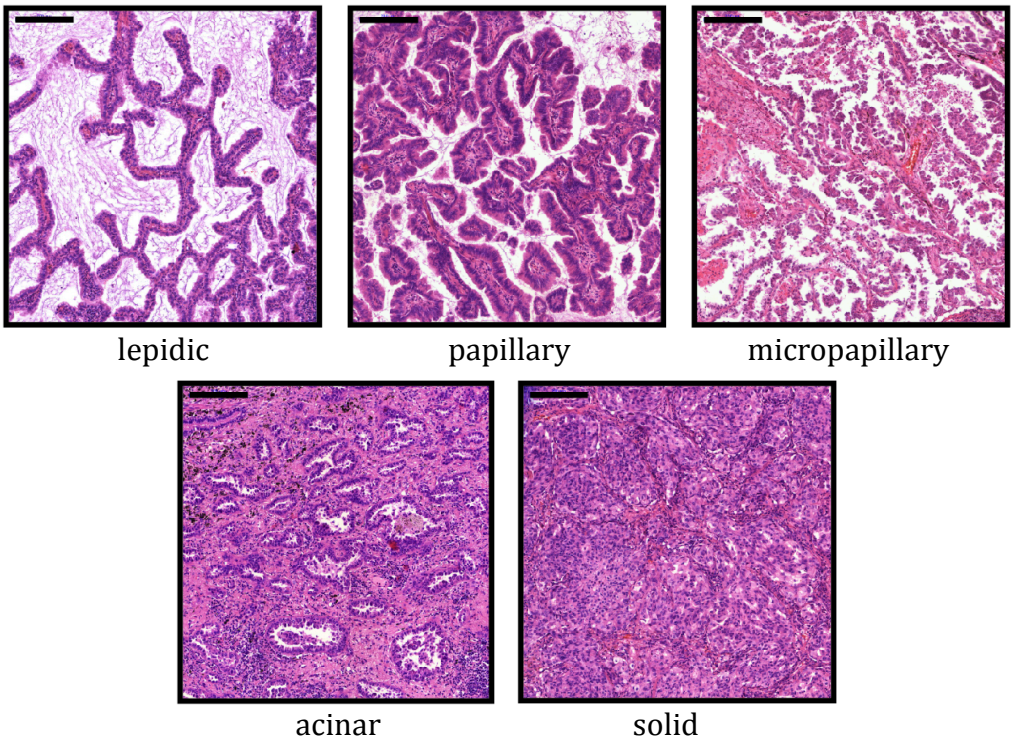

Figure 3 Morphology of histological subtypes of adenocarcinoma. Lepidic: neoplastic cells grow along preexisting alveolar structures, Papillary: glandular cells grow along central fibrovascular cores. Micropapillary: small papillary tufts of tumor cells that lack fibrovascular cores lying apparently free in alveolar spaces, Acinar: tumor cells surround round- to oval-shaped neoplastic glands with a central luminal space, Solid: polygonal tumor cells form sheets without recognizable patterns of adenocarcinoma. (hematoxylin-eosin, black bars correspond to 200 µm)

Large cell carcinoma

Traditionally large cell carcinoma (LCC) is an exclusion diagnosis. The term refers to a barely differentiated, non–small cell cancer, in which neither the characteristics of adenocarcinoma nor those of squamous or small cell carcinoma are detectable. They account for approximately 10% of all lung tumors, of which about one third are large cell carcinoma with neuroendocrine differentiation (LCNEC) that have a poorer prognosis than the other large cell carcinoma. LCNECs are characterized by generally large tumor cells that grow in distinct organoid, trabecular or insular growth patterns, a prominent nucleoli and high mitotic rates (75 mitoses per $2mm^2$). One of the neuroendocrine immunohistochemical markers (CD56, Synaptophysin and Chromogranin) is sufficient for confirmation of this diagnosis. Half of the LCNEC cases express TTF-1.[2]

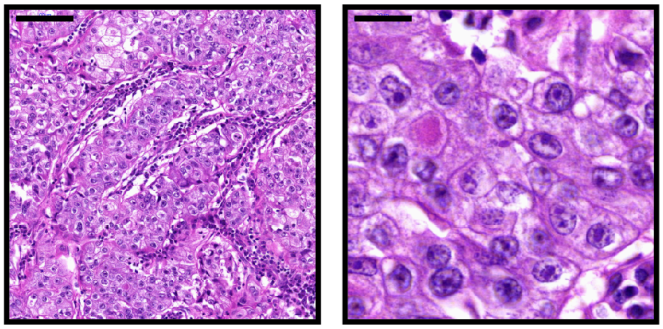

Figure 4 Large cell carcinoma of the lung. This lesion does not exhibit specific differentiation (left). Cells have abundant cytoplasm with large nuclei and prominent nucleoli (right). (hematoxylin-eosin, black bars correspond to 200 µm (left) and 20 µm (right))

Squamous cell carcinoma

Squamous cell carcinomas (SQ) mainly arise centrally in the main stem, lobar and segmental bronchi and are characterized by intra-cytoplasmatic keratinization or intercellular bridges. The majority of squamous cell carcinomas express p63 and cytokeratins 5/6 and very few express thyroid transcription factor-1 (TTF-1) or cytokeratin 7 (CK7).

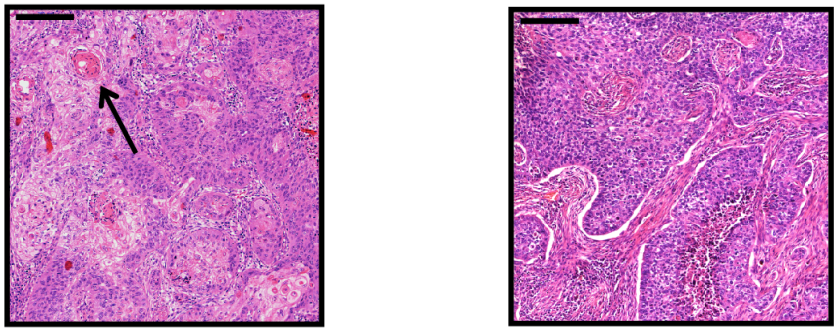

Figure 5 Classical morphological features of squamous cell carcinoma of the lung. Left: keratinizing, arrow indicates keratinized cells forming cell pearls. Right: non-keratinizing: no clear keratinization, pearls or intercellular bridges are present. (hematoxylin-eosin, original magnification 20x)

Small cell carcinoma

Small cell carcinomas (SCLC) are classified into limited (restricted to one hemithorax, LD) or extensive disease (with distant metastasis, ED) with a five-year survival of 5-10% and less than 5%, respectively.[6] The American Joint Commission

on Cancer TNM staging system is less frequently applied for SCLC because treatment options do not differ much between these detailed stages. Without treatment median survival is only 2 to 4 months but can be up to 24 for limited disease and up to 9 months for extensive disease when applying currently available first-line therapies.[56] Approximately 70% of SCLC cases are metastatic at presentation,[56] which limits localized treatment such as surgery or radiation.

Tumor cells of SCLC are characterized by a high nuclear to cytoplasmic rate, granular nuclear chromatin (salt and pepper chromatin pattern), absent nucleoli and nuclear molding, where nuclei conform to one another. In most cases immunohistochemistry is positive for TTF-1, CD56, Synaptophysin, and Chromogranin.[2]

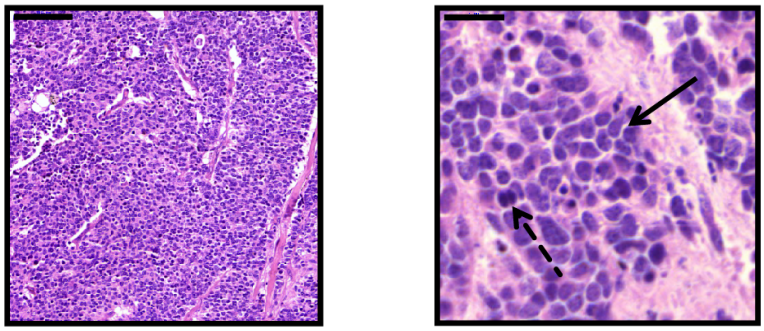

Figure 6 One small cell carcinoma case. Right: A magnified view shows the typical cytological features of small cell carcinoma: no nucleoli, „salt and pepper" chromatin pattern, nuclear molding (arrow), and high mitotic activity (mitotic cell (dashed arrow). (hematoxylin-eosin, black bars correspond to 200 µm (left) and 20 µm (right))

Carcinoid

Carcinoids (CA) are rare neuroendocrine lung tumors that account for approximately 2% of newly diagnosed cases. They arise from neurosecretory cells of bronchial mucosa and are classified into typical and atypical carcinoids with a five-year survival of greater than 90% and 50-60%, respectively. Nine out of 10 newly diagnosed carcinoid cases are typical. These tumors are less aggressively malignant than atypical carcinoids and a five-year survival is observed in almost 100% of the cases. At presentation 10-15% of typical and 40-50% of atypical carcinoids have metastasized to regional lymph nodes and beyond. Carcinoids are

characterized by growth of uniform polygonal cells with granular chromatin, inconspicuous small uniformly round nucleoli and marginal to moderate amount of cytoplasm that are arranged in distinct organoid, trabecular or insular growth patterns. Tumors with typical morphology have less than 2 mitosis per 2 mm^2 and atypical 2 to 10 mitosis per 2 mm^2. Only atypical carcinoids exhibit foci of necrosis.[2] Histological heterogeneity of carcinoids and other subtypes is rare. More than 80% of carcinoids are positive for cytokeratins and for at least one neuroendocrine marker.[2]

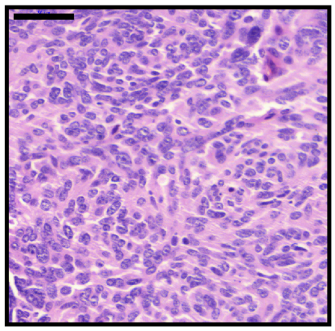

Figure 7 Typical carcinoid. (hematoxylin-eosin, black bars correspond to 50 µm)

Rare lung tumors

Carcinomas with pleomorphic, sarcomatoid or sarcomatous elements that contain a sarcoma (malignant bone, cartilage or skeletal muscle) or sarcoma-like (spindle and/or giant cell) component are globally rare and account for 0.1-0.4% of all lung malignancies. They are highly correlated with smoking history and pursue an aggressive clinical course.[57,58] The latest WHO Classification of the Lung requires a minimum of 10% sarcomatoid pattern (spindle and/or giant cells) for a positive diagnosis.[2] Despite several adaptions in the classification system, diagnosis of such cases remains difficult.

3.2. Treatment options of lung cancer

Lung cancer is mostly diagnosed at an advanced stage when prognosis is poor and curative treatment options are limited.[5,6] Prognosis and treatment decisions

depend on the stage and subtype but also on the patient's age and general health. In early stage lung tumors surgery is the most promising therapy. Adjuvant chemotherapy improves progression-free survival by eliminating remaining cancer cells and metastasized cancer cells in brain, liver or bone.[59] Chemotherapeutics are antineoplastic agents that inhibit DNA replication by various mechanisms and thereby kill the cells. These agents are toxic for every cell and since they are systemically applied, they cause several severe side effects, such as suppression of the immune system, reduction of white and red blood cells, and oral and gastrointestinal mucositis. Tumor therapy can also involve radiotherapy, either by specifically targeting the tumor using high-energy radiation beams or radioactive substances that are delivered to the whole body.

Patients with SCLC are treated with chemotherapy and radiation or chemotherapy alone, if the tumor has spread to other sites. Despite a response rate of more than 80% for limited disease and up to 80% for extensive disease, the cancer almost always returns and patients eventually die within two years.[60] Patients with early stage NSCLC undergo surgery and may receive additional chemotherapy and/or radiation therapy whereas late stage NSCLC patients are only treated with chemotherapy and/or radiation therapy. For therapeutic purposes NSCLC has traditionally been considered as a single disease separated from SCLC. Advances in thoracic medical oncology have challenged the relevance of these diagnostic categories since molecules expressed in tumors have been associated with enhanced response to certain cytotoxic agents. Two new agents, namely bevacizumab (an antibody against the vascular epithelia growth factor) and pemetrexed (an antifolate that inhibits thymidylate synthetase (TS), an enzyme involved in purine and pyrimidine synthesis) caused severe bleeding or had no effect in patients with SQ, respectively. Inhibition of TS with pemetrexed improved PFS and OS in AD and LCC but not in SQ (12.6 vs. 9.4 months).[61] High baseline expression of TS in SQ compared to AD is thought to be the reason for the low efficacy of pemetrexed in SQ.[62]

Through the identification of genetic alterations in *EGFR* and *ALK* in lung AD that confer susceptibility to therapeutic agents and *KRAS* mutations that were associated with resistance to these agents, it became inevitable to identify patient

populations to guide treatment decisions. Today testing for *EGFR* mutations and *ALK* rearrangements is considered to be the standard of care in advanced-stage lung ADs to identify patients suitable for treatment with tyrosine kinase inhibitors.[63] Guidelines recommend testing for these alterations in AD and carcinomas admixed with an AD component, because alterations in *EGFR* and *ALK* segregated with this subtype.[63,64] Pathologists therefore identify ADs - using immunohistochemistry if required. The *KRAS* mutation status is often evaluated to exclude patients from treatment with tyrosine kinase inhibitors (TKI). Almost 100% of mutations identified in *KRAS* in lung cancer occur in exons 2 and 3, thus, sequencing only two exons is sufficient to identify almost one third of AD patients who are not suitable for treatment with TKIs. This strategy has proven to be successful in clinical routine.[65] The importance of genetically stratifying patients with great care has been demonstrated in AD cases where the response rate to erlotinib in unselected patients was 9% and median time to progression was three months.[66] By contrast, advanced NSCLC patients who were selected according to their *EGFR* mutation status showed a response rate of 68% and mean time to progression was 12 months upon treatment with TKI[67]. Furthermore, it could be shown that first-line treatment with TKIs was superior to standard chemotherapy in *EGFR*-mutant AD cases,[4] which had been confirmed in independent randomized clinical trials.[68,69] Patients with *ALK*-rearranged AD who were treated with crizotinib showed a response rate of 57% with a progression free survival of at least six months in 72% of the cases.[70] Similarly promising response rates to targeted therapies were recently reported in patients harboring genetic alterations in *DDR2* or *FGFR* mainly found in lung SQ, as well as in *RET* and *ROS* identified in lung AD.[71-73] In several currently active clinical trials response to new tyrosine kinase inhibitors is tested. For example, BGJ398 that has been developed to selectively target FGFR is tested in a phase I trial in patients with *FGFR1*- or *FGFR2*-amplified and *FGFR3*-mutant solid tumors.[73,74] Several multi-kinase inhibitors that were approved for other molecular targets or other cancer types are tested in phase II studies such as dasatinib (approved in patients with CML) in *DDR2*-mutant SQ patients, sunitinib (approved for treatment of renal cell carcinoma) and vandetanib (approved for treatment of medullary thyroid cancer)

in *RET*-rearranged AD and NSCLC patients, respectively, and crizotinib (approved for *ALK*-rearranged NSCLC) in *ROS1*-rearranged NSCLC patients.[74,75] Many more drugs that show potential as cancer therapies in preclinical models are tested in genetically defined patient cohorts. Because FDA guidelines are adapted to accelerated drug approval (for example approval based on tumor shrinkage and not overall survival, as tumor shrinkage is assumed to reflect improved clinical outcome) an exponential increase of approved targeted therapies for genetically defined subpopulations in (lung) cancer within next few years is likely.

With advanced knowledge with regard to application of chemotherapeutics in selected patient groups[62] and the success of targeted therapy over conventional chemotherapy in genetically selected populations, it became apparent that the traditional distinction between NSCLC and SCLC is no longer sufficient for treatment decisions.

Objective of this study

Genetic mechanisms of carcinogenesis differ between tumor types depending on intrinsic and environmental cues causing the disease as well as on the cell type and site of origin.[7] Treatment of cancer patients has changed within the last decade in which targeted therapies aiming for the functional activity of certain altered proteins has proven to be successful in many tumor types including specific subtypes of lung cancer.[76] Thus, it became urgent to renew the traditional histomorphological-based classification of lung tumors to a genetically informed classification to provide genetically informed medicine.

The major goal of my thesis was to identify genomic alterations in lung cancer that

I) segregate with different histological subtypes,

II) predict clinical phenotype and

III) may serve as new targets for cancer treatment with molecule inhibitors.

Therefore primary lung tumors of all histological subtypes were annotated with genomic, histomorphological and clinical information to define a novel, biologically informed classification of lung tumors. Genomic alterations were detected on chromosomal and gene expression level. Mutations in specific genes (for example, *BRAF, EGFR, ERBB2, FGFR2, KEAP1, KRAS, NFE2L2, PIK3CA, STK11*, and *TP53)* were analyzed using direct DNA sequencing in 1,127 cases; copy number alterations were analyzed using Affymetrix SNP 6.0 arrays in 1,032 cases and gene expression using the Illumina HumanHT-12 BeadChip in 261 cases. In order to functionally characterize newly identified mutations, genes were expressed in interleukin-dependent Ba/F3 cells and NIH3T3 cells and the transforming capacity tested by interleukin-independent or anchorage-independent growth assays in the given cells. In collaboration with WHO pathologists 573 cases were pathological reviewed.

Results

1. Characteristics of the Clinical Lung Cancer Genome Project study population

Primary lung tumors were collected from different institutions from Europe and Australia with the aim to assemble an unbiased collection that authentically reflect the disease in a Caucasian population. The majority of tumors were resected before the *Clinical Lung Cancer Genome Project* (CLCGP) was initiated in 2008 when the era of personalized targeted therapy was just about to evolve. Clinical data for all cases collected for the CLCGP are summarized in Table 1.

Table 1. Summary of clinical characteristics of all specimen collected for this study and 1,255 cases that were suitable for analyses of genetic alterations and gene expression.

		CLCGP all	CLCGP analysis set	
Total number		1,882	1,255	
Age at Diagnosis				
number		1,667	1,225	
median year (min-max)		65 (18-95)	65 (18-94)	
Sex	**Female**	419 (33.5%)	419 (33.5%)	
number* (%)	**Male**	831 (66.5%)	831 (66.5%)	
	Unknown	*632*	*5*	
Histology (WHO 2004)	**Adenocarcinoma**	785 (45.2%)	537 (44.5%)	
number (%)	**Carcinoid**	89 (5.1%)	71 (6%)	
	Large-Cell Carcinoma	169 (9.7%)	129 (10.7%)	
	Small-Cell Carcinoma	97 (5.5%)	65 (5.4%)	
	Squamous-Cell Carcinoma	594 (34.2%)	403 (33.4%)	
	Other/ Unknown	**148**	**50**	
Proportion			female	male
Female and Male across	**Adenocarcinoma**		53.8%	36.5%
subtypes	**Carcinoid**		12.2%	4.1%
	Large-Cell Carcinoma		9.3%	10.7%
	Small-Cell Carcinoma		7.8%	5.4%
	Squamous-Cell Carcinoma		16.9%	43.3%
	Total number		344	647
Stage (UICC)	**IA**	385 (23.5%)	305 (25.4%)	
number* (%)	**IB**	428 (26.1%)	316 (26.3%)	
	IIA	51 (3.1%)	33 (2.8%)	
	IIB	265 (16.1%)	190 (15.8%)	
	IIIA	268 (16.3%)	195 (16.2%)	
	IIIB	127 (7.7%)	89 (7.4%)	
	IV	114 (6.9%)	74 (6.1%)	
	Unknown	**244**	**53**	
Survival	**I**	152 (31)	145 (33)	
median month	**II**	66 (32)	73 (35)	
(median observation	**III**	34 (25)	32 (30)	
time in month)	**IV**	18 (22)	17 (34)	
	Unknown	**58 (23)**	**58 (42)**	
Smoking History	**Current/ Former**	1,161 (86%)	896 (86%)	
number* (%)	**Never**	184 (14%)	147 (14%)	
	Unknown	*537*	*212*	
Proportion	**Female**		73%/ 27%	
Smoker/Never	**Male**		93%/ 7%	

Abbreviations: CLCGP (*Clinical Lung Cancer Genome Project*), WHO (World Health Organization), UICC (Union internationale contre le cancer); * including Large-Cell Neuroendocrine Carcinoma

Clinical information were available for most of the patients and included personal characteristics (age, sex and smoking status), histology, tumor stage, and overall survival. Lung cancer is a disease of the elderly[5] also reflected in our sample set by the median age at diagnosis of 65 years. The majority (66%) of patients with lung cancer were males of whom 93% had a history of smoking whereas lung cancer in women was associated with smoking in only 73% of the cases. Regarding the major histological subtypes, adenocarcinoma (AD) was the most frequently observed subtype overall (45%) and in women (54% vs. 37% in men). Squamous cell carcinoma (SQ) was the second most frequent subtype overall (34.2%) and the most frequent type of lung cancer in men (43% vs. 17% in women). Small cell carcinoma (SCLC) incidence is approximately 15%[5] but was underrepresented in our sample cohort (about 5%). Treatment recommendations for SCLC did not involve surgery therefore most SCLC cases that were included in our sample set were either mistaken for non-small cell lung cancer (NSCLC) at diagnosis or resected post-mortem. Carcinoid (CA) accounted for 5-6% of all tumors.

About 15% of all cases in this study were advanced stage tumors which did not approximate the true incidence of advanced stage lung cancer of about 60%.[6] According to current treatment recommendations lung cancer patients diagnosed in early stages (I and II) by the oncologist are considered candidates for surgery as a first therapeutic option. Surgery for stage IIIA patients with cancer metastasized to lymph nodes in the same lung or chest wall is considered reasonable as first line treatment whereas patients with advanced lung cancer (stage IIIB or IV) are considered unresectable and suitable for concurrent radiation only. Hence lung tumor of advanced stages should be available only in case of SCLC autopsies but these autopsy cases accounted for only 20% of all stage IIIB/IV cases. The remaining 80% of the late stages cases were not SCLC autopsy cases and were available due to the following reasons: Stages used in the analysis were assessed by the pathologist which can differ from those made by the oncologist who might underestimated the true stage but had given treatment recommendations for surgery, also specimen could have been resected for the purpose of diagnosis only without intention to cure by surgery. The median survival is associated with stage

at diagnosis and was longest in stage I (145 months) and lowest in stage IV (17 months) patients. The median observation time for stage I patients was 33 months.

This cohort represents the largest group of lung tumors of all histological subtypes genetically analyzed until now.

2. Central pathological review

(in collaboration with **Elisabeth Brambilla**, CHU Albert Michallon and **William D. Travis**, Memorial Sloan Kettering Cancer Center)

In this study diagnostic accuracy and uniformity was of importance to interpret results obtained from genetic analyses. Diagnostic disagreement between pathologists can have different reasons and can vary widely.[77] Therefore two lung cancer pathologists (EB and WDT) reviewed in total 615 specimens in a blinded fashion to confirm original diagnoses or reclassify tumors. Both pathologists are experts in lung pathology and are editors and authors of the WHO Classification of the Lung.[2] For 583 tumors a final diagnosis was made; 182 supported by immunohistochemistry (Figure 8). For the remaining cases lack of clear morphological characteristics and immunohistochemical staining permit interpretation. Of 535 cases original diagnosis was available.

Eleven percent of originally diagnosed ADs were reclassified to other subtypes: to LCNEC (5%), SQ (2.8%), LCC (2.3%), SCLC or sarcomatoid carcinoma (0.5% each). Reclassification of SQs to other subtypes was 13.7%: to AD, LCC or LCNEC (4.1% each), sarcomatoid carcinoma or SCLC (0.7% each). Eighty three percent of SCLC cases were confirmed (one of 15 cases admixed with LCNEC), 11% were reclassified to LCNEC and 6% to AD. Of 52 LCNECs 83% were confirmed as such (admixture with other entities in about 30% of these cases – admixed with AD (1), SQ (2) or SCLC (9)) and 15% were reclassified to SCLC. One LCNEC case was reclassified as atypical CA. Of 48 original diagnosed LCC cases (excluding LCNEC) 91% were reclassified by morphology and immunohistochemistry as described in the following: 11 cases were reclassified as AD, 13 as SQ, 1 as mixed AD and SQ, 14 as neuroendocrine carcinomas (LCNEC (n=7), LCNEC combined with SQ (n=1), SCLC (n=3), SCLC combined with LCNEC (n=2), SCLC combined with SQ (n=1)), and

4 cases as sarcomatoid carcinoma. Five LCC cases remained LCC (TTF-1 and p63 expression negative by immunohistochemistry).

The majority of AD, SQ, SCLC and LCNEC were confirmed whereas most LCC were reclassified into one of the other subtypes based on immunohistochemistry.

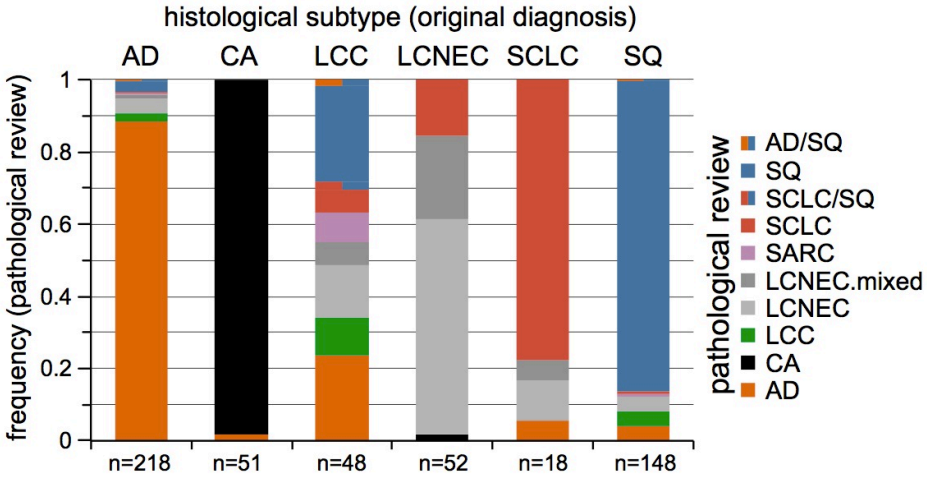

Figure 8 Results of reclassification of 535 lung tumors based on morphology and immunohistochemistry. Frequencies of reclassification are plotted for the major histological subtypes (original diagnosis). LCNEC is presented separately from other LCC. (LCNEC.mixed: LCNEC combined with SCLC (in AD, LCC or SCLC) or combined with SQ (in AD and LCC)).

3. Estimation of specimen mix-ups in this cohort

Sample mix-ups in clinical studies can arise during data management, sample collection and handling. This problem became apparent in rare cases, where predictions of sex from SNP 6.0 arrays did not match the original annotation. Frequency of sample mix-up in the CLCGP was estimated by comparing average copy numbers from SNP 6.0 array data of the gonosomes X and Y. In females only background hybridization on the Y-chromosome should be measured whereas in male signals of both chromosomes must be detected. Therefore copy number ratio X/Y must be higher in females. Calculating this ratio revealed a discrepancy between the prediction using SNP 6.0 copy number data and original annotation in 13 out of 1,032 cases. Based on this calculation the total mix-up rate was computed to 2.8%.

4. Genetic alterations in lung tumors

Known cancer genes are differentially altered in many cancer types. As for example lung tumors frequently harbor mutations in *EGFR*, alterations in this gene are quite rare in breast or skin cancer.[78] To determine frequencies of somatic mutations and chromosomal rearrangements in known cancer genes and chromosomal gains and losses across the genome for lung cancer, tumors were analyzed accordingly. Genome-wide copy number alterations were analyzed in 1,032 cases. In contrast, to assess mutation frequencies specific sites or exons in genes known to be frequently mutated (hotspots) in lung cancer were tested. Thereby in 1,127 cases 327 different mutation sites in the following 26 genes were tested using mass spectrometry: *ABL1*, *AKT2*, *ALK*, *BRAF*, *CDK4*, *DDR2*, *EGFR*, *EPHA3*, *EPHA5*, *ERBB2*, *ERBB4*, *FGFR1*, *FGFR2*, *FGFR3*, *FGFR4*, *FLT3*, *HRAS*, *JAK2*, *KDR*, *KIT*, *KRAS*, *NRAS*, *NTRK1*, *NTRK3*, *PDGFRA*, and *PIK3CA* (Table 3). Furthermore, specific exons were analyzed using Sanger sequencing in 1,127 cases (*EGFR*, *KRAS*, *STK11*, and *TP53*) or 844 cases (*BRAF*, *ERBB2*, *FGFR2*, *KEAP1*, *NFE2L2*, and *PIK3CA*) (Table 4). For 832 lung tumor cases copy number and also mutation data were available. To assess frequencies of chromosomal rearrangements affecting *ALK*, *RET* and *ROS1*, FFPE material of lung tumors was collected in an independent effort as for the primary fresh frozen cases. In total 602 cases were analyzed for rearrangements in *ALK*, 362 in *RET* and 211 in *ROS1* using fluorescence *in situ* hybridization (FISH). *ROS1* rearrangements were analyzed mainly in AD. The overlap of these cases with those that were analyzed for copy number alterations and/or mutations was 79% for *ALK*, 69% for *RET* and 64% for *ROS1*.

4.1. Somatic copy number alterations in lung tumors

Of 1,032 tumor cases with SNP data the histological subtype was available for 992. Segmented copy number data were visualized using the Integrated Genomics Viewer (www.broadinstitute.org) (Figure 9).

