LUND UNIVERSITY

Molecular Characterization of Lung Cancer – Targets for Therapy and Prognostic Markers

Salomonsson, Annette

2022

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Salomonsson, A. (2022). *Molecular Characterization of Lung Cancer – Targets for Therapy and Prognostic Markers*. Lund University, Faculty of Medicine.

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

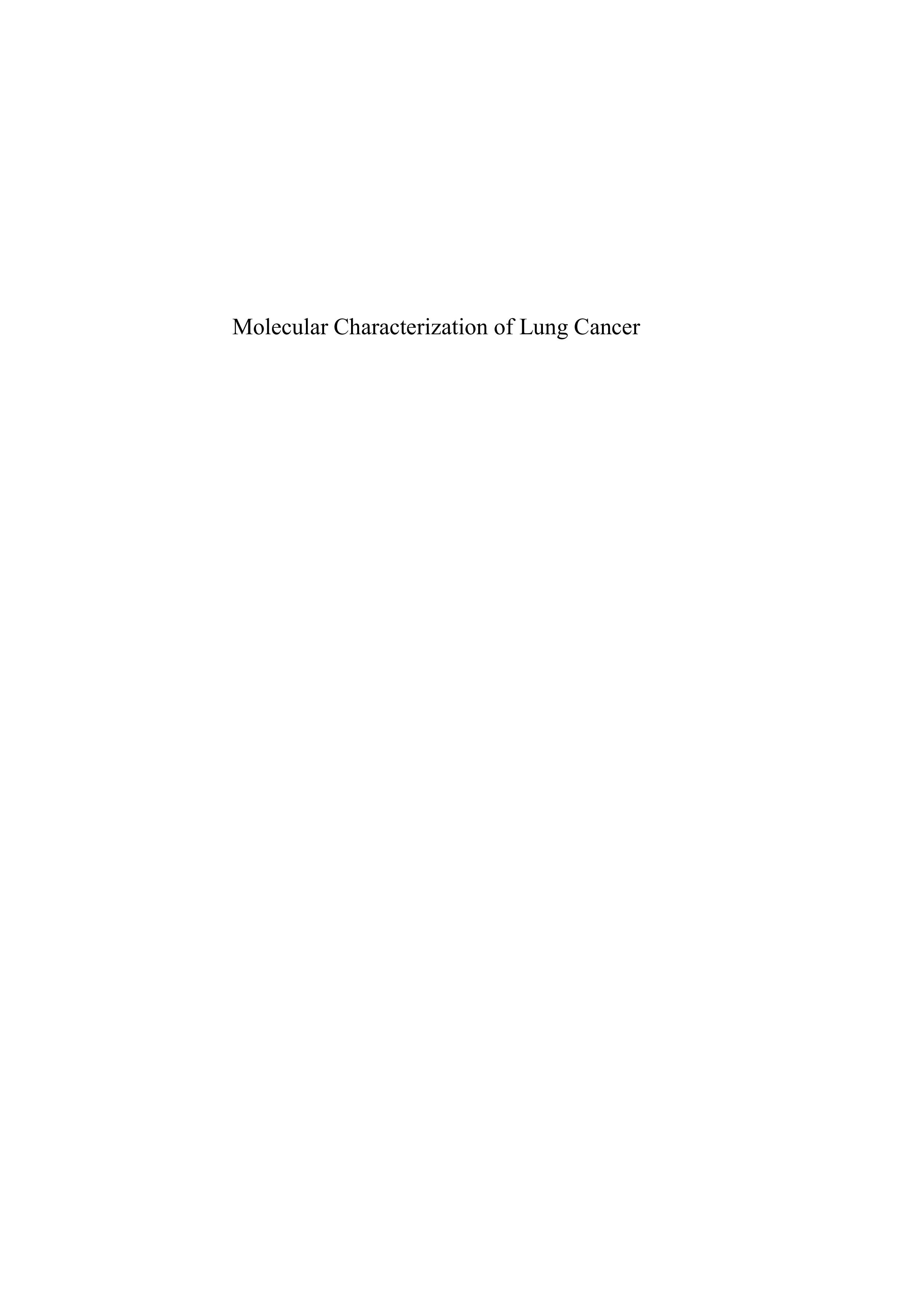
LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Molecular Characterization of Lung Cancer

Targets for Therapy and Prognostic Markers

ANNETTE SALOMONSSON
FACULTY OF MEDICINE | LUND UNIVERSITY

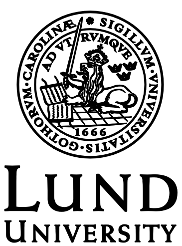


Molecular Characterization of Lung Cancer

Molecular Characterization of Lung Cancer

Targets for Therapy and Prognostic Markers

Annette Salomonsson



LUND UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended in Sharience, The Spark, Medicon Village,
Scheeletorget 1, Lund, Sweden. April 4, 2022 at 09.00 am.

