From Department of Oncology-Pathology
Karolinska Institutet, Stockholm, Sweden

UNDERSTANDING THE RESPONSE TO EGFR TARGETING THERAPY IN NON-SMALL-CELL LUNG CANCER USING -OMICS APPROACH

Yan Zhou Tran

Karolinska Institutet

Stockholm 2020

Published by Karolinska Institutet.
Printed by US-AB

ISBN 978-91-7831-855-1

Understanding the Response to EGFR Targeting Therapy in Non-Small-Cell Lung Cancer Using -omics Approach
THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Yan Zhou Tran

Principal Supervisor:
Lukas Orre, Senior Researcher
Karolinska Institutet
Department of Oncology-Pathology

Co-supervisor(s):
Janne Lehtiö, Professor
Karolinska Institutet
Department of Oncology-Pathology

Petter Brodin, Associate Professor
Karolinska Institutet
Department of Women's and Children's Health
Division of Clinical Pediatrics

Opponent:
Pernilla Wikström, Professor
Umeå University
Department of Medical Biosciences

Examination Board:
Ola Söderberg, Professor
Uppsala University
Department of Pharmaceutical Biosciences

Leopold Ilag, Associate Professor
Stockholm University
Department of Materials and Environmental Chemistry

Charlotte Rolny, Associate Professor
Karolinska Institutet
Department of Oncology-Pathology

The public defence will take place at Nanna Svartz (J3:12), BioClinicum, Solnavägen 30, 171 64 Solna on Friday, August 14th, 2020 at 09:00 AM

ABSTRACT

NSCLC is the leading cause of cancer-related death worldwide. EGFR targeted therapy is used to inhibit the progress of NSCLC, but patients inevitably develop resistance to EGFR-TKIs.

The work presented in this thesis aimed at better understanding response and resistance mechanisms to EGFR-TKIs in NSCLC and to find novel biomarkers as well as drug targets or combination therapies for NSCLC.

Paper I suggested a group of miRNAs with AAGUGC motif as oncomotif-miRNAs. Through the control of their target tumor suppressors, the oncomotif-miRNAs are part of the oncogenic signaling network that regulates NSCLC's and other types of cancer's cell proliferation.

Paper II identified that BCL6, FGFR2, and JAK3 were upregulated after EGFR-TKI treatment. Moreover, BCL6 together with EGFR were confirmed to contribute to treatment escape. Dual targeting of BCL6 and EGFR could therefore be a potential combination therapy for treating NSCLC.

Paper III reported that *CDKN2A* loss is associated with EGFR-TKIs sensitivity in EGFRwt NSCLC, and that *BCL2L1* (encoding for BCL-xL) overexpression existed before EGFR-TKI treatment in a subset of EGFR-TKI responding cell lines. Additionally, EGFRwt/KRASwt, *CDKN2A* deleted NSCLC is sensitive to BCL-xL and EGFR dual inhibition, which could be a potential combination therapy for patients with this profile. These findings suggest that *CDKN2A* deletion can be used as a biomarker to select EGFRwt KRASwt patients for EGFR-TKI based combination therapy.

Paper IV identified the upregulation of AXL and GAS6, EMT, MAPK pathway reactivation, and AXL dependent CDK1 phosphorylation as potential resistance mechanisms of the third-generation EGFR-TKIs.

The work summarized in this thesis add new knowledge to understand the EGFR-TKI response and resistance mechanisms in NSCLC. From our work *CDKN2A* and oncomotif-miRNA have been identified as interesting candidates that can be investigated for use as biomarkers in NSCLC. Moreover, we propose that BCL6 or BCL-xL inhibitors together with EGFR-TKIs should be further investigated as a combination therapy.

LIST OF SCIENTIFIC PAPERS

I. **Zhou Y**, Frings O, Branca RM, Boekel J, le Sage C, Fredlund E, Agami R, *Orre LM*. **microRNAs with AAGUGC seed motif constitute an integral part of an oncogenic signaling network.** *Oncogene*. 2017 Feb 9;36(6):731–745.

II. **Zhou Tran Y**, Minozada R, Cao X, Johansson HJ, Mamede Branca RM, Seashore-Ludlow B, *Orre LM*. **Immediate adaptation analysis implicates BCL6 as an EGFR-TKI combination therapy target in NSCLC.** *Mol Cell Proteomics*. 2020 June 1; 19(6):928-943.

III. *Orre LM*, **Zhou Tran Y**, Minozada R, Seashore-Ludlow B, Östling P, Kallioniemi OP, *Lehtiö J.* **CDKN2A status predicts the response to EGFR targeting therapy in EGFRwt NSCLC.** Manuscript.

IV. **Zhou Tran Y**, Panizza E, Branca RM, Lehtiö J, *Orre LM*. **Integrative proteomics and phosphoproteomics profiling of NSCLC cell lines to explore osimertinib resistance mechanisms.** Manuscript.

Italics – corresponding author

Publication not included in this thesis

Zhu Y, Orre LM, **Zhou Tran Y**, Mermelekas G, Johansson HJ, Malyutina A, Anders S, *Lehtiö J*. DEqMS: a method for accurate variance estimation in differential protein expression analysis. *Mol Cell Proteomics*. 2020 June 1; 19(6):1047-1057.

CONTENTS

LIST OF ABBREVIATIONS

ADC	Adenocarcinoma
ALK	Anaplastic lymphoma kinase gene
ASR	Age-standardized rate
AXL	AXL receptor tyrosine kinase
BRAF	B-Raf proto-oncogene
CDKN2A	Cyclin-dependent kinase inhibitor 2A
CDKN2B	Cyclin-dependent kinase inhibitor 2B
CGC	COSMIC cancer gene census
CK5	Cytokeratin 5
CK6	Cytokeratin 6
DDR2	Discoidin domain receptor tyrosine kinase 2
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
ERBB2	Receptor tyrosine kinase erb-B2
ERBB3	Receptor tyrosine kinase erb-B3
FASP	Filter aided sample preparation
FGFR1	Fibroblast growth factor receptor 1
FGFR2	Fibroblast growth factor receptor 2
FPKM	Fragments per kilobase of transcripts per million fragments mapped
GAS6	Growth arrest specific 6
GO	Gene ontology
GSEA	Gene set enrichment analysis
HGF	Hepatocyte growth factor
HiRIEF	High-resolution isoelectric focusing
IEF	Isoelectric focusing
IPG	Immobilized pH gradient
KRAS	KRAS proto-oncogene
LC	Liquid chromatography
LCC	Large cell carcinoma
MEK1	Mitogen-activated protein kinase kinase 1
MET	Hepatocyte growth factor receptor
miRNA	microRNA
MS	Mass spectrometry
mTOR	Mammalian target of rapamycin
MYC	MYC proto-oncogene
NGS	Next generation sequencing
NSCLC	Non-small-cell lung cancer
PDGFR1	Platelet-derived growth factor receptor-α
PI3KCA	PI3K catalytic subunit-α
PTEN	Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN
RB1	Retinoblastoma-associated protein
RET	Ret proto-oncogene

RNA-seq	RNA sequencing
ROS1	ROS proto-oncogene 1 receptor tyrosine kinase
RPKM	Reads per kilobase per million mapped reads
RTK	Receptor tyrosine kinase
SCLC	Small-cell lung cancer
SOX2	SRY-box 2 transcription factor 2
SP3	Single-pot, solid-phase-enhanced sample preparation
SQCC	Squamous cell carcinoma
TK	Tyrosine kinase
TKI	Tyrosine kinase inhibitor
TNM	Tumor, node, metastasis
TP53	Tumor protein P53
TPM	Transcripts per kilobase million
TTF1	Thyroid transcription factor 1
WHO	World health organization

1 BACKGROUND

Lung cancer is the most commonly diagnosed cancer and the leading cause of cancer-related death worldwide, which causes 1.8 million deaths per year (18.4% of all cancer deaths)[1]. Risk factors of developing lung cancer include cigarette smoke, secondhand smoke (passive smoking), radon, asbestos, diesel exhaust, and ionizing radiation exposure. Smoking is highly associated with lung cancer development, and are estimated to be the cause of 85-90% lung cancer cases[2]. Lung cancer can be classified into two broad histological subtypes based on the tumor cell appearance: non-small-cell lung cancer (NSCLC, accounting for 85% of the lung cancer cases) and small-cell lung cancer (SCLC, accounting for 15% of the cases)[2]. Compared with NSCLC, SCLC has some unique characteristics, including high lethality, rapid tumor growth, early and aggressive metastasis, and near widespread inactivation of TP53 and RB1 tumor suppressor genes[3]. In this thesis, the discussion will focus on NSCLC.

1.1 LUNG CANCER EPIDEMIOLOGY

Lung cancer incidence and mortality are usually highest in developed countries. Among males, lung cancer is the primary cause of cancer-related death in East Europe, West Asia, North Africa, East Asia (China, Japan, and South Korea), and Southeast Asia (Burma, Philippine, Indonesia). Especially in East Asia, the population incidence of lung cancer is more than 40 age-standardized rate (ASR) per 100,000 person per year[4]. Among females, lung cancer is still one of the leading causes of cancer-related death. Countries with the high incidence (> 35 ASR per 100,000 person per year) of lung cancer include United States of America, Europe (Hungary, Serbia, Greece, Belgium, Denmark, Poland, France, Bosnia), Turkey, China, and North Korea[4]. Although smoking is the most notable risk factor of lung cancer, 10-15% of non-smokers developed lung cancer[5].

1.2 CLASSIFICATION

NSCLC is classified into three major histological types: adenocarcinoma (ADC, 50% of NSCLC), squamous cell carcinoma (SQCC, 40%), and large cell carcinoma (LCC, 10)[2,6]. ADC and SQCC are the predominant NSCLC subtypes[2,6]. ADC has glandular histology and it forms in glands that secrete mucus. It arises from the cells located along the alveolar wall or bronchiolar epithelium. ADC often express markers such as thyroid transcription factor 1 (TTF1) that can be detected with immunohistochemistry[6]. SQCC develops in larger, more central airways, progresses through squamous differentiation, and has been associated with smoking. Markers frequently used to identify SQCC include SOX2, p63, cytokeratin 5 (CK5), and cytokeratin 6 (CK6). Cells of LCC do not have a squamous or glandular shape, or do not express ADC or SQCC markers[6]. Since 2003, the significant changes in lung cancer basic and clinical research field reveals that a more detailed histological classification of ADC is associated with patient prognosis and survival. In contrast, SQCC and LCC classification have largely remained the same. The 2015 WHO classification of lung tumors started to use the invasive statuses of ADC to further group the early-stage lung cancer. Early-stage lung cancer is split into three subtypes: adenocarcinoma *in situ*, minimally invasive adenocarcinoma, and invasive adenocarcinoma[2]. Surgery is recommended for non-invasive ADC, as the 5-year survival rate almost reaches 100%. Invasive ADC can be further

classified based on tumor growth type: lepidic, acinar, papillary, solid, and micropapillary. Prognosis of these invasive ADC subtypes rank from good to poor: lepidic ADC > acinar or papillary ADC > solid or micropapillary ADC. In general, the aim of the exact histopathological classification for NSCLC is to guide the treatment and assess the prognosis of patients[7,8].

1.3 MUTATIONS IN NSCLC

NSCLC is a group of diseases that has genetic and cellular heterogeneity. Common driver mutations that have been implicated to influence the oncogenesis of NSCLC have been reported in many studies, including *EGFR* mutation[9–11], *FGFR1/2* amplification or mutation[12,13], *ALK* fusion[14], *MET* amplification[15], *ROS1* fusion[16–18], *RET* fusion[19,20], *ERBB2* amplification or mutation[21,22], *KRAS* mutation[23,24], *BRAF* mutation[25–27], *MEK1* mutation[28], and *CDKN2*A deletion[29–31]. The NSCLC subtypes have different molecular profiles from each other. An overview on key oncogenes, and the main gene alterations along with FDA approved targeted drugs for NSCLC is summarized in Table 1.