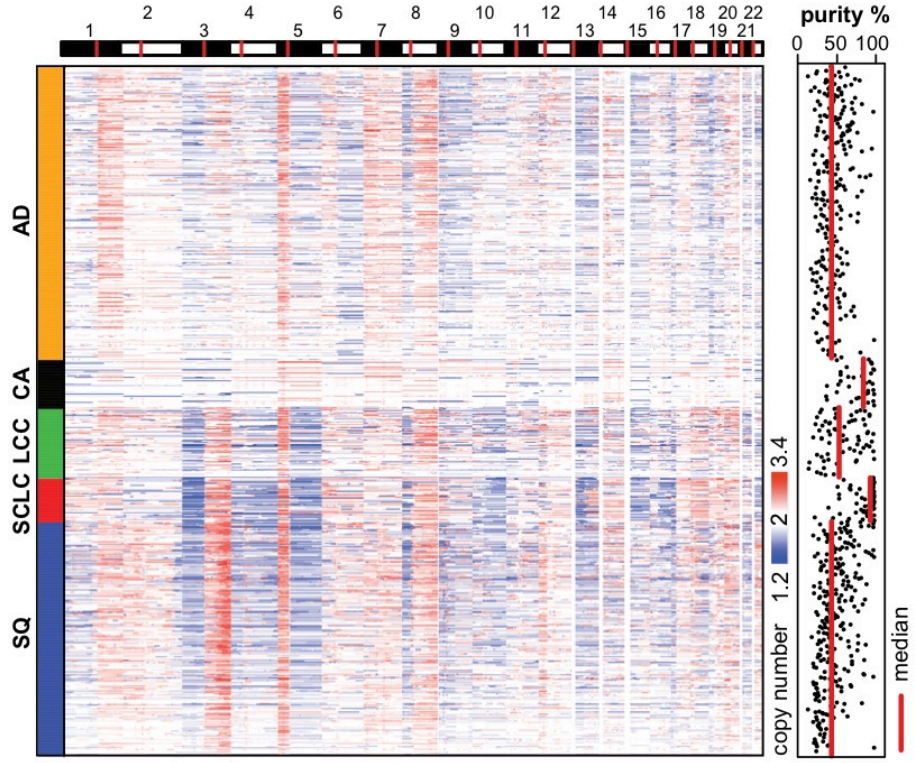

Figure 9 Genome-wide copy number alterations in 992 lung tumors. Cases are sorted according their histological subtype and copy number alterations (amplifications in red, deletions in blue) plotted along the human genome (chromosomes 1 to 22, centromers as red lines). Each row represents one case. On the right estimated purity is given for 667 samples (median calculated per histological subtype in red). AD (n=421), CA (n=69), LCC (n=101), SCLC (n=63), SQ (n=338).

Lung tumor cases sorted according their histological subtype revealed distinct patterns of somatic copy number alterations (SCNA) (Figure 9), where ADs showed less frequently high amplitude alterations compared to SCLCs and SQs. Only few CA cases had copy number changes mainly affecting whole chromosomal arms or chromosomes. CA and SCLC, both neuroendocrine tumor types, showed high purity (median purity 86% and 95%, respectively), meaning that admixture with non-tumor cells (such as fibroblasts and lymphocytes) in the extracted tissue was low. Purity was similar for AD, LCC and SQ (median purity 44%, 54% and 43%, respectively). Within each subtype cases with low purity also showed low

amplitude copy number alterations (mainly in the lower part within each subtype) compared to those that were highly pure in regard to tumor cells.

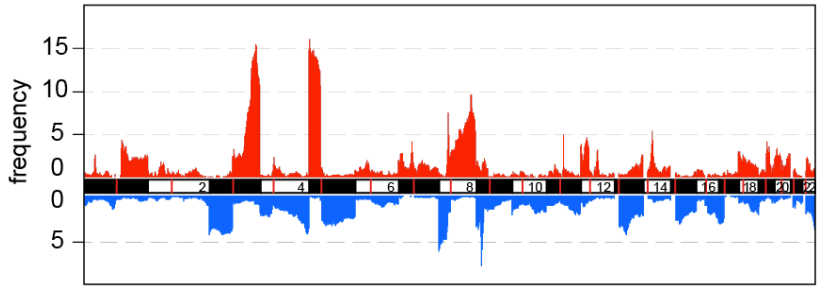

Figure 10. Frequencies of copy number gains (red, cutoff 3.0) and losses (blue, cutoff 1.3) across 1,032 lung tumor cases calculated for adjoining 1 MB fragments using segmented copy number data are represented along the genome (chromosomes 1 to 22 in the middle, centromers in red).

Examining copy number alterations across all tumor samples revealed that certain regions were more often amplified than deleted and vice versa. Copy number gains (cutoff 3) and losses (cutoff 1.3) were calculated using median copy number of adjoining 1 MB fragments across the genome. As illustrated in Figure 10 frequencies of amplifications and deletions differed between chromosomal regions across all tumor samples. For example 1q, 3q, 5p, 8q and focal regions on 11q, 12q and 14q were frequently amplified but infrequently deleted, whereas 3p, 5q, 8p, 9p, 15q, and 16q were mainly deleted but rarely amplified.

In order to identify significantly altered chromosomal regions across lung tumors, a rank sum-based algorithm was applied to segmented copy number data of 1,032 cases. In contrast to other existing methods this algorithm is not sensitive to tumor purity (i.e. admixture with non-cancerous cells) that could otherwise masks the amplitude of copy number alterations and thereby leading to underestimation of true frequencies across lung tumors. In brief, ranks were assigned to genomic positions using raw copy number for each tumor case individually. Upper and lower quantiles of these ranks defined between amplified and deleted regions, respectively (the more narrow the quantiles, the more focal the peak). Further, to identify significantly altered regions, ranks were summed up

for each genomic location. Finally, multiple hypothesis testing was taken into account by using the Benjamini-Hochberg method.

Table 2 Significant copy number alterations identified in 1,032 lung tumors. (abbreviations: Amp=Amplification, Del=Deletion, Chr.=Chromosome)

Region_name	Type	Cytoband	Chr.	Start	End	Size_Mb
1p34 (*MYCL1*)	Amp	1p34.2	1	39649410	40611223	0.96
3q26-29 (*SOX2*)	Amp	3q26.1-3q29	3	162577624	196416603	33.84
5p15 (*PRDM9*)	Amp	5p15.2-5p15.1	5	9268161	36668232	27.40
7p11 (*EGFR*)	Amp	7p11.2	7	54354000	55628997	1.27
8p12-11 (*FGFR1*)	Amp	8p12-8p11	8	36630445	39731149	3.10
8q24 (*MYC*)	Amp	8q24.21	8	127899488	130018973	2.12
11q13 (*CCND1*)	Amp	11q13.2-11q13.3	11	68617234	69944674	1.33
14q13 (*NKX2-1*)	Amp	14q13.2-14q13.3	14	35533908	36619261	1.09
1p13	Del	1p13.3-1p13.2	1	110801228	116635628	5.83
3p26	Del	3p26.3-3p26.2	3	35346	11252648	11.22
4q34-35	Del	4q34.2-4q34.3	4	177144874	191261905	14.12
5q15-21	Del	5q15-5q21.1	5	94179269	97962777	3.78
6p22-21	Del	6p22.1-6p21.33	6	28842983	33440115	4.60
8p23-21	Del	8p23.3-8p21.3	8	21255	32807956	32.79
9p24-22 (*PTPRD*)	Del	9p24.3-9p22.3	9	374058	19436252	19.06
9p21 (*CDKN2A*)	Del	9p21.3	9	21578969	22209276	0.63
16q23	Del	16q23.2-16q23.3	16	78712684	87607274	8.89
18q21	Del	18q21.31-18q21.32	18	52682781	55807278	3.12
19p13	Del	19p13.3	19	41911	3314202	3.27
22q13	Del	22q13.32-22q13.33	22	47039194	49581322	2.54

Using this approach eight amplified and twelve deleted regions were identified across all lung tumors (Table 2). Size of amplified regions ranged from approximately 1 Mb including six genes to almost 34 Mb engulfing 146 genes. The eight amplified regions contained a median of 15 genes. Deleted regions ranged in size from 0.63 Mb (9p21 engulfing the tumor suppressor gene *CDKN2A*) to approximately 32 Mb (182 genes) with a median of 55 genes in the 12 deleted regions. At least one functionally validated oncogene was identified in each of the eight amplified regions, for example *SOX2* in 3p,[79,80] *FGFR1* in 8p[81,82] and *NKX2-1* in 14q[83] (Table 2). Except for 1p and 8p significantly amplified and deleted regions were identified on different chromosomal arms.

The region 9p21 engulfing the tumor suppressor gene *CDKN2A* was the most frequently deleted region across all lung tumors (Figure 11). In cases with a deletion in one of the twelve deleted regions, 9p21 was affected in approximately

40% of the cases. The second and third most frequently deleted regions in cases harboring only one deletion were 8p23-21 and 22q13 respectively (each occurred in approximately 13% of the cases). Deletions in 3p were frequent across all lung tumors but were rarely found as a single deletion event in tumors (< 4%). The most frequently affected region in cases that had multiple deletions was 3p (in 15.5%), followed by 9p21 (14%) and 8p23-21 (12%).

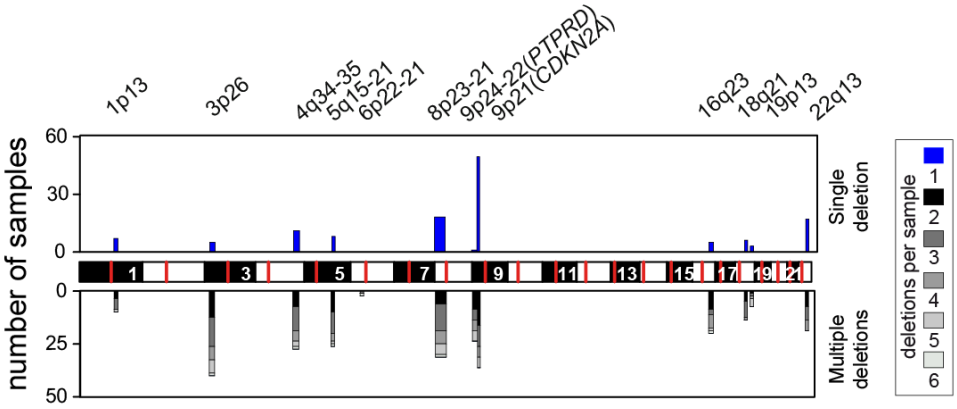

Figure 11 Incidence of deletions in lung tumors within twelve significant deleted regions. Median copy number was calculated for significantly deleted regions (Table 2) per case. Copy number 1.3 was used as the cutoff to determine tumor cases with deletion in the particular region. Cases were counted according occurrence of deletions in one or more regions. Width of the bars illustrates length of the regions on the chromosome.

Overall, chromosomal losses through deletion were more frequent than gains through amplification in lung tumors. Regions that were frequently amplified were rarely deleted and vice versa. Whereas in all significantly amplified regions at least one oncogene was identified, known cancer genes were not found in any of the deleted regions.

4.2. The mutation spectrum in lung tumors

Across all lung tumors *KRAS* was the most frequently mutated gene (16.2%), followed by *EGFR* (7.2%), *KEAP1* (6.7%), *PIK3CA* (5.2%), *NFE2L2* (4.5%), and *BRAF* (2%) (Figure 12). Of note, due to recent findings that suggest KEAP1 (the negative regulator of NFE2L2) to be a tumor suppressor[84] it will be mentioned as such throughout this dissertation. *ROS1* rearrangements were detected in 3 ADs (overall

frequency 1.4%). Two of the three cases with a *ROS1* fusion were male, both with a history of smoking and median age at diagnosis was 68 years (see Appendix, Table 7). Ten cases with an *ALK* fusion were identified. In six out of eight cases that were available for further FISH analysis *EML4* was determined as the fusion partner of *ALK*. The median age at diagnosis in *ALK* positive ADs was 56 vs. 64 years in the total patient cohort (see Appendix, Table 7). In this cohort no case with a *RET* fusion was identified (0/ 362). *TP53*, which is the most frequently mutated TSG in cancer[86], was mutated in 53.6% of all lung tumors. *STK11*, the second most frequently mutated TSG in lung cancer,[87] was mutated in 9.9% of all cases. In Figure 12 mutations in *STK11* and *TP53* are displayed only for cases harboring another mutation in any given oncogene or *KEAP1*.

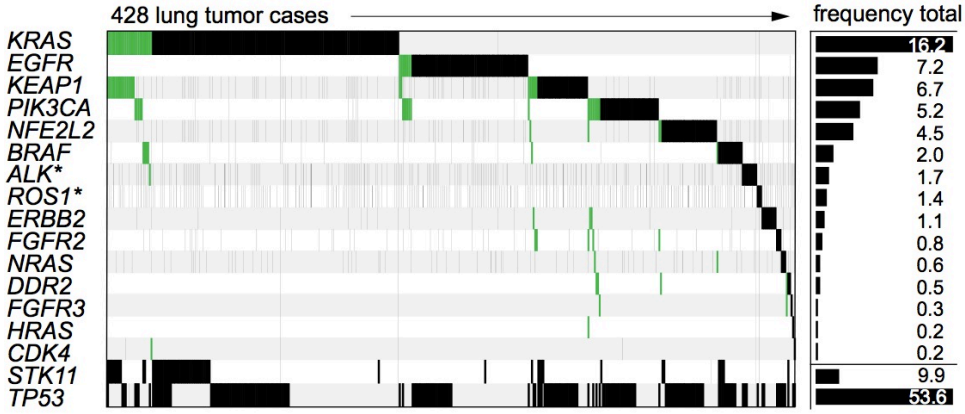

Figure 12 Genetic alterations in oncogenes and tumor suppressors per case and frequencies in % across all lung tumors are presented. Green bars indicate co-occurring alterations. Thin lines represent cases that were not tested in the given gene. Rearrangements are marked with an asterisk.

In Figure 13 the distribution of mutations per protein is presented whereby for deletions and insertions only the first amino acid was included in the calculation. In *EGFR* and *ERBB2* the most common alterations occurred in the intracellular kinase domain: in *EGFR* in-frame deletions within the ELREA motif (amino acid positions 746 to 750) in exon 19 (> 40%) and L858R in exon 21 (28%) and in *ERBB2* exon 20 insertions. *BRAF* mutations were found at five amino acid positions in exon 11 and 15: 464, 466, 469 in exon 11 (62% of all *BRAF* mutations),

600 and 601 in exon 15. In *PIK3CA* 70% of the mutations occurred in codons 542 and 545 in the helical domain both resulting in amino acid change from glutamic acid to lysine. In all *RAS* genes the majority of mutations occurred at codon 12. In *FGFR2* mutations were found in the region coding for the extracellular ligand binding and the intracellular kinase domain. In *FGFR3* only few known mutation sites were analyzed and mutations found in the extracellular domain. In *NFE2L2* two hotspot regions within exon 2 were identified; 63% of the mutations were found in amino acid 20 to 34 and others in 73 to 97. Mutations in TSGs *KEAP1, STK11* and *TP53* were distributed across the whole gene with no specific hotspot regions.

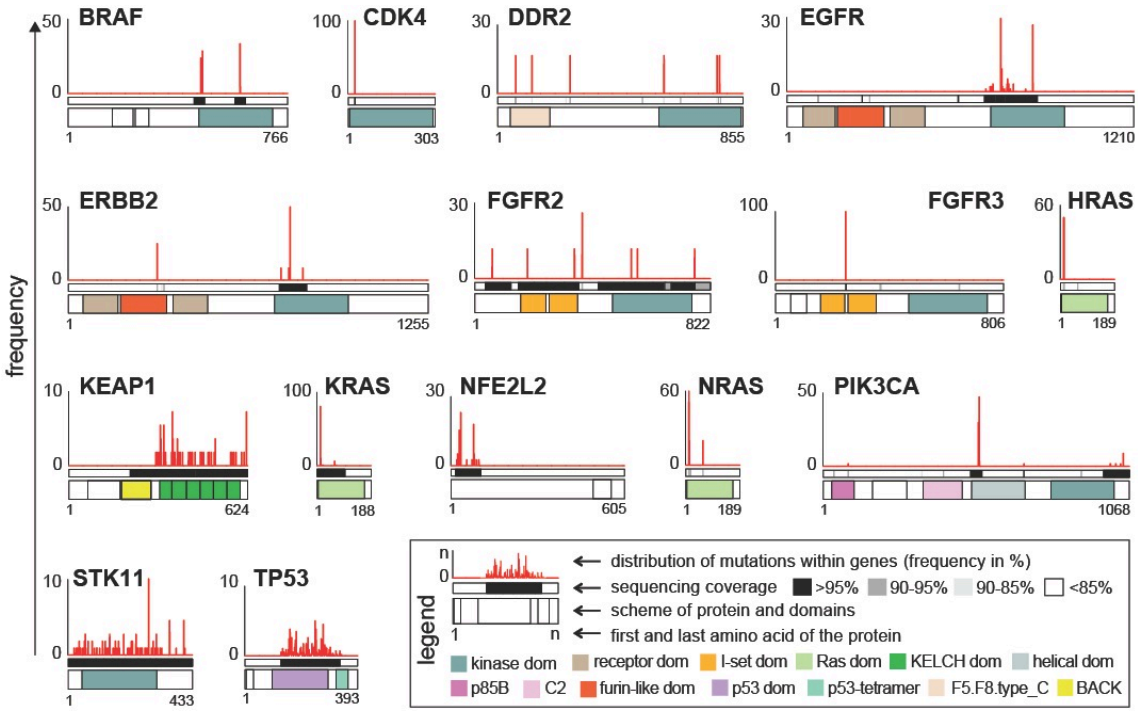

Figure 13 Mutation distribution within proteins. Protein schemata are given with domains (lower part), regions that were analyzed (middle part, grey shades indicate sequencing coverage) and distribution of mutations detected within genes (upper part).

Most oncogene mutations or rearrangements were exclusive to other alterations, consistent with the notion that they are driver alterations (Figure 12;

green lines indicate cases with alterations in multiple genes, Table 8). Of the five most frequent mutated oncogenes (*KRAS, EGFR, PIK3CA, NFE2L2,* and *BRAF*) *EGFR, KRAS* and *NFE2L2* had a co-occurrence rate with mutations in other oncogenes or *KEAP1* of less than 10% (8.6%, 6.6% and 7.9%, respectively). In five *BRAF* mutant cases also *KRAS* or *NRAS* was altered. Of cases mutated in *PIK3CA*, 36% (20/56) also had a mutation in *EGFR, ERBB2*, *DDR2*, *HRAS*, *KRAS, NRAS* or *FGFR*. This promiscuity of *PIK3CA* mutations has been observed before;[88] thus, mutant *PIK3CA* may be necessary but insufficient to drive oncogenic transformation in these cases. In genes that were found mutated in less than ten cases across all samples the co-occurrence rate with mutations in other oncogenes was at least 35% (three of eight *FGFR2*, four of six *DDR2*, two of five *NRAS*, two of three *FGFR3*, one of two *CDK4* and *HRAS* mutated cases harbored at least one additional mutation in another oncogene) with the exception of the three cases with *ROS1* rearrangements - no further alteration in the analyzed oncogenes was identified in those cases. One *KRAS* mutant case harbored a somatic *EGFR* mutation that has not been previously described (substitution of lysine to arginine at position 714, K714N). K714N mutant NIH3T3 cells did not form colonies in soft agar neither did K714N expressing Ba/F3 cells proliferate independent of IL-3 (data not shown). A second case harboring the same *EGFR* K714N mutation also had mutations at positions E709 and G719 on the same allele.[89] It is therefore most likely that the true driver in these tumors is mutant *KRAS* or the compound E709 and G719 *EGFR* mutation, but not K714N.

5. Genetic alterations in histological subtypes of lung cancer

Several cancer genes are frequently altered in specific subtypes of lung cancer, suggesting their important role in developing the specific malignant phenotype. In order to identify subtype-specific genome alterations, oncogenes and TSGs were depicted if they were known to play an important role in tumorigenesis and were either mutated or exhibited significant chromosomal gains or losses in our data set (see below). Mutated genes were included in the analysis if at least three cases harboring an alteration were identified.

To identify frequently amplified or deleted regions per subtype the same rank sum-based algorithm was used as for all lung tumors and applied to samples of the same histological subtype (421 ADs, 69 CAs, 101 LCCs, 63 SCLCs, and 338 SQs) (Figure 14). Significant copy number alterations across all lung tumors summarized in Table 2 could be assigned to histological subtypes and specific genes relevant in cancer were identified within these regions. Amplifications in 17q (*ERBB2*) did not reach significance across all tumors but in the AD subtype.

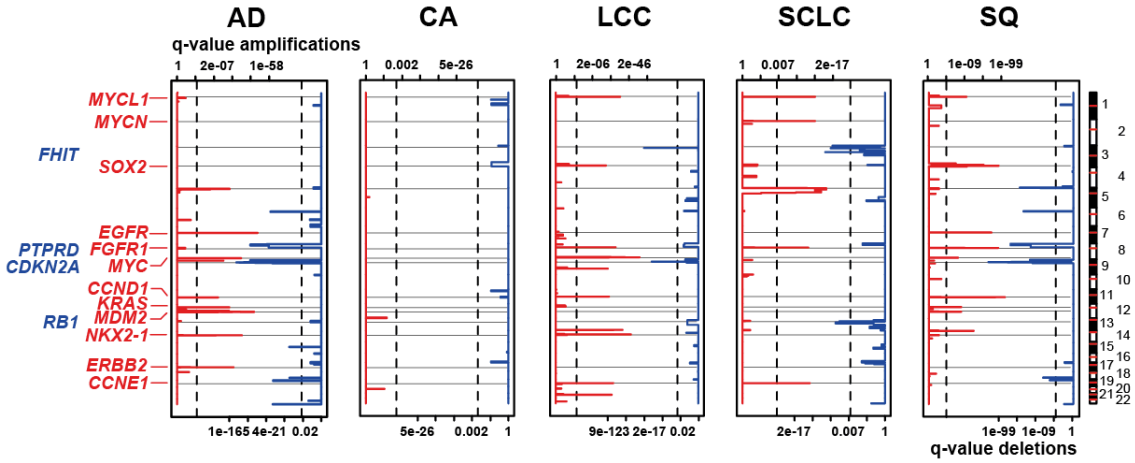

Figure 14 Significantly amplified (red) and deleted (blue) regions plotted along the genome (y-axis) for the five major lung cancer subclasses. Statistical significance, expressed by *q*-values (x-axes: amplification, upper scale; deletion, lower scale; vertical dashed lines indicate q=0.01), is computed for each genomic location. Known or potential oncogenes (red) and tumor suppressor genes (blue) are given at respective locations.

Significantly amplified chromosomal regions in AD affected 5p, 7p (*EGFR*), 8q (*MYC*), 11q (*CCND1*), 12p (*KRAS*), 12q (*MDM2*), 14q (*NKX2-1*), and 17q (*ERBB2*); in SCLC 1p (*MYCL1*), 2p (*MYCN*), 5p, 8p (*FGFR1*), and 19q (*CCNE1*); and in SQ 1p (*MYCL1*), 3q (*SOX2*), 7p (*EGFR*), 8p (*FGFR1*), 11q (*CCND1*), 12p (*KRAS*), and 12q (possibly *FRS2*)[83,90-93] (Figure 14). Several amplified regions were specific for certain lung cancer subtypes, for example 14q and 17q in AD or 3q in SQ. Others were found to be significantly amplified in various subtypes, such as amplifications in 8p (*FGFR1*) in SCLC and SQ or amplifications in 7p (*EGFR*) and 12p (*KRAS*) in AD and SQ. Deletions in AD and SQ affected 6p, 8p, 9p24 (*PTPRD*), 9p21 (*CDKN2A*), 18q, and 19p, in AD also 15q and 22q, in SQ also 4q. In SCLC deletions were significant in 3p (*FHIT*) and 13q (*RB1*). *FHIT* (fragile histidine triad) is discussed as

a potential TSG whose inactivation in lung cancer occurs after initiation of cancer.[94] Due to its location within the most common fragile site of the human genome (where gaps and breaks affect several genes when cells are under replication stress) and the lack of mutations in this gene the relevance of FHIT as a TSG remains unclear.[94] In most significant deleted regions no common TSG could be determined.

LCCs did not exhibit a specific pattern of copy number alterations but showed amplifications and deletions in regions specific of other subtypes, such as amplification of *NKX2-1* (specific of AD) or *MYCL1*, *SOX2* or *FGFR1* (typical of SCLC or SQ) and also amplifications in 19q and deletions in 3p as in SCLC (Figure 14). In 69 CA no significant SCNAs were found.

In order to describe subtype-specific patterns of alterations in more detail, mutated genes and genes identified within significantly amplified or deleted regions were analyzed. The final selection of genes that were frequently altered included 20 oncogenes (amplifications of *CCDN1, CCNE1, ERBB2, EGFR, FGFR1, KRAS*, three *MYC* genes, *NKX2-1, SOX2*; mutations in *BRAF, DDR2, ERBB2, EGFR, FGFR2, FGFR3, KRAS, NFE2L2, NRAS, PIK3CA*; rearrangements affecting *ALK* or *ROS1*) and five tumor suppressor genes (deletions in *CDKN2A* or *RB1*; mutations in *KEAP1*, *STK11* and *TP53*) (Figure 15).

Cases with focal amplification or deletion of the above mentioned genes (Figure 14) were identified by assigning a rank to each case according to the copy number amplitude and segment length for amplification with specific adaptations for cases with low purity. Therefore only copy numbers of the segments spanning the gene of interest were considered for each sample. If the target region had more than one segment the maximum copy number was chosen. The standard deviation was calculated for each gene across all samples to determine thresholds for amplifications and deletions; above the threshold cases were considered amplified, below the threshold they were considered deleted, respectively (Figure 29). Focal amplifications in the tumor genome frequently involve oncogenes that might be important driver alterations is these cases. To identify focal amplifications in this data set, cases were ranked according the length of the segment(s) spanning the

gene of interest and were modeled by an exponential function against the copy number. From this exponential function the critical length of focal amplified regions was derived for each genes. To recover cases with focal amplifications that did not reach the threshold due to low copy number (for example, due to high admixture of non-cancerous cells; see in Figure 29 in the lower left part of the EGFR plot), genome-wide standard deviation was calculated for each case to determine the threshold above which the case was considered amplified.[95]

Histological subtypes display distinct patterns of genetic alterations, i.e., alterations occurred at different frequencies or were exclusive for a certain subtype. For example, in AD mutations in *KRAS* (32.3%)*, STK11* (17.4%) and *EGFR* (15.4%) as well as amplifications of *NKX2-1* (10.5%) were more frequent than in SQ (2.5%, 1.9%, 0.5%, and 4.7%, respectively) and SCLC (two out of 60 SCLC cases harbored a *STK11* mutation but non harbored mutations in *EGFR* and *KRAS* or amplifications of *NKX2-1* were identified) (Fishers exact test per gene for each subtype combination, $p < 0.05$). Amplifications of *FGFR1* and mutations in *TP53* were less frequent in AD (4% and 45%) than in SCLC (12.7% and 70.5%) and SQ (19.5% and 69%). In SQ mutations in *NFE2L2* (10.6%) as well as amplification of *CCND1* (10.4%) and *SOX2* (23.4%) were more frequent than in AD (1.1%, 4.5% and 1.2%, respectively) and amplifications of *SOX2* more frequent than in SCLC (4.8%). In SCLC amplification of *MYCN* (6.3%) and *MYCL1* (6.3%) and loss of *RB1* (63.5%) occurred more often than in AD (0.7%, 0.7% and 12.1%) and SQ (0.6%, 1.8% and 10%). *MYC* was more often amplified in AD (5.5%) than in SCLC and SQ (about 3% in each). Mutations in *ERBB2* occurred more often in AD than in SQ, but did not reach statistical significance. In CA no distinctive genetic alterations were identified. No significant difference between AD, SCLC and SQ was found for mutations in *ERBB2*, *FGFR2/3*, *PIK3CA, KEAP1*, and amplifications of *CCNE1*, *ERBB2*, *KRAS*, *MDM2*, and *MYC* (Fishers exact test, $p > 0.05$). Mutations in *DDR2* and *FGFR3* were found primarily or exclusively in SQ, respectively.

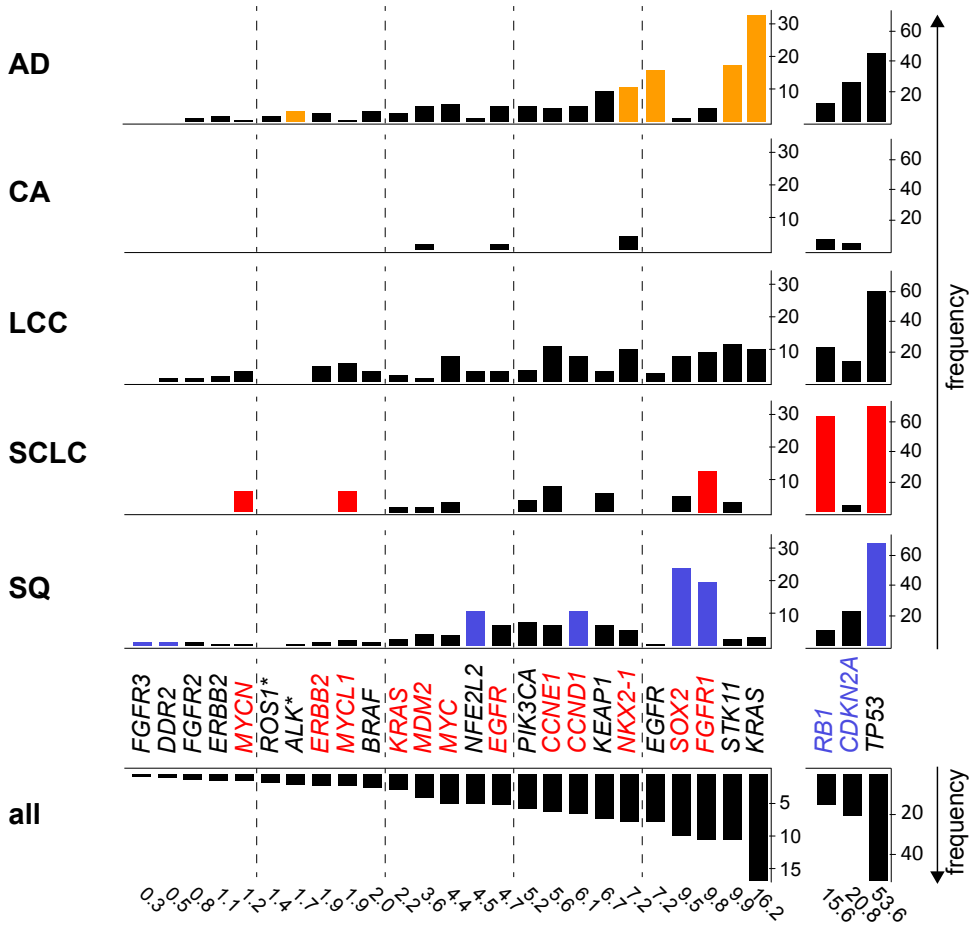

Figure 15 Frequencies of genetic alterations per histological subtype and across all tumors. Colored bars indicate genes typically altered in the respective lung cancer subtype (gene labels: mutated in black, amplified in red, deleted in blue, rearranged marked with an asterisk).