Faculty opponent
Associate Professor Simon Ekman, M.D., Ph.D.
Department of Oncology and Pathology, Karolinska Institutet,
Thoracic Oncology Center, Theme Cancer, Karolinska University Hospital,
Stockholm, Sweden.

Organization LUND UNIVERSITY	**Document name** DOCTORAL DISSERTATION
	Date of issue April 4, 2022
Author Annette Salomonsson	Sponsoring organization

Title and subtitle
Molecular Characterization of Lung Cancer – Targets for Therapy and Prognostic Markers

Abstract
Lung cancer is the most common cause of death from cancer, with a substantial mortality also among patients with localized disease. Smoking is the main risk factor for developing the disease, but 10-15% of all lung cancer patients have never smoked. Despite a great improvement in the understanding and treatment of lung cancer in recent years, the prognosis remains poor. By investigating alterations in targets for therapy and potential prognostic markers, this thesis aims to characterize lung cancer on the molecular level for an improved understanding, treatment and prognostication.

In Study I, we studied the receptor tyrosine kinase KIT and its ligand KITLG on the DNA, RNA and protein levels in 72 surgically treated lung cancer cases of different histological subtypes. Different types of alterations in *KIT* were identified in the histological subtypes. Also, a poor correlation between *KIT* gene copy numbers and expression was identified.

In Study II, we studied mutations and gene fusions in targetable or potentially targetable genes in never smokers with lung cancer. We found a high prevalence of alterations in our nationwide and population-based cohort consisting of over 400 cases.

In Study III, we examined the prognostic role of RBM3 in two lung cancer cohorts, which consisted of 213 and 306 cases. High protein expression of RBM3 was a positive prognostic factor among adenocarcinomas (AC), while mRNA levels of *RBM3* were not clearly associated with patient outcome.

In Study IV, we used six publicly available gene expression data sets, comprising 1,167 lung AC, to identify genes with correlation to overall survival. We identified 19 genes with prognostic potential in all six data sets, and three of these were chosen for validation with immunohistochemistry in two additional lung cancer cohorts consisting of 131 and 194 AC cases. The study demonstrates that potential prognostic markers can be identified by employing our multi-cohort gene expression-based strategy and then further assessed by a technique universally applicable in the clinical practice.

In conclusion, these results bring new insights into potential prognostic markers and alterations in targets for therapy in lung cancer.

Key words
Lung cancer, oncogene, never smoker, prognosis, sequencing, immunohistochemistry, gene expression.

Classification system and/or index terms (if any)

Supplementary bibliographical information		**Language** English
ISSN and key title 1652-8220 Lund University, Faculty of Medicine Doctoral Dissertation Series 2022:47		**ISBN** 978-91-8021-208-3
Recipient's notes	**Number of pages** 112	Price
	Security classification	

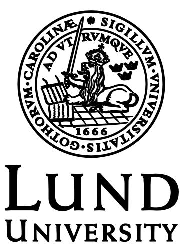
I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature Annette Salomonsson Date 2022-02-25

Molecular Characterization of Lung Cancer

Targets for Therapy and Prognostic Markers

Annette Salomonsson
Faculty of Medicine
Department of Clinical Sciences, Lund
Division of Oncology



LUND
UNIVERSITY

Main supervisor:

Maria Planck, M.D., Ph.D.

Co-supervisors:

Johan Staaf, Ph.D.

Mats Jönsson, Ph.D.

Hans Brunnström, M.D., Ph.D.

Karin Jirström, M.D., Ph.D.

Åke Borg, Ph.D.

Coverphoto by Annette Salomonsson

Copyright pp 1-112 Annette Salomonsson

Study I © 2013 Wiley Periodicals, Inc. Published in Genes, Chromosomes & Cancer by John Wiley & Sons Ltd.

Study II © The Authors (manuscript unpublished)

Study III © The Authors 2020. Published in Cancer Medicine by John Wiley & Sons Ltd.

Study IV © The Authors (manuscript unpublished)

Faculty of Medicine
Department of Clinical Sciences Lund
Division of Oncology, Lund University

ISBN 978-91-8021-208-3
ISSN 1652-8220
Lund University, Faculty of Medicine Doctoral Dissertation Series 2022:47

Printed in Sweden by Media-Tryck, Lund University
Lund 2022

To my father and Alfreda,
who would have loved to read this.