Gene	Status	Frequency (%)		FDA approved targeted drugs
		ADC	SQCC	
RTKs				
EGFR	Amplification	6.8	6.8	Gefitinib, Erlotinib, Afatinib, Osimertinib
	Mutation	20.2	3.2	
FGFR1	Amplification	2.1	17.6	Erdafitinib, Pemigatinib
FGFR2	Amplification	0.3	0.3	
FGFR2	Mutation	1.5	2.7	
ALK	Fusion	1.7	N/A	Alectinib, Crizotinib, Ceritinib
MET	Amplification	2.6	1.4	Crizotinib
ROS1	Fusion	1.7	N/A	Crizotinib
ERBB2	Amplification	2.5	2.6	Trastuzumab, Neratinib, Lapatinib, Tucatinib
	Mutation	2.9	2.3	
RET	Fusion	0.9	N/A	Cabozantinib, Vandetanib
DDR2	Mutation	2.7	2.4	Dasatinib
PDGFRA	Mutation	5.5	4.5	Sunitinib
Signaling				
KRAS	Mutation	29.1	1.3	N/A
BRAF	Mutation	6.6	3.7	Vemurafenib, Dabrafenib + Trametinib
PIK3CA	Mutation	5.7	12.8	Alpelisib
MEK1	Mutation	1.5	1.1	Selumetinib
Transcription factors				
SOX2	Amplification	1.6	43.9	N/A
MYC	Amplification	8	8.4	N/A
Cell cycle				
CDKN2A	Mutation	4.8	15.3	N/A

	Deletion	13.9	27	
CDKN2B	Deletion	13.4	26.8	N/A
Tumor suppressor				
TP53	Mutation	50.3	83	N/A
PTEN	Mutation	2.1	9.4	N/A

Table 1: Overview of common gene alterations in lung ADC and SQCC. This table shows an overview on key oncogenes, and the main gene alterations along with FDA approved targeted drugs for NSCLC. The abbreviations used in this table: *EGFR,* epidermal growth factor receptor; *FGFR1*, fibroblast growth factor receptor 1; *FGFR2*, fibroblast growth factor receptor 2; *ALK*, anaplastic lymphoma kinase; MET: MET proto-oncogene; *ROS1*: ROS proto-oncogene; *ERBB2*: Erb-B2 receptor tyrosine kinase 2; *RET*: Ret proto-oncogene; *DDR2*: discoidin domain receptor tyrosine kinase 2; *PDGFRA*, platelet-derived growth factor receptor-α; *KRAS*: KRAS proto-oncogene; *BRAF*: B-Raf proto-oncogene; *PIK3CA*: PI3K catalytic subunit-α; *MAP2K1*: mitogen-activated protein kinase kinase 1; *SOX2*, SRY-box 2 transcription factor 2; *MYC*: MYC proto-oncogene; *CDKN2A*, cyclin-dependent kinase inhibitor 2A; *CDKN2B*, cyclin dependent kinase inhibitor 2B; *TP53*: tumor protein P53; *PTEN*: phosphatase and tensin homolog. The frequency data from the combined study was collected using cBioPortal[32,33]. Lung ADC frequency data are from the combination of 7 studies (Broad, Cell 2012; MSKCC, Science 2015; TCGA, Firehose Legacy; TCGA, Nature 2014; TCGA, PanCancer Atlas; TSP, Nature 2008; MSKCC, Cancer Discov 2017). Lung SQCC frequency data are from the combination of 3 studies (TCGA, Firehose Legacy; TCGA, Nature 2012; TCGA, PanCancer Atlas). Data in this table partially taken from Chen *et al*., 2014[6].

The rapid development of high throughput sequencing techniques dramatically changed the way of describing NSCLC at the molecular level. Previously, The Cancer Genome Atlas (TCGA) research group released their molecular profiling results of lung ADC and lung SQCC[31,34], which included a thorough analysis of the driver oncogenes and mutations. TCGA identified high rates of somatic mutations and listed 18 significant genetic mutations in lung ADC[34]. Alterations in key pathways of lung ADC are visualized in Fig. 1.

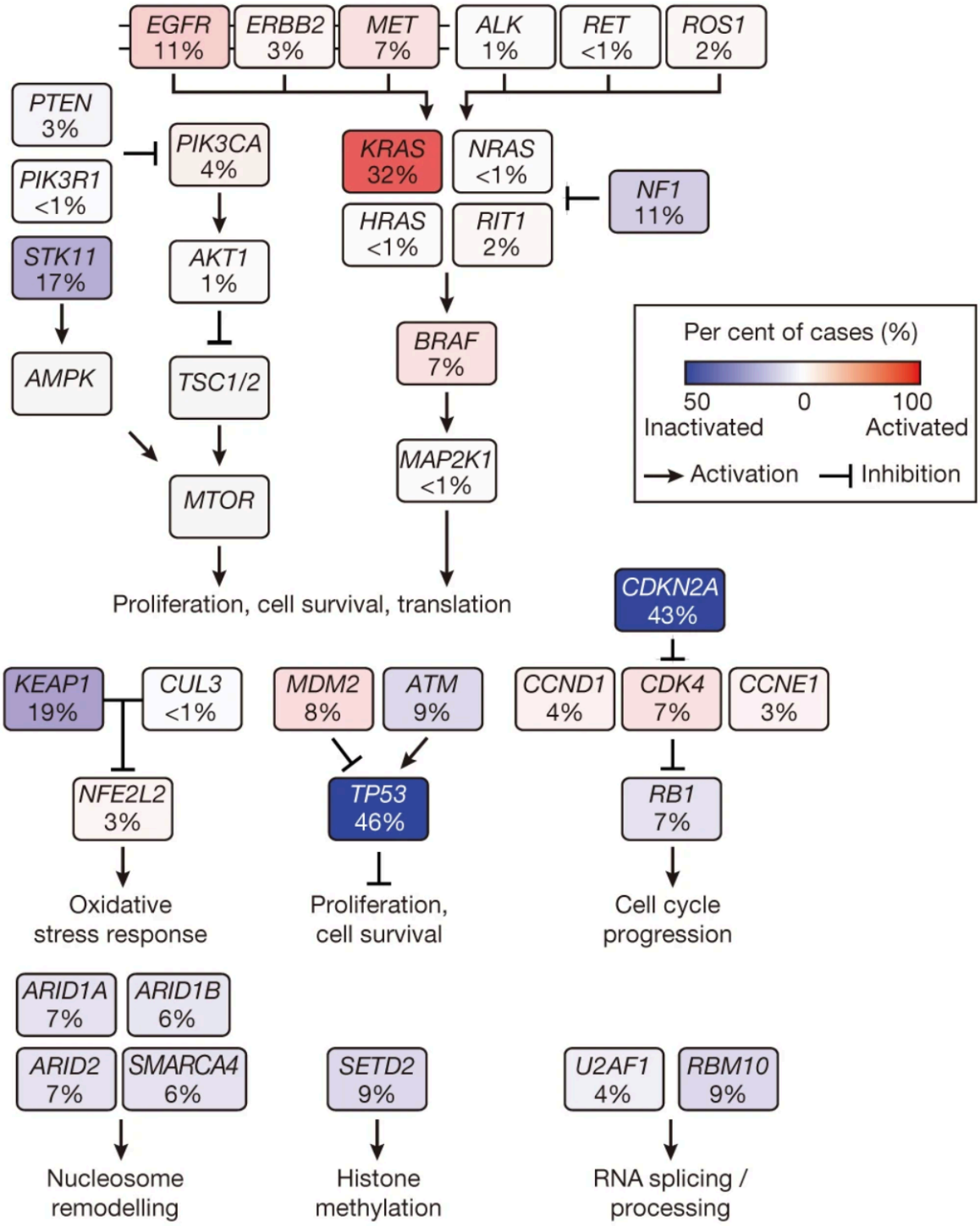

Figure 1: Altered pathways in lung ADC. Alterations in key pathways of lung ADC such as Receptor tyrosine kinase (RTK) signaling, mTOR signaling, cell proliferation, and cell cycle progression. *Reprinted under the Creative Commons Attribution 4.0 International License, http://creativecommons.org/licenses/by/4.0/ from Nature 511, 543-550(2014). The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma.*

TCGA reported 10 significantly mutated gene mutation in SQCC. Pathways that were frequently altered in lung SQCC were related to oxidative stress response and squamous cell differentiation. The paper also provided frequently altered pathways found in SQCC, as visualized in Fig. 2.

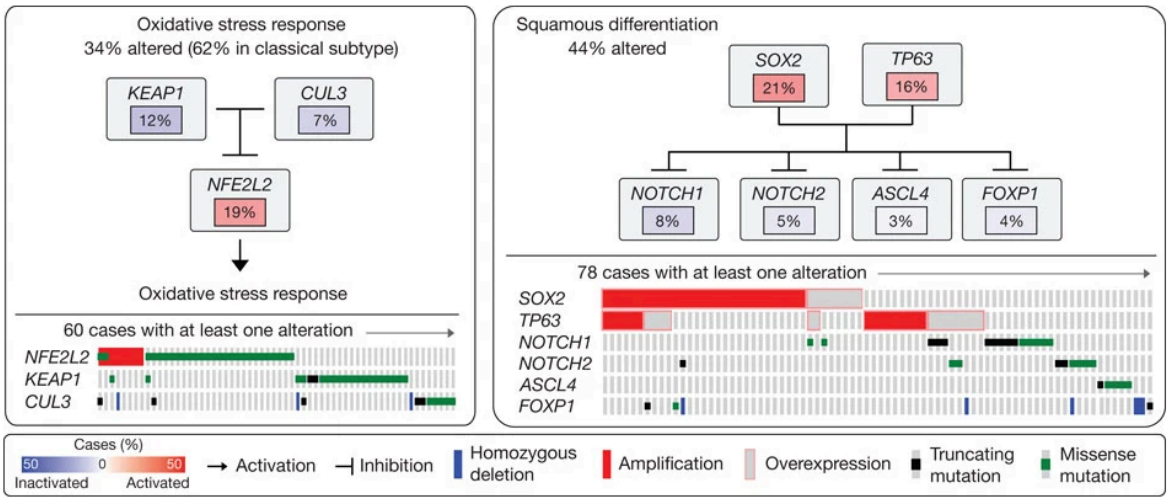

Figure 2: Altered pathways in lung SQCC. The frequently alterated pathways are showed. Left part is oxidative stress response pathway and the right is squamous differentiation pathway. *Reprinted under the Creative Commons Attribution 4.0 International License, http://creativecommons.org/licenses/by/4.0/ from Nature 489, 519-525(2012). The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers.*

1.4 SCREENING

The aim of lung cancer screening is to detect the tumor at an early-stage since it is challenging to get curative treatment for advanced-stage tumors. The National Lung Cancer Screening Trial in the United States proposed that annual screening with low-dose computed tomography could potentially reduce lung cancer mortality by 20% or more in high-risk groups[35]. China did a similar screening experiment in a rural area since 2009 to determine proper screening guidelines for Chinese citizens. Both countries issued guidelines for lung cancer screening that recommend screening healthy patients with certain risk factors, which are quite similar: older age group and history of smoking. The National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology for lung cancer screening (Version 3.2018) set the risk group criteria to be aged > 50 years, at least 20 pack per year smoking history, either quit smoking < 15 years or has additional risk factors to develop lung cancer[36]. The Chinese national lung cancer screening guideline (2018 version) has similar criteria, except a lifted smoke cessation bar to < 5 years and some added high-risk factors (e.g. long history with indoor coal smoke) for cities with high lung cancer incidence[37]. Moreover, patients with a specific size of lung nodules need to be further examined and undergo further diagnostics[36,37].

1.5 DIAGNOSIS

Typical symptoms of NSCLC include cough (> 50% of cases), hemoptysis (25-40% of the cases), chest pain (or other parts of body pain, due to distant metastasis), dyspnea, fever, and wheeze. Suspected patients are recommended to undergo various examinations, including taking tissue biopsy for NSCLC diagnosis and further classification. Regular examinations include serologic testing of NSCLC diagnostic markers (increased level of carcinoembryonic antigen, cytokeratin fragment 21-1, squamous cell carcinoma antigen), imaging examination (e.g., X-ray, computed tomography, magnetic resonance imaging, and positron emission

tomography-computed tomography), endoscopy examination, and histopathological examination. All the examinations contribute to the tumor, node, and metastasis (TNM) classification of NSCLC[2,38,39]. When tissue biopsy is available, an immunohistochemical examination can be performed for diagnosis, using biomarkers (TTF-1, p63, CK5, and CK6) to distinguish between ADC and SQCC[38]. Leftover tissue can then be used for molecular testing (e.g. EGFR mutation, *ALK* fusion gene, and *ROS1* fusion gene detection for advanced-stage NSCLC) to decide therapeutic stratagy[2,38,40].

1.6 TREATMENT

Essential information to determine an appropriate treatment towards individual NSCLC patients is the tumor's histological classification (chapter 1.2), staging, and molecular classification.

Treatment options for NSCLC are mainly surgery, chemotherapy, radiotherapy, targeted therapy, and immunotherapy. The exception being that immunotherapy is not yet available for LCC. LCC occurs rarely but is more malignant and metastasizes faster than other types of NSCLC. SQCC grows slower than other types of NSCLC. Therefore, surgery is more suitable to treat SQCC at an early stage.

The staging classification is useful to determine if the tumor is removable by surgery or not. The staging (clinical, or pathological) of NSCLC classifies into 0, I, II, III, IV, based on the size, location, and metastasis condition of the tumor. The only difference between clinical and pathological staging is that the former is based on the test results (e.g. computed tomography scan) before surgery, and the latter is based on the conclusion from tissue taken during surgery. In general, tumor in stage 0/I/II is surgically removable. In contrast, surgery is not an option for patients with tumors in stage III/IV with metastases to lymph nodes or even to other organs.

The survival rate of NSCLC patients is highly related to the clinical stage (size and location of the tumor); in short, the more tumor metastases, the lower the survival rate is. The 5-year survival rate drops from 54% to 2-15% in NSCLC when the tumor develops from stage I to III/IV[41,42]. The clinical management of each patient differs depending on the tumor characteristics. Detecting the tumor in early-stage is vital for patient survival and increases the applicability of curative treatment (e.g. surgery). However, the reality is that many NSCLC patients are first diagnosed at an advanced stage when the disease is usually incurable.

In the past decade, breakthroughs of molecular diagnostics have improved the clinical management of NSCLC dramatically, as patients can be tested and profiled for the presence of well-known driver mutations. Some of the well-known driver mutations have been used as predictive treatment-biomarkers to select patients who will respond to treatment in NSCLC clinical practice. Successful examples include activating mutations in EGFR, *ALK* chromosomal translocation and fusion, and *ROS1* chromosomal translocation and fusion. As drug targets, these mutations have been used for designing targeted cancer drugs (e.g.

gefitinib). Targeted drugs are selected to treat patients who carry specific mutations and for NSCLC, one of the well-known targets is EGFR activating mutation[43].