In AD mutations in *EGFR*, *KRAS*, *STK11*, and *TP53* were the most frequent alterations and in SQ mutations in *NFE2L2* and *TP53* as well as amplifications of *FGFR1* and *SOX2* (Figure 15). Also rare SQ cases were identified harboring mutations typical for AD whereas the spectrum differed from those seen in AD. *KRAS* was found mutated in less than 2% of SQ cases (vs. 32% in AD). Here the amino acid substitution at position 12 from glycine (G) to aspartic acid (D) (G12D) was the most frequent point mutation identified (four out of seven SQ cases). A substitution at position 15 from glycine to serine (G15S) was identified in one case only diagnosed as SQ, which has not been described in lung cancer before. Its oncogenic significance is unknown.[96] In AD G12C (C, cysteine) was found to be the most common mutation in *KRAS* (39%) followed by G12V (17%; V, valine) and

G12D (16%). Two SQ cases were identified that harbored a mutation in *EGFR*, one with an exon 19 deletion and a second with an exon 18 mutation at AA 694 (substitution from proline to histidine) that has not been described before. *ALK* rearrangements were found in AD except for one case that was SQ, in which *EML4* was not the fusion partner[97]. Three AD cases had *ROS1* rearranged. In one case the split between the two signals of the FISH probe was very narrow, thus, according to the interpretation guidelines this case would be negative for *ROS1* fusion but dependent on the fusion partner, this case could be positive. The fusion in this case was interpreted as positive but not further investigated. In SCLC the most frequently altered genes were *TP53* (70.5%) and *RB1* (63.5%). Only in rare SCLC cases *KEAP1, PIK3CA* or *STK11* was mutated. No mutations in the typical "AD and SQ oncogenes" such as *BRAF*, *EGFR*, *KRAS* or *NFE2L2* were found in SCLC. In LCCs mutations were found in most genes but no particular distinguishable pattern could be identified for this subtype in respect to AD, SQ or SCLC. *TP53* was the most frequently mutated gene in LCC (60.5%). In CA no mutation was found in the analyzed genes but rare cases had amplifications of *EGFR* or *NKX2-1* or deletions of *RB1* or *CDKN2A*.

Overall, while AD, SCLC and SQ exhibit distinct genetic pattern with rare cases exhibiting non-typical alterations of other subtypes, LCC cases show no unique pattern of genetic alterations. In CA only rare cases had genetic alterations that did not segregate specifically with this subtype.

6. Mutually exclusivity and co-occurrence of genetic alterations

Mutations that occur in different genes or signaling pathways may influence each other's phenotypes. For example, mutations in *EGFR* predispose to therapeutic sensitivity to EGFR kinase inhibitors, whereas an additional mutation affecting *PTEN* in the same signaling pathway may lead to resistance.[98] To understand relevant cancer mutations that suggest functional interaction and may predict clinical outcome, it is important to annotate co-occurring genetic alterations in a given tumor. Associations between mutations and copy number alterations were thus calculated using the Fisher's exact test and Benjamini Hochberg method in

1,032 tumors, as well as for each subtype separately (421 ADs, 62 LCCs, 39 LCNECs, 63 SCLCs, and 338 SQs) in order to identify associations of genomic alterations.

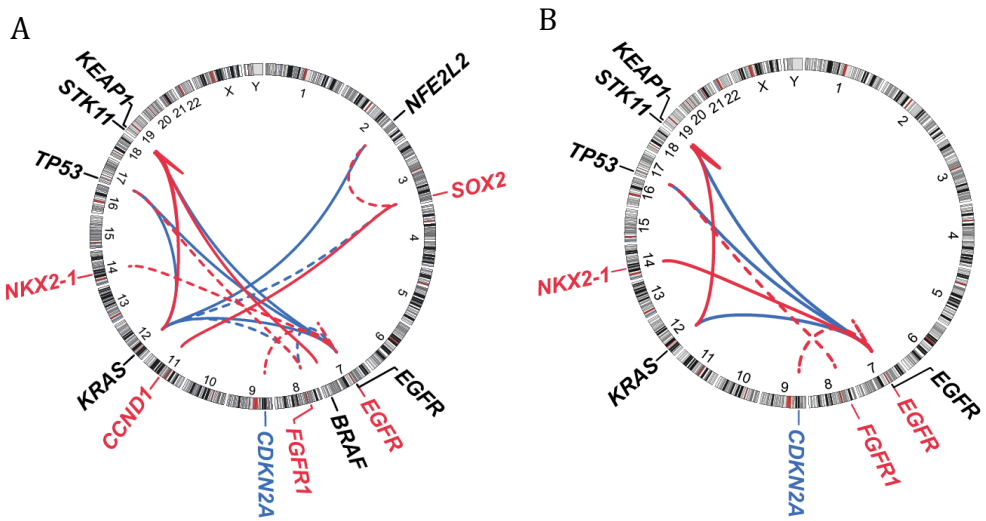

Figure 16 Associations of gene copy number alterations and mutations in lung tumors with the use of Circos plots. **A.** Associations of genetic alterations in all tumor cases are presented. **B.** Associations of genetic alterations found in 380 ADs are shown. (Circos plots (ring: chromosomes 1 to 22; outer ring: genes involved (chromosomal gains in red, deletions in blue, mutations in black); lines within the circle: significant co-occurring (red) and exclusive (blue) events (Bonferroni adjusted Fisher's exact test, $p < 0.05$) between two gene copy number alterations or two frequently mutated genes (solid lines) or between gene copy number alteration and mutation (dashed lines))

In Figure 16 significant co-occurring genetic events are given that were identified in all lung tumors (left) and AD (right). Associations of genetic alterations were not significant in other histological subtypes. In lung tumors *STK11* co-occurred with mutations in *BRAF, KEAP1* or *KRAS* and excluded the occurrence of mutations in *EGFR* (Figure 16.A). Mutations in *KRAS* were mutually exclusive to mutations in *EGFR, NFE2L2, TP53,* and to amplifications of *FGFR1* and *SOX2*. Mutations in *EGFR* were exclusive to mutations in *TP53*. Mutations in *TP53* were only associated with amplifications of *FGFR1*. *EGFR* amplifications were associated with amplifications of *NKX2-1* and *CDKN2A* as well as with mutations in *EGFR*. The latter excluded occurrence of amplification of *FGFR1*. Amplification of *SOX2* co-occurred with mutated *NFE2L2* and amplification of *CCND1*.

In subtype specific analyses significant associations were identified only in AD (Figure 16.B). Associations involving *EGFR* (mutation or amplification) that were identified in the analysis including all lung tumors were also found significant in AD except exclusivity of *EGFR* mutations with *FGFR1* amplifications. Further *NKX2-1* amplifications were associated with *EGFR* amplification but not *EGFR* mutations. *KRAS* mutations were found to be associated with *STK11* mutations and excluded the occurrence of *EGFR* mutations. *FGFR1* amplifications co-occurred with *TP53* mutations and *STK11* mutations with *KEAP1* mutations.

Several co-occurring events did not reach significance after correction for multiple hypothesis testing, such as *PIK3CA* mutations that were frequently found in lung tumors of all histological subtypes harboring another oncogenic alteration, for example in tumors that harbored mutations in *FGFR2* (25% of all *FGFR2* mutant cases had a coexisting mutation in *PIK3CA*), *ERBB2* (20%), *EGFR* (7%), *KRAS* (3%) and amplifications of *FGFR1* (8%) and *SOX2* (7%). In 42% of SCLC cases *TP53* mutation occurred together with *RB1* deletion, whereas in NSCLC only 6% of the cases harbored alterations in both genes.

Overall several cases were identified that had alterations in more than one oncogene or TSG affecting the same or different cellular pathways.

7. Large cell carcinomas of the lung share immunohistochemical, genetic and gene expression characteristics with other histological subtypes

LCC is a diagnostically controversial subtype. It has been the focus of discussions for many years whether these tumors reflect an own entity or if they might be poorly or un-differentiated stages of tumors of other subtypes.[99] Immunohistochemical analysis is now recommended to determine subtypes of lung tumors but application of this approach for LCC that lack any features of differentiation is still not standardized due to limited analyses. To investigate the heterogeneity of LCC, detailed genetic and immunohistochemical analyses were performed in 169 cases of which 64 were LCC with neuroendocrine differentiation (LCNEC). Mutation data were available for 119, genome-wide copy number data for 101 and gene expression data for 31 LCC cases. Gene expression was assessed by

gene expression array and immunohistochemistry. In order to identify tumor types, immunohistochemical profiles were assessed using AD-specific markers (TTF-1 and CK7), SQ-specific markers (p63 and CK5/6) and markers for neuroendocrine differentiation (CD56, Synaptophysin and Chromogranin A).

As shown above tumors of histological subtypes AD, SCLC and SQ displayed distinct patterns of genetic alterations (Figure 15). In contrast, for LCC no specific pattern could be identified but alterations occurred in genes typically altered in other subtypes. LCC harbored for example alterations in *EGFR*, *KRAS*, *STK11* as typically found in AD, and had amplifications of *FGFR1* and *SOX2* as typical for SQ (see Results: Genetic alterations in histological subtypes of lung cancer). Immunohistochemical analysis revealed similar results. One hundred and twelve LCC cases were pathologically reviewed (58 LCC without neuroendocrine differentiation and 54 LCNEC) and immunohistochemical stainings performed for 60 cases (45 LCCs and 15 LCNECs). For some cases immunohistochemical results were available only for diagnostically relevant marker (Figure 17). Final pathological diagnosis was available for 100 LCC cases (48 LCC and 52 LCNEC).

The diagnosis LCNEC was confirmed in 43 of 52 cases (83%), of which 12 were admixed with AD, SQ or in the majority of cases with SCLC components. Eight cases (15%) were reclassified as SCLC and one case as CA. On the contrary 23% (11/48) and 27% (13/48) of LCC cases without neuroendocrine differentiation (further referred to in this section as LCC) were reclassified as AD ($TTF1^{+}$, $p63^{-}$) and SQ ($TTF1^{-}$, $p63^{+}$), respectively. Fourteen cases were tested positive for at least one neuroendocrine marker by IHC and were thereby reclassified either to LCNEC (in three out of ten cases with admixture of SCLC or SQ components) or SCLC (4/14, one with SQ component). Four cases were sarcomatoid (SARC). Three LCC cases were reclassified as AD (2) or SARC (1) without the need of immunohistochemistry, assuming misclassification by the primary pathologist. In only five of 48 cases the diagnosis remained LCC due to lack of glandular or squamous expression profiles.

Only 38 of the 48 LCC cases described above (11 AD-like, 13 SQ-like, 14 LCNEC) were considered *true* LCCs. One can assume that cases reclassified as CA, SARC or SCLC were misclassified by the primary pathologist since SCLC is morphological clearly distinct from LCC by cell size and fine granular chromatin[2]

and diagnosis of SARC can be challenging[100]. Therefore these cases were excluded from further analyses. Considering only those 38 cases, the use of immunohistochemistry allowed sub-classification of 87% of the cases (33/38) into AD, SQ or neuroendocrine tumors (see also Central pathological review, Figure 8). AD differentiation was supported by positive TTF-1 and/or CK7 and no p63, CK5/6 expression (only one case was negative for TTF-1 but exhibited CK7 expression[2]) and SQ differentiation by positive p63, CK5/6 and no TTF-1 and CK7 expression. In Figure 17 results of immunohistochemistry and genetic analysis of 69 LCCs are illustrated. One case was positive for TTF-1 and p63 expression and was reclassified as ADSQ. Cases with expression of at least one neuroendocrine marker were reclassified into LCNEC (CD56 was positive in 50%). CK7 was expressed in all LCNEC cases, TTF-1 in 40%. LCC cases reclassified into AD based on their immunohistochemical profile harbored AD-typical alterations such as mutations in *KRAS* (44%) and *STK11* (22%) and amplification of *NKX2-1* (33%) and cases reclassified as SQ harbored amplifications of *SOX2* typical for SQ (28%) but no mutations in *EGFR* or *KRAS*. Thus, reclassifying cases based on immunohistochemistry is in line with genetic findings.

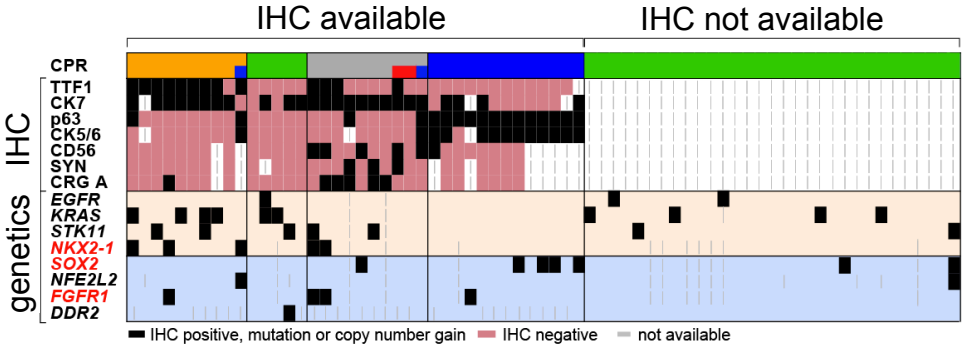

Figure 17 Results of central pathological review (CPR), immunohistochemistry (IHC) and genetic alterations of 69 LCC cases. Genes are arranged according mutation frequency in histological subtypes (background: orange=AD, blue=SQ). Histology (CPR) color code: orange=AD, green=LCC, grey=LCNEC, red=SCLC, blue=SQ, mixed types colored accordingly.

Further sub-classification was not possible for 62 originally diagnosed LCC cases as for five cases no clear immunohistochemical profile was observed, for ten cases only H&E staining was available but definite diagnosis without IHC was not possible or pathological review has not been done. Of these 62 cases, 36 were tested for genetic alterations. In 16% of these *EGFR* or *KRAS* was mutated, typical alterations observed in AD, one case had a *DDR2* mutation, two cases *SOX2* amplification of which one had also a mutation in *NFE2L2*, typical for SQ. In single cases *NRAS* or *PIK3CA* was mutated or *ERBB2* amplified. In total, 45% of LCCs that were not further sub-classified based on immunohistochemical profiles, harbored genetic alterations in at least one gene that segregate with a certain histological subtype and might be useful for classification or predict response to targeted therapies (Figure 18).

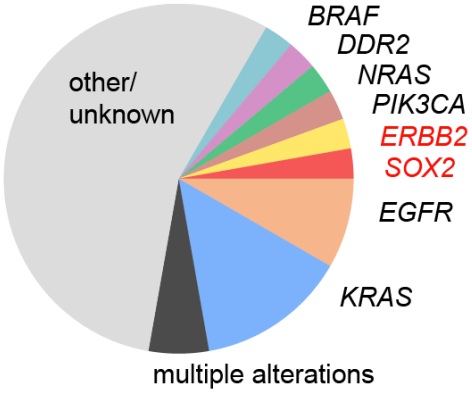

Figure 18 Distribution of subtype specific genetic alterations identified in 36 not further classified large cell carcinomas.

Typical morphology (H&E) and immunohistochemical staining of the three commonly used markers in lung cancer diagnostics TTF-1, p63 and CD56 are shown in Figure 19 for four LCC cases, where one case was assigned to AD based on TTF-1 expression, another to SQ based on p63 expression and a third based on its expression of CD56 to neuroendocrine LCC. The lowest panel in this figure presents one case without a distinctive marker profile that could not be further sub-classified and remained LCC (Figure 19, NOS, not otherwise specified). In cases reclassified as AD (AD-like), SQ (SQ-like) or neuroendocrine (NEC) typical genetic alterations for these subtypes were identified supporting the validity of the immunochemical interpretation, such as mutation in *KRAS* in the AD case, amplification of *FGFR1* in the SQ case, and *RB1* loss in the NEC case. The NOS case

harbored a mutation in *KRAS* and the non-oncogenic *EGFR* K714N mutation typical for AD as well as amplification of *SOX2* typical for non-AD (Figure 19, right).

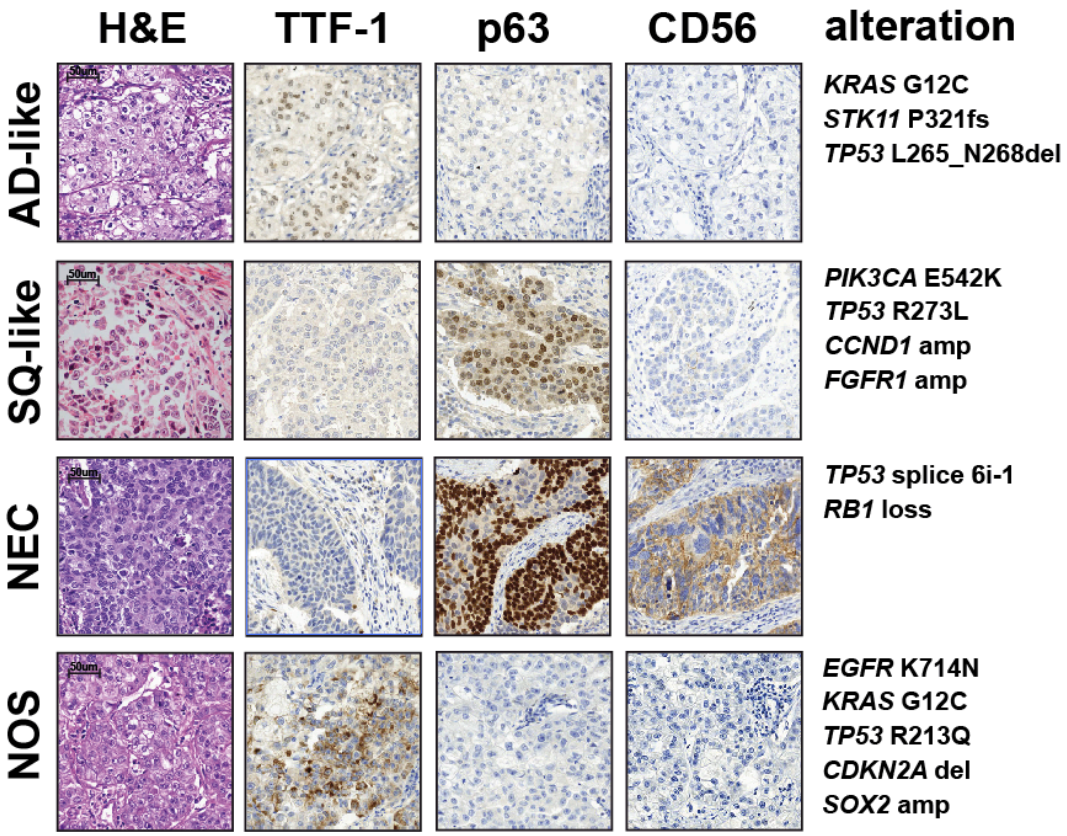

Figure 19 Typical immunohistochemistry of LCC cases exhibiting characteristics of AD, SQ, NEC or no distinct pattern. Respective genetic alterations are depicted on the right.

Overall more than 90% of LCC cases could be further sub-classified based on their immunohistochemical profile. Notably, the genetic alterations in these reassigned cases were in line with the newly assigned diagnostic category. However, in an additional 8% of cases, in which IHC was inconclusive, genome alterations were present that were specific of any one of the other subtypes (i.e., AD, SQ, etc.). Thus, in order to sub-classify LCCs into clinical relevant subgroups detailed immunohistochemical stains and genomic analyses are needed.

To further investigate genetic associations of LCC with tumors of other subtypes, copy numbers in 20 chromosomal regions (see Table 2, regions that were significantly altered across all lung tumors) were compared between the major

histological subtypes AD, SCLC and SQ with LCC and LCNEC using the Student's t test (Figure 20). Significant differences between LCC and LCNEC were seen in chromosomal regions 3p26, 5p15 and 18q21. LCCs were similar to AD and SQ and exhibited differences in copy number patterns only for amplifications on 3q26-29 (*SOX2*) compared to AD and SQ and deletions on 5q15-21 for AD. LCNEC on the contrary differed from AD and SQ in several regions. Amplifications on 3q26-29 and deletions on 3q26 and 18q21 differed in LCNEC from AD and SQ. Additionally, differences between LCNEC and AD were identified for amplifications on 5p15 and deletions on 5q15-21 and 19p13 and between LCNEC and SQ for deletions on 1p13, 6p22-21 and 22q13. Comparing SCLC with LCC eight of the 12 deleted regions were found to be significantly different whereas no difference in the eight amplified and twelve deleted regions was observed comparing LCNEC with SCLC (Figure 20, lane LCNEC-SCLC).

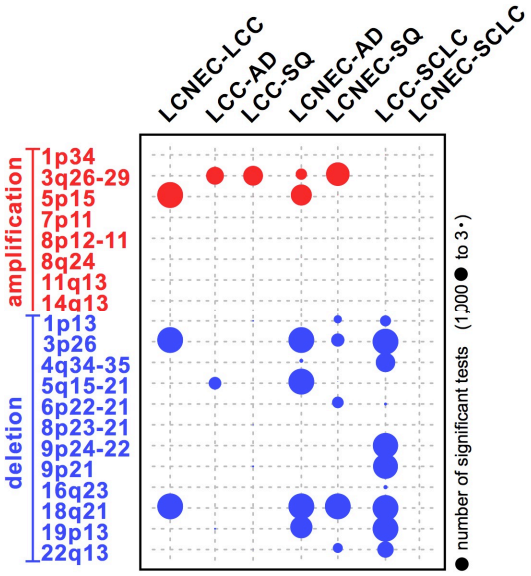

Figure 20 Comparison of significantly amplified and deleted regions between histological subtypes (405 AD, 62 LCC, 39 LCNEC, 63 SCLC, 316 SQ) given for pairs. P-values were calculated by permuting each sample set 1,000 times (according to the smallest observed subtype combination) using the student's t-test. Dots represent statistical significance (size correlates with the number of significant events (p-value < 0.05).

Similarities between LCNEC and SCLC as well as high agreement of LCC with AD and SQ were found when comparing copy number alterations. To further investigate the genetic heterogeneity, unsupervised hierarchical clustering using gene expression data of the 294 most variably expressed genes was applied to 261 lung tumor cases comprising all histological subtypes (Figure 21). Most cases of the same histological subtype clustered together (92% of 13 CA in cluster I, 84% of 25

SCLC in cluster II, 85% of 95 AD in cluster III and 76% of 86 SQ cases in cluster IV) whereas LCCs did not form a distinct cluster but were grouped together with either SCLC (22%), AD (39%) or SQ (39%) (Figure 21, colored triangles). Of the seven LCC that clustered together with SCLC three were histologically LCNECs (Figure 21, grey triangles). Two of these had *RB1* loss, typical for SCLC. Except for one case that clustered together with SQ, LCC harbored genetic alterations typical for the respective histological subtype, for example four LCC cases that had mutations in either *BRAF* (one), *ERBB2* (one), or *KRAS* (two) clustered with AD (Figure 21, orange triangles) and two LCC cases that harbored either a *DDR2* or *NFE2L2* mutation clustered with SQ (Figure 21, blue triangles).

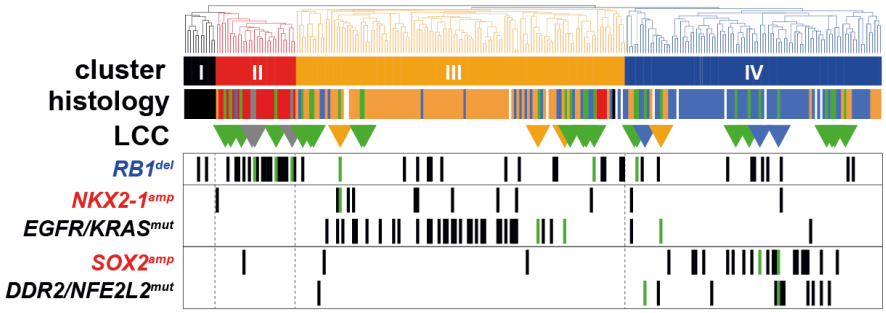

Figure 21 Unsupervised hierarchical clustering using 294 genes. (color code, histology: AD orange, CA black, LCNEC grey, LCC green, SCLC red, SQ blue). LCC cases are indicated as triangles at corresponding positions (in orange with AD-typical alterations, blue with SQ-typical alterations, grey if this case was initially diagnosed as a LCNEC, and green with no known alteration). Genetic alterations as vertical lines (LCC in green; others in black).

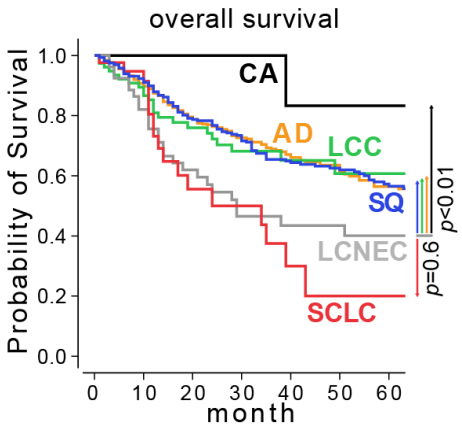

Figure 22 Kaplan Meier Curve for overall survival for histological subtypes with LCNEC seperated from LCC. Survial was compared using a logrank test between LCNEC and SCLC (red, *p*-value=0.6), as well as LCNEC and CA (black), AD (orange), LCC (green) and SQ (blue) (*p*-values<0.01).

In summary, tumors that were originally classified as LCC presented a variety of genetic alterations, heterogeneous immunohistochemical and gene expression profiles similar to other histological subtypes rather than a distinctive pattern. LCNEC are genetically and clinically (Figure 22) more similar to SCLC than to other LCC tumors. In contrast, LCC without neuroendocrine differentiation show characteristics of either AD or SQ and could be further classified into one of them based on their immunohistochemical or genetic profile. Typical alterations of AD or SQ such as mutations in *KRAS* or *DDR2* respectively were identified in several cases where immunohistochemistry was not available or revealed no clear marker profile. Thus, in order to reclassify LCC cases into clinical relevant subgroups combined immunohistochemical and genomic analyses is necessary.

8. Lung tumors are frequently altered in a limited set of oncogenes

In this sample set genetic alterations were identified in genes encoding for proteins of different classes and involved in different signaling pathways. Several oncogenes were frequently mutated or amplified but also rare events were identified in this study. To genetically sub-classify lung tumors genes were chosen based their current clinical significance (targeted therapies already approved or clinical trials ongoing), important functional role in tumorigenesis and thereby potential targets for new targeted therapies as well as recurrence. In total (*ALK, ROS, RET;*[101] *DDR2;*[72] *ERBB2;*[34,102] *BRAF;*[103,104] *EGFR;*[4,69] *FGFR;*[82,105] *CDK4, CCNE1, CCND1* their role in the cell cycle[106]; *MDM2* that inhibits wild-type p53 transactivation function;[92] *NFE2L2* involved in oxidative stress response,[107] and *SOX2*[93]) (Figure 23). In total 56% of the tumor cases had a mutation and copy number alteration in at least one of these oncogenes. Thus, more than half of the patients were genetically characterized using this limited set of genes.

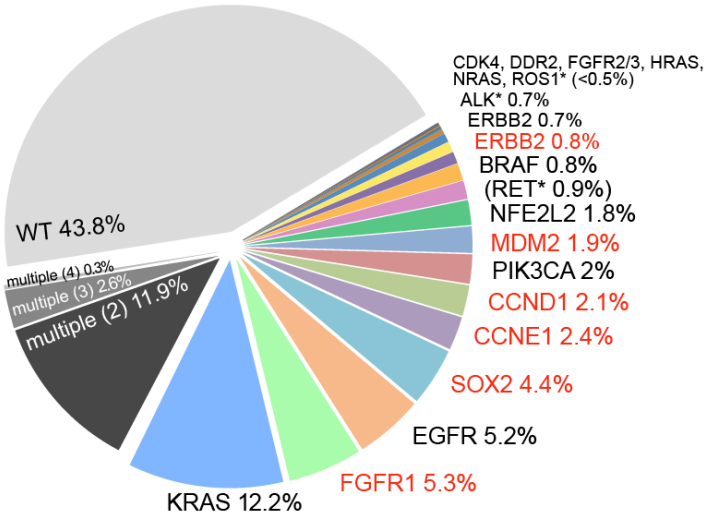

Figure 23 Overview of genetic alterations in lung tumors. Tumors that harbor alterations in at least two different genes were combined in the "multiple" portion (number of co-occurring alterations given in brackets). Gene labels: mutated genes in black, amplified in red, rearranged marked with an asterisk. Frequency of *RET* rearrangements were adapted from Takeuchi *et al* (2012)[101]. WT (wild type) in the given genes.