Table of Contents

Thesis at a Glance **11**

Studies Included in the Thesis **12**

Study Contributions **13**

Related Studies Not Included in the Thesis **14**

Abbreviations **15**

Populärvetenskaplig sammanfattning **18**

Introduction and Background **21**

- Epidemiology and Etiology 21
- Histology 23
  - Adenocarcinoma 24
  - Squamous Cell Carcinoma 25
  - Large Cell Carcinoma 26
  - Neuroendocrine Tumors 26
- Molecular Alterations 26
  - Oncogene Drivers in Non-Small Cell Lung Cancer 29
- Diagnostics 32
- Staging 34
- Screening 36
- Treatment 36
  - Surgery 37
  - Radiotherapy 38
  - Chemotherapy 39
  - Immunotherapy 40
  - Targeted Therapies 40
- Follow-up and Prognosis 44
- Lung Cancer in Never Smokers 47

Aims **49**

- Overall Aim 49
- Specific Aims 49

Materials......50
Patient Cohorts......50
Study I......51
Study II......51
Studies III and IV......52
External Data Sets......53
Methods......54
Tumor Tissue Handling......54
Tissue Microarrays......55
Immunohistochemistry......56
Sequencing......57
Real-time Quantitative Reverse Transcription PCR......59
NanoString......61
Microarrays......63
Statistical Methods......64
Study I......65
Study II......65
Study III and IV......65
Results and Discussion......67
Targets for Therapy......67
Study I......67
Study II......71
Prognostic Markers......76
Study III......76
Study IV......81
Conclusions......89
Studies I and II......89
Studies III and IV......89
Future Perspectives......91
Acknowledgements......93
References......95

Thesis at a Glance

Study	Question	Materials and Methods	Results and Conclusions
		Targets for Therapy	
I	Which alterations in *KIT* and *KITLG* are present in lung cancer?	• 72 lung tumors of different histological subtypes • External microarray-based data sets - Real-time quantitative reverse transcription PCR - Immunohistochemistry - Sanger sequencing - Analysis of microarray-based data generated from previous studies	Different histological subtypes harbored different types of *KIT* alterations. A poor correlation between *KIT* gene copy numbers and expression of mRNA or protein was identified. The role of *KIT* in lung cancer deserves further investigation, particularly in neuroendocrine lung tumors.
II	How frequent are alterations in targetable or potentially targetable genes in lung cancer among never smokers?	• 431 lung tumors from never smokers - NanoString technology - Next generation sequencing	Mutations or gene fusions in targetable or potentially targetable genes were identified in 72% of tumors from Swedish never smokers. The frequent occurrence of oncogene driver alterations in never smokers illustrates the importance of accurate treatment prediction in this group.
		Prognostic Markers	
Study	**Question**	**Materials and Methods**	**Results and Conclusions**
III	Is RBM3 a prognostic marker in lung cancer?	• Two independent lung cancer cohorts (213 and 306 cases) • External microarray-based gene expression data sets - Immunohistochemistry - Analysis of gene expression data generated from previous studies	High protein expression of RBM3 was associated with improved outcome in surgically treated lung adenocarcinoma cases and may become a useful prognostic marker.
IV	Can potential prognostic markers be identified by employing a multi-cohort, gene expression-based strategy, and further assessed by a technique universally applicable in clinical practice?	• External microarray-based gene expression data sets • Two independent lung cancer cohorts (213 and 194 cases) - Analysis of gene expression data generated from previous studies - Immunohistochemistry	In total, 19 potential prognostic markers were identified by employing our multi-cohort, gene expression-based approach. This may be a useful research strategy to identify candidate markers for further testing with, for example, immunohistochemistry.

Studies Included in the Thesis

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. Histological Specificity of Alterations and Expression of *KIT* and *KITLG* in Non-Small Cell Lung Carcinoma.

Salomonsson A, Jönsson M, Isaksson S, Karlsson A, Jönsson P, Gaber A, Bendahl PO, Johansson L, Brunnström H, Jirström K, Borg Å, Staaf J, Planck M. Genes Chromosomes Cancer. 2013 Nov;52(11):1088-96.

II. Lung Cancer in Never Smokers: A Nationwide Population-Based Mapping of Targetable Alterations.

Salomonsson A, Jönsson M, Rosengren F, Veerla S, Karlsson A, Reuterswärd C, Behndig A, Bergman B, Botling J, Brandén E, De Petris L, Helenius G, Hussein A, Johansson M, Kentson M, Koyi H, Lamberg K, Lewensohn R, Mager U, Micke P, Ortiz-Villalon C, Sundh J, Thurfjell V, Vikström A, Wagenius G, Monsef N, Patthey A, Brunnström H, Staaf J, Planck M. Manuscript 2022.

III. Comprehensive Analysis of RNA Binding Motif Protein 3 (RBM3) in Non-Small Cell Lung Cancer.

Salomonsson A, Micke P, Mattsson JSM, La Fleur L, Isaksson J, Jönsson M, Nodin B, Botling J, Uhlén M, Jirström K, Staaf J, Planck M, Brunnström H. Cancer Med. 2020 Aug;9(15):5609-5619.

IV. Gene Expression-Based Identification of Prognostic Markers in Lung Adenocarcinoma.

Salomonsson A, Ehinger D, Jönsson M, Botling J, Micke P, Brunnström H, Staaf J, Planck M. Manuscript 2022.