1.7 EGFR TARGETED THERAPY

1.7.1 EGFR

The *EGFR* gene is located on chromosome 7, coding for a matured 1186-residue, 170-kDa transmembrane protein EGFR. From the N- to C-terminal, EGFR comprises an extracellular domain that controls the ligand binding, a transmembrane and juxtamembrane segment, an intracellular tyrosine kinase domain, and a C-terminal tail domain[44]. It belongs to the ERBB family, which includes four members with similar molecular structures (EGFR/ERBB1, ERBB2(HER2), ERBB3, ERBB4). Upon binding with its ligands (e.g. EGF, TGFα) EGFR dimerizes with another EGFR or ERBB and autophosphorylates its tyrosine kinase domain. The activated EGFR can interact with adaptor molecules and regulates the intracellular signaling through RAS-RAF-MEK-MAPK, PI3K-PTEN-AKT, and JAK-STAT pathways. In this way, the EGFR downstream signaling controls cell proliferation, survival, and apoptosis[45,46].

The studies of EGFR and its pathway began with pioneer work initiated by 1986's Nobel prize winner Stanley Cohen. In the 1960s, he isolated the epidermal growth factor (EGF) and named it based on the observation that it stimulates the epidermal cell growth *in vitro* and *in vivo*[47–49]. In 1977, Manjusri Das and colleagues identified a cell surface receptor for EGF in mouse cells[50]. In the early 1980s, Cohen reported that in A431 cells, EGF binds to its receptor, a 170 KDa membrane protein named EGF receptor (EGFR)[51,52]. Later on, Ushiro and Cohen discovered that the EGF-enhanced phosphorylation reaction in EGFR mainly happened on its tyrosine residues, which suggested that the EGF signaling involved protein tyrosine kinase activity[53]. In the mid of 1980s there were studies that reported the sequencing of the *EGFR* gene, and that the *EGFR* gene was amplified and rearranged in cancers[54–61]. The expression level of EGFR in normal human fibroblasts is estimated to be 40,000-100,000 receptors per cell, whereas in the EGFR overexpressed epidermoid carcinoma cell line A431 there was $2*10^6$ receptors per cell[58,59]. In the late 1980s, scientists suggested that the EGFR status could be used as a predictor of tumor size, relapse-free survival, and patient prognosis in breast cancer[60]. EGFR overexpression associates with poor patient prognostics and progression of many cancers, including head and neck, ovarian, cervical, bladder, breast cancer, vulva cancer, and NSCLC[59,61]. EGFR was the first among the RTK family to be linked with cancer when the receptor expression level is abnormal.

Previous meta-analysis reported that the EGFR mutation frequency in NSCLC patient is 32.3%, and among the Asian patients, mutation frequency was significantly higher (38.8%) than Caucasian patients (17.4%)[62]. Other factors that associate with high EGFR mutation include smoking status, histology and gender[62]. EGFR activating mutations frequently occur in kinase-coding regions (exon 18-21), clustered around the ATP-binding pocket of EGFR tyrosine kinase domain[63]. Reported EGFR activating mutations include exon 19 deletions, point mutation exon 18 G719X (X indicates A or C), and exon 21 L858R[9,10]. In total, exon 21 L858R and exon 19 deletions account for ~80-90% of EGFR tyrosine kinase mutations in NSCLC (data from COSMIC cancer browser using their NSCLC samples[64]). The activating

mutation results in the constant activation of EGFR kinase activity, which results in constant activation of the downstream signaling and uncontrolled cancer cell proliferation and survival[65]. EGFR is therefore a prominent drug target for NSCLC patients with EGFR activating mutations.

1.7.2 EGFR-TKIs

Developing the EGFR targeted drugs includes two different approaches. One is to design an antibody that targets the EGFR extracellular domain, while the other is to design a small-molecular EGFR tyrosine kinase inhibitor (TKI) that targets intracellular tyrosine kinase domain of EGFR. This section will focus on EGFR-TKIs. Currently, three generations of EGFR-TKIs for treating NSCLC are approved by FDA for treating NSCLC, and the fourth generation of EGFR-TKIs are undergoing the drug development process.

First-generation of FDA approved EGFR-TKIs are gefitinib (Iressa®, AstraZeneca) and erlotinib (Tarceva®, Genetech/Roche). Both gefitinib and erlotinib are reversible small molecular inhibitors based on a 4-anilinoquinazoline scaffold, and are designed to target exon 19 deletion or L858R mutant EGFR. They share similar binding mechanisms: inhibiting kinase function by mimicking the ATP molecule to competitively bind to the ATP binding pocket of the EGFR TK domain, hence reduces the downstream signaling that regulates cell proliferation and survival. During the clinical trials, researchers noticed that patients with EGFR L858R mutation in exon 21 or deletions in exon 19 responded well towards gefitinib or erlotinib[9,10,66]. Yun and colleagues reported that the potency of gefitinib to $EGFR^{L858R}$ is 100 times fold higher than wild type EGFR, which explains why patients carrying $EGFR^{L858R}$ or EGFR exon19 in-frame deletion respond to the first generation of EGFR-TKIs[67]. Unfortunately, the first-generation EGFR-TKIs brought epithelial toxicity as a side effect due to the inhibition of wild type EGFR in the skin and GI tract. Later on, researchers discovered that not all NSCLC patients could benefit from the first-generation of EGFR-TKIs because of intrinsic and acquired resistances (see 1.8.2). For instance, EGFR T790M mutation makes the tumor resistant to first generation EGFR-TKIs[68–70].

The second-generation of EGFR-TKIs approved by the FDA includes afatinib (Gilotrif®, Boehringer Ingelheim) and dacomitinib (Vizimpro®, Pfizer). Both drugs share a 4-anilinoquinazoline structure and are irreversible pan-ERBB TKIs (EGFR, ERBB2, ERBB4). They are not only ATP mimics that can covalently bind to Cys797 of EGFR, but can also irreversibly bind to other ERBB family member receptors and block their enzyme activity. For instance afatinib, it binds with wild type EGFR through forming a hydrogen bond between Met793 and the quinazoline core of afatinib, as well as a covalent bond between Cys797 and the acrylamide group of afatinib[71]. Moreover, it inhibits other ERBB family members by forming covalent bonds also with these kinases[71]. The irreversible binding of afatinib to ERBB2 inactivates the preferred dimerization of ERBB2 with EGFR and hence prevent downstream signaling. Although the second-generation EGFR-TKIs showed potential to overcome T790M resistance by irreversible covalent binding with the target kinase, the drug efficacy in patients was insufficient, as demonstrated by the LUX-Lung 1 clinical trial result of 7% of responding rate and 3.3 months of progression-free survival[72].

Irreversible binding towards EGFRwt caused increased toxicity, which further limited the clinically achievable concentration due to side effects such as skin rash and diarrhea.

The design of the third-generation of EGFR-TKIs considered several aspects to be able to target $EGFR^{T790M}$ more successfully. The drug candidates needed to be mutant-specific in order to avoid binding to EGFRwt and hence lessen side effects. The drug candidates also needed to have a substituent to bind with Cys797 and avoid steric conflict with Met790 in order to stabilize binding[73]. The FDA approved drug osimertinib (Tagrisso®, AstraZeneca) managed to satisfy the three mentioned aspects[73]. Osimertinib is a pyrimidine-based EGFR-TKI, which contains an acrylamide group for Cys797 binding. Cross and colleagues showed that the binding of osimertinib to $EGFR^{T790M}$ took place via two hydrogen bonds to the hinge region (Met793), and a covalent bond between the acrylamide group of osimertinib and Cys797[73]. Osimertinib selectively and irreversibly target sensitizing mutants of EGFR ($EGFR^{L858R}$ and EGFR ex19del) as well as the resistance mutant $EGFR^{T790M}$. The affinity of osimertinib to EGFRwt is much lower compared to any of the following: T790M, L858R, and ex19del[73]. Less binding to EGFRwt reduced the toxicity of osimertinib, supported by the fact that the skin rash occurred in 58% of patients with osimertinib treatment compared to 78% with first-generation of EGFR-TKIs[74]. Osimertinib showed better progression free survival (PFS) of 9.6 months in $EGFR^{T790M}$ harboring patients and significantly reduced toxicity due to lower unspecific binding to wild type EGFR[75]. Later on, $EGFR^{C797S}$ was identified as a leading acquired resistance mechanism to third-generation of EGFR-TKIs, which guided the development of fourth-generation EGFR-TKIs[76,77].

Currently, no fourth-generation of EGFR-TKIs are approved. The design of fourth-generation EGFR-TKIs was based on the acquired mutation $EGFR^{C797S}$ after treatment with third-generation EGFR-TKIs. The first fourth-generation EGFR-TKIs reported was EAI045[78]. Different from the previous three generations of EGFR-TKIs that target the ATP site of EGFR, it is an allosteric, non-ATP competitive inhibitor with high potency and selectivity for $EGFR^{T790M}$/$EGFR^{L858R}$ mutation. When combined with EGFR targeting antibody cetuximab, EAI045 was able to induce tumor regression in mouse models with $EGFR^{L858R/T790M/C797S}$ triple mutation[78]. Fig.3 below shows a history of EGFR-TKI related discoveries and FDA approvals. The next section will focus on introducing the resistance mechanisms of EGFR-TKIs.

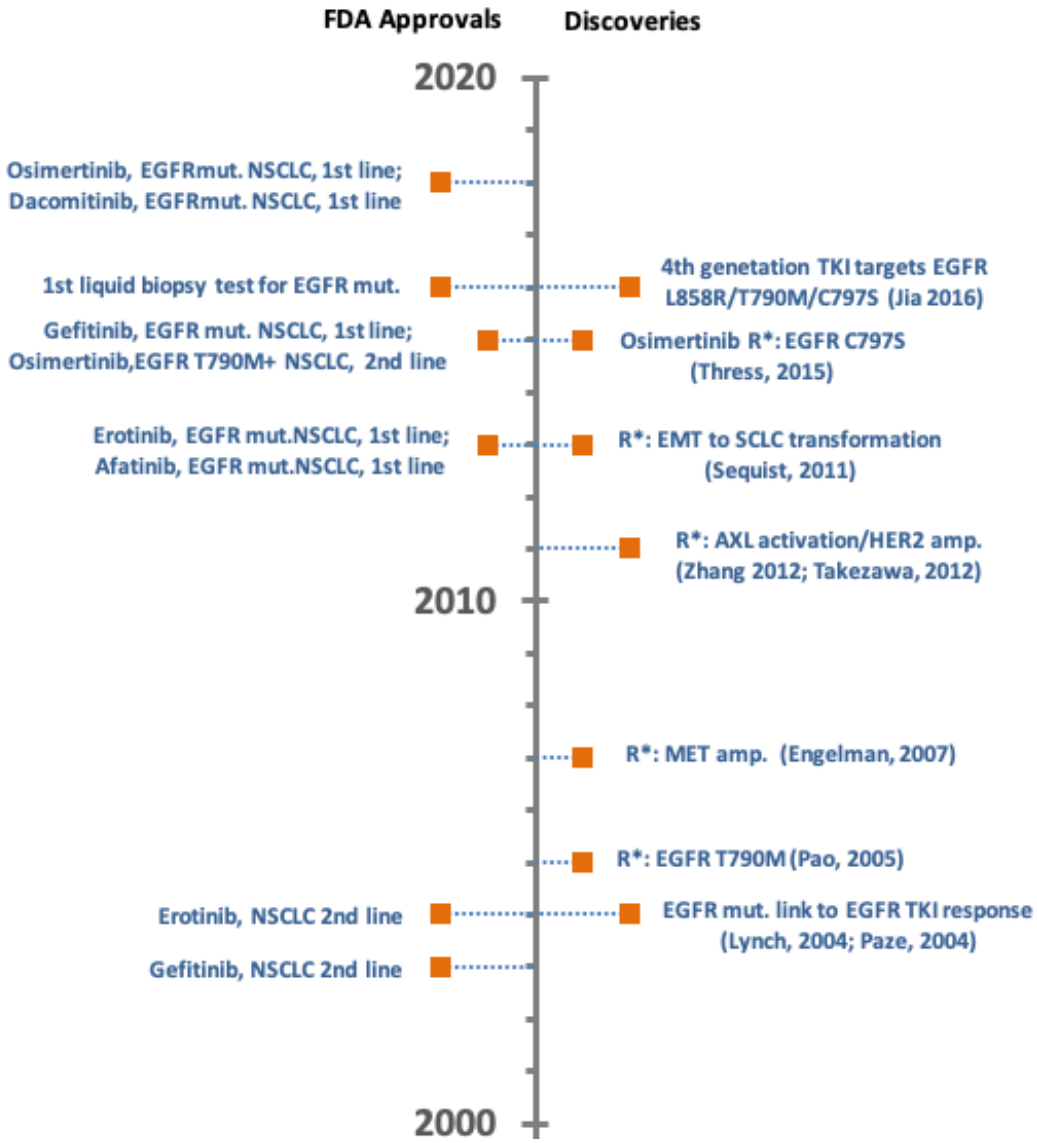

Figure 3: A history of EGFR targeted therapy and FDA approved drugs for NSCLC. The left side of the timeline listed the relevant FDA approvals of EGFR-TKIs. The right side of the timeline listed the critical scientific discoveries regarding EGFR targeted therapy. All the events have been placed chronologically, with the most recent events closer to the top. R*: resistance mechanisms to EGFR-TKIs. Mut.: EGFR activating mutations (exon 19 deletion or exon 21 L858R mutation).