In cases with multiple alterations (14.8%) mutations in *KRAS, NFE2L2, PIK3CA* or amplifications of *CCND1, CCNE1, FGFR1,* or *SOX2* were involved in 80% (*SOX2* amplifications with about 16% being the most frequent co-occurring event). With the exception of *ROS1*, for every gene at least one case was identified with a co-occurring alteration in another oncogene. In less than 3% of the cases more than two alterations were identified within the same tumor. Mutations occurred rarely in *CDK4, DDR2, FGFR2, FGFR3, HRAS,* or *NRAS* (< 0.5%) across all lung tumors (Figure 23). Frequency for rearrangements of *RET* is adapted from Takeuchi *et al* in 2012[101], who identified 13 AD cases with RET fusion in 1,529 NSCLC cases. In our cohort no case with *RET* rearrangement was identified. The most commonly altered genes were *KRAS* (15.2%), *SOX2* (10%), *FGFR1* (9.8%), and *EGFR* (6.4%). Mutation frequencies vary slightly from the before mentioned since only cases were included in this analysis that had genetic information available for at least 17 of the here depicted 20 genes or had co-occurring alterations and were therefore included in the section with multiple alterations.

Several of the above mentioned genetic alterations predict sensitivity to therapeutics that are currently in clinical use and segregate with specific histological subtypes in lung cancer. Genes were chosen if I) drugs are already approved for targeting the gene product in cancer patients (for example crizotinib against ALK in *ALK*-rearranged NSCLC; erlotinib and gefitinib against mutant EGFR in lung AD),[75] II) for which current clinical trials investigate the effect of drugs that are approved for other molecular targets and cancer types or the effect of new medicinal products in genetically selected patient cohorts (e.g., crizotinib against ROS1; GSK2118436 against BRAF V600 mutant NSCLC; dasatinib in *DDR2*-mutant SQ of the lung; trastuzumab against Her2 in breast cancer; NVP-BGJ398 against FGFR)[74] or III) that serve as exclusion criteria for certain therapies (*KRAS* mutations predict resistance to tyrosine kinase inhibitors[108]). This lead to a selection of the following genes: *ALK, BRAF, DDR2, EGFR, ERRB2, KRAS, FGFR1, PIK3CA*, and *ROS1*.

In 60% of ADs at least one genetic alteration could be identified, in 30% of SQs and 14% of SCLCs (Figure 24). In AD a variety of genetically defined subpopulations were identified, whereas in SCLC amplifications of *FGFR1* was the most prominent alteration. Only rare SCLC cases were identified harboring a *PIK3CA* mutation. Similar in SQ *FGFR1*-amplified cases represented the largest genetically distinctive group. In addition *DDR2* (about 1%) and *KRAS* (2.5%) mutations were identified in SQ cases but only rare cases were altered in *ALK, BRAF, EGFR* or *ERBB2*.

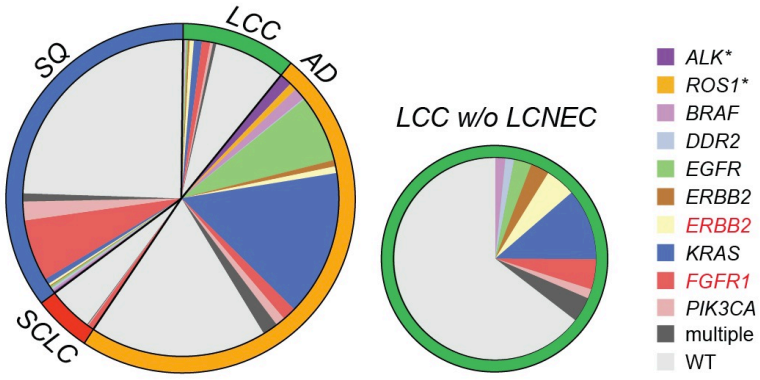

Figure 24 Frequencies of genetic alterations in genes per histological subtype (LCC without LCNEC on the right) (gene labels: mutated in black, amplified in red).

9. Mutant FGFR3: a new potential therapeutic target in lung tumors

Most therapeutic targets have been identified mainly in the AD subtype. Only recently potential new targets were identified in the SQ subtype and drugs directed against these targets already tested in clinical trials. Here recurrent mutations in the fibroblast growth factor receptor 3 (FGFR3) that might be a promising therapeutic target in SQ are described for the first time in lung cancer. A single nucleotide alteration in *FGFR3* that results in a change of an arginine to cysteine at the AA position 248 was identified in three SQ lung tumor cases. Similar to the well studied S249C mutation the AA change to cysteine at position 248 causes formation of stable homodimers through cysteine disulfide bridges in a ligand-independent manner leading to constitutively tyrosine phosphorylation and thereby downstream signaling activation.[109] In this data set *FGFR3* mutations excluded occurrence of mutations in *FGFR2*. One of the three mutated cases also had an amplification of *FGFR1* as well as the *PIK3CA* E545K mutation and a missense mutation in *TP53* and a second case harbored a mutation in the tyrosine kinase domain of *DDR2*. All cases were early stage squamous cell carcinomas (Figure 25).

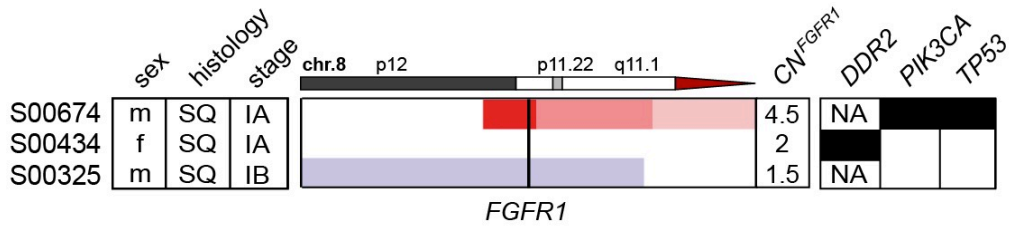

Figure 25 Summary of three lung cancer cases harboring the *FGFR3* R248C mutation. Clinical characteristics are listed on the left. In the middle copy numbers (CN^{FGFR1}) are displayed for the locus on chromosome 8 (chromosomal location: 30,897,037 to 45,848,014) encompassing *FGFR1* (red=amplified, blue=deleted) and mutations in *DDR2, PIK3CA* or *TP53* indicated as black squares per case on the right.

The *FGFR3* R248C mutation has been described in bladder cancer[110] but not in lung cancer. Furthermore, data on the oncogenic properties of this mutation as well as its possible association with sensitivity to FGFR inhibition has so far been limited. In order to assess its role as a potential target of cancer therapeutics mutant and wild-type *FGFR3* were cloned into the retroviral pBabe-puro backbone

and expressed stably in NIH3T3 cells. These immortalized murine fibroblasts exhibit a hypertriploid karyotype but do not show a transformed phenotype.[111] Plated in soft agar non-transformed NIH3T3 cells stop dividing whereas oncogene-expressing cells continue to grow and form colonies.

The transforming ability of the *FGFR3* R248C mutation could be shown in cell line models using NIH3T3 cells. This mutation was able to transform NIH3T3 cells to anchorage independence as assayed by colony formation in soft agar whereas cells expressing wild type *FGFR3* failed to induce colony formation after four weeks (*t*-test, $p=1.07^{-07}$) (Figure 26). NIH3T3 cells expressing the fusion *EML4-ALK* (used as a positive control for this assay) showed anchorage-independent growth. NIH3T3 transduced with an empty vector (negative control) formed isolated colonies comparable to *FGFR3* wild type expressing cells ($average_{15wells} = 0.7$). Expression levels of mutant and wild type *FGFR3* and *EML4-ALK* were evaluated by Western blot and shown to be approximately equal in the cell lines (data not presented).

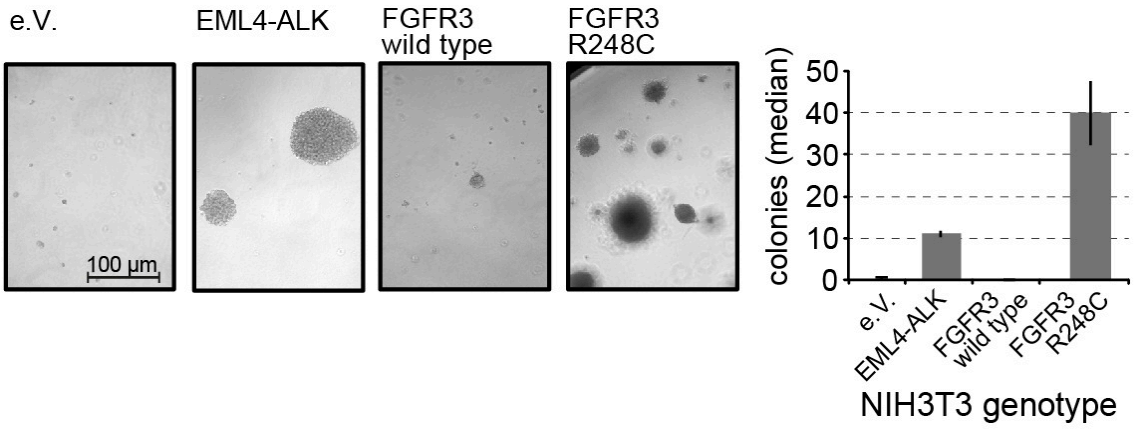

Figure 26 Clonal NIH3T3 cells expressing *EML4-ALK*, wild type *FGFR3* or R248 mutant *FGFR3* and empty vector control (e.V.) were plated in soft agar and colonies counted after 4 weeks of incubation using the Zeiss Vert.A1 microscope (10x original magnification). Median number of colonies per cell line is given on the right. Error bars indicate standard deviation of median determined by bootstrap (10,000 sample drawn).

Sensitivity to the selective FGFR kinase inhibitor NVP-BGJ398 was tested in NIH3T3 cells expressing R248C mutant *FGFR3* using the same approach and adding

increasing concentrations of the compound to the cells. NVP-BGJ398 binds into the cytosolic ATP-binding pocket of FGFR and has been shown to inhibit proliferation in cell lines with significant association to FGFR genetic alterations.[105,112] Currently NVP-BGJ398 is tested in phase I clinical trials in patients with advanced solid tumors harboring alterations in *FGFR.*[113] Anchorage-independent growth in R248C mutant cells was inhibited at a concentration of 10 nM NVP-BGJ398 (Figure 27). At the concentration of 1 nM NVP-BGJ398 no significant difference of growth compared to untreated cells was observed. Growth of the cancer cell line A549 harboring the oncogenic *KRAS* G12S mutation and no alterations in the *FGFR* genes was not inhibited at a concentration of 1 µM BGJ398 (data not shown), ruling out possible toxic side effects of the compound at the given concentrations in this cell line model.

In summary, the *FGFR3* mutation R248C has been identified in rare SQ cases. This mutation transformed cells *in vitro* and confers sensitivity to FGFR inhibition.

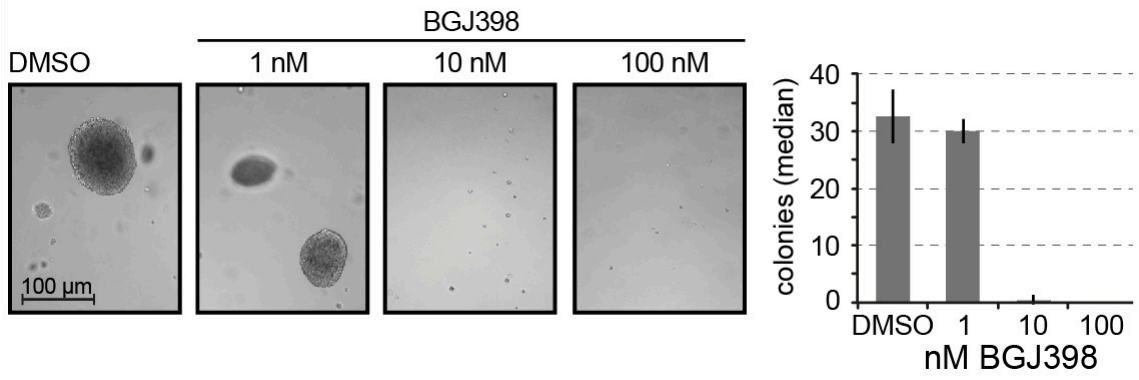

Figure 27 NIH3T3 cells expressing R248 mutant FGFR3 treated the pan-FGFR inhibitor NVP-BGJ398 plated in soft agar (10x original magnification). Median number of colonies for each concentration is given on the right. Error bars indicate standard deviation.

Discussion

1. Genetic characterization of lung tumors

This is the first study, where distribution and frequencies of genetic alterations have been comprehensively characterized across all lung cancer subtypes. Mutations in *BRAF, EGFR, ERBB2, KRAS, STK11*, chromosomal rearrangements affecting *ALK* and *ROS1*, as well as amplifications of *NKX2-1* and *MDM2* were associated with the AD subtype confirming previous findings.[18,38,101,114] Mutations in *DDR2* and *NFE2L2*, amplifications of *CCND1, FGFR1* and *SOX2* segregated with SQ,[72,82,115] and *RB1* and the genes of the *MYC* family were typically altered in SCLC[81]. LCC on the contrary did not reveal a distinct genetic or histopathologic pattern. Instead the majority of LCC cases retained expression of differentiation tumor markers supporting their biological similarity to AD, SCLC and SQ, which is further supported by the presence of genetic alterations typical for these subtypes in many cases. LCCs that exhibited neuroendocrine differentiation were found to be biologically and clinically more similar to SCLC than to other LCCs.

With the increasing understanding of cancer biology, treatment strategies shift towards a personalized approach based on genomic and molecular alterations rather than the "one fits all" chemotherapeutic approach. Thus, the findings of this study could motivate the development of new diagnostic strategies for patient stratification into genetically defined subgroups.

1.1. Therapeutically important signature alterations in AD, SCLC and SQ are rare events in other lung cancer subtypes

Most genetic alterations that are or might be suitable for targeted therapies were identified in AD.[76,83,114] In SQ only recently therapeutic targets were identified by others and us, showing that the genes *DDR2* and *FGFR1* are recurrently mutated or amplified.[72,82] *DDR2* mutations were associated with sensitivity to dasatinib (a multi-target kinase inhibitor) in xenograft models and in patients.[72,116] FGFR inhibition (for instance using the selective FGFR inhibitors PD173074 or NVP-BGJ398) lead to shrinkage of tumors that harbored *FGFR* alterations (i.e. amplifications or mutations affecting one of the FGFR family members).[105] In this study the oncogenic *FGFR3* R248C mutation has been identified in rare SQ cases

and sensitivity to the FGFR inhibitor BGJ398 identifies this alterations as a novel therapeutic target in lung cancer. The most frequently altered pathway in SQ was the Nrf2-KEAP1 pathway.[115] Mutations in the transcription factor Nrf2 and its negative regulator KEAP1 (the E3 ubiquitin ligase) were also found in AD and *KEAP1* in few SCLC cases. Nrf2 up-regulation has been associated with resistance to chemotherapeutic agents (for example to cisplatin).[117,118] Therefore, alterations in the Nrf2-KEAP1 pathway might serve as predictive markers for the response to chemotherapy.

Several therapeutic targets were found in AD and SQ that could already be translated into clinical application. By contrast, only few such targets have been identified in SCLC. Most SCLCs did not exhibit mutations in genes that are typically altered in AD (such as *ALK, BRAF, EGFR, ERBB2 or KRAS*) and SQ (*NFE2L2*). However, few SCLC cases harbored alterations that were also found in AD and also in SQ, such as mutations in *KEAP1* or *PIK3CA*. The genes *FGFR1*, *MYC* and *SOX2* were frequently amplified in SCLC,[80,81] but the design of potent inhibitors is still challenging for the transcription factors *MYC* and *SOX2* because of their ubiquitous functional relevance in cells. Typical genetic alterations in SCLC were chromosomal loss of *RB1* and mutations in *TP53* that frequently co-occur[119] and predict for worse outcomes in patients harboring such alterations.

In line with other studies, rare lung cancer cases were identified that harbored alterations typical for other tumor subtypes. These rare cases revealed a spectrum of genetic alterations that is different from those subtypes, where such signature alterations were most frequently found. For instance, AD-typical alterations were found in SQ cases, such as *EGFR* (<1%), *BRAF* (0.5%)[104] and *KRAS* (<3%) mutations or *ALK* rearrangement (<0.5%)[115,120]. One of the two *EGFR* mutations has been associated with sensitivity to erlotinib and gefitinib[121]. Currently, EGFR inhibitors are approved for *EGFR*-mutant AD only. Even though clinical outcome in SQ patients has not been tested in a prospective study, *EGFR*-mutant non-AD patients might also benefit from such treatment. Similarly, one SQ case had an *ALK* rearrangement; those cases might be sensitive to crizotinib. The two *BRAF*-mutant SQ cases harbored the V600E (E, glutamate) mutation that

has been associated with increased overall survival and PFS in patients with advanced melanoma treated with vemurafenib compared to standard chemotherapy[122]. Promising results have recently been published also for a lung AD patient.[125] However, preclinical data show that acquired resistance can be caused by EGFR activation and formation of RAF dimers that in turn are insensitive to vemurafenib.[123] EGFR overexpression is frequently observed in lung carcinoma and might cause primary resistance to BRAF inhibition.[124,125]

Similarly, typical SQ alterations occur also at low frequencies in other subtypes, such as mutations in *NFE2L2* in AD and amplifications of *FGFR1* in SCLC. Other than in SQ where focal, high copy amplifications pinpoint *FGFR1* as one of the main oncogenes in this region, gains in SCLC affect more frequently large chromosomal regions. Amplification of genes adjacent to *FGFR1* implies that *FGFR1* may not be the driver gene in these cases. Relatively low level amplification of *FGFR1* in most SCLC cases compared to SQ further puts its importance as a driver oncogene in SCLC tumorigenesis in question. Nonetheless, evaluation of such alterations as molecular targets might be reasonable in certain SCLC cases.[74]

To date, diagnostic screening is recommended for *ALK* and *EGFR* in AD only. Since many (potential) molecular targets in lung cancer segregate with a specific histological subtype, clinical trials are conducted for these subtypes accordingly. Thus, response rates to specific therapeutics are not available in tumors of a different histological subtype that harbor the same oncogenic driver alteration. However, it is possible that for instance SQ tumors harboring activating *EGFR* mutations or *ALK* rearrangements might be sensitive to EGFR inhibitors or crizotinib, respectively. With the increasing understanding of the diversity of molecular mechanisms underlying tumor development and drug resistance in different subtypes, applicability of therapeutic that are successful in a certain tumor subtype to other subtypes will most likely be better predictable in the near future.

Thus, in order to make well-informed treatment decisions clinically relevant alterations should be tested in all lung tumors irrespective of the histological subtype.

1.2. The revocation of large cell carcinoma of the lung

To date, no classification system is available for the clinically and morphologically heterogeneous group of LCC that describes therapeutically relevant subgroups. Discussions about a re-evaluation of this subtype are ongoing since immunohistochemistry became a standard tool in cancer diagnostics and certain LCC were found to retain immunohistochemical marker profiles of AD (TTF-1^{+}, p63^{-}, neuroendocrine markers^{-}) and SQ (TTF-1^{-}, p63^{+}, neuroendocrine markers^{-}).[126-128] LCC with neuroendocrine differentiation were grouped with other LCC in 1999 based on morphology but the detailed analysis rather suggests that LCNEC is biologically more similar to other neuroendocrine tumors than to other LCCs.[129-131]

In our sample set 90% of LCCs were reassigned based on their immunohistochemical profile to AD, SQ or neuroendocrine tumors (neuroendocrine markers^{+}) comparable to other studies.[128,132] Remaining cases were not further specified due to inconclusive IHC profile. In a study analyzing gene expression in LCCs, SQ lineage was correlated with high expression of thymidylate synthase, the enzyme that is associated with reduced sensitivity to pemetrexed.[133] Thus, marker-based classification of LCC might be useful to select patients for approved therapies. LCC tumors were reassigned to other subtypes with high concordance of signature alterations. Mutations in *KRAS* or *STK11* were found in LCCs that were reassigned to AD, and mutations in *DDR2* and *NFE2L2* as well as amplification of *FGFR1* were found in cases that were reclassified as SQ (Figure 21)[99,134]. In total, one third of LCC cases that were reassigned to either AD or SQ based on immunohistochemistry (potential) therapeutically relevant alterations could be identified (Figure 17). Furthermore, in almost half of the LCC cases where immunohistochemistry was not available (for example due to insufficient tumor tissue) genetic alterations were identified that might be useful for patient stratification into clinically relevant subgroups, such as mutations in *BRAF*, *EGFR* and *PIK3CA* as well as amplifications of *ERBB2* and *SOX2*. In cases where immunohistochemical profiles revealed inconclusive results and thus diagnosis remained to be LCC, subtype specific genetic alterations were identified in most of these cases.

LCCs with neuroendocrine differentiation were found to be genetically and clinicopathologically different from other LCCs. Instead these tumors share characteristics with those of the SCLC subtype. Retrospective studies show that the response rates of LCNEC to cisplatin-based chemotherapeutic regimes are comparable to SCLC[135-137] underlining biological similarities of both lung cancer subtypes. Sun *et al* described in 2009 that immunohistochemical staining of three neuroendocrine markers (CD56, Synaptophysin, Chromogranin A) was not significantly different between LCNEC and SCLC but these results remain controversial.[129,130,138,139] In our study statistical similarity was observed for the typical immunomarkers except CD56 (Appendix, Figure 31). Furthermore typical alterations found in AD or SQ such as mutations in *KRAS* and *NFE2L2* were also found in rare LCNEC cases but not in SCLC. Nonetheless striking similarities between LCNEC and SCLC were observed on the genomic and on the expression level. *RB1* and *TP53* were found significantly altered in SCLC[81] and LCNEC when compared to AD[114] and SQ[115]. Also, LCNEC and SCLC showed the highest concordance with regard to significantly amplified and deleted regions in comparison to all the other lung cancer subtypes (Figure 20, Figure 21). Survival of LCNEC patients was comparable with SCLC but not with other LCC cases (Figure 22).

Overall, LCCs seem to be undifferentiated forms of other subtypes and thus might benefit from therapy options already approved for AD or SQ. LCNEC showed more similarities on genomic and expression level with other neuroendocrine tumors in particular with SCLC and should therefore be separated from other LCC in the WHO classification. Of cases initially diagnosed as LCC less than 1% remained LCC in our setting after immunohistochemical and genetic re-assignment. Our findings suggest that patients with LCC should also be included for routinely molecular testing to identify patients that are eligible for targeted therapies. Immunophenotyping might be a useful diagnostic approach to further classify tumors where no clinically relevant alteration is identified. Clinical relevance of the stratification into genetically or immunohistochemically defined subgroups needs to be investigated in a prospective patient cohort where AD-like and SQ-like LCC cases receive treatment accordingly.

1.3. A comprehensive profile of alterations in cellular signaling pathways is required for optimal treatment strategies

In tumor cells the complexity and redundancy of signaling pathways involved in cell proliferation and survival complicate successful inhibition of either of the pathways.[140] Initial responses to single-agent therapeutics can be spectacular, as for example shown for the use of erlotinib in *EGFR*-mutant lung tumors, but all tumors reappear within few months and typically they are resistant to the initially applied regimen.[141] Several resistance mechanisms have been described for most cancer types; typically the signaling pathway that is specifically targeted can be re-activated through signaling of bypass effectors or acquired resistance of the target protein to the inhibitor (e.g., acquired mutation in the binding site of the crizotinib in *ALK* rearranged tumors[142]), or alternative signaling pathways can be activated. For example, tumor cells with deregulated FGFR exhibit activation of the PI3K-AKT pathway and also the RAS-RAF-MAPK pathway. These tumors develop some degree of addiction to signaling through the downstream component ERK and become sensitive to ERK signaling inhibitors.[143] In order to overcome signaling through the PI3K-AKT pathway, inactivation of the upstream driver oncoprotein (in this case FGFR) or simultaneous inhibition of the PI3K-AKT-mTOR and the RAS-RAF-MEK-ERK pathway might be more effective.[144]

Also tumors with alterations in *FGFR* frequently had additional alterations in other oncogenes: *PIK3CA* or *HRAS* mutations were found in *FGFR2*-mutant cases, *PIK3CA* or *DDR2* mutations in *FGFR3*-mutant cases, and *KRAS* mutations were identified in tumors that had *FGFR1* amplified. In cell line models *KRAS* predicted insensitivity to BGJ398 in *FGFR1*-amplified cases.[105] Thus, tumors that harbor additional oncogenic alterations might be insensitive to FGFR inhibition through bypass activation of downstream components in signaling pathways.[105] Therefore, combinatory inhibition of FGFR and PI3K in case of additional mutations in *PIK3CA* or *PTEN*, or combinatory inhibition of FGFR and MEK in case of additional alterations in components of the RAS-RAF-MAPK pathway might be effective.

Furthermore, in this study lung tumors have been identified that harbored co-occurring oncogenic alterations affecting signaling molecules in the same or a different signaling pathway that might cause resistance to targeted therapies. In

contrast to the confirmed mutually exclusivity of *BRAF* with *KRAS*,[103,145,146] cases with alterations in both genes were found in our sample set. In such cases inhibiting BRAF may lead to the activation of other members of the RAF family in a RAS-dependent manner that mediate transduction of cellular signals to ERK and thus lead to the activation of the MEK-ERK pathway.[147] Instead, inhibition of all RAF molecules using a pan-RAF inhibitor or inhibition of MEK, downstream of RAS/RAF might be successful therapy options in double mutant cases.

These findings underline the importance of simultaneously genotyping all clinically relevant genes, instead of testing single genes, because complex results might necessitate advanced treatment strategies.

2. Genetically informed diagnosis of lung tumors becomes pivotal

Lung cancer is one of the genomically most diverse of all cancers. On the one hand this complicates treatment strategies, on the other hand offers many opportunities to classify patients into genetically defined subclasses for targeted therapies. Making treatment decisions based on genomic alterations is relatively new but became standard of care in advanced-stage lung AD in case of activating *EGFR* mutations and *ALK* rearrangements.[63] Several new molecular targets were identified within the past few years in AD, SQ and SCLC.[63,65,148] In this study approximately 70% of AD, 50% of SQ and 20% of SCLC cases could be assigned to genetically defined subgroups; most of which had important implications in current clinical settings. The clinical significance of re-assignments from histomorphologic AD or SQ to other subtypes according to their genetic make-up needs to be assessed in a prospective study where treatment decisions are based solely on genetics instead of morphology. In cases with multiple alterations that predict for different histological subtypes, individual assessment by the treating physician would be required. Here, immunohistochemistry could add further information. Especially in cases with controversial results, treatment decisions will strongly depend on the updated knowledge of the physician on current treatment options.

Successful translation of molecular screening into clinical routine has been demonstrated in a collaboration of the Center for Integrated Oncology (CIO), the

joint comprehensive cancer center of the University Hospitals of Cologne and Bonn in North Rhine Westphalia. This network has been initiated in 2010 to offer molecular diagnostics to lung cancer patients in the catchment area of the cancer center. Based on alterations identified in the CLCGP a screening strategy was devised for centralized genotyping. Advanced stage AD tumors were analyzed in a single gene-based and consecutive approach first testing for alterations in *BRAF, KRAS* and *PIK3CA*. If these genes were found to be wild type then *EGFR* and *ERBB2* were tested and if these were wild type as well, *ALK* was tested for chromosomal rearrangements. Advanced stage SQ tumors were tested simultaneously for alterations in *DDR2* and *FGFR1*. To test for mutations in *BRAF, DDR2, EGFR, KRAS,* and *PIK3CA* conventional technics such as the high resolution melting curve approach or Sanger sequencing were used. Genome copy number alterations and chromosomal rearrangements in *ALK, ERBB2* and *FGFR* were analyzed using FISH. Overall 70% of new cases were genotyped accordingly and in almost 40% of the cases alterations in at least one of the genes has been identified. Treatment recommendations were given to the treating physicians and patients received targeted therapeutics, if applicable. *EGFR*-mutant lung tumor patients routinely received approved EGFR inhibitors, whereas patients with *ALK* rearrangements were enrolled in clinical trials (www.lungcancergroup.de) before crizotinib was approved in summer 2012 for the European market for advanced lung AD. Treatment with targeted therapeutics significantly increased overall survival when compared to standard chemotherapy (*EGFR*-mutant AD: median OS of 31.5 vs. 9.6 months, $p < 0.001$; ALK-rearranged AD: median OS of 23 vs. 11 months, $p = 0.024$).[95] Here the feasibility and significance of molecular diagnostics in clinical routine has been impressively shown. Major limitation in this setting were first, the consecutive approach of genotyping for alterations in several genes, which could take up to one month for completion and second, the selection of tumors for genetic testing based on histology. A multiplex approach would minimize labor costs and time until results are available. It would also require less tumor material and would leave enough tissue for immunohistochemistry to provide a complete profile for diagnosis.

Application of such genomics-based classification will be the future but is still far from mainstream. High costs of appropriate sequencing equipment, infrastructural limitations and lack of adequate training of medical staff are still factors that limit the ubiquitous application of genomic medicine. Also, the rapid pace of target discovery and drug approvals for single agent and combined therapies are factors that might seem daunting to physicians and still hinder them from applying genomic medicine to patients.