Study Contributions

My contribution to the studies included in this doctoral thesis were as follows:

I. I collected the clinical data, performed the real-time qRT-PCR experiment and Sanger sequencing, did the immunohistochemical evaluations, performed the statistical analyses, and wrote the manuscript.

II. I collected the clinical data for the patients in the southern region and assembled clinical and histopathological data collected by our collaborators in other regions. I collected survival data, interpreted results from molecular analyses, and wrote the manuscript.

III. I participated in the planning of the study, collected original clinical data for cohort I, complemented clinical data for cohort II, did the immunohistochemical evaluations, took part in the analyses of gene-expression data, performed the statistical analyses, and wrote the manuscript.

IV. I participated in the planning of the study, analyzed gene expression data to generate the candidate prognostic markers, collected original clinical data for cohort I, complemented clinical data for cohort II, did the immunohistochemical evaluations, performed the statistical analyses, and wrote the manuscript.

Related Studies Not Included in the Thesis

The following studies are not included in this thesis but are of relevance to the field:

Relation Between Smoking History and Gene Expression Profiles in Lung Adenocarcinomas. *Staaf J, Jönsson G, Jönsson M, Karlsson A, Isaksson S,* ***Salomonsson A****, Pettersson HM, Soller M, Ewers SB, Johansson L, Jönsson P, Planck M.* BMC Med Genomics. 2012 Jun 7;5:22.

Detecting EGFR Alterations in Clinical Specimens – Pitfalls and Necessities. *Isaksson S, Bendahl PO,* ***Salomonsson A****, Jönsson M, Haglund M, Gaber A, Jirström K, Jönsson P, Borg A, Johansson L, Staaf J, Planck M.* Virchows Arch. 2013 Dec;463(6):755-64.

A Gene Expression-Based Single Sample Predictor of Lung Adenocarcinoma Molecular Subtype and Prognosis. *Liljedahl H, Karlsson A, Oskarsdottir GN,* ***Salomonsson A****, Brunnström H, Erlingsdottir G, Jönsson M, Isaksson S, Arbajian E, Ortiz-Villalón C, Hussein A, Bergman B, Vikström A, Monsef N, Branden E, Koyi H, de Petris L, Patthey A, Behndig AF, Johansson M, Planck M, Staaf J.* Int J Cancer. 2021 Jan 1;148(1):238-251

Abbreviations

AC	Adenocarcinoma
aCGH	Array-based comparative genomic hybridization
ALK	Anaplastic lymphoma kinase
ASPM	Abnormal spindle-like microcephaly-associated protein
ATP	Adenosine triphosphate
BRAF	Rapidly accelerated fibrosarcoma kinase B
BTG2	B-cell translocation gene 2
BUB1B	Budding uninhibited by benzimidazoles 1 homolog beta
CCNA2	Cyclin A2
CCNE1	Cyclin E1
cDNA	Complementary deoxyribonucleic acid
CIRBP	Cold inducible RNA binding protein
COL4A3	Collagen, type IV, alpha 3
CSS	Cancer-specific survival
CT	Computed tomography
Ct	Cycle threshold
CTTN	Cortactin
ddNTP	Dideoxynucleoside triphosphate
DLGAP5	Discs large homolog associated protein 5
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
E	Efficiency
EBUS	Endobronchial ultrasound
ECT2	Epithelial cell transforming sequence 2 oncogene
EGFR	Epidermal growth factor receptor
EML4	Echinoderm microtubule-associated protein-like 4
ERK	Extracellular signal-regulated kinase
FDA	Food and Drug Administration
FDG PET	Fluorodeoxyglucose positron emission tomography
FFPE	Formalin-fixed paraffin embedded
GDF10	Growth differentiation factor 10
GTP	Guanosine triphosphate
HER2	Human epidermal growth factor receptor 2
HLF	Hepatic leukemia factor
HPV	Human papillomavirus

HRP	Horseradish peroxide
IHC	Immunohistochemistry
Ki67	Marker of proliferation Kiel 67
KIF14	Kinesin family member 14
KRAS	Kirsten rat sarcoma viral oncogene homolog
LCC	Large cell carcinoma
LCNEC	Large cell neuroendocrine carcinoma
MAPK	Mitogen-activated protein kinase
MCM4	Minichromosome maintenance complex component 4
MET	Mesenchymal epithelial transition factor
mRNA	Messenger ribonucleic acid
NF	Nuclear fraction
NGS	Next generation sequencing
NI	Nuclear intensity
NLCR	National Lung Cancer Registry
NRG1	Neuregulin 1
NS	Nuclear score
NSCLC	Non-small cell lung cancer
NTRK	Neurotrophic tropomyosin receptor kinase
NUSAP1	Nucleolar and spindle-associated protein 1
OS	Overall survival
P	Probability value
PCR	Polymerase chain reaction
PD-1	Programmed cell death protein 1
PD-L1	Programmed death ligand 1
PFS	Progression-free survival
PI3K	Phophoinositide 3-kinase
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
POLR2A	RNA polymerase II subunit A
PRC1	Protein regulator of cytokinesis 1
PS	Performance status
qRT	Quantitative reverse transcription
RACGAP1	Rac GTPase activating protein 1
RAF	Rapidly accelerated fibrosarcoma
RAS	Rat sarcoma virus
RBM3	RNA-binding motif protein 3
RBP	RNA-binding proteins
RET	Rearranged during transfection
RFI	Recurrence-free interval
RNA	Ribonucleic acid
ROS1	ROS proto-oncogene 1
RTK	Receptor tyrosine kinase