1.8 MECHANISMS OF RESISTANCE TO EGFR-TKIS

EGFR targeted therapy using EGFR-TKIs improves the prognosis of NSCLC patients, but the response is usually temporary, patients inevitably develop resistance to EGFR-TKIs. The resistance mechanisms to EGFR-TKIs can be classified into intrinsic resistance and acquired resistance mechanisms. These resistance mechanisms of EGFR-TKIs are summarized in Fig.4.

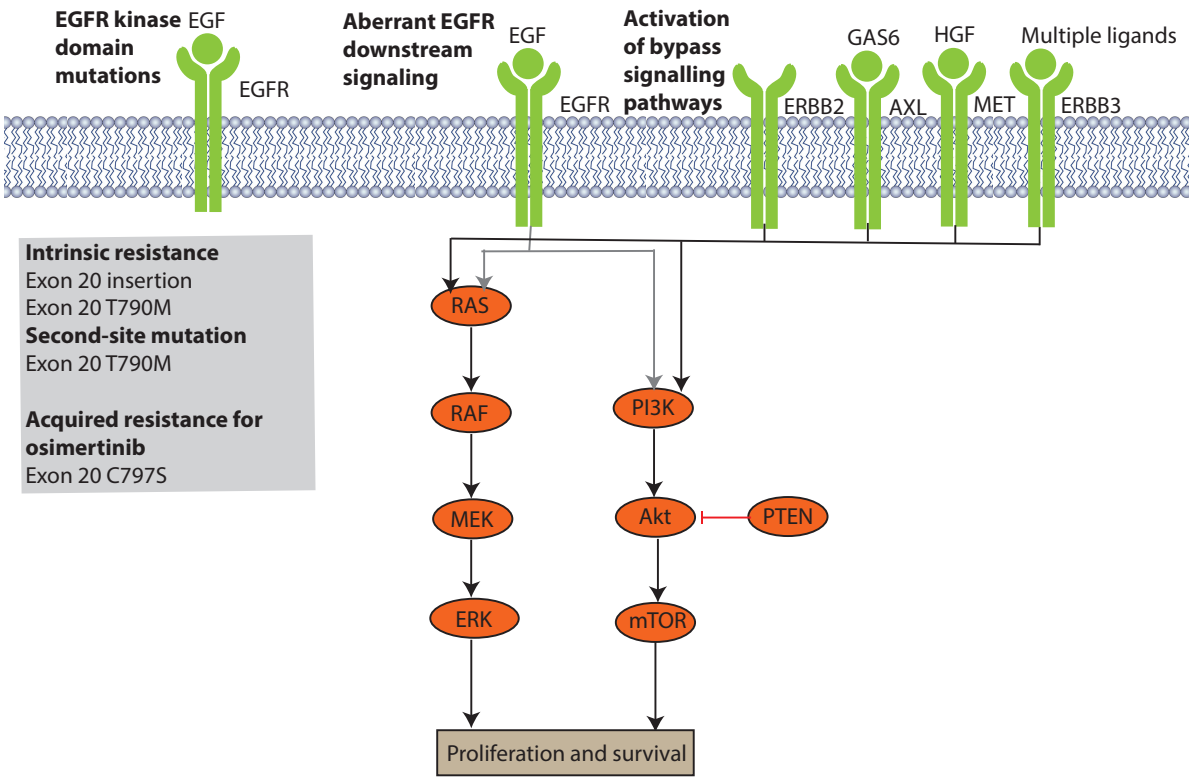

Figure 4: Resistance mechanisms towards EGFR-TKIs in NSCLC. The resistance mechanisms towards EGFR-TKIs include primary resistance, second-site mutations on EGFR, failure in EGFR downregulations and bypass signaling pathways. Abbreviations: GAS6: growth arrest specific 6; AXL: AXL receptor tyrosine kinase; HGF: hepatocyte growth factor; MET: hepatocyte growth factor receptor; ERBB3: receptor tyrosine kinase erb-B3; mTOR: mammalian target of rapamycin; PTEN: phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN. Adapted from Rotow J. 2017[79].

1.8.1 Intrinsic resistance mechanisms

Intrinsic resistance to EGFR-TKIs is defined as immediate inefficacy when receiving EGFR-TKI treatment.

1.8.1.1 Resistant EGFR mutations

The most frequently observed intrinsic EGFR-TKI resistant mutation is exon 20 insertion, which accounts for 4-10% of EGFR mutation[80,81]. Wu and colleagues first reported the association of inadequate gefitinib response with EGFR exon 20 deletion in 2008[80]. ~90% of the exon 20 insertions lie near the end of the C-helix between amino acid 768-774, with 1-4 amino acids[82]. The crystal structure of EGFR with an exon 20 insertion shows that the insertion sits at the end of C-helix to lock the helix in its activating position and blocks the conformational change that is required for EGFR-inactivation. In this manner, the activation of the EGFR mutant can no longer be inhibited by EGFR-TKIs[81]. Exon 20 insertion mutations are commonly observed in Asian non-smoker patients with adenocarcinoma histology[83].

Another intrinsic resistance mutation on the EGFR target is T790M mutation. $EGFR^{T790M}$ is occasionally detected in germline and somatic cells of patients even before EGFR-TKI treatment has been prescribed. For instance, germline T790M mutation accounts for 0.54% of

never-smoker patients[84]. In an East Asian cohort, patients who carried T790M mutation before TKI treatment had been received were associated with a shorter PFS than those with wild type EGFR[85]. The EGFR T790M mutation can also be an acquired resistance mutation and is more frequently reported as such. It is therefore also explained in the context of an acquired resistance mechanism in section 1.8.2.1.

1.8.1.2 EGFR-independent resistance mutations

Intrinsic resistance can be induced by molecular or genetic alterations that are independent from EGFR mutations, but provide a similar uncontrolled downstream pathway activation. These alterations can be referred to as EGFR-independent resistance mechanisms.

Mutations affecting EGFR downstream signaling are EGFR-independent resistance mechanisms, such as *PTEN* deletion and *PIK3CA* mutation, which continuously activates the PI3K/AKT pathway regardless of EGFR activation status[86,87]. Bypass signaling such as HGF overexpression and ALK rearrangement were also reported as intrinsic resistance mechanisms to EGFR-TKIs, which also are EGFR-independent resistance mutations[14,88,89].

Similar intrinsic resistance mechanisms were identified to the third-generation of EGFR-TKI drug osimertinib. For example, it was reported that AXL receptor tyrosine kinase (AXL) interacts with EGFR and ERBB3 to maintain cell survival and induce the osimertinib resistance[90]. In addition, ERBB2 amplification, MET amplification or SCLC transformation were all suggested to be potential intrinsic resistance mechanisms to osimertinib[91,92].

1.8.2 Acquired resistance mechanisms

Acquired resistance occurs in tumors that initially respond to therapy but later on continue to progress.

1.8.2.1 Second-site mutation on EGFR

The most common acquired resistance mechanism of EGFR-TKIs is the T790M mutation, observed in ~50% of NSCLC patients with resistance to EGFR-TKIs[93]. T790 is a gatekeeper residue of EGFR located in the ATP binding pocket. It controls the entrance of EGFR-TKIs to the ATP binding pocket. When the T790M mutation occurs, the ATP affinity to EGFR increases, and binding ability of EGFR towards first-generation EGFR-TKIs reduces[94]. The increased ATP affinity to EGFR leads to the kinase phosphorylation, hence activates the downstream pathways to promote tumor progression. Other less reported secondary mutations on EGFR include D761Y on exon 19, L747S on exon 19, and T854A on exon 21[95–97].

Unsurprisingly, the resistance to third-generation of EGFR-TKIs involves a second-site mutation. EGFR C797S mutation is a common third-generation EGFR-TKI resistance mechanism in patients with resistance to osimertinib[76,98,99]. Cys797 is located at the ATP binding site and the C797S mutation leads to covalent bond loss between EGFR and third-generation EGFR-TKIs. Other rare reported resistance mutations to third-generation of EGFR-TKIs include C797G, G796X(R/S), and L792H[100,101].

1.8.2.2 Aberrant EGFR downstream signaling

The uncontrolled activation of EGFR downstream signaling can also be caused by acquired resistance to EGFR-TKIs. In RAS / MAPK pathway, BRAF V600E mutation is a common acquired resistance mechanism that was observed following treatment with any of the three generations of EGFR-TKIs[25,102,103]. BRAF V600E mutant induces activation of BRAF, hence activates MEK-ERK. The reactivation of MAPK pathway is then independent from EGFR activation and results in lack of response to EGFR inhibition. ERK2 amplification, was also reported as a third-generation EGFR-TKI resistance mechanism by also reactivating the MAPK pathway independent of EGFR[104]. In the PI3K/AKT pathway, *PI3KCA* mutation or *PTEN* loss was reported to induce EGFR-TKI resistance and are suggested to be able to predict the poor response to EGFR-TKI in NSCLC patients[86,105,106].

1.8.2.3 Activation of bypass signaling pathways

The EGFR-TKI resistance can be mediated by the activation of bypass signaling pathways that are independent of EGFR. Many of the bypass signaling are linked to the activity of other RTK family members. For instance, the amplification of RTK family member MET, was detected in 5-22% of patients with acquired resistance to first-generation of EGFR-TKIs[15,107,108]. Once activated by its ligand HGF, MET activates ERBB3 tyrosine phosphorylation, and subsequently PI3K/AKT pathway to regulate cell proliferation and survival. The whole process is independent from EGFR activation[15]. The overexpression of HGF could also induce resistance via HGF mediated MET activation to restore the PI3K/Akt pathway. HGF overexpression was detected in 61% of patients with EGFR-TKI acquired resistance[108]. Other overexpressed RTKs or their ligands such as ERBB2, AXL, and GAS6, were reported to decrease sensitivity towards osimertinib, which confirmed its capacity to induce acquired resistance to EGFR-TKIs[91,109–112]. In general, activation of other RTK family members regulates downstream PI3K or MAPK signaling to promote cell proliferation and survival independent of EGFR activation status.

1.8.3 Strategies to overcome EGFR-TKI resistance

To overcome the resistance driven by EGFR mutations, the primary choice is to use appropriate EGFR targeting drugs. Resistance related to EGFR T790M mutation can be circumvented with osimertinib treatment. When it comes to EGFR exon 20 insertion mutation, it is possible to try Poziotinib, a covalent, irreversible EGFR/ERBB2 inhibitor that was reported to be 65 times more potent in inhibiting cell lines with EGFR exon 20 insertion than those harboring the EGFR T790M mutation. The initial reported response rate to poziotinib is 64% in NSCLC patients with EGFR exon 20 insertion (ongoing clinical trial NCT03066206)[113].

To overcome the resistance driven by EGFR-independent mechanisms, a suitable combination therapy using EGFR-TKI and other targeted drugs could lead to a good treatment outcome. Take tumors with ERBB2 amplification as an example, although there still being no approved combination therapy for EGFR and ERBB2 dual inhibition in NSCLC, approved targeted drugs for ERBB2 (trastuzumab, lapatinib, and pertuzumab) are

available for clinical trials. Another example is MET amplification. A clinical trial (NCT02468661) that combines erlotinib and FDA approved MET inhibitor capmatinib, has recently been finalized although no result published yet. Similar combination strategy can be investigated in other bypass signaling mechanisms such as BRAF mutation and KRAS mutation.

2 THE PRESENT WORK

2.1 AIMS

The general aim of the thesis was to investigate the response and resistance mechanisms of EGFR-TKIs in NSCLC, discover predictive biomarkers for drug response, and find novel drug targets to improve the current NSCLC therapy. Consistently throughout the work presented in this thesis, -omics methods (see 2.2.1 and 2.2.2) have been central in our investigations.

The specific aims for each paper were as following:

- Paper I: To identify and explore the oncogenic microRNAs (miRNA) function in NSCLC in relation to EGFR-TKI response and cellular signalling
- Paper II: To explore the immediate adaptive response to EGFR inhibition and find drug targets for EGFR-TKI combination therapy in NSCLC
- Paper III: To find novel biomarkers and new drug targets for EGFR-TKI combination therapy in NSCLC with wild type EGFR
- Paper IV: To explore the resistance mechanisms of third-generation EGFR-TKIs in EGFR mutated NSCLC

2.2 METHODS

Introduction to the main experimental and data analysis methods used in this thesis are summarized below.

2.2.1 RNA sequencing and data analysis

A transcriptomics study's focus is to identify and quantify mRNAs, but sometimes also other species of transcripts such as long non-coding RNAs or miRNAs. Classical technologies include hybridization-based or sequencing-based methods. Sequencing-based methods directly determine the cDNA sequence by tag-based Sanger sequencing of cDNA libraries. This includes the first-generation high throughput sequencing (commonly referred to as Sanger dideoxy sequencing, the automated Sanger method). First-generation sequencing generates sequences length of 600-1000 bases, and a standard run yields 100,000 bases. RNA sequencing methods are derived from next-generation sequencing (NGS) technology that is performed in a massively parallel manner (can consist of 6 billion sequencing reactions in a single run). A single run of NGS can generate 600 billion bases of sequence data[114]. As a comparison, the total human genome is about 3 billion base pairs.