In summary, in this study a comprehensive genomic characterization of lung tumors has been provided. Genetically distinct subgroups were described for the major lung cancer subtypes AD, SQ and SCLC that could serve as a blueprint for new diagnostic strategies. Tumors of the LCC subtype could be re-assigned based on immunohistochemical profiles and genetic alterations to other subtypes and thus to clinically relevant groups. In this study it became clear that morphology should not be used as a criterion to preselect patients for genotyping. Rare cases harbored alterations in genes that are untypical for this subtype and which would be missed if tumors were preselected for genotyping in limited gene sets. Thus, in order to assure the best care for each patient, all genes that are clinically relevant should be tested in every tumor.

Outlook

Discovery of new targetable drivers seems to have reached a plateau, because only small numbers of cases harboring new oncogenic alterations are identified in current genomic studies.[65,141] Previous screening efforts mainly focused on analyzing coding regions. Since not in all tumors alterations were identified, it can be assumed that the spectrum of alterations that are associated with malignant phenotypes will increase. It is likely that alterations affecting non-coding genomic regions and regulatory elements, such as promoters, enhancers and silencers, play an important role in tumorigenesis. Since Hanahan and Weinberg first published the *Hallmarks of Cancer* in 2000 a more complex picture of cancer biology emerged where tumor microenvironment and immune system were found to play critical roles in tumorigenesis.[22] Also the role of DNA methylation in transcriptional regulation is currently investigated. Cancer is much more complex than initially thought. The straightforward inhibition of oncogenic drivers inevitable leads to recurrence of the tumor and resistance in almost all cases.[141] Several resistance mechanisms – pre-existing in a minor tumor cell population or acquired - have already been identified and second generation drugs are designed with the aim to avoid or overcome such resistances.[149,150] Many clinical trials are already initiated to evaluate which single and combined agents in selected patient groups have the highest potential and should be further investigated.[74] Many more strategies to specifically inhibit tumor drivers and critical pathways will probably be found. With this vision in mind, it is now time to adapt traditional diagnostic and more general treatment strategies to tumor-specific and thus individualized molecular diagnostic and treatment.

Materials & Methods

1. Workflow

An overview of the project experimental workflow is given in Figure 28. Fresh frozen lung tumor samples were pathologically reviewed regarding tumor content and tissue quality. If suitable, DNA and RNA were extracted for mutation, copy number and gene expression analysis. Tumor samples embedded in paraffin were processed for detection of rearrangements in *ALK, RET* and *ROS1* using fluorescence *in situ* hybridization. Further, two lung cancer pathologists reviewed cases histomorphological and immunohistochemical.

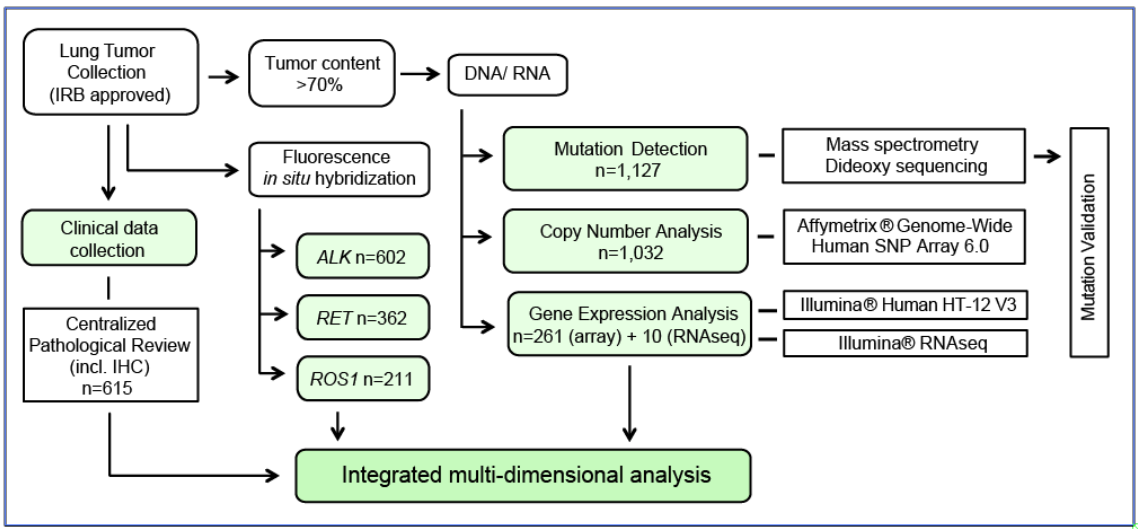

Figure 28 The experimental workflow of the CLCGP.

2. Source of patient material and cell lines

Tissue from resected primary lung tumors of all histologies and stages was provided by multiple institutions (Lung Cancer Group Cologne, Cologne, Merheim; Jena University Hospital, Friedrich-Schiller-University, Jena, Germany; Thoraxklinik-Heidelberg gGmbH, Heidelberg, Germany; VU University Medical Center Amsterdam, Amsterdam, The Netherland; University Medical Centre Groningen, Groningen, The Netherlands; Department of Medicine, Institute Gustave Roussy, Villejuif, France; Université Joseph Fourier, Grenoble, France; University Hospital Zürich, Zürich, Switzerland; The University of Liverpool, Liverpool, UK; Norwegian Radium Hospital, Oslo, Norway; University Hospital Bologna, Bologna,

Italy; Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy and St Vincent's Hospital & The Peter MacCallum Cancer Center, Melbourne, Australia) in form of fresh frozen biopsy samples, frozen sections or extracted genomic DNA depending on the individual institutional review board approval decisions. Tissue was snap-frozen within 30 min post surgery and stored at -80°C. For autopsy cases, tumors were derived within a few hours after death. Cases were staged according the TNM system version 6.

Normal HapMap cell lines were purchased from the Coriell Institute (GM12144C, GM18504B, GM18508B, GM18517C, GM18857B, GM18863B, GM18870B, GM18871B, GM18912B, GM19099B, GM19137B, GM19142B, GM19143B, GM19153B, GM19194B).

NIH3T3 and HEK293T cells were purchased from ATCC, Ba/F3 cells were a kind gift from Nikolas von Bubnoff.

3. Pathology review

1,860 primary lung tumors were submitted for the CLCGP. If paraffin embedded tumor material was available hematoxylin and eosin-stained sections were histomorphologically reviewed by a team of expert lung cancer pathologists (Elisabeth Brambilla, CHU Albert Michallon and William D. Travis, Memorial Sloan Kettering Cancer Center (MSKCC)) to confirm the initial diagnosis or reclassify the tumor. Immunohistochemical staining for TTF-1, CK7, p63, CK5/6, CD56, Synaptophysin, and Chromogranin A was utilized to reclassify cases where a final diagnosis based on morphology alone was not possible. FFPE sections were mounted on X-tra® slides (Leica) and routinely stained using the following clones: 8G7G3/1 (TTF-1), 4A4 (p63), D5&16B4 (CK5/6), OV-TL 12/30 (CK7), N-CAM (CD56), polyclonal (Synaptophysin), DAK-A3 (Chromogranin A) in the pathology department of the University Hospital of Cologne. Stained sections were scanned using the Pannoramic 250 Flash (3DHistech, Hungary) and reviewed using the Pannoramic Viewer.

4. Sample inclusion criteria

From 1,438 out of 1,860 collected tumor specimens histologic sections were obtained and reviewed by a board-certified pathologist (**Prof. Sven Perner,** Pathology of the University of Bonn, Germany) to identify areas with high content of tumor nuclei. 1,320 lung tumor cases had a confirmed minimum of 70% tumor cell nuclei in sufficient quantity for further analysis.

In total 1,255 samples were morphologically confirmed as lung tumors, with a minimum of 70% tumor cell nuclei, with less then 10% necrosis required to meet further inclusion criteria for mutation, copy number and/ or gene expression analysis (e.g., sufficient DNA, RNA quantity and quality, affirmed agreement of the patient) and were therefore suitable for data analysis.

5. Estimation of specimen mix-up frequency

The ratio between the average copy number of X- and Y-chromosome was calculated. In order to predict the sex of each sample, this ratio was clustered into two partitions using the K-means algorithm. Cases with ambiguity were adjusted manually. The total mix-up rate was computed using

$$\mu = \frac{\mu_m, f}{2\mu_m\mu_f}$$

where μ_m was the fraction of male samples, μ_f the fraction of female samples and $\mu_{m,f}$ was the observed mix-up rate between male and female samples.

6. DNA and RNA procedures

Standard procedures including polymerase chain reaction (PCR), whole genome amplification, gel electrophoresis, restriction enzyme digestion, DNA ligations and bacterial transformations were carried out according to the manufacturers protocol. Preparation of plasmid DNA was performed using the NucleoSpin mini-kit (Machery Nagel).

DNA and RNA preparation

If the whole surface of the frozen tumor sample comprised a minimum of 70% tumor cells, 15 and 30 sections each 20 µm thick were prepared using the cryostat (Leica) at -20°C for DNA and RNA extraction. Otherwise, tumor rich areas were punched using sterile disposable biopsy punches (1.5mm in diameter) not deeper than 1.5 mm into the frozen tissue to avoid lower layers that might be admixed with higher content of fibrotic, necrotic or other non-tumor cells. Punched material was used for DNA extraction only, whereas sections were used for DNA and RNA extraction. HapMap normal cell lines were cultured at 37°C in RPMI 1640 medium supplemented with 10% fetal calf serum. For harvesting the cells were washed two times in phosphate-buffered saline and $2x10^7$ cells collected for DNA extraction.

DNA was extracted using Puregene Extraction Kit (Qiagen, Germany). DNA was eluted in 75 µl 1xTE buffer (pH 8.0). Sections prepared for RNA extraction were disrupted and homogenized for 2 min at 20 Hz with the Tissue Lyser (Qiagen, Germany) and centrifuged for 3 min at maximum speed. RNA was extracted using Qiagen RNeasy Mini Kit (Qiagen, Germany).

Quality and quantity control of DNA and RNA

High molecular weight of genomic DNA (>10kb) was confirmed by electrophoresis using a 0.7% agarose gel. DNA concentration was measured in 1 µl of the molecular analyte using the Quant-iT™ Picogreen® Assay (Invitrogen) and fluorescence emission intensity measured at 520 nm using a multimode reader Mithras LB940 (Berthold Technologies). DNA was further diluted in 1xTE buffer (pH 8.0) to a concentration between 75 and 150 ng/µl. If DNA concentration was below 75 ng/µl precipitation was performed as recommended (Puregene Extraction Kit, Qiagen).

RNA quality was assessed using the Bioanalyzer 2100 DNA Chip 7500 (Agilent) and samples with a RIN>8 were processed for gene expression analysis. RNA and DNA stocks were stored at -80°C.

Phi29-based whole genome amplification (WGA)

Genomic DNA (100 ng) was amplified using the Repli-g Midi Kit (Qiagen). Amplified DNA was diluted 1:10 in 1xTE buffer (pH 8.0) and hydrated for 24 h at room temperature and further diluted to a working concentration of 5 ng/µl in water.

7. Mutation detection and validation

Mutation detection

Mutations were detected in whole genome-amplified DNA using a mass spectrometry-based single base extension technique (Sequenom, Inc.) as described previously[18]. Primer sequences of detection assays were applied from Thomas et al[18]. Additional assays detecting mutations that were described as somatic in several cancer studies from 2008 to 2010[16,83,106] (Table 3) were designed using the Sequenom Assay Design Software leading to a set of 327 mutations in 26 genes (primer sequences for new assays are not provided here, since table extended reasonable size). Primers were purchased from Integrated DNA Technologies (Belgium). Following target site amplification, phosphatase treatment and mutation site specific probe elongation analytes were spotted on SpectroCHIPs I and masses detected using a Bruker matrix-assisted laser desorption/ionization–time of flight mass spectrometer (Sequenom). Mass signals of mutant and wild type alleles were analyzed using the Typer4.0 software (Sequenom).

Table 3 Mutation sites by which the Oncomap panel[18] for mass spectrometry-based mutation detection has been extended.

Gene	Protein_change	Nucleotide_change
ALK	p.A1234T	c.3700G>A
ALK	p.D1091N	c.3271G>A
ALK	p.F1174I	c.3520T>A
ALK	p.F1174L	c.3522C>A
ALK	p.F1174V	c.3520T>G
ALK	p.F1245C	c.3734T>G
ALK	p.F1245V	c.3733T>G
ALK	p.I1171N	c.3512T>A
ALK	p.I1250T	c.3749T>C
ALK	p.M1166R	c.3497T>G
ALK	p.R1275Q	c.3824G>A
ALK	p.V1135E	c.3404T>A
DDR2	p.C580Y	c.1739G>A
DDR2	p.G253C	c.757G>T
DDR2	p.G505A	c.1514G>C
DDR2	p.G774E	c.2321G>A
DDR2	p.I120M	c.360C>G
DDR2	p.I638F	c.1912A>T
DDR2	p.L239R	c.716T>A
DDR2	p.L63V	c.187C>G
DDR2	p.T765P	c.2293A>C
EPHA3	p.A435S	c.1303G>T
EPHA3	p.D446Y	c.1336G>T
EPHA3	p.D678E	c.2034C>A
EPHA3	p.G518L	c.1552_1553GG>TT
EPHA3	p.G766E	c.2297G>A
EPHA3	p.K761N	c.2283G>T
EPHA3	p.M269I	c.807G>A
EPHA3	p.N379K	c.1137T>G
EPHA3	p.R728L	c.2183G>T
EPHA3	p.S229Y	c.686C>A
EPHA3	p.T166N	c.497C>A
EPHA3	p.T166N	c.497C>A
EPHA3	p.T393K	c.1178C>A
EPHA3	p.W250R	c.748T>A
EPHA5	p.D493Y	c.1477C>A
EPHA5	p.M1034I	c.3102C>T
EPHA5	p.P1036A	c.3106G>C
EPHA5	p.R1007Q	c.3020C>T
EPHA5	p.S566Y	c.1697G>T
EPHA5	p.S810I	c.2429C>A
ERBB2	p.C334S	c.1000T>A
ERBB2	p.D326G	c.977A>G
ERBB2	p.E321G	c.962A>G
ERBB2	p.L49H	c.146T>A
ERBB2	p.M774_A775insAYVM	c.2322_2322insGCATACGTGATG
ERBB2	p.N319D	c.955A>G
ERBB2	p.S310F	c.929C>T
ERBB2	p.T216S	c.646A>T
ERBB2	p.T216S	c.646A>T
ERBB2	p.V750E	c.2249T>A
ERBB2	p.W906*	c.2718G>A
ERBB2	p.W906*	c.2718G>A
ERBB4	p.D931Y	c.2791C>A
ERBB4	p.H618P	c.1853T>G
ERBB4	p.V348L	c.1042G>T
ERBB4	p.Y285C	c.854T>C
FGFR1	p.V664L	c.1990G>T
FGFR2	p.C382R	c.1144T>C
FGFR2	p.D283N	c.847G>A

Gene	Protein_change	Nucleotide_change
FGFR2	p.G272V	c.815G>T
FGFR2	p.H544Q	c.1632C>A
FGFR2	p.N211I	c.632A>T
FGFR2	p.N211I	c.632A>T
FGFR2	p.N549K	c.1647T>A
FGFR2	p.P253R	c.758C>G
FGFR2	p.Q212K	c.634C>A
FGFR2	p.R203C	c.607C>T
FGFR2	p.R496T	c.1487G>C
FGFR2	p.S252W	c.755C>G
FGFR2	p.W290C	c.870G>C
FGFR2	p.W290C	c.870G>C
FGFR2	p.Y375C	c.1124A>G
FGFR3	p.T79S	c.235A>T
FGFR4	p.A729G	c.2186C>G
FGFR4	p.E681K	c.2041G>A
FGFR4	p.H192fs	c.576delT
FGFR4	p.P672T	c.2014C>A
FGFR4	p.R183S	c.547C>A
FGFR4	p.R411fs	c.1230_1240delCCGCTTCCCTC
FGFR4	p.R616G	c.1846C>G
FGFR4	p.S732N	c.2195G>A
FGFR4	p.V510M	c.1528G>A
KDR	p.A248G	c.743C>G
KDR	p.G1145E	c.3434C>T
KDR	p.G1308*	c.3922C>A
KDR	p.L1140M	c.3418G>T
KDR	p.Q2R	c.5A>G
KDR	p.S984T	c.2951C>G
KIT	p.554-559delEVQWKV	c.1660_1678delGAAGTACAGTGGAAGGTT
KIT	p.A829P	c.2485G>C
KIT	p.K558N	c.1674G>C
KIT	p.N495I	c.1484A>T
KIT	p.N566D	c.1696A>G
KIT	p.N567K	c.1701T>A
KIT	p.Y823D	c.2467T>G
NTRK1	p.A107V	c.320C>T
NTRK1	p.D776E	c.2328T>G
NTRK1	p.G368C	c.1102G>T
NTRK1	p.Q80*	c.238C>T
NTRK1	p.R119H	c.356G>A
NTRK1	p.S326R	c.978C>G
NTRK1	p.V422L	c.1264G>T
NTRK3	p.H677Y	c.2029C>T
NTRK3	p.I783N	c.2348A>T
NTRK3	p.L152I	c.454G>T
NTRK3	p.L270M	c.808G>T
NTRK3	p.R678Q	c.2033G>A
NTRK3	p.R721F	c.2161_2162CG>TT
NTRK3	p.T283K	c.848G>T
PDGFR	p.A210V	c.629C>T
PDGFR	p.A210V	c.629C>T
PDGFR	p.D480E	c.1440C>G
PDGFR	p.E338Q	c.1012G>C
PDGFR	p.L580M	c.1738C>A
PDGFR	p.M133K	c.398T>A
PDGFR	p.V193I	c.577G>A
PDGFR	p.W549*	c.1647G>A
PIK3CA	p.M1043V	c.3127A>G

In addition, selected exons in *BRAF* (exons 11 and 15), *EGFR* (exons 18 to 21), *ERBB2 (*exons 19 and 20), *FGFR2* (exons 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, and 17), *KRAS* (exons 2 and 3), *KEAP1* (exons 3 to 6), *NFE2L2* (exon 2), *PIK3CA* (10 and 21), *STK11* (exons 1 to 10), and *TP53* (exons 5 to 9) were analyzed using bidirectional Sanger sequencing (Beckman Coulter Genomics, USA). Primers were

designed by Beckman Coulter (primer sequences listed in Table 4). Either 1,141 WGA samples (*EGFR, KRAS, STK11, TP53*) or 844 WGA samples (*BRAF, ERBB2, FGFR2, KEAP1, NFE2L2, PIK3CA*) were sent for analysis. Sequencing results were manually revised using the Mutation Surveyor® v.2.61 (Soft Genetics).

Results of both sequencing methods (mass spectrometry and dideoxy sequencing) per gene were combined and a tumor sample was considered to be: mutated, if a mutation was detected by mass spectrometry or dideoxy sequencing; wild type, if at least one method covered all mutation sites given in the Oncomap panel v.1x but no mutation was called by neither of these two methods; failed, if no mutation was called by neither of these two methods and coverage of the analyzed region was incomplete for both methods.

Table 4 Primers used for sequencing specific exons in *BRAF, EGFR, ERBB2, FGFR2, KEAP1, NFE2L2, PIK3CA, STK11*, and *TP53*.

Gene	Exon	Primer Sequence_forward	Primer Sequence_reverse
BRAF	11	GTAAAACGACGGCCAGTTGTATCCCTCTCAGGCATAAGG	CAGGAAACAGCTATGACCACAAAATAAAAGTTGTTAAACATATCC
BRAF	15	GTAAAACGACGGCCAGTTTCATAATGCTTGCTCTGATAGG	CAGGAAACAGCTATGACCAGTAACTCAGCAGCATCTCAGG
EGFR	18	GTAAAACGACGGCCAGTAGGCGTACATTTGTCCTTCC	CAGGAAACAGCTATGACCGCCTTTGGTCTGTGAATTGG
EGFR	19	GTAAAACGACGGCCAGTCGCAGCATCATTAAATTCTGG	CAGGAAACAGCTATGACCGGCCAGTGCTGTCTCTAAGG
EGFR	20	GTAAAACGACGGCCAGTCCCTGTGCTAGGTCTTTTGC	CAGGAAACAGCTATGACCTTGATGAGAGTTTCCACATGC
EGFR	21	GTAAAACGACGGCCAGTGCTCCTGCTCTTCTTTGTCC	CAGGAAACAGCTATGACCACCCAGAATGTCTGGAGAGC
ERBB2	19	GTAAAACGACGGCCAGTGTGAAGGTGCTTGGATCTGG	CAGGAAACAGCTATGACCGGAGTCATATCTCCCCAAACC
ERBB2	20	GTAAAACGACGGCCAGTAGGGTTTCACCATGTTGTCC	CAGGAAACAGCTATGACCACTCTTGACCAGCACGTTCC
FGFR2	3	GTAAAACGACGGCCAGTCAGAGCAAGAACAGCTTCTCC	CAGGAAACAGCTATGACCTTAATGATCGGCCTTTCTGG
FGFR2	5	GTAAAACGACGGCCAGTCATGAACCATCCCTCTCTGG	CAGGAAACAGCTATGACCTGAAAGCTTAATTCTACCTTGTAGCC
FGFR2	6	GTAAAACGACGGCCAGTCCTTTCTTGCCTCCTTCAGC	CAGGAAACAGCTATGACCCCATGAGGATCATGCAAAGC
FGFR2	7	GTAAAACGACGGCCAGTAGGCTTTTCTGGCATGAGG	CAGGAAACAGCTATGACCGAGAATCATCCTCTCTCAACTCC
FGFR2	8	GTAAAACGACGGCCAGTTTTCCATGCGTTTGATTGC	CAGGAAACAGCTATGACCTTCCAAGGCAGTTTTCTTATCC
FGFR2	10	GTAAAACGACGGCCAGTTCTGGTACTGCATGGACTGG	CAGGAAACAGCTATGACCCCTGCTGACATCATCACACC
FGFR2	11	GTAAAACGACGGCCAGTACCGCACTAGCAAGGATACC	CAGGAAACAGCTATGACCCCGAGATGCTGATTTATACCG
FGFR2	12	GTAAAACGACGGCCAGTGAGGCAGTTCCCTGTGTAGG	CAGGAAACAGCTATGACCATAAGGAGGCTGCCTTTTCC
FGFR2	13	GTAAAACGACGGCCAGTAAAAATGTTTTGCTGAATTGC	CAGGAAACAGCTATGACCGGTTTATGAGGCTGCTTTGG
FGFR2	14	GTAAAACGACGGCCAGTAACATCAGCTATATTTTCTATCTGC	CAGGAAACAGCTATGACCAGTCAAAAGAACGGGAATCG
FGFR2	15	GTAAAACGACGGCCAGTTTTTATTAACGGCCCAGACC	CAGGAAACAGCTATGACCAGCAGTTCCACCTTCTGTGC
FGFR2	16	GTAAAACGACGGCCAGTGGCAGGAAAGAGCACATAGG	CAGGAAACAGCTATGACCAAGTAATGGTTGTCGGTGTCG
FGFR2	17	GTAAAACGACGGCCAGTCCCCCTAATCTAGTTGCTTGG	CAGGAAACAGCTATGACCTTCCCACTCTCCCACATACC
KEAP1	3	GTAAAACGACGGCCAGTCTGCAGAAGTGCGAGATCC	CAGGAAACAGCTATGACCTCCCTGAAGACAGGAAGAGG
KEAP1	4	GTAAAACGACGGCCAGTAAGTATTCCACGAAGGTCAGC	CAGGAAACAGCTATGACCGGGGAGAGAGAGAAGCTTGG
KEAP1	5	GTAAAACGACGGCCAGTCACAGCAATGAACACCATCC	CAGGAAACAGCTATGACCAGCAAAAGCAAAAGCAGTCC
KEAP1	6	GTAAAACGACGGCCAGTCAATCCTCCTGCCTCAGC	CAGGAAACAGCTATGACCCAGGCTGTCTTGGACACTCC
KRAS	2	GTAAAACGACGGCCAGTTTTTTCTTAAGCGTCGATGG	CAGGAAACAGCTATGACCGCACAGAGAGTGAACATCATGG

Gene	Exon	Primer Sequence_forward	Primer Sequence_reverse
KRAS	3	GTAAAACGACGGCCAGTTTTCAAGTCCTTTGCCCATT	CAGGAAACAGCTATGACCTCTAAAAAGTTTAAAGTCTTGCTTTT
NFE2L2	2	GTAAAACGACGGCCAGTTTTGTAATCTCCCCACTTCC	CAGGAAACAGCTATGACCATCAGGAGGCTGAGGTTGG
PIK3CA	10	GTAAAACGACGGCCAGTCCTGTCTCTGAAAATAAAGTCTTGC	CAGGAAACAGCTATGACCAAGCATTTAATGTGCCAACTACC
PIK3CA	21	GTAAAACGACGGCCAGTAAAGGGAATCAAAAGATGTTGG	CAGGAAACAGCTATGACCATGCTGTTCATGGATTGTGC
STK11	1	GTAAAACGACGGCCAGTAAGTCGGAACACAAGGAAGG	CAGGAAACAGCTATGACCCCTTGCCTGAGTGAAAGTCC
STK11	2	GTAAAACGACGGCCAGTGGCCGATGACAGACTAGAGG	CAGGAAACAGCTATGACCGCAGACCGTGGCTACACC
STK11	3	GTAAAACGACGGCCAGTCCAAGAGTCAGCCCTGTCC	CAGGAAACAGCTATGACCCCTTCATGTCAATGAATATCAGG
STK11	4	GTAAAACGACGGCCAGTCTGCTGGACCTAGCCTTTCC	CAGGAAACAGCTATGACCAGTGTGCGTGTGGTGAGTGC
STK11	5	GTAAAACGACGGCCAGTGCCTGGAGTACCTGCATAGC	CAGGAAACAGCTATGACCCTCTGTGAAGATCCCTGACG
STK11	6	GTAAAACGACGGCCAGTGGTGTCCTTGAGTCCACAGG	CAGGAAACAGCTATGACCCCCCAACCCTACATTTCTGC
STK11	7	GTAAAACGACGGCCAGTGTCCTGCTGCACTTCCTACG	CAGGAAACAGCTATGACCCTCACTCAGACCCCAGTTCG
STK11	8	GTAAAACGACGGCCAGTACATGGCTGAGCTTCTGTGG	CAGGAAACAGCTATGACCGAGAGACGCCTGAGGAACC
STK11	9	GTAAAACGACGGCCAGTGATACACCTGGGCCTGACC	CAGGAAACAGCTATGACCACACGTAGGCCTCCATGACC
STK11	10	GTAAAACGACGGCCAGTCCCCAGGAGTCCGGTAGC	CAGGAAACAGCTATGACCATGCATGGCGGGGTCAGG
TP53	5	GTAAAACGACGGCCAGTTGAGGTGTAGACGCCAACTCT	CAGGAAACAGCTATGACCACACGCAAATTTCCTTCCAC
TP53	6	GTAAAACGACGGCCAGTTTGCTTTATCTGTTCACTTGTGC	CAGGAAACAGCTATGACCCACATCTCATGGGGTTATAGGG
TP53	7	GTAAAACGACGGCCAGTCTGCTTGCCACAGGTCTCC	CAGGAAACAGCTATGACCGGCTCCATCTACTCCCAACC
TP53	8+9	GTAAAACGACGGCCAGTAGGCTCCAGAAAGGACAAGG	CAGGAAACAGCTATGACCTGTCTTTGAGGCATCACTGC

Mutation validation

Putative mutations detected in WGA samples were validated in genomic DNA utilizing exon-wise dideoxy sequencing. Primers spanning the mutation site were designed using Primer3 v.0.4.0 software (http://frodo.wi.mit.edu/primer3/) and synthesized by Eurofins MWG Operon (Germany). Sequencing was performed at the Cologne Center for Genomics (CCG) using standard dideoxy-sequencing (BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems)) and results manually revised using the Mutation Surveyor®.

Alterations detected by mass spectrometry were considered technically validated if they were detected in at least two independent analyses of which at least one was performed in genomic DNA. Alterations found in sequenced genes using Sanger sequencing were filtered for variants using dbSNP. For a small set of alterations that were not described as somatic in cancer (http://www.sanger.ac.uk/genetics/CGP/cosmic/), normal tissue was requested from collaborators and validation performed as described above. The majority of *TP53* mutations found were not validated[81].

8. Fluorescence in situ hybridization

To determine genomic rearrangements affecting *ALK*, *RET* or *ROS1* a subset of the specimens were analyzed using Fluorescence *in situ* hybridization (FISH).

Tissue microarrays were constructed as previously described.[151] In brief, sections from formalin-fixed paraffin-embedded (FFPE) primary lung carcinoma samples were stained with hematoxylin and eosin to determine normal and tumor areas. Three representative cylindrical tissue cores (each 0.6 mm in diameter) per patient were taken to assemble a tissue microarray (TMA) using the semiautomatic Tissue Arrayer MTA-1 (Beecher Instruments, USA). Identically, for a subset of cases, benign tissue samples were included in the TMA as controls. 4 µm thick sections were cut from these TMA blocks and mounted onto positively charged glass slides.

ALK and *ROS1* FISH

(in collaboration with **Toni-Maree Rogers** and **Ben Solomon**, Peter MacCallum Cancer Centre, Melbourne, Australia and **Hans-Ulrich Schildhaus**, Institute of Pathology, Cologne)

In brief, TMA sections were de-paraffinized (124°C for 2min), treated with protease (30 min at room temperature), washed in distilled water and dehydration performed in graded alcohol (70%, 85% and 100%). After pretreatment the sections were denatured in the presence of 10 µl probe for 5 min at appropriate temperature and hybridized at 37°C over night. After hybridization sections were washed (2xSSC), counterstained with 4,6-Diamidino-2-phenylindole, dihydrochlorid (DAPI) and mounted. Sections were hybridized with the Vysis LSI ALK Dual Color Break Apart Rearrangement Probe (Abbott Molecular, USA) or Cytocell Aquarius ROS1 Break Apart probe (Cytocell, UK). To determine the *ALK* and *ROS1* fusion status 50 tumor cells were counted per specimen.