SCLC	Small cell lung cancer
SBRT	Stereotactic body radiotherapy
SMIL	Swedish Molecular Initiative against Lung cancer
SqCC	Squamous cell carcinoma
TKI	Tyrosine kinase inhibitor
TMA	Tissue microarray
TNM	Tumor node metastasis
TP53	Tumor protein 53
TYMS	Thymidylate synthetase
VATS	Video-assisted thoracoscopic surgery
WHO	World Health Organization

Populärvetenskaplig sammanfattning

Lungcancer är en komplex sjukdom som förekommer i många olika former och med olika bakomliggande tumöregenskaper. Rökning är den vanligaste orsaken till sjukdomen men 10–15% av lungcancerpatienterna har aldrig rökt. Trots de stora framsteg som skett inom lungcancerforskningen på senare tid, och som bland annat lett till effektivare och mer individualiserade behandlingar, är lungcancer den cancerform som orsakar flest dödsfall, både i Sverige och globalt. Detta beror delvis på att lungcancer ofta upptäcks för sent. Hälften av alla patienter med lungcancer får sin diagnos när sjukdomen redan är spridd och det inte längre är möjligt att ge behandling i botande syfte (kurativt syftande behandling). Efter fem år lever endast fem av 100 inom denna patientgrupp. För patienter där sjukdomen ännu inte hunnit sprida sig är prognosen bättre och ungefär hälften lever efter fem år.

Kurativt syftande behandling i tidiga stadier av lungcancer kan utgöras av kirurgi eller strålbehandling (eventuellt kombinerat med cellgifter). Lungcancerkirurgi innebär att delar av lungan, eller i vissa fall hela lungan, opereras bort. Trots operation är det dessvärre vanligt att sjukdomen kommer tillbaka efter en tid och då är mer svårbehandlad. Idag är det främst lungcancerns utbredning och patientens allmäntillstånd som ligger till grund för att förutsäga prognosen och anpassa behandlingsstrategin. Detta är dock otillräckligt och det behövs fler sätt att bedöma vilka patienter som har störst risk för återfall efter genomgången operation och som därmed kanske skulle ha störst nytta av särskild tilläggsbehandling eller intensifierad uppföljning.

För patienter med spridd lungcancer har det skett stora forskningsframgångar de senaste decennierna. Tidigare var cellgifter (cytostatika) det enda alternativet för dessa patienter. Cytostatikabehandling kan ge förlängd överlevnad, lindring av symptom och bättre livskvalitet, men det är även vanligt med biverkningar, till exempel illamående, trötthet och infektionskänslighet. Idag kan vissa undergrupper av lungcancerpatienter istället erbjudas så kallad målriktad behandling, som i jämförelse med cytostatika ger både förlängd överlevnad och mindre biverkningar. Hos cancerceller kan så kallade mutationer eller fusioner ha skett i cancercellernas arvsmassa (DNA) så att proteiner av betydelse för normal celltillväxt och celldelning blivit ständigt aktiverade och driver cancerutvecklingen. Kring millennieskiftet upptäcktes det att målriktad behandling kunde blockera vissa sådana felaktigt aktiverade proteiner.

Den första målriktade behandlingen som började användas vid lungcancer var ett läkemedel som hämmar proteinet EGFR (Epidermal Growth Factor Receptor), som finns på cellytan och har betydelse för cellernas tillväxt- och delningsförmåga. Med tiden upptäckte man att det är just patienter vars lungcancer uppvisar mutationer i *EGFR*-genen som har bättre effekt av EGFR-hämmare än av cytostatika. Upptäckterna kring EGFR-hämmarna följdes av nya upptäckter kring förbättrade läkemedel och kartläggning av andra lungcancerdrivande gener där mutationer eller fusioner gett upphov till felaktigt aktiverade proteiner mot vilka man kan ge målriktad behandling. Såväl *EGFR*-mutationer som de flesta av de senare upptäckta förändringarna är vanligare i tumörer hos lungcancerpatienter som aldrig rökt. En annan grupp läkemedel som på senare år förbättrat behandlingsmöjligheterna vid spridd lungcancer är immunterapi, där man utnyttjar kroppens eget immunförsvar till att bekämpa cancerceller. Under kommande decennier kommer sannolikt ytterligare framsteg att ske inom området, både med nya sätt att använda och kombinera befintliga läkemedel samt upptäckter av nya mål för behandling.