A typical workflow of an RNA sequencing (RNA-seq) experiment includes isolation of RNAs, quality control of RNAs, library construction, high-throughput sequencing, data processing, and data analysis. Depending on the material and the type of RNA, the selection of RNA isolation methods varies. For instance, commercially available products like RNeasy (Qiagen) can be used to isolate mRNA from cells or tissue samples. However, small RNAs (such as miRNA, siRNA, or small nuclear RNA) require methods with specialized RNA-isolation columns that enrich for small RNAs (10-200 nucleotides). Before converting RNAs to a cDNA library used in sequencing, a quality check is performed on the RNA samples to check for degradation, purity, and quantity. In the next step, library preparation of RNAs to cDNAs is performed because RNA is more labile than DNA and hence not suitable for direct sequencing. The cDNA library should be representative of the original RNA sample.

When the cDNA library is created, it can be used in a sequencing platform such as Illumina or SOLID (Applied Biosystems), among others. Depending on the platform, the tagging methods of cDNA molecules vary, for instance, Illumina ligates sequencing adaptors to targeted DNA molecules, and SOLID (Applied Biosystems) uses magnetic beads to tag DNA. After the sequencing step, the reads can be obtained. Quality control steps and pre-processing of the sequencing data is also required to prevent possible problems for mapping to reference genomes later. Unprocessed sequences can, for example, contain low confidence bases or untrimmed adaptors (for Illumina data). Those problems can be solved by the FastQC program[115]. Data analysis begins after obtaining the reads from sequencers. The RNA-sequencing data can be used to explore multiple questions such as differential expression, splice variant expression, and novel fusion genes. For differential expression analysis, a general workflow could be as follows: First, the raw data from the sequencer is pre-processed and aligned to a reference genome. Here the reads should be appointed to annotated genes or transcripts. Once genes or transcript models are generated, the sequencing depth is normalized across the samples. Then the expression levels of each transcript are

calculated. Commonly used expression units are “reads per kilobase per million mapped reads” (RPKM), “fragments per kilobase of transcripts per million fragments mapped” (FPKM), and “transcripts per kilobase million” (TPM). From here, the gene expression across experimental groups can be compared with the help of statistics[116].

Paper I/III performed standard RNA sequencing. Total RNA was extracted, quality checked, and progressed to RNA library preparation. The Sequencing was performed by HiSeq2500. FPKM values for genes and transcripts were reported after removing missing values. **Paper II/III** performed differential expression analysis on total normalized read count values using DESeq2 R package[117].

2.2.2 MS-based Proteomics and data analysis

The term proteomics was first introduced as ‘the protein complement’ of the genome in 1990s and referred to the identification and sometimes quantification of all proteins present in an organism’s cell or organism’s cellular sub-compartment[118]. Proteomics strategies include bottom-up, middle-down, and top-down, which refer to the analysis of small peptides, large peptides (resulting from limited digestion or more selective protease), or whole proteins. Recent advances over the past decade in mass spectrometry (MS) have enabled the characterization of the proteome.

To get a representative proteome, cell, or tissue lysis steps must ensure complete solubilization of proteins. This step usually needs optimization to be compatible with MS analysis. For example, when extracting membrane proteins, many protocols require solubilization based on detergents that interfere with the MS analysis and such detergents need to be replaced. Proteins are digested into peptides before the MS analysis. After enzyme digestion, the abundance of different peptides can span more than 7 orders of magnitude, making it difficult to identify low abundant peptides. In many cases, peptide pre-fractionation steps are applied before LC-MS analysis to reduce the sample complexity. Pre-fractionation methods available can separate peptides according to e.g., charge, hydrophobicity, or isoelectric point. A recently published method called HiRIEF (high-resolution isoelectric focusing) describes a solution to reduce sample complexity[119]. In this method, peptide mixtures are separated by isoelectric focusing (IEF), using immobilized pH gradient (IPG) gel strips. Typically, every sample is fractionated into 72 largely non-overlapping fractions of peptides that are then analyzed individually by mass spectrometry, resulting in dramatically increased analytical depth. Mass spectrometry (MS) is an analytical method to ionize molecules such that they become gaseous and charged, enabling them to travel towards a detector. On the path to the detector, the ionized particles are accelerated by a magnetic field, whereby ions are separated according to their mass-to-charge (m/z) ratios by the time they reach the detector. Traditionally, the ionization techniques generate high amount of heat. This causes proteins and larger peptides to fragment before ionizing and becoming gaseous, which alters the chemical composition and hence makes it impossible to derive the protein’s chemical composition or identity. A milestone of MS applications in protein chemistry, however, solved this with the introduction of soft-ionization techniques. This enabled proteins and larger peptides to become enclosed in aerosolized droplets (e.g. electrospray ionization) or matrix-media (e.g.

Matrix-assisted laser desorption/ionization) carrying a charge resulting in no decomposition in the ionization process such that they could be accelerated intact towards the detector.

A typical shotgun (bottom-up) proteomics workflow looks as follows: All proteins from samples of interest (such as cells or tissues) are extracted and digested with one or several proteases (such as trypsin alone or combined with Lys-C) to generate a set of peptides. Depending on research question and sample, enrichment (for a specific set of proteins or peptides), labelling, and fractionation steps can be introduced to enhance detection, identification, and quantification. The peptides are then separated by liquid chromatography (LC), ionized by electrospray ionization, and analyzed by mass spectrometry. Multiple peptide ions entering the MS at the same time are analyzed with respect to their mass to charge ratio (m/z), resulting in an MS1 spectrum. In parallel, selected peptide ions are isolated, fragmented, and analyzed again to generate a fragment spectrum (MS2 or MS/MS) containing m/z information related to the different fragments from the selected peptide. The peptide and peptide fragment mass information generated in MS1 and MS2 steps of the analysis are then used to identify peptides through database matching towards a database containing all theoretical peptides from the organism under investigation. Through the identified peptides, the protein identifications are then inferred. The output of a proteomics analysis is commonly a list of identified proteins and peptides with related quantitative information. This data usually needs to be cleaned up and pre-processed before performing any statistical analysis and biological interpretation. This step includes removing missing values and outliers in the dataset as well as normalizing the data. After this step, statistical analysis is applied to identify a subset of proteins that differ between sample groups in the experiment. More importantly, the data may need to be cross-referenced with other -omics data (e.g. transcriptomics) for context. Various bioinformatics tools are available for enrichment, pathway, and network analysis. As examples, Gene Ontology (GO) enrichment analysis, and Gene Set Enrichment Analysis (GSEA) provides methods to highlight genes or pathways based on their previous annotations and reveals differentially regulated biological pathways. Additional tools include analysis, visualization, and interpretation of data commonly performed using R, Python, or other programming languages.

In this thesis, protein cleanup was performed using slightly modified filter-aided sample preparation (FASP) protocol (**paper I/III/IV**) or single-pot, solid-phase-enhanced sample-preparation (SP3) protocol (**paper II**)[120,121]. Proteins were digested into peptides using trypsin alone (**paper I/III**) or together with Lys-C (**paper II/IV**). Additional phospho-peptide enrichment step was performed in **paper IV** using TiO2 beads in phosphoproteomics samples. Peptides (**paper I-III)**, or enriched phosphopeptides (**paper IV**) were labelled with TMT10plex isobaric label reagent. Labelled peptides were pooled together and pre-fractionated using HiRIEF fractionation method[119] (**paper I-IV**). Proteomic and phosphoproteomic analysis by LC-MS/MS were performed using Q Exactive mass spectrometer (**paper I-IV**). DEqMS R package was used to perform differential protein expression analysis in **paper II**[122]. Pathway enrichment was performed using WebGestalt in **paper II/IV**[123].

2.3 RESULTS AND DISCUSSION

2.3.1 Paper I: microRNAs with AAGUGC seed motif constitute an integral part of an oncogenic signaling network

It has been previously shown that miRNA dysregulation is involved in carcinogenesis and could be used to improve cancer therapy[124,125]. How miRNAs regulate cancer progression and the response to EGFR-TKI remains largely unknown. This paper was initiated based on the general interest towards the function of oncogenic miRNAs' in NSCLC and in response to EGFR-TKIs in particular.

The main findings in this paper include:

- The identification of a group of miRNAs with the same AAGUGC motif (oncomotif) in their seed sequence that has an impact on cell proliferation and EGFR-TKI sensitivity.
- Well-known tumor suppressors were identified and validated as miR-372-3p targets in the NSCLC cell line model.
- High expression of miR-372-3p increased the sensitivity to EGFR-TKI in NSCLC cell lines.
- The expression of oncomotif-miRNAs in lung ADC is associated with E2F driven cell proliferation, TP53 mutation, MYC amplification, and shorter relapse-free survival.
- Oncomotif-miRNAs are part of the oncogenic signaling network in NSCLC and multiple other cancer types by regulating cancer cell proliferation through controlling their target tumor suppressors.

The target analysis of miRNAs revealed that our suggested oncomotif-miRNAs regulate several previously reported tumor suppressors in various cancer types[126–128]. Notably, the selection of miR-372-3p targets investigated in this paper were chosen by the results provided by a miRNA target prediction algorithm together with a literature search of currenly known tumor suppressors. The miRNA target prediction algorithm may introduce false-positive and false-negative predictions, which when taken together with current tumor suppressor knowledge from the literature search could be a limiting factor by influencing our perspective on which miRNA are true as well as relevant for further investigation. Therefore, important miRNA targets that contribute to cell proliferation and EGFR-TKI sensitivity may still be among the false-negatives and hence unfortunately left uninvestigated. Nevertheless, since the miRNA targets that were identified and validated contained previously known tumor supressors, it would provide a relevant justification to investigate the clinical prevalence of our identified oncomotif-miRNAs in NSCLC patients, but perhaps improved diagnostic kits would have to be developed to achieve such a study.

2.3.2 Paper II: Immediate adaptation analysis implicates BCL6 as an EGFR-TKI combination therapy target in NSCLC

In **paper I**, oncomotif-miRNAs were shown to impact EGFR-TKI sensitivity, but miRNAs are not commonly used as drug targets. Instead, it is more common to target a protein. Previously, many studies were focusing on finding the proteins involved in intrinsic resistance or acquired resistance of EGFR-TKIs. However, rapid adaptative response to targeted therapies is equally interesting because it could be responsible for short-lived clinical response[129]. Immediate adaptative response to EGFR-TKI treatment is less reported in NSCLC. We therefore initiated the study of the immediate adaptive response to EGFR-TKI treatment in NSCLC to identify and investigate the involved proteins, which in turn would also be anticipated potential drug targets in combination with EGFR-TKI treatment.

The main findings in this paper include:

- Short term EGFR-TKI treatment already results in dramatic changes in the cellular signaling as immediate adaptative response.
- EGFR-TKI treatment results in the upregulation of BCL6, ERBB3, FGFR2, JAK3, which indicate a role for them in the immediate adaptive response and could be essential regulators in treatment escape.
- The drug combination screen indicates that dual inhibition combinations of EGFR and FGFR2 or JAK3 are synergetic.
- Dual targeting of BCL6 and EGFR results in increased apoptosis and synergistic cell killing in various NSCLC cell lines.

The observation of BCL6, ERBB3, FGFR2, and JAK3 as critical components as immediate adaptive response to EGFR-TKI treatment in NSCLC is exciting, as some of the targets/pathways were previously reported to be TKI resistance mechanisms[15,130–133]. Our primary focus, BCL6, was reported as a drug resistance mechanism to the BCR-ABL1 targeting TKI imatinib in leukemia cells[130]. The combination therapy targeting BCL6 and phosphor-STAT3 was suggested for NSCLC treatment due to its ability to reduce tumor growth[134]. This was in line with our observation that BCL6 and JAK regulated STAT signaling contribute to EGFR-TKI treatment escape.

Notably, there was a limitation in our initial observation of immediate adaptive response to EGFR-TKI treatment, as it was performed in the A431 cell line but not in an NSCLC cell line. Our rational for initially using A431 was based on its high expression level of EGFR and commonly used as a model system for EGFR signaling. However, this study was able to show in NSCLC cells that EGFR-TKI treatment resulted in significant increase in cell killing upon dual targeting of BCL6 and EGFR as opposed to targeting only either one of the proteins. Thus, at least indicating that BCL6 in NSCLC cells may still contribute to survival and treatment escape following EGFR-TKI treatment. Another limitation regarded the selection of upregulated mRNAs and proteins to further explore, as this was partially based on the overlap between our profiling data and the COSMIC cancer gene census (CGC) data. Since the CGC does not cover all cancer related mutated genes, this added a bias in the selection of candidates to further explore. It is therefore important to note that transcripts and

proteins not present in the overlap between our profiling data and the CGC's data, may still hold important roles in EGFR-TKI response. Furthermore, the work in this paper is performed using cell lines as model systems, but the heterogeneity of tumors and the toxicity of combination therapy could introduce more complexity to combination treatment. It is therefore recognized that further investigation would be needed in an *in vivo* and/or clinical setting to validate these findings.