Cells were considered FISH-positive if orange (*ALK*) or red (*ROS1*) and green signals are at least two signal widths apart or at least one set of orange and green signals are two or more signal widths apart and there was a single orange signal (3' end) without a corresponding green signal in addition to fused and/or split signals.

A case was considered *ALK* or *ROS1* fusion positive if ≥15% of the tumor cells showed a positive signal pattern.

ALK fusion status was analyzed for additional 70 lung tumor samples as described in Perner et al 2008.[152]

ALK FISH was repeated for four *ALK*-rearranged cases (patient IDs: S00006, S00054, S00092, S01339) to identify *EML4-ALK* fusions. FFPE sections were hybridized with the Zytolight® SPEC ALK/EML4 TriCheck™ Probe (ZytoVision, Bremerhaven, Germany). Sections were analyzed using 63x and 100x objectives (DM5500 fluorescent microscope, Leica). In S00006 and S00054 *EML4* was confirmed as the fusion partner. In three additional cases (patient IDs: S01122, S01124, S01320) *EML4* was identified as the fusion partner in *ALK* positive cases using RNAseq (by **Lynnette Fernandez Cuesta**). *EML4-ALK* fusions were confirmed in five out of seven tested cases.

RET FISH

(in collaboration with **Sven Perner**, Pathology of the University of Bonn, Germany)
In brief, *RET* FISH probes were denatured at 73°C for 5 min and immediately placed on ice. After de-paraffinization and protease treatment (Digest-All III (1:2) at 37°C for 14 min) sections were hybridized with the biotin labeled centromeric BAC clones CTD2105E16 and the digoxigenin labeled telomeric BAC clone CTD2348B12 (Invitrogen, CA, USA) overnight at 37°C. After hybridization sections were washed (2xSSC) and fluorescence detection was carried out using streptavidin-Alexa-594 conjugates (dilution 1:200) and anti-digoxigenin-FITC (dilution 1:200). Sections were counterstained with DAPI. Samples were analyzed under a 63x oil immersion objective using a fluorescence microscope (Zeiss, Jena, Germany). At least 100 nuclei per case were evaluated.[153] No case with a *RET* rearrangement could be identified in the samples tested.

9. Copy number analysis

Copy number analysis was performed using data of 1,032 cases. Genomic DNA of tumor samples and HapMap or matched normal DNA were hybridized to Affymetrix Genome Wide Human SNP array 6.0 following manufacturers' instructions. Per 79 tumor cases 15 normal cases were analyzed in the same batch as a reference for following data processing. Raw signal intensities were processed as described previously[82]. Segmented copy number data were visualized in the integrated genome viewer (http://www.broadinstitute.org/igv/). SNP, gene and cytogenetic band locations are based on the hg18 (March 2006) genome build.

To determine significantly amplified and deleted regions a rank-sum based algorithm was applied as described previously.[81] To further quantify significant copy number aberrations an algorithm was implemented to identify samples altered in these regions. In brief, a representative copy number segment for each region of interest was selected: the segment of the highest copy number for amplifications and the lowest in case of deletions. Then the standard deviation from location-representative copy numbers was calculated across all samples to define a region-dependent copy number threshold of 1.96 times standard deviation above copy number two for amplifications and 1.65 times standard deviation below copy number two in case of deletions. To further distinguish between high-level focal amplification events a criterion that allowed discrimination between focal and broad amplifications was defined. This criterion is based on the observation that segment-length and copy number is correlated[81] and fitted an exponentially decaying curve to segment-length rank ordered copy numbers. The segment-length rank where the exponential reaches 5% of its initial value was then taken as focality threshold. Finally, to sensitize the algorithm to impure samples (e.g., due to the admixture of non-tumor cells), the copy number of each region was screened if it was sticking out of the genome-wide total variation using the same procedure and parameters as within the location. All samples that were identified as being amplified had to fulfill the focality criterion.

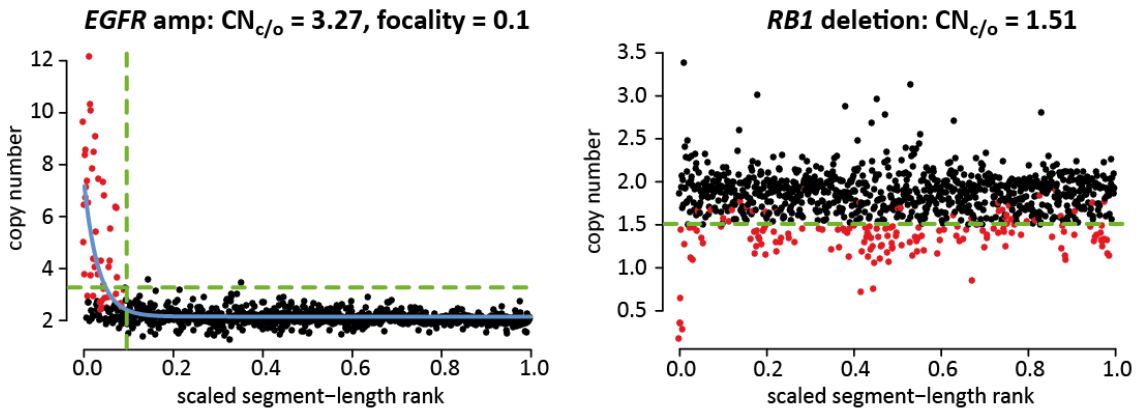

Figure 29 Examples for selection of amplified (*EGFR*, left) and deleted (*RB1*, right) cases. Each dot represents a case. Green lines indicate copy number cutoffs ($CN_{c/o}$) and in case of amplifications also the focality cutoff. Cases in red are determined amplified or deleted by the defined criteria.

To assess the purity and absolute copy numbers in a subset of 920 samples, a recently published algorithm was applied to the SNP 6.0 data[1]. Assessing these values was technically possible for 696 samples using this approach.

To determine the copy number of specific genomic regions (genes and larger regions) the median copy number was calculated using segmented data. If one region comprises more than one segment, the fractions were weighed according their lengths (Figure 20).

$$CN_m = \frac{\sum_i (length_i \times CN_i)}{\sum_i (length_i)}$$

The cutoff used to specify a copy number amplification was 2.7 and 1.3 for a deletion (loss) based on these calculated median copy numbers per gene/ region.

10. Exclusivity of genomic alterations

Significant associations between copy number alterations and frequently mutated genes across all lung tumors and histological subtypes were determined applying the Fisher's exact test (p-value ≤ 0.05) and the Benjamini and Hochberg algorithm to correct for multiple comparisons (q-value ≤ 0.05).

11. Expression analysis

Expression arrays

Array-based gene expression analysis of 261 tumor samples was performed on Illumina® Human HT-12 V3 Expression BeadChips according standard protocols (Illumina). For data collection, we used Illumina BeadStudio 3.1.1.0 software.

Data processing and gene selection for hierarchical clustering

To identify variable genes the standard deviation of expression data was calculated per gene across 261 tumor samples and the data normalized by standardizing each gene to the mean. Using the cutoff of 2.1 resulted in 294 genes (Appendix Table 5) that were further used for unsupervised hierarchical clustering using the dCHIP software (http://biosun1.harvard.edu/complab/dchip/). Values were analyzed using an average-linkage hierarchical algorithm that resulted in four major expression clusters.

Table 5 294 genes selected for hierarchical clustering. Genes are listed that show high variance of gene expression across 261 cases of all lung cancer subtypes that were used for unsupervised hierarchical clustering.

ABCC8	*CXCL6*	*GNRHR*	*LOC728910*	*PGC*	*SPANXB2*
ADCYAP1	*CYP4F11*	*GP2*	*LOC730994*	*PHACTR3*	*SPANXC*
ADH1A	*DCT*	*GPC5*	*LOR*	*PHOX2B*	*SPANXE*
ADH1C	*DDC*	*GRIA2*	*LY6H*	*PI3*	*SPHKAP*
ADH7	*DEFA5*	*GRM3*	*MAGEA8*	*PIGR*	*SPINK1*
AGT	*DEFB1*	*GRM8*	*MAPK4*	*PIP*	*SPINK5*
ALB	*DEFB4*	*GRP*	*MEG3*	*PLA2G10*	*SPRR1A*
ALDH3A1	*DHRS2*	*GSTA1*	*MMP1*	*PLA2G1B*	*SPRR2A*
AMY1B	*DIRAS3*	*GUCA2B*	*MMP10*	*PLUNC*	*SPRR2C*
ANXA10	*DLK1*	*H19*	*MMP3*	*POMC*	*SPRR2D*
APCDD1L	*DLX5*	*HAND1*	*MS4A8B*	*POU4F1*	*SPRR2E*
APOH	*DNER*	*HLA-DQB2*	*MSMB*	*POU4F2*	*SPRR2F*
ART3	*DRD1IP*	*HMGCS2*	*MT1G*	*PPY*	*SPRR2G*
ASCL1	*DSCR6*	*HOXD1*	*MT1H*	*PRPH*	*SPRR3*
AVP	*EBF3*	*HPGD*	*MUC13*	*PRSS2*	*SRL*
BHMT	*ECEL1*	*IGFBP1*	*MUC5AC*	*PRSS3*	*SST*
BPIL1	*ELMOD1*	*IL13RA2*	*MUCL1*	*PTHLH*	*STATH*
C20ORF114	*F5*	*IL1F7*	*NDST4*	*PTPRN*	*STMN2*
C4ORF7	*FABP4*	*IL1F9*	*NEFL*	*RALYL*	*SYT13*
C6ORF15	*FABP7*	*INSM1*	*NEFM*	*RCVRN*	*SYT4*
C6ORF205	*FAIM2*	*INSM2*	*NELL1*	*RDH12*	*TAC1*
C7ORF16	*FBN2*	*ISL1*	*NEUROG2*	*REG4*	*TAGLN3*
C8ORF46	*FEV*	*KCNS1*	*NGB*	*RNASE7*	*TCEAL2*
CABP7	*FGA*	*KIR2DL3*	*NKD1*	*RPRM*	*TCN1*

CALCA	*FGB*	*KIR2DL4*	*NKX2-2*	*S100A7*	*TF*
CARTPT	*FGFBP2*	*KIR2DL5A*	*NNAT*	*S100A7A*	*TFF1*
CBLN2	*FGG*	*KIR2DS5*	*NOL4*	*SBSN*	*TFF2*
CCL17	*FGL1*	*KLK12*	*NOS2A*	*SCG2*	*TFPI2*
CD7	*FLJ23834*	*KLK5*	*NOTUM*	*SCG5*	*THBS4*
CDH10	*FMO3*	*KRT13*	*NPAS4*	*SCGB1A1*	*TKTL1*
CDH17	*FRMPD1*	*KRT14*	*NPPA*	*SCGB2A1*	*TM4SF4*
CEL	*GABBR2*	*KRT15*	*NPTX2*	*SCGB3A1*	*TM4SF5*
CGA	*GABRB1*	*KRT34*	*NRSN1*	*SCGN*	*TMED6*
CHGA	*GABRP*	*KRT6C*	*NT5E*	*SERPINA3*	*TMEFF2*
CHGB	*GAGE12G*	*KRT81*	*NTS*	*SERPINB3*	*TMEM100*
CHODL	*GAGE12I*	*KRTDAP*	*OLFM4*	*SERPINB4*	*TPRX1*
CLCA4	*GAGE2A*	*LBP*	*OTP*	*SFRP1*	*TRH*
CLDN11	*GAGE4*	*LECT1*	*OTX2*	*SHISA3*	*TRPM8*
CLDN18	*GAGE5*	*LEFTY1*	*PAEP*	*SLC18A1*	*TSPY2*
CLDN2	*GAGE6*	*LGALS4*	*PAGE2*	*SLC1A7*	*TTR*
CPB2	*GAGE7*	*LGALS7*	*PCK1*	*SLC35D3*	*UPK1A*
CPNE4	*GAGE8*	*LOC284422*	*PCSK1*	*SLC8A2*	*VGF*
CPS1	*GAL*	*LOC645037*	*PCSK1N*	*SNAP25*	*VSIG1*
CRABP1	*GC*	*LOC651898*	*PDE6H*	*SNAP91*	*VSIG2*
CRYBA2	*GFRA3*	*LOC652102*	*PDK4*	*SOST*	*VTN*
CSN1S1	*GJB6*	*LOC652683*	*PDYN*	*SOX10*	*WIF1*
CST1	*GKN2*	*LOC653178*	*PEG10*	*SOX8*	*WT1*
CTCFL	*GNAT2*	*LOC728403*	*PEG3*	*SPANXA2*	*ZCCHC12*
CXCL13	*GNGT1*	*LOC728454*	*PENK*	*SPANXB1*	*ZIC2*

12. Statistical analysis

Survival was calculated from the day of diagnosis until death or the last follow-up visit. Associations between clinical characteristics were analyzed with the use of the Chi-square test or Fisher's exact test. Survival curves were constructed by the Kaplan–Meier method and compared with the use of the log-rank test. Statistical tests were performed using SPSS version 21.

13. Cell culture techniques and required molecular biological methods

Standard procedures including restriction enzyme digestion and bacterial transformations were carried out according to the manufacturers protocol.

Site directed mutagenesis

Site directed mutagenesis (SDM) was used to introduce mutations at specific sites into the pBabe-puro (pBp) plasmid harboring the *EGFR wild type* cDNA. Mutagenic oligonucleotides were designed using the QuikChange Primer Design (www.agilent.com/genomics/qcpd) (Primers listed in Table 6) and reactions

performed using the QuikChange II Site-Directed Mutagenesis Kit according manufacturer's instructions (Stratagene). Emanating from the *EGFR wild type* pBp plasmid the plasmids pBp *EGFR K714N,* pBp *EGFR E709K+V717E+G719R* (further referred to as pBp *EGFR+3*), and pBp *EGFR E709K+K714N+V717E+G719R* (further referred to as pBp *EGFR+4*) were generated.

Bacterial transformation

Molecules containing *EGFR* mutant variants produced by SDM were transformed into bacterial XL10-Gold ultracompetent cells. Plasmids with *FGFR3IIIB* wild type and *FGFR3IIIB* harboring the R248C mutation (*FGFR3* plasmids were received as a kind gift from Margaret Knowles, Professor of Experimental Cancer Research of the University of Leeds, England) were transformed into DH5α cells. In brief, 45 µl of cells and 2 µl Dpn I-treated DNA or plasmid were incubated for 30 min on ice followed by a 30 sec heat-pulse in a 42°C water bath. Transformed cells were plated on LB-agar plates containing Ampicillin.

Genotypes were verified in all plasmids using standard dideoxy-sequencing (BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems)). For this purpose five clones per plasmid were picked from LB agar plates 24 h after transformation and cells transferred into the reaction mix for target amplification using cDNA specific primers (primers are listed in Table 6).

Table 6 Primers used for site directed mutagenesis to introduce mutations into the wild type *EGFR* pBabe-puro plasmid and cDNA sequencing of *EGFR* and *FGFR3*.

gene	oligoname	Primer Sequence
SDM	K714N	GGATCTTGAAGGAAACTGAATTCAAAAATATCAAAGTGCTGGGC
SDM	K714N_anti	GCCCAGCACTTTGATATTTTTGAATTCAGTTTCCTTCAAGATCC
SDM	E709K	CAAGCTCTCTTGAGGATCTTGAAGAAAACTGAATTCAAAAATATCAAAG
SDM	E709K_anti	CTTTGATATTTTTGAATTCAGTTTTCTTCAAGATCCTCAAGAGAGCTTG
SDM	V717E_G719R	TGAATTCAAAAATATCAAAGAGCTGCGCTCCGGTGCGTTCG
SDM	V717E_G719R_anti	CGAACGCACCGGAGCGCAGCTCTTTGATATTTTTGAATTCA
Sanger	EGFR_ATG-73_F	GTAAAACGACGGCCAGGCCTCGATCCTCCCTTTATC
Sanger	EGFR_1_90_F	GTAAAACGACGGCCAGCTGGAGGAAAAGAAAGTTTGC
Sanger	EGFR_1_630_R	CAGGAAACAGCTATGACCTGGCACAGATGATTTTGGTCA
Sanger	EGFR_2_590_F	GTAAAACGACGGCCAGGTGATCCAAGCTGTCCCAAT
Sanger	EGFR_2_1150_R	CAGGAAACAGCTATGACCTTGTGGATCCAGAGGAGGAGT
Sanger	EGFR_3_1090_F	GTAAAACGACGGCCAGCAAAAACTGCACCTCCATCA
Sanger	EGFR_3_1770_R	CAGGAAACAGCTATGACCTGGCAGGTCTTGACGCAGT
Sanger	EGFR_4_1740_F	GTAAAACGACGGCCAGGACGGGGACCAGACAACTG
Sanger	EGFR_4_2400_R	CAGGAAACAGCTATGACCTGTCTTTGTGTTCCCGGACAT

gene	oligoname	Primer Sequence
Sanger	EGFR_5_2380_F	GTAAAACGACGGCCAGAGCTCATCACGCAGCTCAT
Sanger	EGFR_5_3030_R	CAGGAAACAGCTATGACCTATGAGGTACTCGTCGGCATC
Sanger	EGFR_6_3000_F	GTAAAACGACGGCCAGACTTCTACCGTGCCCTGATG
Sanger	EGFR_6_3600_R	CAGGAAACAGCTATGACCTTGCTCCAATAAATTCACTGCTT
Sanger	EGFR_TGA+137_R	CAGGAAACAGCTATGACACTTTCCACACCTGGTTGCT
Sanger	FGFR3_1_ATG-93_F	CCACTGCTTACTGGCTTATCG
Sanger	FGFR3_1_228_R	ATCCTTGACCCAGACAGTGG
Sanger	FGFR3_2_233_F	ACTGTCTGGGTCAAGGATGG
Sanger	FGFR3_2_672_R	CCACGACGCAGGTGTAGTT
Sanger	FGFR3_3_618_F	CATTGGAGGCATCAAGCTG
Sanger	FGFR3_3_1135_R	GGATGCCTGCATACACACTG
Sanger	FGFR3_4_1028_F	CTCTGTCGAGCCACCAATTT
Sanger	FGFR3_4_1772_R	ACACCAGGTCCTTGAAGGTG
Sanger	FGFR3_5_1702_F	AAGGGTAACCTGCGGGAGT
Sanger	FGFR3_5_2361_R	GTGGGCAAACACGGAGTC
Sanger	FGFR3_6_2209_F	ACGAGTACCTGGACCTGTCG
Sanger	FGFR3_6_TGA+77_R	TCTCTCCATGTGCAGTGAGTCT

Midi-Prep

Positive clones were cultured in 100 ml LB medium containing 100 ng/ml Ampicillin at 37°C for 18 h and midi-preparations were performed using the NucleoBond Xtra Midi EF kit (Machery Nagel) according to manufacturer's instructions.

Virus production and transduction of Ba/F3 and NIH3T3 cells

HEK293T cells cultured in DMEM medium with 10% fetal calf serum (FCS) were transfected with retroviral plasmids (pBp empty vector, pBp *EML4-ALK* variant1 (produced by Johannes Heuckmann), pBp *EGFR wild type*, pBp *EGFR K714N,* pBp *EGFR+3*, and pBp *EGFR+4*, and pBp *FGFR3 IIIb wild type* and pBp *FGFR3 IIIb R248C* each mixed with packaging plasmid pCL-eco in Opti-Mem) using *Trans*IT®LTI Transfection Reagent (Mirus, USA). After 24 h medium was changed to 3 ml DMEM + 30% FCS + 1% Penicillin/Streptavidin (P/S). Medium was collected after 24 h, 3 ml new medium added to the cells and collected after additional 24 h. Combined supernatants were filtered through 0.45 µm filters.

NIH3T3 cells, cultured in DMEM + 10% FCS + 1% P/S and Ba/F3 cells, cultured in RPMI + 20% FCS + 1%P/S, were transduced at 50% confluence in a 60 mm dish with 1 ml virus using the cross linker Polybren®. IL-3 was added to the Ba/F3 cells and cells cultured for two days, before centrifuging and transferring

cells in IL-3 free medium. NIH3T3 cells were transferred one day after transfection into a 25 cm culture bottle and puromycin (2.5 µg/ml) added for selection.

Protein quantification and Western blot

Cells were lysed in lysis buffer (20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM Na2EDTA, 1 mM EGTA, 1% TritonX-100 2.5 mM sodium pyrophosphate, 1 mM beta-glycerophosphate, 1 mM Na3VO4, 1 µg/ml leupeptin, 1 mM PMSF, complete Protease Inhibitors Cocktail and Phosphatase Inhibitors Cocktail Set II) and protein concentration measured using the BCA Protein Assay (Thermoscientific) using 5 µl of the lysate.

A mixture of 40 µg proteins and 5x Laemmli buffer (5x SDS sample buffer: 250 mM Tris (pH 6.8), 50% glycerol, 5% β-mercaptoethanol, 10% SDS and 0.05% bromophenol blue) was incubated at 95°C for 10 min. Proteins were separated on Novex® 4-12% Tris-Glycin polyacrylamide gel (Invitrogen) in SDS running buffer (192 mM Glycin, 25 mM Tris-HCl, 0.1% SDS) at 100V in XCell SureLock Gel Chambers (Invitrogen) by electrophoresis. As a standard PageRulerTM Plus Prestained Protein Ladder (Fermentas) was used.

Soft Agar Anchorage-Independent Growth Assay

Agarose Type IX ultra low (Sigma Aldrich) was prepared using ddH2O as a 2% and 1.2% stock. Agar was heated to 95 °C for 30 minutes using a thermo block. After cooling down to approximately 60°C 25 µl of the 2% agar and 25 µl of pre-warmed 2x DMEM medium with 20% FCS and 2% P/S were prepared for each well of a 96-well plate. The 1% agar was added to the wells and hardened at room temperature.

For the top layer the 1.2% agar was further cooled to 40°C. NIH3T3 cells were detached by trypsin and cell numbers determined using the Z2-Coulter counter (Beckman Coulter). For each well of a 96-well plate 1,000 cells were mixed with 25 µl of 1.2% agar and pre-warmed 2x DMEM medium with 20% FCS and 2% P/S to a total volume of 50 µl. The cell-agar-mix was plated on the bottom agar layer and solidified for 10 min at 4°C. DMEM medium with 10% FCS + 1% P/S (and

appropriate volumes of compound) was added the next day and exchanged when necessary.

NIH3T3 colonies were counted microscopically (Zeiss Vert.A1 microscope) and pictures taken using the AxioVision 4.2 software.

Appendix

Table 7 Patient characteristics for *ALK* and *ROS1* fusion positive cases.

PatID	Gene	Histology	Sex	Age	Smoking_Hx	Stage (UICC)	Rearrangement comment
S01124	*ALK*	AD	female	52	never	IA	EML4-ALK
S00300	*ALK*	AD	female	59	never	IA	ALK_rearrangement
S01320	*ALK*	AD	female	62	never	IB	EML4-ALK
S00462	*ALK*	AD	female	55	current	IIB	ALK_rearrangement (no EML4-ALK)
S00006	*ALK*	AD	female	79	never	IIB	EML4-ALK
S01122	*ALK*	AD	female	42	never	IIIB	EML4-ALK
S01339	*ALK*	AD	male	71	former	IB	ALK_rearrangement (no EML4-ALK)
S00388	*ALK*	AD	male	56	current	IIIA	EML4-ALK
S00054	*ALK*	AD	male	52	never	IIIA	EML4-ALK
S00092	*ALK*	SQ	male	79	former	IB	ALK_rearrangement (no EML4-ALK)
S00545	*ROS1*	AD	female	68	never	IIIB	ROS1 rearrangement
S01345	*ROS1*	AD	male	74	former	IB	ROS1 rearrangement
S00310	*ROS1*	AD	male	66	former	IB	narrow split (signals <2 signal widths apart)

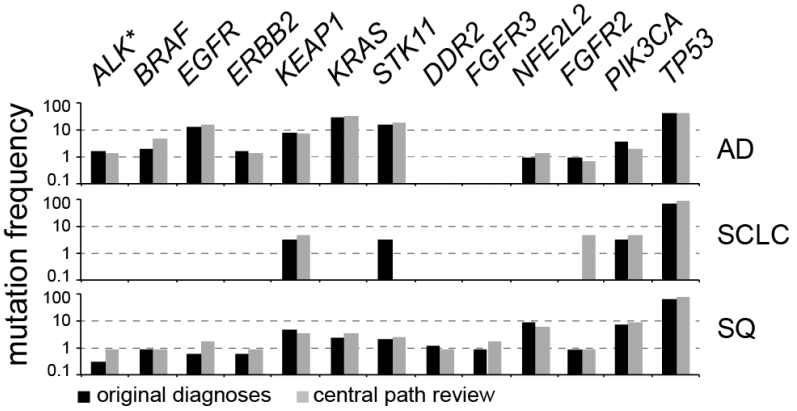

Figure 30 Mutation frequencies for major lung cancer subtypes (AD, SCLC and SQ) in sample sets defined by the original histology (black) or the reclassified histology (grey) plotted per subtype and gene. Mutation frequencies in both groups are highly similar (t-test for each gene per histological subtype, $p>0.05$).

Mutation frequencies described in the results section were calculated based on the original histological subtypes as these reflect best the real distribution in todays clinical diagnostics routine where diverse reasons might lead to misclassifications (see Figure 8). To rule out major misinterpretations that might affect mutation frequencies and thereby bias analysis of signature alterations mutation frequencies were compared between original and reviewed diagnosis

(Figure 30, Fishers Exact test, $p<0.05$). No significant differences of mutation frequencies between these two groups were found.

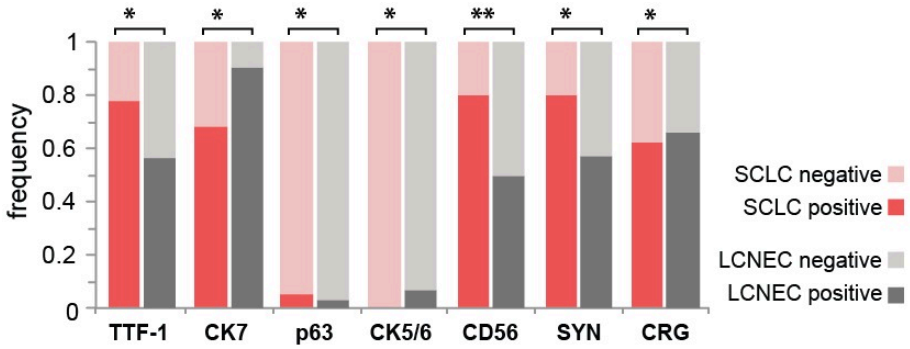

Figure 31 Comparison of expression of seven typical immunohistochemical markers in 32 LCNEC with 22 SCLC cases. Relationship between LCNEC and SCLC was calculated using the Pearson's Chi-square test (*p-value > 0.05, **p-value < 0.05).

Table 8 Mutations per case.