Denna avhandling utgörs av fyra delarbeten med det övergripande målet att öka kunskapen om olika prognostiska markörer i lungcancer och om olika tumördrivande gener som kan vara viktiga vid målriktad behandling av lungcancer. I arbetena har vi nästan uteslutande studerat den vanligaste formen av lungcancer, icke-småcellig lungcancer, och tumörmaterialet kommer i de flesta fall från patienter som har genomgått lungcancerkirurgi i kurativt syfte.

Delarbete I och II fokuserar på förändringar i tumördrivande gener.

I delarbete I studerade vi en molekyl som heter KIT. Från studier på andra tumörformer vet man att förändringar i KIT kan driva tumörutveckling, men det är oklart hur viktig denna molekyl är i just lungcancer. Vi kartlade förändringar på flera olika nivåer; DNA, RNA (ett mellanled vid proteintillverkning) och protein. Vi använde oss av tumörvävnad från 72 opererade lungcancerpatienter. Tumörerna var av olika undergrupper av icke-småcellig lungcancer, samt tre fall av småcellig lungcancer. Vi använde även resultat från andra forskares studier för att bekräfta våra fynd. Vi kunde se att olika typer av lungcancer hade olika typer av förändringar inom *KIT*. I en viss undergrupp, så kallad storcellig neuroendokrin lungcancer, var det vanligt med ett högt uttryck av *KIT* (protein och RNA), medan det i en annan undergrupp, skivepitelcancer, var vanligare med ökat antal DNA-kopior. Det var alltså dålig korrelation mellan antal DNA-kopior och uttrycksnivåerna av RNA och protein. Vi undersökte även förekomsten av mutationer inom *KIT* i undergruppen storcellig neuroendokrin lungcancer men fann där inga mutationer.

I delarbete II genomförde vi en kartläggning av förändringar i tumördrivande gener hos lungcancerpatienter som aldrig rökt, d v s en patientgrupp där det är vanligare att tumörerna uppvisar förändringar mot vilka man kan ge målriktad behandling. Genom att söka i Nationella Lungcancerregistret kunde vi identifiera samtliga patienter som genomgått kirurgi för lungcancer i Sverige under åren 2005–2014 och

som enligt registeruppgifterna aldrig hade rökt. Tumörvävnad från dessa patienter samlades in. Vi gick även igenom journalerna för att bekräfta att patienterna aldrig rökt, samt för att samla in uppgifter kring behandling och uppföljning. Vi kunde då inkludera 431 tumörer i studien. Tumörvävnaden analyserades för att finna mutationer och fusioner. Vi kunde se att det var mycket vanligt med förändringar inom gener mot vilka man idag kan ge målriktad behandling och även i gener där det forskas kring nya läkemedel. I 307 av de 428 tumörer (72%) som kunde analyseras kunde vi påvisa sådana mutationer eller fusioner. Av de 121 tumörer där vi inte kunde påvisa någon förändring var 73 tumörer ofullständigt analyserade på grund av att provet var av för dålig kvalitet. Det är alltså sannolikt att även en del av dessa tumörer hade sådana förändringar.

Delarbete III och IV fokuserar på prognostiska faktorer vid lungcancer.

I delarbete III undersökte vi om ett protein, RBM3, kan användas vid lungcancer för att hitta patientgrupper med bättre respektive sämre prognos. I tidigare studier har RBM3 visat sig ha prognostisk betydelse i andra tumörformer. Vi använde oss av tumörvävnad från patienter som opererats för lungcancer i Lund under åren 2005–2011 (213 patienter) eller i Uppsala under åren 2006–2010 (306 patienter). Tumörernas uttryck av RBM3 analyserades med hjälp av immunhistokemi (infärgning av specifika proteiner i vävnadsprover). Vi kunde se att patienter med en viss typ av lungcancer, adenocarcinom, som hade ett högt uttryck av RBM3 hade en bättre prognos än de som hade ett lågt uttryck.

I delarbete IV ville vi bredda vår strategi för att identifiera och utvärdera nya potentiellt prognostiska markörer. Vi använde oss av resultat från sex redan publicerade studier (tillgängliga i offentliga databaser) som rapporterat uttrycksnivåerna av tusentals gener i ett mycket stort antal tumörer. Med hjälp av olika statistiska beräkningar fann vi 19 gener som var kopplade till prognos i alla sex studierna. Eftersom metoden för att analysera genuttrycksnivåer inte rutinmässigt används i sjukvården valde vi ut tre av de 19 generna och gjorde immunhistokemiska färgningar av tumörmaterial från en del av patienterna från delarbete III. Även om immunhistokemin inte säkert kunde bekräfta att just de tre gener vi valt ut för det sista steget i studien var prognostiska, så visar studien att vårt tillvägagångssätt som helhet fungerar och kan bli värdefullt i fortsatt forskning för förbättrad prognostik vid lungcancer.