2.3.3 Paper III: *CDKN2A* status predicts the response to EGFR targeting therapy in EGFRwt NSCLC

Current approved EGFR-TKIs are limited and focused to treat NSCLC patients with EGFR mutations. For patients with wild type EGFR, only a subpopulation of them will benefit from EGFR-TKI treatment[135,136]. It is therefore desirable to be able to identify the subpopulation of NSCLC patients with wild type EGFR (EGFRwt) that could benefit from EGFR-TKI treatment[136]. No biomarker for estimating EGFR-TKI response in EGFRwt NSCLC is currently available, making it difficult for clinicians to decide the best treatment for the patients with EGFRwt. In this paper, we wanted to find novel biomarkers and drug targets for EGFR-TKI based combination therapy in EGFRwt NSCLC.

The main findings in this paper include:

- *CDKN2A* loss is associated with EGFR-TKI sensitivity in EGFRwt NSCLC.
- Epithelial cell lineage is associated with EGFR-TKI sensitivity, and mesenchymal cell lineage is associated with resistance to EGFR-TKI in *CDKN2A* deleted EGFRwt NSCLC.
- *BCL2L1* (coding for BCL-xL) overexpression exist before EGFR-TKIs in EGFR-TKI responding cell lines.
- EGFRwt, KRASwt, and *CDKN2A* deleted NSCLC is sensitive to BCL-xL and EGFR dual inhibition.

CDKN2A encodes for both p16 and p14, where p16 regulates G1/S transition to control proliferation and p14 regulates p53 to control apoptosis[137]. The cell lines in our study which were defined as molecular responders, all harbor *CDKN2A* deletion and are sensitive to EGFR-TKIs. Considering that *CDKN2A* deletion is frequently detected in 13.9% of ADC and 27% of SQCC (Table 1), the finding in this paper has the potential to better select patients who can benefit from EGFR-TKIs.

The limitations of this study include:

- *CDKN2A* loss as a biomarker requires further validation *in vivo* and clinical settings.
- Currently, all the BCL-xL targeting drugs are still under investigation, limiting our understanding of their suitability and toxicity in a clinical setting.

2.3.4 Paper IV: Integrative proteomics and phosphoproteomics profiling of NSCLC cell lines to explore osimertinib resistance mechanisms

Paper I-III studied the response or resistance mechanisms using first-generation of EGFR-TKIs. Recently, the third-generation of EGFR-TKI osimertinib received quick approval from FDA to be used as first-line treatment. However, tumors change and develop new resistance mechanisms also to third-generation of EGFR-TKIs. In this study we investigated the osimertinib resistance mechanisms using osimertinib resistant NSCLC cell models.

The main findings in this paper include:

- Identification of upregulation of AXL and its ligand GAS6, and EMT as potentioal resistance mechanisms to osimertinib by proteomics profiling
- Identification of reactivation of MAPK pathway, and AXL dependent CDK1 phosphorylation as potential resistance mechanisms to osimertinib by phosphoproteomics profiling
- A proteomic and phosphoproteomic profiling data resource of osimertinib resistant cell line models for further investigation

The osimertinib resistance mechanisms identified via our -omics profiling have been previously reported by others, which indicate that our -omics approach identified relevant resistance mechanisms. For instance, the upregulation of AXL and GAS6 were reported in various studies as a resistance mechanisms to first-generation of EGFR-TKIs or osimertinib[90,111,138]. Moreover, Kim and colleagues recently reported that the degradation of AXL and EGFR-TKI treatment overcame the acquired resistance in NSCLC[139].

Constraints in time and funding led us to only initiate a minor pilot investigation. The proteomic and phosphoproteomic profiling was performed only with duplicate samples, which limited the statistical analysis. Consequently, the suggested resistance mechanisms to osimertinib in this paper require further validatory experiments with more replicates and statistical evaluation.

2.4 GENERAL CONCLUSIONS AND FUTURE PERSPECTIVE

The work presented in this thesis aimed at better understanding response and resistance mechanisms to EGFR-TKIs in NSCLC and to find novel biomarkers as well as drug targets or combination therapies for NSCLC. This has been achieved and summarized with this thesis.

Presence of cancer specific miRNA fingerprints were identified in various cancers including lung cancer[140–142]. miR-373 and miR-520c have been observed to be metastasis promoting miRNAs in breast cancer cells[143]. In lung cancer, the overexpression of miR-17-92 cluster was observed to enhance cell proliferation[144]. **Paper I** described onco-motif miRNAs as part of oncogenic signaling in NSCLC and other types of cancers, which included miR-372 that is a member of the miR-371-373 cluster located on chromosome 19q13. This type of study led us to further understand NSCLC cellular signaling in response to EGFR-TKI. Moreover, it would be interesting to explore in NSCLC patients the prevalence of the oncomotif-miRNAs that we identified and the subsequent application of these oncomotif-miRNAs as a predictive biomarker for sensitivity to EGFR-TKI treatment.

A natural question that arises then is how well miRNAs can be detected in patients. Previously, Weber and colleagues reported that the miRNAs were detected in 12 biological fluids, which included serum, plasma, saliva, and urine[145]. Later on Roa and colleagues reported the high sensitivity and specificity of detecting miRNA panel in NSCLC patients using quantitative real-time polymerase chain reaction (RT-qPCR)) to analyze patient sputum[146]. It is therefore conceivable that miRNAs could be clinically used as biomarkers to determine which EGFR-TKI treatment is suitable for a given NSCLC patient. Especially since using body fluid is much less invasive than using tissue biopsy (easier sampling for NSCLC patients) and RT-qPCR is already widely used in hospitals (easy and cost-effective detection method).

When we observed the upregulation of BCL6, ERBB3, FGFR2, and JAK3 in our molecular profiling in response to EGFR-TKIs in **Paper II**, we also noticed that three out of four were protein kinases. Protein kinases are widely used as drug targets because their function is frequently altered in cancer associated signaling transduction networks and 48 protein kinase inhibitor drugs have been approved by FDA, 35 of which are TKIs[147]. Therefore, the three protein kinases that we identified could be proposed as targets for EGFR-TKI drug combination for NSCLC treatment.

Previous studies revealed that protein kinases ERBB3 and FGFR2 were associated to EGFR-TKI resistance in NSCLC[15,131]. Dual targeting of EGFR and ERBB3 was shown to overcome the acquired resistance to EGFR-TKIs[148]. A more recent study observed EGFR and ERBB3 expression in circulating tumor cells and tumor tissues from NSCLC patients, which made the ERBB3 driven resistance to NSCLC more easily diagnosed[149]. There are, however, no currently ERBB3 targeting drugs approved by FDA. Our drug combination therapy screen data showed a multi-targeting drug nintedanib (Vargatef®) targeting FGFR1-3 that is synergistic with EGFR-TKIs. In EU it is approved for treating advanced

lung ADC patients. However, it is not currently approved as a combination therapy together with an EGFR-TKI. Although, this would be an interesting avenue to explore.

Upregulation of JAK3 was only observed in our mRNA level molecular profiling but the inhibition of JAK by multi-kinase targeting drug momelotinib together with EGFR-TKI was shown to be synergistic in our drug combination therapy screen. The investigation of JAK members and their relation to EGFR-TKI resistance in NSCLC mainly focused on JAK1/2, since JAK2 inhibition was previously shown to restore EGFR-TKI sensitivity in EGFR-TKI resistant NSCLC in both cell lines and xenograft models[150]. The combination of an approved JAK1/2 inhibitor ruxolitinib (Jakafi®) with first-generation EGFR-TKI erlotinib has already been investigated in a clinical trial, but the patients' response to this combination were poor. Another clinical trial of osimertinib and ruxolitinib combination was planned, according to Yu and colleagues[151]. The role of JAK3 in resistance mechanisms of EGFR-TKI in NSCLC remains unclear, but likely would be relying on JAK-STAT signaling, hence inhibitors of this pathway could be used to further investigate this resistance mechanism. As a potential JAK3 inhibitor, momelotinib recently received FDA fast track designation for targeting JAK1/2, but is not yet approved. Its JAK3 inhibition ability was reported by Selleckchem with cell-free assay although the selectivity to JAK3 compared with JAK1/2 was weaker. Momelotinib showed synergistic effect together with EGFR-TKIs in our drug screen data. It would be interesting to continue to study JAK3 and EGFR inhibitor combination in NSCLC with JAK3 specific inhibitors instead of momelotinib, which target JAK1-3.

BCL6 has not been previously associated with EGFR-TKI response in NSCLC, but BCL6 and STAT3 inhibitor combination was previously reported to inhibit cell growth in NSCLC cell lines and xenografted tumor[134]. Notably, the combination of BCL6 inhibitor RI-BPI and imatinib (BCR-ABL1 TKI drug) has synergistic effects to rapidly induce cell death in cancer cells and promote longer survival for host mice in a xenograft model[130]. However, the dual targeting of EGFR and BCL6 in NSCLC have not previously been reported. In **Paper II** we observed that NSCLC cell sensitivity to EGFR-TKI is increased when BCL6 is also inhibited and hence this would be an interesting combination therapy to evaluate.

CDKN2A homozygous deletion was observed in 29% of the lung SQCC cases[31]. BCL2 or BCL-xL reactivation was observed in EGFR mutated NSCLC cell line with TKI resistance and [152]. No report has previously associated CDKN2A deletion as a biomarker for EGFR-TKI sensitivity in NSCLC with wild type EGFR, hence in **Paper III** we were excited to have observed such a connection. From this observation, we suggested CDKN2A loss as a biomarker to predict EGFR-TKI sensitivity and proposed BCL-xL as a target for EGFR-TKI combination therapy in EGFRwt *CDKN2A* deletion NSCLC.

Due to limitations in time and available materials, we could only observe but not yet fully validate several potential osimertinib resistance mechanisms in **Paper IV**. The potential resistance mechanisms observed were via the upregulation of AXL or its ligand GAS6, MAPK pathway reactivation, and AXL dependent CDK1 phosphorylation. Further validatory experiments with more replicates could generate better insight into which

potential resistance mechanisms would be worth targeting as a means to improve EGFR-TKI treatment.

The recommendations for future research from the work presented in this thesis can be summarized as follows:

- Finding and validating an efficient method to detect oncomotif-miRNAs from body fluids or tumor tissue and test the oncogmotif-miRNAs' potential to be used as a biomarker for selecting patients for targeted therapy.
- Validating the potential of dual targeting BCL6 and EGFR as a combination therapy in NSCLC
- Validating *CDKN2A* deletion as a biomarker for predicting EGFR-TKI sensitivity in EGFRwt NSCLC related clinical setting.
- Validating BCL-xL inhibitor and EGFR-TKI combination therapy in EGFRwt CDKN2Adel NSCLC.

3 ACKNOWLEDGEMENTS

I would like to thank for all of you who have contributed to the completion of this thesis.

Special thanks to:

Lukas Orre, my main supervisor, for your patient guidance during the PhD years. I have learnt a lot from you.

Janne Lehtiö, my co-supervisor, for providing the excellent proteomics platform and the wonderful colleagues.

Petter Brodin, my co-supervisor, for kindly introducing me to your fantastic group and sharing some equipments.

Yan Xiong, my mentor, for the scientific discussions.

All the co-authors of the papers, for the effort that you put in to generate the papers and manuscripts included in this thesis.

All the colleagues in SciLifeLab and department of Oncology-pathology, for the scientific support, friendship, and fun activities after work.

All my friends who have been supportive during my PhD journey.

Finally, **Martin Tran**, my husband, and **Anthony Tran**, my son, for your endless love, encouragement and support during my years at KI.

4 REFERENCES

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* (2018). doi:10.3322/caac.21492

2. Duma, N., Santana-Davila, R. & Molina, J. R. Non–Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. *Mayo Clinic Proceedings* (2019). doi:10.1016/j.mayocp.2019.01.013

3. Gazdar, A. F., Bunn, P. A. & Minna, J. D. Small-cell lung cancer: What we know, what we need to know and the path forward. *Nature Reviews Cancer* **17**, 725–737 (2017).

4. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, B. F. Global Cancer Observatory: Cancer Today. *Lyon, France: International Agency for Research on Cancer.* (2018). Available at: https://gco.iarc.fr/today. (Accessed: 6th June 2020)

5. Samet, J. M. *et al.* Lung cancer in never smokers: Clinical epidemiology and environmental risk factors. *Clinical Cancer Research* **15**, 5626–5645 (2009).

6. Chen, Z., Fillmore, C. M., Hammerman, P. S., Kim, C. F. & Wong, K. K. Non-small-cell lung cancers: A heterogeneous set of diseases. *Nature Reviews Cancer* (2014). doi:10.1038/nrc3775

7. Yang, X. & Lin, D. Changes of 2015 WHO histological classification of lung cancer and the clinical significance. *Chinese J. Lung Cancer* **19**, 332–336 (2016).