PatID	Histology	Gene	Aminoacid_change
S00006	AD	ALK	EML4-ALK
S00054	AD	ALK	EML4-ALK
S00300	AD	ALK	ALK
S00388	AD	ALK	EML4-ALK
S00462	AD	ALK	ALK
S01122	AD	ALK	EML4-ALK
S01124	AD	ALK	EML4-ALK
S01320	AD	ALK	EML4-ALK
S01339	AD	ALK	no EML4-ALK
S00092	SQ	ALK	no EML4-ALK
S00406	AD	BRAF	p.G464V
S00423	AD	BRAF	p.K601N
S00427	AD	BRAF	p.G466V
S00576	AD	BRAF	p.G469V
S00586	AD	BRAF	p.V600E
S00591	AD	BRAF	p.V600E
S00651	AD	BRAF	p.G469A
S00698	AD	BRAF	p.G466A
S01057	AD	BRAF	p.G469V
S01159	AD	BRAF	p.G464V
S01201	AD	BRAF	p.V600E
S01259	AD	BRAF	p.G464V
S01707	AD	BRAF	p.V600E
S00438	ADSQ	BRAF	p.G469V
S00496	LCC	BRAF	p.G464V
S00613	LCNEC	BRAF	p.V600E
S00713	LCNEC	BRAF	p.G469V
S00073	SQ	BRAF	p.V600E
S00117	SQ	BRAF	p.G464V
S00997	SQ	BRAF	p.V600E
S00402	UNKNOWN	BRAF	p.G469V
S00593	AD	CDK4	p.R24H
S00754	AD	CDK4	p.R24H
S00727	AD	DDR2	p.L63V
S01571	LCC	DDR2	p.G253C
S00069	SQ	DDR2	p.G774E
S00145	SQ	DDR2	p.C580Y
S00281	SQ	DDR2	p.I120M
S00434	SQ	DDR2	p.T765P
S00005	AD	EGFR	p.E746_A750del_1A
S00027	AD	EGFR	p.L747P
S00082	AD	EGFR	p.exon20indel
S00106	AD	EGFR	p.L858R
S00113	AD	EGFR	p.L858R
S00116	AD	EGFR	p.L747_S752>S
S00123	AD	EGFR	p.E746_A750del_1A
S00142	AD	EGFR	exon 20 insertion
S00155	AD	EGFR	p.E746_A750del_1B
S00156	AD	EGFR	p.(R776H(+)L858R)
S00299	AD	EGFR	p.E746_A750del_1B
S00302	AD	EGFR	p.(T790M(+)L858R)
S00304	AD	EGFR	p.L858R
S00308	AD	EGFR	p.(S768I(+)V774M)
S00398	AD	EGFR	p.T751_I759>N
S00403	AD	EGFR	exon 20 insertion
S00411	AD	EGFR	p.D770_N771insSVD
S00413	AD	EGFR	p.E746_A750del;insQP
S00416	AD	EGFR	p.E746_A750>A
S00430	AD	EGFR	exon 20 insertion
S00436	AD	EGFR	p.L858R
S00464	AD	EGFR	p.(L747_P753del>S(+)T790M)
S00465	AD	EGFR	p.L747_A750>P
S00561	AD	EGFR	p.E746_A750del_1A
S00565	AD	EGFR	p.G719S
S00590	AD	EGFR	p.E746_A750del_1A
S00610	AD	EGFR	p.E746_A750del_1B
S00616	AD	EGFR	p.E746_A750del_1B
S00623	AD	EGFR	exon 20 insertion
S00624	AD	EGFR	p.E746_S752>V
S00633	AD	EGFR	p.E746_ S752>V
S00652	AD	EGFR	p.(S768I(+)V774M)
S00654	AD	EGFR	p.L858R
S00659	AD	EGFR	p.E746_S752>V

PatID	Histology	Gene	Aminoacid_change
S00660	AD	EGFR	p.L747_T751>N
S00662	AD	EGFR	p.L858R
S00665	AD	EGFR	p.L858R
S00667	AD	EGFR	p.L858R
S00669	AD	EGFR	p.L747P
S00670	AD	EGFR	p.L858R
S00678	AD	EGFR	p.E709A
S00692	AD	EGFR	p.E746_A750del_1A
S00730	AD	EGFR	p.L858R
S00735	AD	EGFR	p.L858R
S00741	AD	EGFR	p.L858R
S00745	AD	EGFR	p.L858R
S01048	AD	EGFR	p.L747_T751del
S01058	AD	EGFR	p.L858R
S01078	AD	EGFR	p.(L833F(+)L858R)
S01087	AD	EGFR	p.G779F
S01096	AD	EGFR	p.L858R
S01098	AD	EGFR	p.L858R
S01125	AD	EGFR	p.E746_A750del_1A
S01134	AD	EGFR	p.E746_P753>VS
S01136	AD	EGFR	p.E746_A750del_1B
S01139	AD	EGFR	p.E746_A750del_1A
S01157	AD	EGFR	p.E746_A750delB
S01166	AD	EGFR	p.(E709K[+]K714N[+]V717E[+]G719R)
S01172	AD	EGFR	p.L858R
S01186	AD	EGFR	p.L858R
S01217	AD	EGFR	p.E746_A750del_1B
S01218	AD	EGFR	p.T751_I759>N
S01227	AD	EGFR	p.E746_A750del_1A
S01244	AD	EGFR	p.E746_A750del_1A
S01328	AD	EGFR	p.(T790M(+)L858R)
S01362	AD	EGFR	p.E746_A750del_1A
S01373	AD	EGFR	p.E746_S752>V
S01414	AD	EGFR	p.E746_A750del_1A
S01507	AD	EGFR	p.E746_A750del_1A
S01634	AD	EGFR	p.L747_S752>S
S01645	AD	EGFR	p.E746_S752>V
S01646	AD	EGFR	p.L858R
S01718	AD	EGFR	p.(G719A(+)S768I)
S00066	ADSQ	EGFR	p.L858R
S00498	ADSQ	EGFR	p.E746_A750del_1B
S00195	LCC	EGFR	p.K714N
S00215	LCC	EGFR	p.V774M
S00824	LCC	EGFR	p.N771>TT
S00510	SQ	EGFR	p.E746_A750del_1B
S00769	SQ	EGFR	p.P694H
S00393	UNKNOWN	EGFR	exon 20 insertion
S00058	AD	ERBB2	p.D742N
S00405	AD	ERBB2	p.D769N
S00684	AD	ERBB2	p.M774_A775insAYVM
S00734	AD	ERBB2	p.M774_A775insAYVM
S00750	AD	ERBB2	p.M774_A775insAYVM
S00752	AD	ERBB2	p.M774_A775insAYVM
S00757	AD	ERBB2	p.M774_A775insAYVM
S01341	AD	ERBB2	p.M774_A775insAYVM
S00084	LCC	ERBB2	p.S310F
S01168	LCC	ERBB2	p.S310F
S00030	SQ	ERBB2	p.S310F
S01402	SQ	ERBB2	p.G818A
S00197	AD	FGFR2	p.N767S
S00286	AD	FGFR2	p.N184S
S00900	AD	FGFR2	p.I348F
S01190	AD	FGFR2	p.Y375C
S01586	LCNEC	FGFR2	p.A568G
S00045	SQ	FGFR2	p.N549T
S01112	SQ	FGFR2	p.Y375C
S01245	SQ	FGFR2	p.R61C
S00325	SQ	FGFR3	p.R248C
S00434	SQ	FGFR3	p.R248C
S00674	SQ	FGFR3	p.R248C
S01581	LCNEC	HRAS	p.G13R
S01245	SQ	HRAS	p.G12D
S00003	AD	KEAP1	p.Q620del

PatID	Histology	Gene	Aminoacid_change
S00071	AD	KEAP1	p.R362W
S00085	AD	KEAP1	p.R459X
S00099	AD	KEAP1	p.I425F
S00120	AD	KEAP1	p.Y584fs
S00161	AD	KEAP1	p.R362Q
S00165	AD	KEAP1	p.G332R
S00205	AD	KEAP1	p.A322fs
S00206	AD	KEAP1	p.N382D
S00240	AD	KEAP1	p.G332C
S00267	AD	KEAP1	p.T309A
S00286	AD	KEAP1	p.G332C
S00292	AD	KEAP1	p.G364C
S00425	AD	KEAP1	p.G379V
S00459	AD	KEAP1	p.E441X
S00488	AD	KEAP1	p.G364C
S00495	AD	KEAP1	p.G511V
S00499	AD	KEAP1	p.S431F
S00694	AD	KEAP1	p.G423V
S00900	AD	KEAP1	p.R362P
S01057	AD	KEAP1	p.G464V
S01152	AD	KEAP1	p.V418fs
S01235	AD	KEAP1	p.S508R
S01241	AD	KEAP1	p.Q620del
S01266	AD	KEAP1	p.Q315fs
S01306	AD	KEAP1	p.N397del
S01368	AD	KEAP1	p.G462W
S01398	AD	KEAP1	p.G367C
S01404	AD	KEAP1	p.D389Y
S01409	AD	KEAP1	p.R320P
S01608	AD	KEAP1	p.G603R
S01646	AD	KEAP1	p.W497C
S00469	ADSQ	KEAP1	p.Y490X
S00084	LCC	KEAP1	p.G509V
S00799	LCNEC	KEAP1	p.R362Q
S01318	LCNEC	KEAP1	p.?
S00188	SCLC	KEAP1	p.V370M
S01524	SCLC	KEAP1	p.K323fs
S00013	SQ	KEAP1	p.G570E
S00041	SQ	KEAP1	p.V359L
S00160	SQ	KEAP1	p.T576M
S00223	SQ	KEAP1	p.S486T
S00280	SQ	KEAP1	p.G511C
S00316	SQ	KEAP1	p.Y567X
S00317	SQ	KEAP1	p.P384L
S00344	SQ	KEAP1	p.S338L
S00437	SQ	KEAP1	p.Q620del
S00475	SQ	KEAP1	p.P322H
S00720	SQ	KEAP1	p.R470H
S00721	SQ	KEAP1	p.R320L
S00769	SQ	KEAP1	p.G379F
S00785	SQ	KEAP1	p.Q620del
S01054	SQ	KEAP1	p.R565X
S01189	SQ	KEAP1	p.I304M
S01472	SQ	KEAP1	p.P612_K615del
S01651	SQ	KEAP1	p.R320W
S00002	AD	KRAS	p.G12D
S00003	AD	KRAS	p.G12V
S00015	AD	KRAS	p.G12D
S00018	AD	KRAS	p.G12V
S00019	AD	KRAS	p.G12S
S00025	AD	KRAS	p.G12C
S00029	AD	KRAS	p.G12D
S00048	AD	KRAS	p.G12C
S00057	AD	KRAS	p.G12C
S00063	AD	KRAS	p.G12V
S00071	AD	KRAS	p.G12D
S00074	AD	KRAS	p.G12C
S00077	AD	KRAS	p.G12F
S00078	AD	KRAS	p.Q61H
S00080	AD	KRAS	p.G12C
S00085	AD	KRAS	p.G12V
S00091	AD	KRAS	p.Q61L
S00099	AD	KRAS	p.G12C

PatID	Histology	Gene	Aminoacid_change
S00100	AD	KRAS	p.G12V
S00103	AD	KRAS	p.G12D
S00114	AD	KRAS	p.G12C
S00120	AD	KRAS	p.G12V
S00124	AD	KRAS	p.G12C
S00133	AD	KRAS	p.G12C
S00135	AD	KRAS	p.G12C
S00138	AD	KRAS	p.G12C
S00146	AD	KRAS	p.Q61H
S00153	AD	KRAS	p.G12V
S00170	AD	KRAS	p.G13D
S00171	AD	KRAS	p.G12C
S00172	AD	KRAS	p.G12C
S00178	AD	KRAS	p.G12C
S00180	AD	KRAS	p.G12C
S00190	AD	KRAS	p.G12C
S00191	AD	KRAS	p.Q61H
S00205	AD	KRAS	p.Q61H
S00206	AD	KRAS	p.G12A
S00207	AD	KRAS	p.G12C
S00211	AD	KRAS	p.G12C
S00240	AD	KRAS	p.G12V
S00256	AD	KRAS	p.G12A
S00261	AD	KRAS	p.G12C
S00275	AD	KRAS	p.G13C
S00278	AD	KRAS	p.G12R
S00282	AD	KRAS	p.G12C
S00284	AD	KRAS	p.G12D
S00288	AD	KRAS	p.G12D
S00290	AD	KRAS	p.G12C
S00301	AD	KRAS	p.G12C
S00305	AD	KRAS	p.G12D
S00307	AD	KRAS	p.G12F
S00309	AD	KRAS	p.G12C
S00351	AD	KRAS	p.G12V
S00352	AD	KRAS	p.G12C
S00353	AD	KRAS	p.G12V
S00357	AD	KRAS	p.G12C
S00359	AD	KRAS	p.G13C
S00361	AD	KRAS	p.G13C
S00397	AD	KRAS	p.G13C
S00418	AD	KRAS	p.G12D
S00425	AD	KRAS	p.G12D
S00431	AD	KRAS	p.G12L
S00432	AD	KRAS	p.G12C
S00435	AD	KRAS	p.G12V
S00445	AD	KRAS	p.G12C
S00452	AD	KRAS	p.G12C
S00462	AD	KRAS	p.G12D
S00483	AD	KRAS	p.G12C
S00484	AD	KRAS	p.G12V
S00486	AD	KRAS	p.Q61H
S00488	AD	KRAS	p.G12C
S00489	AD	KRAS	p.G12D
S00490	AD	KRAS	p.G12D
S00492	AD	KRAS	p.Q22K
S00493	AD	KRAS	p.G12C
S00494	AD	KRAS	p.G12C
S00499	AD	KRAS	p.G12C
S00562	AD	KRAS	p.G12C
S00569	AD	KRAS	p.G12V
S00573	AD	KRAS	p.G13D
S00574	AD	KRAS	p.G13C
S00576	AD	KRAS	p.G13C
S00583	AD	KRAS	p.G13C
S00588	AD	KRAS	p.G12D
S00589	AD	KRAS	p.G12C
S00592	AD	KRAS	p.Q61H
S00593	AD	KRAS	p.G12C
S00597	AD	KRAS	p.G12C
S00603	AD	KRAS	p.G12V
S00604	AD	KRAS	p.G12C
S00605	AD	KRAS	p.G12D

PatID	Histology	Gene	Aminoacid_change
S00607	AD	KRAS	p.G12D
S00618	AD	KRAS	p.G13C
S00619	AD	KRAS	p.G12V
S00621	AD	KRAS	p.G12D
S00631	AD	KRAS	p.G12V
S00642	AD	KRAS	p.G12C
S00643	AD	KRAS	p.G12C
S00644	AD	KRAS	p.G12V
S00645	AD	KRAS	p.G12A
S00653	AD	KRAS	p.G12C
S00657	AD	KRAS	p.G12D
S00677	AD	KRAS	p.G12C
S00694	AD	KRAS	p.G12C
S00699	AD	KRAS	p.G12C
S00702	AD	KRAS	p.G12V
S00704	AD	KRAS	p.G12V
S00724	AD	KRAS	p.G12D
S00740	AD	KRAS	p.G12D
S00742	AD	KRAS	p.G12D
S01056	AD	KRAS	p.G12V
S01074	AD	KRAS	p.Q61H
S01108	AD	KRAS	p.G12C
S01119	AD	KRAS	p.G12C
S01147	AD	KRAS	p.G12D
S01163	AD	KRAS	p.G13C
S01171	AD	KRAS	p.Q61H
S01232	AD	KRAS	p.G12D
S01234	AD	KRAS	p.G13D
S01235	AD	KRAS	p.G13V
S01241	AD	KRAS	p.G12C
S01250	AD	KRAS	p.G12C
S01253	AD	KRAS	p.G12V
S01258	AD	KRAS	p.G12C
S01259	AD	KRAS	p.G13D
S01290	AD	KRAS	p.G12C
S01316	AD	KRAS	p.L19F
S01329	AD	KRAS	p.Q61L
S01331	AD	KRAS	p.G12C
S01342	AD	KRAS	p.G12C
S01350	AD	KRAS	p.R68M
S01354	AD	KRAS	p.G12V
S01357	AD	KRAS	p.G13C
S01359	AD	KRAS	p.G12A
S01365	AD	KRAS	p.G12C
S01380	AD	KRAS	p.(G12V(+)G13C)
S01387	AD	KRAS	p.G12C
S01403	AD	KRAS	p.G12C
S01404	AD	KRAS	p.G12C
S01409	AD	KRAS	p.G12C
S01413	AD	KRAS	p.G12C
S01455	AD	KRAS	p.G12C
S01467	AD	KRAS	p.G12C
S01470	AD	KRAS	p.G12V
S01482	AD	KRAS	p.G12D
S01486	AD	KRAS	p.G12V
S01498	AD	KRAS	p.G12C
S01500	AD	KRAS	p.Q61H
S01609	AD	KRAS	p.G12V
S01610	AD	KRAS	p.G12V
S01632	AD	KRAS	p.G12C
S01647	AD	KRAS	p.G12V
S01729	AD	KRAS	p.G12F
S01906	AD	KRAS	p.G12C
S00909	ADSQ	KRAS	p.G12C
S01192	ADSQ	KRAS	p.G12V
S00046	LCC	KRAS	p.A59E
S00125	LCC	KRAS	p.G13E
S00195	LCC	KRAS	p.G12C
S00336	LCC	KRAS	p.G12C
S00627	LCC	KRAS	p.G12V
S01097	LCC	KRAS	p.G12V
S01141	LCC	KRAS	p.G12V
S01353	LCC	KRAS	p.G12C

PatID	Histology	Gene	Aminoacid_change
S01575	LCC	KRAS	p.G12S
S01654	LCC	KRAS	p.G12C
S00570	LCNEC	KRAS	p.G13C
S00713	LCNEC	KRAS	p.G12C
S01120	SARC	KRAS	p.G12C
S00117	SQ	KRAS	p.G13D
S00149	SQ	KRAS	p.G12A
S00223	SQ	KRAS	p.G12D
S00314	SQ	KRAS	p.G12C
S00318	SQ	KRAS	p.G12D
S00338	SQ	KRAS	p.G15S
S00412	SQ	KRAS	p.G12D
S00513	SQ	KRAS	p.G12D
S01447	SQ	KRAS	p.G12D
S00269	UNKNOWN	KRAS	p.G12C
S00272	UNKNOWN	KRAS	p.G12D
S00366	UNKNOWN	KRAS	p.G12C
S00517	UNKNOWN	KRAS	p.G12D
S00453	AD	NFE2L2	p.L23I
S01031	AD	NFE2L2	p.P88L
S01142	AD	NFE2L2	p.R34Q
S01252	AD	NFE2L2	p.D77Y
S01277	ADSQ	NFE2L2	p.D29N
S01100	LCC	NFE2L2	p.S97G
S01687	LCC	NFE2L2	p.E79K
S00799	LCNEC	NFE2L2	p.G31A
S00045	SQ	NFE2L2	p.L23V
S00064	SQ	NFE2L2	p.G81V
S00139	SQ	NFE2L2	p.R34P
S00144	SQ	NFE2L2	p.T80K
S00173	SQ	NFE2L2	p.R34Q
S00200	SQ	NFE2L2	p.I28T
S00222	SQ	NFE2L2	p.D29H
S00224	SQ	NFE2L2	p.R34G
S00247	SQ	NFE2L2	p.R34G
S00281	SQ	NFE2L2	p.G81V
S00355	SQ	NFE2L2	p.E79K
S00424	SQ	NFE2L2	p.D29N
S00509	SQ	NFE2L2	p.E79K
S00512	SQ	NFE2L2	p.R34Q
S00705	SQ	NFE2L2	p.E79K
S00760	SQ	NFE2L2	p.E79K
S00983	SQ	NFE2L2	p.G31A
S01106	SQ	NFE2L2	p.Q73H
S01109	SQ	NFE2L2	p.I20F
S01110	SQ	NFE2L2	p.D29N
S01236	SQ	NFE2L2	p.D29N
S01245	SQ	NFE2L2	p.V32P
S01282	SQ	NFE2L2	p.E79Q
S01385	SQ	NFE2L2	p.R34P
S01475	SQ	NFE2L2	p.W24C
S01481	SQ	NFE2L2	p.I28T
S01598	SQ	NFE2L2	p.D29H
S01635	SQ	NFE2L2	p.R34P
S01650	SQ	NFE2L2	p.(E55Q(+)E79K)
S01655	SQ	NFE2L2	p.R34P
S00635	AD	NRAS	p.Q61L
S00438	ADSQ	NRAS	p.G12A
S01450	LCC	NRAS	p.G13V
S00506	SQ	NRAS	P.G12C
S01085	SQ	NRAS	p.G12D
S00027	AD	PIK3CA	p.E545K
S00065	AD	PIK3CA	p.H701P

PatID	Histology	Gene	Aminoacid_change
S00074	AD	PIK3CA	p.F1002L
S00105	AD	PIK3CA	p.E545K
S00299	AD	PIK3CA	p.E545K
S00348	AD	PIK3CA	p.E545K
S00386	AD	PIK3CA	p.H1047L
S00398	AD	PIK3CA	p.E545K
S00459	AD	PIK3CA	p.H1047Q
S00561	AD	PIK3CA	p.E545K
S00562	AD	PIK3CA	p.E545K
S00590	AD	PIK3CA	p.E545K
S00598	AD	PIK3CA	p.E545K
S00604	AD	PIK3CA	p.H1047Y
S00662	AD	PIK3CA	p.E545K
S00750	AD	PIK3CA	p.R88Q
S00770	AD	PIK3CA	p.H1047L
S01111	AD	PIK3CA	p.Y1038F
S01190	AD	PIK3CA	p.E545K
S01334	AD	PIK3CA	p.E545Q
S01473	AD	PIK3CA	p.E545K
S00047	ADSQ	PIK3CA	p.E542K
S00909	ADSQ	PIK3CA	p.E545K
S01000	ADSQ	PIK3CA	p.E545K
S00102	LCC	PIK3CA	p.E545K
S00110	LCC	PIK3CA	p.E542K
S01066	LCNEC	PIK3CA	p.E545K
S01410	LCNEC	PIK3CA	p.E545K
S00213	SCLC	PIK3CA	p.S1003L
S00538	SCLC	PIK3CA	p.E542K
S00008	SQ	PIK3CA	p.E542K
S00030	SQ	PIK3CA	p.(E542K(+)E545K)
S00044	SQ	PIK3CA	p.E545Q
S00068	SQ	PIK3CA	p.E542K
S00069	SQ	PIK3CA	p.E545K
S00093	SQ	PIK3CA	p.E545K
S00145	SQ	PIK3CA	p.E542K
S00157	SQ	PIK3CA	p.E545K
S00198	SQ	PIK3CA	p.E542K
S00199	SQ	PIK3CA	p.E545K
S00252	SQ	PIK3CA	p.E542K
S00321	SQ	PIK3CA	p.Q546K
S00330	SQ	PIK3CA	p.E542K
S00408	SQ	PIK3CA	p.E542Q
S00480	SQ	PIK3CA	p.H1047R
S00674	SQ	PIK3CA	p.E545K
S00743	SQ	PIK3CA	p.Y1021C
S00758	SQ	PIK3CA	p.E542K
S01085	SQ	PIK3CA	p.E542K
S01164	SQ	PIK3CA	p.E542K
S01178	SQ	PIK3CA	p.E545K
S01229	SQ	PIK3CA	p.E542K
S01245	SQ	PIK3CA	p.E542K
S01299	SQ	PIK3CA	p.E545K
S01457	SQ	PIK3CA	p.E542K
S00366	UNKNOWN	PIK3CA	p.Q546H
S00310	AD	ROS1	narrow split
S00545	AD	ROS1	ROS1
S01345	AD	ROS1	ROS1

References

1 Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol 30, 413-421, doi:10.1038/nbt.2203

2 Travis, W., Brambilla, E., Mueller-Hermelink, H. & Harris, C. World Health Organization Classification of Tumours. Pathology and Genetics: Tumours of the Lung, Pleura, Thymus and Heart., (2004).

3 Lynch, T. J. et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 350, 2129-2139, doi:10.1056/NEJMoa040938 (2004).

4 Mok, T. S. et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med 361, 947-957, doi:10.1056/NEJMoa0810699 (2009).

5 WHO. http://www.who.int/en/ (2013).

6 National Cancer Institute, <http://www.cancer.gov/> (2012).

7 Weinberg, R. A. The Biology of Cancer. (Garland Science, Taylor & Francis Group, 2007).

8 Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A 100, 3983-3988, doi:10.1073/pnas.0530291100 (2003).

9 Heppner, G. H. Tumor heterogeneity. Cancer Res 44, 2259-2265 (1984).

10 Boveri, T. Zur Frage der Entstehung maligner Tumoren. Gustav Fischer (1914).

11 Nowell, P. C. & Hungerford, D. A. Chromosome studies on normal and leukemic human leukocytes. J Natl Cancer Inst 25, 85-109 (1960).

12 TCGA. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455, 1061-1068, doi:10.1038/nature07385 (2008).

13 Beroukhim, R. et al. The landscape of somatic copy-number alteration across human cancers. Nature 463, 899-905, doi:10.1038/nature08822 (2010).

14 Taylor, B. S. et al. Integrative genomic profiling of human prostate cancer. Cancer Cell 18, 11-22, doi:10.1016/j.ccr.2010.05.026 (2010).

15 Molenaar, J. J. et al. Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. Nature 483, 589-593, doi:10.1038/nature10910 (2012).

16 Greenman, C. et al. Patterns of somatic mutation in human cancer genomes. Nature 446, 153-158, doi:10.1038/nature05610 (2007).

17 Mullighan, C. G. et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature 446, 758-764, doi:10.1038/nature05690 (2007).

18 Thomas, R. K. et al. High-throughput oncogene mutation profiling in human cancer. Nat Genet 39, 347-351, doi:10.1038/ng1975 (2007).

19 Hahn, W. C. & Weinberg, R. A. Rules for making human tumor cells. N Engl J Med 347, 1593-1603, doi:10.1056/NEJMra021902 (2002).

20 Baudot, A., de la Torre, V. & Valencia, A. Mutated genes, pathways and processes in tumours. EMBO Rep 11, 805-810, doi:10.1038/embor.2010.133 (2010).

21 Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. Cell 100, 57-70 (2000).

22 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. Cell 144, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).

23 Leary, R. J. et al. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. Proc Natl Acad Sci U S A 105, 16224-16229, doi:10.1073/pnas.0808041105 (2008).

24 Bert Vogelstein, K. W. K. The genetic basis of Human Cancer. McGraw-Hill Companies, Inc. (2002).

25 de Ridder, J. et al. Co-occurrence analysis of insertional mutagenesis data reveals cooperating oncogenes. Bioinformatics 23, i133-141, doi:10.1093/bioinformatics/btm202 (2007).

26 Bruce Alberts, A. J., Peter Walter, Julian Lewis, Martin Raff, Keith Roberts. Molecular Biology of the Cell. (Taylor & Francis, 2007).

27 Press, M. F. et al. HER-2/neu gene amplification characterized by fluorescence in situ hybridization: poor prognosis in node-negative breast carcinomas. J Clin Oncol 15, 2894-2904 (1997).

28 Soda, M. et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature 448, 561-566, doi:10.1038/nature05945 (2007).

29 Turner, N. & Grose, R. Fibroblast growth factor signalling: from development to cancer. Nat Rev Cancer 10, 116-129, doi:10.1038/nrc2780 (2010).

30 Gazdar, A. F. Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. Oncogene 28 Suppl 1, S24-31, doi:10.1038/onc.2009.198 (2009).

31 Sharma, S. V., Bell, D. W., Settleman, J. & Haber, D. A. Epidermal growth factor receptor mutations in lung cancer. Nat Rev Cancer 7, 169-181, doi:10.1038/nrc2088 (2007).

32 Chen, Y. R. et al. Distinctive activation patterns in constitutively active and gefitinib-sensitive EGFR mutants. Oncogene 25, 1205-1215, doi:10.1038/sj.onc.1209159 (2006).

33 Yarden, Y. Biology of HER2 and its importance in breast cancer. Oncology 61 Suppl 2, 1-13, doi:55396 (2001).

34 Olayioye, M. A. Update on HER-2 as a target for cancer therapy: intracellular signaling pathways of ErbB2/HER-2 and family members. Breast Cancer Res 3, 385-389 (2001).

35 Horn, L. & Pao, W. EML4-ALK: honing in on a new target in non-small-cell lung cancer. J Clin Oncol 27, 4232-4235, doi:10.1200/JCO.2009.23.6661 (2009).

36 Lin, E. et al. Exon array profiling detects EML4-ALK fusion in breast, colorectal, and non-small cell lung cancers. Mol Cancer Res 7, 1466-1476, doi:10.1158/1541-7786.MCR-08-0522 (2009).

37 Knudson, A. G., Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 68, 820-823 (1971).

38 Weir, B., Zhao, X. & Meyerson, M. Somatic alterations in the human cancer genome. Cancer Cell 6, 433-438, doi:10.1016/j.ccr.2004.11.004 (2004).

39 Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of gene expression in Escherichia coli. Science 324, 255-258, doi:10.1126/science.1170160 (2009).

40 James D. Watson, T. A. B., Stephen P. Bell. Molecular Biology of the Gene. 5th edn, Vol. Auflage: 5 Har/Cdr 768 (Addison-Wesley Longman, Amsterdam, 2003).

41 Mort, M., Ivanov, D., Cooper, D. N. & Chuzhanova, N. A. A meta-analysis of nonsense mutations causing human genetic disease. Hum Mutat 29, 1037-1047, doi:10.1002/humu.20763 (2008).

42 Mitsudomi, T. & Yatabe, Y. Epidermal growth factor receptor in relation to tumor development: EGFR gene and cancer. Febs J 277, 301-308, doi:10.1111/j.1742-4658.2009.07448.x (2010).

43 Hyman, E. et al. Impact of DNA amplification on gene expression patterns in breast cancer. Cancer Res 62, 6240-6245 (2002).

44 Tang, Y. C. & Amon, A. Gene copy-number alterations: a cost-benefit analysis. Cell 152, 394-405, doi:10.1016/j.cell.2012.11.043 (2013).

45 Gainor, J. F. & Shaw, A. T. Novel targets in non-small cell lung cancer: ROS1 and RET fusions. Oncologist 18, 865-875, doi:10.1634/theoncologist.2013-0095theoncologist.2013-0095 [pii] (2013).

46 Heuckmann, J. M. et al. Differential protein stability and ALK inhibitor sensitivity of EML4-ALK fusion variants. Clin Cancer Res 18, 4682-4690, doi:10.1158/1078-0432.CCR-11-32601078-0432.CCR-11-3260 [pii] (2012).

47 WHO. International Agency for Research on Cancer: Globocan. (2008).

48 Support, M. C. The Cancer Survival Lottery, < http://www.macmillan.org.uk/Aboutus/News/Latest_News/TheCancerSurvivalLottery.aspx > (2011).

49 Thun, M. J. et al. Cigarette smoking and changes in the histopathology of lung cancer. J Natl Cancer Inst 89, 1580-1586 (1997).

50 Noguchi, M. et al. Small adenocarcinoma of the lung. Histologic characteristics and prognosis. Cancer 75, 2844-2852 (1995).

51 Motoi, N. et al. Lung adenocarcinoma: modification of the 2004 WHO mixed subtype to include the major histologic subtype suggests correlations between papillary and

micropapillary adenocarcinoma subtypes, EGFR mutations and gene expression analysis. Am J Surg Pathol 32, 810-827, doi:10.1097/PAS.0b013e31815cb162 (2008).

52 Russell, P. A. et al. Does lung adenocarcinoma subtype predict patient survival?: A clinicopathologic study based on the new International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary lung adenocarcinoma classification. J Thorac Oncol 6, 1496-1504, doi:10.1097/JTO.0b013e318221f701 (2011).

53 Shim, H. S., Lee da, H., Park, E. J. & Kim, S. H. Histopathologic characteristics of lung adenocarcinomas with epidermal growth factor receptor mutations in the International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society lung adenocarcinoma classification. Arch Pathol Lab Med 135, 1329-1334, doi:10.5858/arpa.2010-0493-OA (2011).

54 Travis, W. D. et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society: international multidisciplinary classification of lung adenocarcinoma: executive summary. Proc Am Thorac Soc 8, 381-385, doi:10.1513/pats.201107-042ST (2011).