Sammantaget har dessa fyra arbeten givit nya inblickar i några av alla de komplexa molekylära förändringar som finns i lungcancer. Med ökad kunskap om olika tumördrivande gener och prognostiska markörer hoppas jag att studierna i denna avhandling kan bidra till det enorma pussel som nu håller på att läggas inom lungcancerområdet och som förhoppningsvis kan leda till att fler patienter i framtiden kan erbjudas en mer individualiserad och bättre behandling.

Introduction and Background

Lung cancer is not *one* but many diseases. Albeit all tumors rely on alterations in their genetic apparatus, the molecular landscape of lung cancer is diverse and complex. Many tumors share common pathways, but the molecular details, and how to exploit these alterations in the battle against cancer, have only just begun to be unfolded.

Considerable advancement has taken place in the past decades when it comes to understanding the biology of lung cancer, and improvements in detection and treatment have led to an improvement in the survival rates[1, 2]. Through further understanding of the molecular alterations in lung cancer, additional improvement can hopefully take place in the coming decades.

Epidemiology and Etiology

Due to a high incidence and poor survival rates, lung cancer is the worldwide leading cause of cancer-related death[3]. Globally, over 2 million cases were diagnosed in 2020[3]. In Sweden, about 4200 cases are diagnosed annually, and the median age at diagnosis is 69 years[1]. The incidence of lung cancer varies between different countries and regions, as illustrated in Figure 1. Sweden has a relatively low incidence, especially when comparing the incidence among men[1]. In the United States, a decline in the age standardized incidence has been observed in the past decades[4]. However, within the large population of China, a sharp increase in lung cancer incidence is observed and will likely continue to rise in the coming decades[5]. In Sweden, there has been a decline in the male lung cancer incidence in the past decades, but an increase among women, and today slightly more women than men are diagnosed with the disease[1]. These variations in incidence over time mainly reflect the smoking habits in the population.

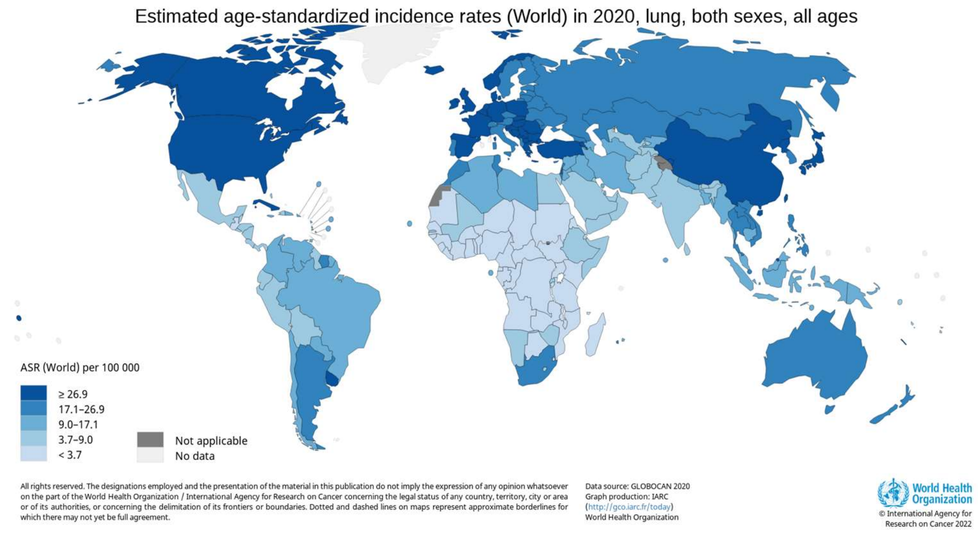


Figure 1. A world map illustrating the estimated age-standardized incidence rates (ASR) of lung cancer in 2020 for men and women.
Reprinted with permission from Sung et al.[3] and Ferlay et al.[6], Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. Available from: https://gco.iarc.fr/today, accessed [29 01 2022].

It is estimated that 85% of all lung cancer cases in men, with a lower fraction in women, are attributable to smoking[7]. Tobacco smoke contains over 70 carcinogenic compounds which can promote cancer development by several different pathways[8]. Environmental exposure to tobacco smoke, for example from a smoking spouse or workplace exposure, also increases the risk of developing lung cancer[9]. Smoking not only increases the risk of lung cancer, but also that of several other types of cancer[10]. There is a latency period of 10-30 years from start of smoking to development of lung cancer, and the risk remains elevated for up to 30 years after cessation[1, 11].

Radon is a radioactive gas from the ground that can enter houses. Indoor radon gas concentrations are dependent on the local geology and on how houses are built and ventilated. The radon gas is inhaled and further decay in the respiratory airways results in radioactive emission, resulting in an increased risk for lung cancer development[12]. Furthermore, due to synergistic effects, the risk from radon exposure is even greater among smokers[12].