8. Langer, C. J., Besse, B., Gualberto, A., Brambilla, E. & Soria, J. C. The evolving role of histology in the management of advanced non - small-cell lung cancer. *Journal of Clinical Oncology* (2010). doi:10.1200/JCO.2010.28.8126

9. 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.* (2004). doi:10.1056/NEJMoa040938

10. Paez, J. G. *et al.* EGFR mutations in lung, cancer: Correlation with clinical response to gefitinib therapy. *Science (80-.).* (2004). doi:10.1126/science.1099314

11. Cappuzzo, F. *et al.* Epidermal growth factor receptor gene and protein and gefitinib sensitivity in non-small-cell lung cancer. *J. Natl. Cancer Inst.* (2005). doi:10.1093/jnci/dji112

12. Weiss, J. *et al.* Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. *Sci. Transl. Med.* (2010). doi:10.1126/scitranslmed.3001451

13. Liao, R. G. *et al.* Inhibitor-sensitive FGFR2 and FGFR3 mutations in lung squamous cell carcinoma. *Cancer Res.* (2013). doi:10.1158/0008-5472.CAN-12-3950

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

15. Engelman, J. A. *et al.* MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science (80-.).* (2007). doi:10.1126/science.1141478

16. Bergethon, K. *et al.* ROS1 rearrangements define a unique molecular class of lung cancers. *J. Clin. Oncol.* (2012). doi:10.1200/JCO.2011.35.6345

17. Rikova, K. *et al.* Global Survey of Phosphotyrosine Signaling Identifies Oncogenic Kinases in Lung Cancer. *Cell* (2007). doi:10.1016/j.cell.2007.11.025

18. Davies, K. D. *et al.* Identifying and targeting ROS1 gene fusions in non-small cell lung cancer. *Clin. Cancer Res.* (2012). doi:10.1158/1078-0432.CCR-12-0550

19. Kohno, T. *et al.* KIF5B-RET fusions in lung adenocarcinoma. *Nat. Med.* (2012). doi:10.1038/nm.2644

20. Takeuchi, K. *et al.* RET, ROS1 and ALK fusions in lung cancer. *Nat. Med.* (2012). doi:10.1038/nm.2658

21. Stephens, P. *et al.* Lung cancer: intragenic ERBB2 kinase mutations in tumours. *Nature* (2004).

22. Wang, S. E. *et al.* HER2 kinase domain mutation results in constitutive phosphorylation and activation of HER2 and EGFR and resistance to EGFR tyrosine kinase inhibitors. *Cancer Cell* (2006). doi:10.1016/j.ccr.2006.05.023

23. Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.* (2001). doi:10.1101/gad.943001

24. Mascaux, C. *et al.* The role of RAS oncogene in survival of patients with lung cancer: A systematic review of the literature with meta-analysis. *British Journal of Cancer* (2005). doi:10.1038/sj.bjc.6602258

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

26. Paik, P. K. *et al.* Clinical characteristics of patients with lung adenocarcinomas harboring BRAF mutations. *J. Clin. Oncol.* (2011). doi:10.1200/JCO.2010.33.1280

27. Nieto, P. *et al.* A Braf kinase-inactive mutant induces lung adenocarcinoma. *Nature* (2017). doi:10.1038/nature23297

28. Marks, J. L. *et al.* Novel MEK1 mutation identified by mutational analysis of epidermal growth factor receptor signaling pathway genes in lung adenocarcinoma. *Cancer Res.* (2008). doi:10.1158/0008-5472.CAN-08-0099

29. Sanchez-Cespedes, M. *et al.* Inactivation of the INK4A/ARF locus frequently coexists with TP53 mutations in non-small cell lung cancer. *Oncogene* **18**, 5843–5849 (1999).

30. Seike, M. *et al.* Increase in the frequency of p16(INK4) gene inactivation by hypermethylation in lung cancer during the process of metastasis and its relation to the status of p53. *Clin. Cancer Res.* (2000).

31. Hammerman, P. S. *et al.* Comprehensive genomic characterization of squamous cell lung cancers. *Nature* (2012). doi:10.1038/nature11404

32. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* (2012). doi:10.1158/2159-8290.CD-12-0095

33. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* (2013). doi:10.1126/scisignal.2004088

34. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* (2014). doi:10.1038/nature13385

35. Aberle, D. R. *et al.* Reduced lung-cancer mortality with low-dose computed tomographic screening. *N. Engl. J. Med.* **365**, 395–409 (2011).

36. Wood, D. E. *et al.* Lung cancer screening, version 3.2018. *JNCCN Journal of the National Comprehensive Cancer Network* (2018). doi:10.6004/jnccn.2018.0020

37. Zhou, Q. *et al.* China national lung cancer screening guideline with low-dose computed tomography (2018 version). *Chinese J. Lung Cancer* **21**, 67–75 (2018).

38. Health Commission of PRC, N. Chinese guidelines for diagnosis and treatment of primary lung cancer 2018 (English version). *Chinese J. Cancer Res.* **31**, 1–28 (2019).

39. Goldstraw, P. *et al.* The IASLC lung cancer staging project: Proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM Classification for lung cancer. *J. Thorac. Oncol.* **11**, 39–51 (2016).

40. Dietel, M. *et al.* Diagnostic procedures for non-small-cell lung cancer (NSCLC): Recommendations of the European Expert Group. *Thorax* **71**, 177–184 (2016).

41. Torre, L. A., Siegel, R. L. & Jemal, A. Lung cancer statistics. *Adv. Exp. Med. Biol.* **893**, 1–19 (2016).

42. Five-Year Survival Rates | SEER Training. Available at: https://training.seer.cancer.gov/lung/intro/survival.html. (Accessed: 8th June 2020)

43. Heist, R. S. & Engelman, J. A. SnapShot: Non-Small Cell Lung Cancer. *Cancer Cell* **21**, (2012).

44. Arkhipov, A. *et al.* Architecture and membrane interactions of the EGF receptor. *Cell* (2013). doi:10.1016/j.cell.2012.12.030

45. Yarden, Y. & Sliwkowski, M. X. Untangling the ErbB signalling network. *Nature Reviews Molecular Cell Biology* (2001). doi:10.1038/35052073

46. Avraham, R. & Yarden, Y. Feedback regulation of EGFR signalling: Decision making by early and delayed loops. *Nature Reviews Molecular Cell Biology* (2011). doi:10.1038/nrm3048

47. COHEN, S. Isolation of a mouse submaxillary gland protein accelerating incisor eruption and eyelid opening in the new-born animal. *J. Biol. Chem.* (1962).

48. Cohen, S. The stimulation of epidermal proliferation by a specific protein (EGF). *Dev. Biol.* **12**, 394–407 (1965).

49. COHEN, S. & ELLIOTT, G. A. The stimulation of epidermal keratinization by a protein isolated from the submaxillary gland of the mouse. *J. Invest. Dermatol.* **40**, 1–5 (1963).

50. Das, M. *et al.* Specific radiolabeling of a cell surface receptor for epidermal growth factor. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 2790–2794 (1977).

51. Cohen, S., Carpenter, G. & King, L. Epidermal growth factor-receptor-protein kinase interactions. Co-purification of receptor and epidermal growth factor-enhanced phosphorylation activity. *J. Biol. Chem.* (1980).

52. Cohen, S., Ushiro, H., Stoscheck, C. & Chinkers, M. A native 170,000 epidermal growth factor receptor-kinase complex from shed plasma membrane vesicles. *J. Biol. Chem.* (1982).

53. Ushiro, H. & Cohen, S. Identification of phosphotyrosine as a product of epidermal growth factor-activated protein kinase in A-431 cell membranes. *J. Biol. Chem.* (1980).

54. Lin, C. *et al.* Expression cloning of human EGF receptor complementary DNA: gene amplification and three related messenger RNA products in A431 cells. *Science (80-.).* (1984). doi:10.1126/science.6326261

55. Merlino, G. *et al.* Amplification and enhanced expression of the epidermal growth factor receptor gene in A431 human carcinoma cells. *Science (80-.).* (1984). doi:10.1126/science.6200934

56. Libermann, T. A. *et al.* Amplification, enhanced expression and possible rearrangement of EGF receptor gene in primary human brain tumours of glial origin. *Nature* (1985). doi:10.1038/313144a0

57. Ullrich, A. *et al.* Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. *Nature* (1984). doi:10.1038/309418a0

58. Carpenter, G. & Cohen, S. Epidermal Growth Factor. *Annu. Rev. Biochem.* **48**, 193–216 (1979).

59. Bobrow, L., Marsden, J. J., Whittle, N. & Waterfield, M. D. Expression of Epidermal Growth Factor Receptors on Human Cervical, Ovarian and Vulva! Carcinomas. *Cancer Res.* (1986).

60. Richard, J. *et al.* EPIDERMAL-GROWTH-FACTOR RECEPTOR STATUS AS PREDICTOR OF EARLY RECURRENCE OF AND DEATH FROM BREAST CANCER. *Lancet* **329**, 1398–1402 (1987).

61. Veale, D., Ashcroft, T., Marsh, C., Gibson, G. J. & Harris, A. L. Epidermal growth factor receptors in non-small cell lung cancer. *Br. J. Cancer* **55**, 513–516 (1987).

62. Zhang, Y. L. *et al.* The prevalence of EGFR mutation in patients with non-small cell lung cancer: A systematic review and meta-analysis. *Oncotarget* **7**, 78985–78993 (2016).

63. Pines, G., Köstler, W. J. & Yarden, Y. Oncogenic mutant forms of EGFR: Lessons in signal transduction and targets for cancer therapy. *FEBS Letters* (2010). doi:10.1016/j.febslet.2010.04.019

64. Tate, J. G. *et al.* COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* (2019). doi:10.1093/nar/gky1015

65. Sordella, R., Bell, D. W., Haber, D. A. & Settleman, J. Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. *Science (80-.).* (2004). doi:10.1126/science.1101637

66. Pao, W. *et al.* EGF receptor gene mutations are common in lung cancers from 'never smokers' and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. U. S. A.* (2004). doi:10.1073/pnas.0405220101

67. Yun, C. H. *et al.* Structures of Lung Cancer-Derived EGFR Mutants and Inhibitor Complexes: Mechanism of Activation and Insights into Differential Inhibitor Sensitivity. *Cancer Cell* (2007). doi:10.1016/j.ccr.2006.12.017

68. Shih, J. Y., Gow, C. H. & Yang, P. C. EGFR mutation conferring primary resistance to gefitinib in non-small-cell lung cancer [2] (multiple letters). *New England Journal of Medicine* (2005). doi:10.1056/NEJM200507143530217

69. Kobayashi, S. *et al.* EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* (2005). doi:10.1056/NEJMoa044238

70. Pao, W. *et al.* Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med.* (2005). doi:10.1371/journal.pmed.0020073

71. Solca, F. *et al.* Target binding properties and cellular activity of afatinib (BIBW 2992), an irreversible ErbB family blocker. *J. Pharmacol. Exp. Ther.* (2012). doi:10.1124/jpet.112.197756

72. Miller, V. A. *et al.* Afatinib versus placebo for patients with advanced, metastatic non-small-cell lung cancer after failure of erlotinib, gefitinib, or both, and one or two lines of chemotherapy (LUX-Lung 1): A phase 2b/3 randomised trial. *Lancet Oncol.* (2012). doi:10.1016/S1470-2045(12)70087-6

73. Cross, D. A. E. *et al.* AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. *Cancer Discov.* (2014). doi:10.1158/2159-8290.CD-14-0337

74. Green, J. A. Osimertinib—first or second line for epidermal growth factor (EGFR) mutation-positive non-small cell lung cancer? *Journal of Thoracic Disease* (2018). doi:10.21037/jtd.2018.09.52

75. Jänne, P. A. *et al.* AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. *N. Engl. J. Med.* (2015). doi:10.1056/NEJMoa1411817

76. Thress, K. S. *et al.* Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. *Nat. Med.* (2015). doi:10.1038/nm.3854

77. Ercan, D. *et al.* EGFR mutations and resistance to irreversible pyrimidine-based EGFR inhibitors. *Clin. Cancer Res.* (2015). doi:10.1158/1078-0432.CCR-14-2789

78. Jia, Y. *et al.* Overcoming EGFR(T790M) and EGFR(C797S) resistance with mutant-selective allosteric inhibitors. *Nature* (2016). doi:10.1038/nature17960

79. Rotow, J. & Bivona, T. G. Understanding and targeting resistance mechanisms in NSCLC. *Nature Reviews Cancer* (2017). doi:10.1038/nrc.2017.84

80. Wu, J. Y. *et al.* Lung cancer with epidermal growth factor receptor exon 20 mutations is associated with poor gefitinib treatment response. *Clin. Cancer Res.* (2008). doi:10.1158/1078-0432.CCR-07-5123

81. Yasuda, H. *et al.* Structural, biochemical, and clinical characterization of epidermal growth factor receptor (EGFR) exon 20 insertion mutations in lung cancer. *Sci. Transl. Med.* (2013). doi:10.1126/scitranslmed.3007205

82. Yasuda, H., Kobayashi, S. & Costa, D. B. EGFR exon 20 insertion mutations in non-small-cell lung cancer: Preclinical data and clinical implications. *The Lancet Oncology* (2012). doi:10.1016/S1470-2045(11)70129-2

83. Oxnard, G. R. *et al.* Natural history and molecular characteristics of lung cancers harboring egfr exon 20 insertions. *J. Thorac. Oncol.* (2013). doi:10.1097/JTO.0b013e3182779d18

84. Girard, N. *et al.* Analysis of genetic variants in never-smokers with lung cancer facilitated by an internet-based blood collection protocol: A preliminary report. *Clin. Cancer Res.* (2010). doi:10.1158/1078-0432.CCR-09-2437

85. Su, K. Y. *et al.* Pretreatment Epidermal Growth Factor Receptor (EGFR) T790M mutation predicts shorter EGFR tyrosine kinase inhibitor response duration in patients with non-small-cell lung cancer. *J. Clin. Oncol.* (2012). doi:10.1200/JCO.2011.38.3224

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

87. Carcereny, E. *et al.* Mutations of the catalytic subunit a of PI3K (PIK3CA) in erlotinib-treated non-small cell lung cancer (NSCLC) patients (p) with epidermal growth factor receptor (EGFR) mutations. *J. Clin. Oncol.* **29**, 7588–7588 (2011).