55 Travis, W., Brambilla, E., Noguchi, M. &

56 Noda, K. et al. Irinotecan plus cisplatin compared with etoposide plus cisplatin for extensive small-cell lung cancer. N Engl J Med 346, 85-91, doi:10.1056/NEJMoa003034 (2002).

57 Ishida, T. et al. Carcinosarcoma and spindle cell carcinoma of the lung. Clinicopathologic and immunohistochemical studies. J Thorac Cardiovasc Surg 100, 844-852 (1990).

58 Fishback, N. F. et al. Pleomorphic (spindle/giant cell) carcinoma of the lung. A clinicopathologic correlation of 78 cases. Cancer 73, 2936-2945 (1994).

59 Crino, L., Weder, W., van Meerbeeck, J. & Felip, E. Early stage and locally advanced (non-metastatic) non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 21 Suppl 5, v103-115, doi:10.1093/annonc/mdq207 (2010).

60 Jackman, D. M. & Johnson, B. E. Small-cell lung cancer. Lancet 366, 1385-1396, doi:10.1016/S0140-6736(05)67569-1 (2005).

61 Scagliotti, G. V. et al. Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. J Clin Oncol 26, 3543-3551, doi:10.1200/JCO.2007.15.0375 (2008).

62 Einhorn, L. H. First-line chemotherapy for non-small-cell lung cancer: is there a superior regimen based on histology? J Clin Oncol 26, 3485-3486, doi:10.1200/JCO.2008.17.2056 (2008).

63 Lindeman, N. I. et al. Molecular Testing Guideline for Selection of Lung Cancer Patients for EGFR and ALK Tyrosine Kinase Inhibitors: Guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. J Mol Diagn 15, 415-453, doi:10.1016/j.jmoldx.2013.03.001 (2013).

64 e.V., D. G. f. H. u. M. O. Leitlinie: Lungenkarzinom, nicht-kleinzellig (NSCLC), <http://www.dgho-onkopedia.de/de/onkopedia/leitlinien/lungenkarzinom-nicht-kleinzellig-nsclc> (2013).

65 Buettner, R., Wolf, J. & Thomas, R. K. Lessons learned from lung cancer genomics: the emerging concept of individualized diagnostics and treatment. J Clin Oncol 31, 1858-1865, doi:10.1200/JCO.2012.45.9867 (2013).

66 Garassino, M. C. et al. Biological and clinical features in predicting efficacy of epidermal growth factor receptor tyrosine kinase inhibitors: a systematic review and meta-analysis. Anticancer Res 29, 2691-2701 (2009).

67 Shepherd, F. A. et al. Erlotinib in previously treated non-small-cell lung cancer. N Engl J Med 353, 123-132, doi:10.1056/NEJMoa050753 (2005).

68 Mitsudomi, T. et al. Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. Lancet Oncol 11, 121-128, doi:10.1016/S1470-2045(09)70364-X (2010).

69 Rosell, R. et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a

multicentre, open-label, randomised phase 3 trial. Lancet Oncol 13, 239-246, doi:10.1016/S1470-2045(11)70393-X (2012).
70 Kwak, E. L. et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. N Engl J Med 363, 1693-1703, doi:10.1056/NEJMoa1006448 (2010).
71 Bos, M., Gardizi, M., Schildhaus, H.-U., Buettner, R. & Wolf, J. Activated RET and ROS: two new driver mutations in lung adenocarcinoma. Translational lung cancer research 2, 10 (2013).
72 Hammerman, P. S. et al. Mutations in the DDR2 kinase gene identify a novel therapeutic target in squamous cell lung cancer. Cancer Discov 1, 78-89, doi:10.1158/2159-8274.CD-11-0005 (2011).
73 Brooks, A. N., Kilgour, E. & Smith, P. D. Molecular pathways: fibroblast growth factor signaling: a new therapeutic opportunity in cancer. Clin Cancer Res 18, 1855-1862, doi:10.1158/1078-0432.CCR-11-06991078-0432.CCR-11-0699 [pii] (2012).
74 ClinicalTrials.gov. (2013).
75 Administration, U. S. F. a. D. Drug Approvals, <http://www.fda.gov> (2013).
76 Pao, W., Iafrate, A. J. & Su, Z. Genetically informed lung cancer medicine. J Pathol 223, 230-240, doi:10.1002/path.2788 (2011).
77 Renshaw, A. A. & Gould, E. W. Comparison of disagreement and error rates for three types of interdepartmental consultations. Am J Clin Pathol 124, 878-882 (2005).
78 Institute, W. T. S. The Catalogue Of Somatic Mutations In Cancer (COSMIC), <http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/> (
79 Bass, A. J. et al. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nat Genet 41, 1238-1242, doi:10.1038/ng.465 (2009).
80 Rudin, C. M. et al. Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. Nat Genet 44, 1111-1116, doi:10.1038/ng.2405 (2012).
81 Peifer, M. et al. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. Nat Genet 44, 1104-1110, doi:10.1038/ng.2396 (2012).
82 Weiss, J. et al. Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. Sci Transl Med 2, 62ra93, doi:10.1126/scitranslmed.3001451 (2010).
83 Weir, B. A. et al. Characterizing the cancer genome in lung adenocarcinoma. Nature 450, 893-898, doi:10.1038/nature06358 (2007).
84 Yoo, N. J., Kim, H. R., Kim, Y. R., An, C. H. & Lee, S. H. Somatic mutations of the KEAP1 gene in common solid cancers. Histopathology 60, 943-952, doi:10.1111/j.1365-2559.2012.04178.x (2012).
85 Bergethon, K. et al. ROS1 rearrangements define a unique molecular class of lung cancers. J Clin Oncol 30, 863-870, doi:10.1200/JCO.2011.35.6345 (2012).
86 Bert Vogelstein, S. S., Carol Prives. p53 : The Most Frequently Altered Gene in Human Cancers. Nature education 3(9):6 (2010). <http://www.nature.com/scitable/topicpage/p53-the-most-frequently-altered-gene-in-14192717 - >.
87 Blanco, R. et al. A gene-alteration profile of human lung cancer cell lines. Hum Mutat 30, 1199-1206, doi:10.1002/humu.21028 (2009).
88 Oxnard, G. R., Binder, A. & Janne, P. A. New targetable oncogenes in non-small-cell lung cancer. J Clin Oncol 31, 1097-1104, doi:10.1200/JCO.2012.42.9829JCO.2012.42.9829 [pii] (2013).
89 Chen, Z. et al. EGFR somatic doublets in lung cancer are frequent and generally arise from a pair of driver mutations uncommonly seen as singlet mutations: one-third of doublets occur at five pairs of amino acids. Oncogene 27, 4336-4343, doi:10.1038/onc.2008.71 (2008).
90 Atchley, W. R. & Fitch, W. M. Myc and Max: molecular evolution of a family of proto-oncogene products and their dimerization partner. Proc Natl Acad Sci U S A 92, 10217-10221 (1995).
91 Wagner, P. L. et al. In situ evidence of KRAS amplification and association with increased p21 levels in non-small cell lung carcinoma. Am J Clin Pathol 132, 500-505, doi:10.1309/AJCPF10ZUNSOLIFG (2009).
92 Iwakuma, T. & Lozano, G. MDM2, an introduction. Mol Cancer Res 1, 993-1000 (2003).
93 Lu, Y. et al. Evidence that SOX2 overexpression is oncogenic in the lung. PLoS One 5, e11022, doi:10.1371/journal.pone.0011022 (2010).
94 Zochbauer-Muller, S., Wistuba, II, Minna, J. D. & Gazdar, A. F. Fragile histidine triad (FHIT) gene abnormalities in lung cancer. Clin Lung Cancer 2, 141-145 (2000).

95 Seidel, D. et al. A Genomics-Based Classification of Human Lung Tumors. ScienceTranslationalMedicine 5 (2013).

96 Xu, B. et al. Spindle epithelial tumor with thymus-like differentiation of the thyroid: a case report with pathological and molecular genetics study. Hum Pathol 34, 190-193, doi:10.1053/hupa.2003.29 (2003).

97 Mano, H. ALKoma: a cancer subtype with a shared target. Cancer Discov 2, 495-502, doi:10.1158/2159-8290.CD-12-0009 (2012).

98 Sos, M. L. et al. PTEN loss contributes to erlotinib resistance in EGFR-mutant lung cancer by activation of Akt and EGFR. Cancer Res 69, 3256-3261, doi:10.1158/0008-5472.CAN-08-4055 (2009).

99 Rekhtman, N. et al. Distinct profile of driver mutations and clinical features in immunomarker-defined subsets of pulmonary large-cell carcinoma. Mod Pathol 26, 511-522, doi:10.1038/modpathol.2012.195modpathol2012195 [pii] (2013).

100 Tan, D. & Zander, D. S. Immunohistochemistry for assessment of pulmonary and pleural neoplasms: a review and update. Int J Clin Exp Pathol 1, 19-31 (2008).

101 Takeuchi, K. et al. RET, ROS1 and ALK fusions in lung cancer. Nat Med 18, 378-381, doi:10.1038/nm.2658 (2012).

102 Herter-Sprie, G. S., Greulich, H. & Wong, K. K. Activating Mutations in ERBB2 and Their Impact on Diagnostics and Treatment. Front Oncol 3, 86, doi:10.3389/fonc.2013.00086 (2013).

103 Brose, M. S. et al. BRAF and RAS mutations in human lung cancer and melanoma. Cancer Res 62, 6997-7000 (2002).

104 Marchetti, A. et al. Clinical features and outcome of patients with non-small-cell lung cancer harboring BRAF mutations. J Clin Oncol 29, 3574-3579, doi:10.1200/JCO.2011.35.9638 (2011).

105 Guagnano, V. et al. FGFR genetic alterations predict for sensitivity to NVP-BGJ398, a selective pan-FGFR inhibitor. Cancer Discov 2, 1118-1133, doi:10.1158/2159-8290.CD-12-0210 (2012).

106 Ding, L. et al. Somatic mutations affect key pathways in lung adenocarcinoma. Nature 455, 1069-1075, doi:10.1038/nature07423 (2008).

107 Sporn, M. B. & Liby, K. T. NRF2 and cancer: the good, the bad and the importance of context. Nat Rev Cancer 12, 564-571, doi:10.1038/nrc3278 (2012).

108 Massarelli, E. et al. KRAS mutation is an important predictor of resistance to therapy with epidermal growth factor receptor tyrosine kinase inhibitors in non-small-cell lung cancer. Clin Cancer Res 13, 2890-2896, doi:10.1158/1078-0432.CCR-06-3043 (2007).

109 Knowles, M. A. Novel therapeutic targets in bladder cancer: mutation and expression of FGF receptors. Future Oncol 4, 71-83, doi:10.2217/14796694.4.1.71 (2008).

110 Cappellen, D. et al. Frequent activating mutations of FGFR3 in human bladder and cervix carcinomas. Nat Genet 23, 18-20, doi:10.1038/12615 (1999).

111 Todaro, G. J. & Green, H. Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. J Cell Biol 17, 299-313 (1963).

112 Selleck. Compare IC50 - FGFR Inhibitors, <http://www.selleckchem.com/FGFR.html> (2013).

113 ClinicalTrialsFeeds. 3 Clinical Trials for: BGJ398, <http://clinicaltrialsfeeds.org/clinical-trials/results/term=BGJ398> (2013).

114 Imielinski, M. et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell 150, 1107-1120, doi:10.1016/j.cell.2012.08.029 (2012).

115 Hammerman, P. S. et al. Comprehensive genomic characterization of squamous cell lung cancers. Nature 489, 519-525, doi:10.1038/nature11404 (2012).

116 Pitini, V., Arrigo, C., Di Mirto, C., Mondello, P. & Altavilla, G. Response to dasatinib in a patient with SQCC of the lung harboring a discoid-receptor-2 and synchronous chronic myelogenous leukemia. Lung Cancer 82, 171-172, doi:10.1016/j.lungcan.2013.07.004S0169-5002(13)00313-9 [pii] (2013).S0169-5002(13)00313-9 [pii] (2013).

117 Homma, S. et al. Nrf2 enhances cell proliferation and resistance to anticancer drugs in human lung cancer. Clin Cancer Res 15, 3423-3432, doi:10.1158/1078-0432.CCR-08-2822 (2009).

118 Wang, X. J. et al. Nrf2 enhances resistance of cancer cells to chemotherapeutic drugs, the dark side of Nrf2. Carcinogenesis 29, 1235-1243, doi:10.1093/carcin/bgn095 (2008).

119 Olivier, M., Hollstein, M. & Hainaut, P. TP53 mutations in human cancers: origins, consequences, and clinical use. Cold Spring Harb Perspect Biol 2, a001008, doi:10.1101/cshperspect.a001008 (2010).

120 Rekhtman, N. et al. Clarifying the spectrum of driver oncogene mutations in biomarker-verified squamous carcinoma of lung: lack of EGFR/KRAS and presence of PIK3CA/AKT1 mutations. Clin Cancer Res 18, 1167-1176, doi:10.1158/1078-0432.CCR-11-2109 (2012).

121 Jackman, D. M. et al. Exon 19 deletion mutations of epidermal growth factor receptor are associated with prolonged survival in non-small cell lung cancer patients treated with gefitinib or erlotinib. Clin Cancer Res 12, 3908-3914, doi:10.1158/1078-0432.CCR-06-0462 (2006).

122 Chapman, P. B. et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med 364, 2507-2516, doi:10.1056/NEJMoa1103782 (2011).

123 Yang, H. et al. Antitumor activity of BRAF inhibitor vemurafenib in preclinical models of BRAF-mutant colorectal cancer. Cancer Res 72, 779-789, doi:10.1158/0008-5472.CAN-11-2941 (2012).

124 Gautschi, O. et al. A patient with BRAF V600E lung adenocarcinoma responding to vemurafenib. J Thorac Oncol 7, e23-24, doi:10.1097/JTO.0b013e3182629903 (2012).

125 Peters, S., Michielin, O. & Zimmermann, S. Dramatic Response Induced by Vemurafenib in a BRAF V600E-Mutated Lung Adenocarcinoma. J Clin Oncol 31, e341-344, doi:10.1200/JCO.2012.47.6143 (2013).

126 Hou, J. et al. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. PLoS One 5, e10312, doi:10.1371/journal.pone.0010312 (2010).

127 Takeuchi, T. et al. Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. J Clin Oncol 24, 1679-1688, doi:10.1200/JCO.2005.03.8224 (2006).

128 Pardo, J. et al. Large cell carcinoma of the lung: an endangered species? Appl Immunohistochem Mol Morphol 17, 383-392, doi:10.1097/PAI.0b013e31819bfd59 (2009).

129 Sun, L. et al. High-grade neuroendocrine carcinoma of the lung: comparative clinicopathological study of large cell neuroendocrine carcinoma and small cell lung carcinoma. Pathol Int 59, 522-529, doi:10.1111/j.1440-1827.2009.02402.x (2009).

130 Takeuchi, T. et al. Characteristics of loss of heterozygosity in large cell neuroendocrine carcinomas of the lung and small cell lung carcinomas. Pathol Int 56, 434-439, doi:10.1111/j.1440-1827.2006.01985.x (2006).

131 Peng, W. X., Sano, T., Oyama, T., Kawashima, O. & Nakajima, T. Large cell neuroendocrine carcinoma of the lung: a comparison with large cell carcinoma with neuroendocrine morphology and small cell carcinoma. Lung Cancer 47, 225-233, doi:S0169-5002(04)00304-6 [pii]10.1016/j.lungcan.2004.06.018 (2005).

132 Natasha Rekhtman, L. J. T., Jamie E Chaft, Lu Wang, Maria E Arcila, Agnes Colanta, Andre L Moreira, Maureen F Zakowski, William D Travis, Camelia S Sima, Mark G Kris, Marc Ladanyi. Distinct profile of driver mutations and clinical features in immunomarker-defined subsets of pulmonary large-cell carcinoma. Mod Pathol 2012 Nov 30 (2012).

133 Monica, V. et al. Differential Thymidylate Synthase Expression in Different Variants of Large-Cell Carcinoma of the Lung. Clin Cancer Res 15, 7547-7552, doi:1078-0432.CCR-09-1641 [pii]10.1158/1078-0432.CCR-09-1641 (2009).

134 Barbareschi, M. et al. Heterogeneity of large cell carcinoma of the lung: an immunophenotypic and miRNA-based analysis. Am J Clin Pathol 136, 773-782, doi:10.1309/AJCPYY79XAGRAYCJ136/5/773 [pii] (2011).

135 Iyoda, A. et al. Prospective study of adjuvant chemotherapy for pulmonary large cell neuroendocrine carcinoma. Ann Thorac Surg 82, 1802-1807, doi:10.1016/j.athoracsur.2006.05.109 (2006).

136 Rossi, G. et al. Role of chemotherapy and the receptor tyrosine kinases KIT, PDGFRalpha, PDGFRbeta, and Met in large-cell neuroendocrine carcinoma of the lung. J Clin Oncol 23, 8774-8785, doi:10.1200/JCO.2005.02.8233 (2005).

137 Yamazaki, S. et al. Clinical responses of large cell neuroendocrine carcinoma of the lung to cisplatin-based chemotherapy. Lung Cancer 49, 217-223, doi:10.1016/j.lungcan.2005.01.008 (2005).

138 Hiroshima, K. et al. Distinction of pulmonary large cell neuroendocrine carcinoma from small cell lung carcinoma: a morphological, immunohistochemical, and molecular analysis. Mod Pathol 19, 1358-1368, doi:10.1038/modpathol.3800659 (2006).

139 Iyoda, A. et al. Pulmonary large cell carcinomas with neuroendocrine features are high-grade neuroendocrine tumors. Ann Thorac Surg 73, 1049-1054 (2002).

140 Yoon, Y. K. et al. Combination of EGFR and MEK1/2 inhibitor shows synergistic effects by suppressing EGFR/HER3-dependent AKT activation in human gastric cancer cells. Mol Cancer Ther 8, 2526-2536, doi:10.1158/1535-7163.MCT-09-0300 (2009).

141 Garraway, L. A. & Lander, E. S. Lessons from the cancer genome. Cell 153, 17-37, doi:10.1016/j.cell.2013.03.002 (2013).

142 Garraway, L. A. & Janne, P. A. Circumventing cancer drug resistance in the era of personalized medicine. Cancer Discov 2, 214-226, doi:10.1158/2159-8290.CD-12-00122159-8290.CD-12-0012 [pii] (2012).

143 Tomlinson, D. C., Lamont, F. R., Shnyder, S. D. & Knowles, M. A. Fibroblast growth factor receptor 1 promotes proliferation and survival via activation of the mitogen-activated protein kinase pathway in bladder cancer. Cancer Res 69, 4613-4620, doi:10.1158/0008-5472.CAN-08-2816 (2009).

144 Jokinen, E., Laurila, N. & Koivunen, J. P. Alternative dosing of dual PI3K and MEK inhibition in cancer therapy. BMC Cancer 12, 612, doi:10.1186/1471-2407-12-612 (2012).

145 Schmid, K. et al. EGFR/KRAS/BRAF mutations in primary lung adenocarcinomas and corresponding locoregional lymph node metastases. Clin Cancer Res 15, 4554-4560, doi:10.1158/1078-0432.CCR-09-0089 (2009).

146 Davies, H. et al. Mutations of the BRAF gene in human cancer. Nature 417, 949-954, doi:10.1038/nature00766 (2002).

147 Heidorn, S. J. et al. Kinase-dead BRAF and oncogenic RAS cooperate to drive tumor progression through CRAF. Cell 140, 209-221, doi:10.1016/j.cell.2009.12.040 (2010).

148 Travis, W. D., Brambilla, E. & Riely, G. J. New pathologic classification of lung cancer: relevance for clinical practice and clinical trials. J Clin Oncol 31, 992-1001, doi:10.1200/JCO.2012.46.9270 (2013).

149 Minuti, G., D'Incecco, A. & Cappuzzo, F. Targeted therapy for NSCLC with driver mutations. Expert Opin Biol Ther 13, 1401-1412, doi:10.1517/14712598.2013.827657 (2013).

150 Solomon, B., Wilner, K. D. & Shaw, A. T. Current status of targeted therapy for anaplastic lymphoma kinase (ALK) rearranged non-small cell lung cancer. Clin Pharmacol Ther, doi:10.1038/clpt.2013.200clpt2013200 [pii] (2013).

151 Goke, F. & Perner, S. [Translational research and diagnostics in lung cancer.]. Pathologe, doi:10.1007/s00292-012-1664-y (2012).

152 Perner, S. et al. EML4-ALK fusion lung cancer: a rare acquired event. Neoplasia 10, 298-302 (2008).

153 Perner, S. et al. TMPRSS2:ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer. Cancer Res 66, 8337-8341, doi:10.1158/0008-5472.CAN-06-1482 (2006).

Publications

Osteoblastic response in patients with non-small cell lung cancer with activating EGFR Mutations and bone metastases during treatment with EGFR kinase inhibitors.
Ansén S, Bangard C, Querings S, Gabler F, Scheffler M, **Seidel D**, Saal B, Zander T, Nogová L, Töpelt K, Markert E, Stoelben E, Ernestus K, Thomas RK, Wolf J.
J Thorac Oncol. 2010 Mar;5(3):407-9.

Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer.
Weiss J*, Sos ML*, **Seidel D***, Peifer M, Zander T, Heuckmann JM, Ullrich RT, Menon R, Maier S, Soltermann A, Moch H, Wagener P, Fischer F, Heynck S, Koker M, Schöttle J, Leenders F, Gabler F, Dabow I, Querings S, Heukamp LC, Balke-Want H, Ansén S, Rauh D, Baessmann I, Altmüller J, Wainer Z, Conron M, Wright G, Russell P, Solomon B, Brambilla E, Brambilla C, Lorimier P, Sollberg S, Brustugun OT, Engel-Riedel W, Ludwig C, Petersen I, Sänger J, Clement J, Groen H, Timens W, Sietsma H, Thunnissen E, Smit E, Heideman D, Cappuzzo F, Ligorio C, Damiani S, Hallek M, Beroukhim R, Pao W, Klebl B, Baumann M, Buettner R, Ernestus K, Stoelben E, Wolf J, Nürnberg P, Perner S, Thomas RK.
Sci Transl Med. 2010 Dec 15;2(62):62ra93.
* These authors contributed equally

Benchmarking of mutation diagnostics in clinical lung cancer specimens.
Querings S, Altmüller J, Ansén S, Zander T, **Seidel D**, Gabler F, Peifer M, Markert E, Stemshorn K, Timmermann B, Saal B, Klose S, Ernestus K, Scheffler M, Engel-Riedel W, Stoelben E, Brambilla E, Wolf J, Nürnberg P, Thomas RK.
PLoS One. 2011 May 5;6(5):e19601.

Mutations in the DDR2 kinase gene identify a novel therapeutic target in squamous cell lung cancer.
Hammerman PS*, Sos ML*, Ramos AH, Xu C, Dutt A, Zhou W, Brace LE, Woods BA, Lin W, Zhang J, Deng X, Lim SM, Heynck S, Peifer M, Simard JR, Lawrence MS, Onofrio RC, Salvesen HB, **Seidel D**, Zander T, Heuckmann JM, Soltermann A, Moch H, Koker M, Leenders F, Gabler F, Querings S, Ansén S, Brambilla E, Brambilla C, Lorimier P, Brustugun OT, Helland A, Petersen I, Clement JH, Groen H, Timens W, Sietsma H, Stoelben E, Wolf J, Beer DG, Tsao MS, Hanna M, Hatton C, Eck MJ, Janne PA, Johnson BE, Winckler W, Greulich H, Bass AJ, Cho J, Rauh D, Gray NS, Wong KK, Haura EB, Thomas RK, Meyerson M.
Cancer Discov. 2011 Jun;1(1):78-89.
* These authors contributed equally

Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing.
Imielinski M*, Berger AH*, Hammerman PS*, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, Sougnez C, Auclair D, Lawrence MS, Stojanov P, Cibulskis K, Choi K, de Waal L, Sharifnia T, Brooks A, Greulich H, Banerji S, Zander T, **Seidel D**, Leenders F, Ansén S, Ludwig C, Engel-Riedel W, Stoelben E, Wolf J, Goparju C, Thompson K, Winckler W, Kwiatkowski D, Johnson BE, Jänne PA, Miller VA, Pao W, Travis WD, Pass HI, Gabriel SB, Lander ES, Thomas RK, Garraway LA, Getz G, Meyerson M.
Cell. 2012 Sep 14;150(6):1107-20
* These authors contributed equally

Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer.
Peifer M*, Fernández-Cuesta L*, Sos ML, George J, **Seidel D**, Kasper LH, Plenker D, Leenders F, Sun R, Zander T, Menon R, Koker M, Dahmen I, Müller C, Di Cerbo V, Schildhaus HU, Altmüller J, Baessmann I, Becker C, de Wilde B, Vandesompele J, Böhm D, Ansén S, Gabler F, Wilkening I, Heynck S, Heuckmann JM, Lu X, Carter SL, Cibulskis K, Banerji S, Getz G, Park KS, Rauh D, Grütter C, Fischer M,

Pasqualucci L, Wright G, Wainer Z, Russell P, Petersen I, Chen Y, Stoelben E, Ludwig C, Schnabel P, Hoffmann H, Muley T, Brockmann M, Engel-Riedel W, Muscarella LA, Fazio VM, Groen H, Timens W, Sietsma H, Thunnissen E, Smit E, Heideman DA, Snijders PJ, Cappuzzo F, Ligorio C, Damiani S, Field J, Solberg S, Brustugun OT, Lund-Iversen M, Sänger J, Clement JH, Soltermann A, Moch H, Weder W, Solomon B, Soria JC, Validire P, Besse B, Brambilla E, Brambilla C, Lantuejoul S, Lorimier P, Schneider PM, Hallek M, Pao W, Meyerson M, Sage J, Shendure J, Schneider R, Büttner R, Wolf J, Nürnberg P, Perner S, Heukamp LC, Brindle PK, Haas S, Thomas RK.
Nat Genet. 2012 Oct;44(10):1104-10
* These authors contributed equally

High-throughput mutation profiling of primary and metastatic endometrial cancers identifies KRAS, FGFR2 and PIK3CA to be frequently mutated
Krakstad C*, Birkeland E*, **Seidel D**, Kusonmano K, Petersen K, Mjøs S, Hoivik EA, Wik E, Halle MK, Oyan AM, Kalland KH, Werner HM, Trovik J, Salvesen H.
PLoS One. 2012;7(12):e52795
* These authors contributed equally

Results presented in this dissertation are included in a publication accepted for publication in *Science* Translational Medicine. Some figures shown here are adopted from this publication.

A genomics-based classification of human lung tumors
Seidel D*, Zander T*, Heukamp LC*, Peifer M*, Bos M* *et al*.
* These authors contributed equally

Erklärung

Hiermit erkläre ich, Danila Seidel, dass ich die Dissertation mit dem Titel “Epidemiological and functional characterization of oncogenic drivers in lung cancer“ selbständig und ohne fremde Hilfe verfasst habe. Andere als die von mir angegebenen Quellen und Hilfsmittel habe ich nicht benutzt. Die den herangezogenen Werken wörtlich oder sinngemäß entnommenen Stellen sind als solche gekennzeichnet.

Ich erkläre hiermit, dass die von mir vorgelegte Dissertation bisher nicht im In- oder Ausland in dieser oder ähnlicher Form in einem anderen Promotionsverfahren vorgelegt wurde.

Köln, 24 April 2014

_____

Danila Seidel

Acknowledgment

I would like to thank Roman *Nini* Thomas for the opportunity to work in his laboratory, for his patience and constant support. It was a valuable experience to work with Nini and in his group, which shaped me personally and scientifically. I thank him for the unique opportunity to work as part of the CLCGP where I have been exposed to an international and intercultural scientific environment.

I would like to thank Peter Nürnberg and Bernd Wollnik for mentoring me during my thesis within the Interdisciplinary Program Molecular Medicine of the University of Cologne. I thank them for stimulating discussions.

I thank Thomas Zander for his enthusiasm for this project and his invaluable scientific contributions. I highly appreciate the kind and helpful attitude and I am endlessly thankful for his help in completing this thesis. I also thank Martin Peifer, Frauke Leenders, Xin Lu, Lynnette Fernandez-Cuesta, Christian Müller, and Ilona Dahmen for their help and encouragement.

Julie George, Johannes Heuckmann, Jonathan Weiß, Martin Peifer, Sandra Ortiz Cuarán, Silvia Querings, Franziska Gabler, Miriam Koker, Florian Malchers, Ilona Dahmen, Hyatt Balke-Want, Xin Lu, Christian Müller, Dennis Plenker, Lynnette Fernandez-Cuesta, Anne Raspel, and Graziella Bosco - thank you for the scientific help, for the great atmosphere at work and outside the lab. Thanks for the great time!

I would like to extend my thanks to all the collaboration partners of the University Hospital Cologne, especially Jürgen Wolf, Reinhard Büttner, Sascha Ansén, Lukas Heukamp, Marc Bos, Nadia Massoudi, Alex Florin, Uschi Rommerscheidt-Fuß, Katharina König, and Kerstin Albus. Thanks for the efficient and uncomplicated collaboration.

In this context I would also like to thank all the outstanding scientists I have been collaborating with throughout these years. I would like to express my appreciation for Elisabeth Brambilla, Sven Perner, Bill Travis, Gavin Wright, Zoe Wainer, Prue Russell, Chau Nguyen, Johannes Berg, Erik Thunnissen, Daniëlle Heidemann, Roopika Menon, Wim Timens, Yuan Chen, Ben Solomon, Toni-Maree Rogers, William Pao, Russell Hyde, and Hans-Ulrich Schildhaus, and thank them for their enthusiasm and constant efforts within the CLCGP.

I thank my family and close friends for their emotional support, patience and understanding especially in times when I have been busy and unavailable. Thank you all for your encouragement in all these years!