88. Yano, S. *et al.* Hepatocyte growth factor induces gefitinib resistance of lung adenocarcinoma with epidermal growth factor receptor-activating mutations. *Cancer Res.* (2008). doi:10.1158/0008-5472.CAN-08-1643

89. Shaw, A. T. *et al.* Clinical features and outcome of patients with non-small-cell lung cancer who harbor EML4-ALK. *J. Clin. Oncol.* (2009). doi:10.1200/JCO.2009.22.6993

90. Taniguchi, H. *et al.* AXL confers intrinsic resistance to osimertinib and advances the emergence of tolerant cells. *Nat. Commun.* (2019). doi:10.1038/s41467-018-08074-0

91. Ortiz-Cuaran, S. *et al.* Heterogeneous mechanisms of primary and acquired resistance to third-generation EGFR inhibitors. *Clin. Cancer Res.* **22**, 4837–4847 (2016).

92. Minari, R. *et al.* Primary resistance to osimertinib due to SCLC transformation: Issue of T790M determination on liquid re-biopsy. *Lung Cancer* **115**, 21–27 (2018).

93. Oxnard, G. R. *et al.* New strategies in overcoming acquired resistance to epidermal growth factor receptor tyrosinekinase inhibitors in lung cancer. *Clinical Cancer Research* (2011). doi:10.1158/1078-0432.CCR-10-2571

94. Yun, C. H. *et al.* The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc. Natl. Acad. Sci. U. S. A.* (2008). doi:10.1073/pnas.0709662105

95. Balak, M. N. *et al.* Novel D761Y and common secondary T790M mutations in epidermal growth factor receptor-mutant lung adenocarcinomas with acquired

resistance to kinase inhibitors. *Clin. Cancer Res.* (2006). doi:10.1158/1078-0432.CCR-06-1570

96. Costa, D. B. *et al.* BIM mediates EGFR tyrosine kinase inhibitor-induced apoptosis in lung cancers with oncogenic EGFR mutations. *PLoS Med.* (2007). doi:10.1371/journal.pmed.0040315

97. Bean, J. *et al.* Acquired resistance to epidermal growth factor receptor kinase inhibitors associated with a novel T854A mutation in a patient with EGFR-mutant lung adenocarcinoma. *Clin. Cancer Res.* (2008). doi:10.1158/1078-0432.CCR-08-0151

98. Niederst, M. J. *et al.* The allelic context of the C797S mutation acquired upon treatment with third-generation EGFR inhibitors impacts sensitivity to subsequent treatment strategies. *Clin. Cancer Res.* (2015). doi:10.1158/1078-0432.CCR-15-0560

99. Mok, T. S. *et al.* Osimertinib or platinum-pemetrexed in EGFR T790M-Positive lung cancer. *N. Engl. J. Med.* (2017). doi:10.1056/NEJMoa1612674

100. Zhang, Q. *et al.* EGFR L792H and G796R: Two Novel Mutations Mediating Resistance to the Third-Generation EGFR Tyrosine Kinase Inhibitor Osimertinib. *J. Thorac. Oncol.* (2018). doi:10.1016/j.jtho.2018.05.024

101. Yang, Z. *et al.* Investigating novel resistance mechanisms to third-generation EGFR tyrosine kinase inhibitor osimertinib in non–small cell lung cancer patients. *Clin. Cancer Res.* (2018). doi:10.1158/1078-0432.CCR-17-2310

102. Ho, C. C. *et al.* Acquired BRAF V600E Mutation as Resistant Mechanism after Treatment with Osimertinib. *J. Thorac. Oncol.* (2017). doi:10.1016/j.jtho.2016.11.2231

103. Kim, T. M. *et al.* Mechanisms of Acquired Resistance to AZD9291: A Mutation-Selective, Irreversible EGFR Inhibitor. *J. Thorac. Oncol.* (2015). doi:10.1097/JTO.0000000000000688

104. Ercan, D. *et al.* Reactivation of ERK signaling causes resistance to EGFR Kinase inhibitors. *Cancer Discov.* (2012). doi:10.1158/2159-8290.CD-12-0103

105. Ludovini, V. *et al.* Phosphoinositide-3-kinase catalytic alpha and KRAS mutations are important predictors of resistance to therapy with epidermal growth factor receptor tyrosine kinase inhibitors in patients with advanced non-small cell lung cancer. *J. Thorac. Oncol.* (2011). doi:10.1097/JTO.0b013e31820a3a6b

106. Shen, H. *et al.* TGF-β1 induces erlotinib resistance in non-small cell lung cancer by down-regulating PTEN. *Biomed. Pharmacother.* (2016). doi:10.1016/j.biopha.2015.10.018

107. Yu, H. A. *et al.* Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. *Clin. Cancer Res.* (2013). doi:10.1158/1078-0432.CCR-12-2246

108. Yano, S. *et al.* Hepatocyte growth factor expression in EGFR mutant lung cancer with intrinsic and acquired resistance to tyrosine kinase inhibitors in a japanese cohort. *J. Thorac. Oncol.* (2011). doi:10.1097/JTO.0b013e31823ab0dd

109. Takezawa, K. *et al.* HER2 amplification: A potential mechanism of acquired resistance

to egfr inhibition in EGFR -mutant lung cancers that lack the second-site EGFR T790M mutation. *Cancer Discov.* (2012). doi:10.1158/2159-8290.CD-12-0108

110. Byers, L. A. *et al.* An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. *Clin. Cancer Res.* (2013). doi:10.1158/1078-0432.CCR-12-1558

111. Zhang, Z. *et al.* Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer. *Nat. Genet.* (2012). doi:10.1038/ng.2330

112. Scaltriti, M., Elkabets, M. & Baselga, J. Molecular pathways: AXL, a membrane receptor mediator of resistance to therapy. *Clin. Cancer Res.* (2016). doi:10.1158/1078-0432.CCR-15-1458

113. Robichaux, J. P. *et al.* Mechanisms and clinical activity of an EGFR and HER2 exon 20-selective kinase inhibitor in non-small cell lung cancer. *Nat. Med.* (2018). doi:10.1038/s41591-018-0007-9

114. Barba, M., Czosnek, H. & Hadidi, A. Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses* **6**, 106–136 (2013).

115. Simon, A. FastQC: a quality control tool for high throughput sequence data. *Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc* (2010). Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc%0A%0A.

116. Eija Korpelainen, Tuimala, J., Somervuo, P., Huss, M. & Wong, G. *RNA-Seq Data Analysis: A practical approach*. *RNA-Seq Data Analysis: A practical approach* (2015).

117. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* (2014). doi:10.1186/s13059-014-0550-8

118. Wasinger, V. C. *et al.* Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. *Electrophoresis* (1995). doi:10.1002/elps.11501601185

119. Branca, R. M. M. *et al.* HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. *Nat. Methods* (2014). doi:10.1038/nmeth.2732

120. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* (2009). doi:10.1038/nmeth.1322

121. Hughes, C. S. *et al.* Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* (2019). doi:10.1038/s41596-018-0082-x

122. Zhu, Y. *et al.* DEqMS: A Method for Accurate Variance Estimation in Differential Protein Expression Analysis. *Mol. Cell. Proteomics* (2020). doi:10.1074/mcp.TIR119.001646

123. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* (2019). doi:10.1093/nar/gkz401

124. Croce, C. M. Causes and consequences of microRNA dysregulation in cancer. *Nature Reviews Genetics* (2009). doi:10.1038/nrg2634

125. Esquela-Kerscher, A. & Slack, F. J. Oncomirs - MicroRNAs with a role in cancer. *Nature Reviews Cancer* (2006). doi:10.1038/nrc1840

126. Wang, G. *et al.* Zbtb7a suppresses prostate cancer through repression of a Sox9-dependent pathway for cellular senescence bypass and tumor invasion. *Nat. Genet.* (2013). doi:10.1038/ng.2654

127. Li, Y. *et al.* Lats2, a putative tumor suppressor, inhibits G1/S transition. *Oncogene* **22**, 4398–4405 (2003).

128. Datto, M. B. *et al.* Transforming growth factor β induces the cyclin-dependent kinase inhibitor p21 through a p53-independent mechanism. *Proc. Natl. Acad. Sci. U. S. A.* (1995). doi:10.1073/pnas.92.12.5545

129. Vasan, N., Baselga, J. & Hyman, D. M. A view on drug resistance in cancer. *Nature* (2019). doi:10.1038/s41586-019-1730-1

130. Duy, C. *et al.* BCL6 enables Ph+ acute lymphoblastic leukaemia cells to survive BCR-ABL1 kinase inhibition. *Nature* (2011). doi:10.1038/nature09883

131. Ware, K. E. *et al.* Rapidly Acquired Resistance to EGFR Tyrosine Kinase Inhibitors in NSCLC Cell Lines through De-Repression of FGFR2 and FGFR3 Expression. *PLoS One* (2010). doi:10.1371/journal.pone.0014117

132. Buchert, M., Burns, C. J. & Ernst, M. Targeting JAK kinase in solid tumors: Emerging opportunities and challenges. *Oncogene* (2016). doi:10.1038/onc.2015.150

133. Yu, H., Lee, H., Herrmann, A., Buettner, R. & Jove, R. Revisiting STAT3 signalling in cancer: New and unexpected biological functions. *Nature Reviews Cancer* (2014). doi:10.1038/nrc3818

134. Deb, D. *et al.* Combination therapy targeting BCL6 and phospho-STAT3 defeats intratumor heterogeneity in a subset of non-small cell lung cancers. *Cancer Res.* (2017). doi:10.1158/0008-5472.CAN-15-3052

135. Garassino, M. C. *et al.* Erlotinib versus docetaxel as second-line treatment of patients with advanced non-small-cell lung cancer and wild-type EGFR tumours (TAILOR): A randomised controlled trial. *Lancet Oncol.* **14**, 981–988 (2013).

136. Lee, S. M. *et al.* First-line erlotinib in patients with advanced non-small-cell lung cancer unsuitable for chemotherapy (TOPICAL): A double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* **13**, 1161–1170 (2012).

137. Kim, W. Y. & Sharpless, N. E. The Regulation of INK4/ARF in Cancer and Aging. *Cell* (2006). doi:10.1016/j.cell.2006.10.003

138. Namba, K. *et al.* Activation of AXL as a preclinical acquired resistance mechanism against osimertinib treatment in EGFR-mutant non-small cell lung cancer cells. *Mol. Cancer Res.* (2019). doi:10.1158/1541-7786.MCR-18-0628

139. Kim, D. *et al.* AXL degradation in combination with EGFR-TKI can delay and overcome acquired resistance in human non-small cell lung cancer cells. *Cell Death Dis.* (2019). doi:10.1038/s41419-019-1601-6

140. Yanaihara, N. *et al.* Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **9**, 189–198 (2006).

141. Takamizawa, J. *et al.* Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res.* (2004). doi:10.1158/0008-5472.CAN-04-0637

142. Calin, G. A. & Croce, C. M. MicroRNA-cancer connection: The beginning of a new tale. *Cancer Research* **66**, 7390–7394 (2006).

143. Huang, Q. *et al.* The microRNAs miR-373 and miR-520c promote tumour invasion and metastasis. *Nat. Cell Biol.* (2008). doi:10.1038/ncb1681

144. Hayashita, Y. *et al.* A polycistronic MicroRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer Res.* (2005). doi:10.1158/0008-5472.CAN-05-2352

145. Weber, J. A. *et al.* The microRNA spectrum in 12 body fluids. *Clin. Chem.* (2010). doi:10.1373/clinchem.2010.147405

146. Roa, W. H. *et al.* Sputum microRNA profiling: A novel approach for the early detection of non-small cell lung cancer. *Clin. Investig. Med.* (2012). doi:10.25011/cim.v35i5.18700

147. Roskoski, R. Properties of FDA-approved small molecule protein kinase inhibitors. *Pharmacological Research* **144**, 19–50 (2019).

148. Huang, S. *et al.* Dual targeting of EGFR and HER3 with MEHD7945A overcomes acquired resistance to EGFR inhibitors and radiation. *Cancer Res.* (2013). doi:10.1158/0008-5472.CAN-12-1611

149. Scharpenseel, H. *et al.* EGFR and HER3 expression in circulating tumor cells and tumor tissue from non-small cell lung cancer patients. *Sci. Rep.* **9**, 1–9 (2019).

150. Gao, S. P. *et al.* JAK2 inhibition sensitizes resistant EGFR-mutant lung adenocarcinoma to tyrosine kinase inhibitors. *Sci. Signal.* (2016). doi:10.1126/scisignal.aac8460

151. Yu, H. A. *et al.* A Phase 1/2 Trial of Ruxolitinib and Erlotinib in Patients with EGFR-Mutant Lung Adenocarcinomas with Acquired Resistance to Erlotinib. in *Journal of Thoracic Oncology* (2017). doi:10.1016/j.jtho.2016.08.140

152. Fan, W. *et al.* MET-independent lung cancer cells evading EGFR kinase inhibitors are therapeutically susceptible to BH3 mimetic agents. *Cancer Res.* (2011). doi:10.1158/0008-5472.CAN-10-2668