Occupational exposure to carcinogens that increase the risk of lung cancer includes, among others, asbestos and other silica compounds, chromium, heavy metals and diesel engine exhaust[1, 13]. As for radon, smoking and asbestos have a synergistic effect in increasing the risk for lung cancer.

Exposure to indoor and outdoor air pollution, such as exhaust fumes from vehicles, particles from vehicle tires, factory emissions, cooking oil vapors, and indoor coal burning, also increase the risk of lung cancer[9, 14].

An increased risk of developing the disease has been observed for individuals with a family history of lung cancer, indicating that hereditary factors may play a role in the carcinogenesis of lung cancer[15]. Polymorphisms in deoxyribonucleic acid (DNA) repair genes and in enzyme genes that govern the metabolism of carcinogens can contribute to differences in the susceptibility to environmental risk factors[16]. The T790M mutation in the epidermal growth factor receptor (*EGFR*) gene, which is typically a mechanism of acquired resistance during treatment with tyrosine kinase inhibitors (TKI), also occurs as a germline mutation and confers an increased risk of lung cancer[17]. In addition, within families with known hereditary cancer syndromes caused by germline mutations in, e.g., the tumor protein 53 (*TP53*) or breast cancer 2 (*BRCA2*) genes, there is an increased risk also of lung cancer among the mutation carriers[18, 19].

The human papillomaviruses (HPV) are risk factors for several types of cancer[20]. Whereas some studies have reported the presence of HPV DNA in lung tumors, a study conducted on a subset of the tumors included in Study II found no evidence for HPV or human polyomaviruses (HPyV) in the etiology of lung cancer in Swedish never smokers[21, 22].

The significance of gender for the risk of developing lung cancer has been studied. Higher incidence rates have been noted among young women compared to young men in some countries, and the results could not be fully explained by differences in tobacco use[23]. Furthermore, the lung cancer incidence among never smokers has been noted to be higher among women than men[9, 24]. These findings suggest that there may exist gender-dependent differences in susceptibility or exposure to risk factors for lung cancer other than smoking. For example, the role of hormonal factors has been studied, and estrogen has been suggested to be a carcinogenic factor, but further studies are warranted[25].

As other lung diseases lead to inflammation, and inflammation promotes cancer development, diseases such as chronic obstructive pulmonary disease (COPD), pneumonia and tuberculosis have been suggested to increase the risk of lung cancer, irrespective of smoking habits[26].

Histology

Lung cancer is traditionally divided into small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). Non-small cell lung cancer is the more common entity, accounting for approximately 80-85% of all lung cancer cases[27, 28]. The

histological type is often determined based on a tumor biopsy or cytology specimen, and the workup includes morphological evaluation and staining with immunohistochemical markers. Non-small cell lung cancer is further divided into the main subtypes adenocarcinoma (AC), squamous cell carcinoma (SqCC), and large cell carcinoma (LCC).

The studies included in this thesis focus on NSCLC, except for a few SCLC cases in Studies I and II.

In Study I, the 2004 World Health Organization (WHO) classification, which was current at the time of the study, was used[29]. In Studies II, III and IV, the tumors were histologically classified according to the 2015 WHO classification[30]. The 2015 WHO classification was more adapted to meet the new demands of more exact histopathological subtyping, since this became the basis for therapeutic decision making as certain drugs were proved more effective for specific subtypes. Hence, some of the differences between these two editions include a more prominent role for immunohistochemistry (IHC), a greater importance to further classify NSCLC into specific subtypes, and a restriction of the use of the diagnosis LCC[31]. The WHO classification for lung tumors was updated in 2021, where a few new tumor types were added and a formal grading system for lung AC was introduced[32].

Adenocarcinoma

Adenocarcinoma is the most common subtype, accounting for approximately 50% of all lung cancer cases[27]. Adenocarcinoma is less associated with smoking than SqCC and SCLC[33]. The tumorigenesis of AC often relies on alterations in single driver oncogenes[34]. Morphological features of AC include glandular formations or mucin inclusions, as illustrated in Figure 2. Furthermore, AC of the lung can be discriminated by immunopositivity of the diagnostic markers thyroid transcription factor 1 (TTF1) and napsin A.

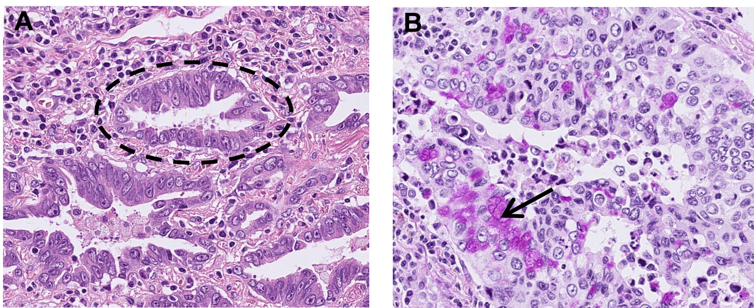


Figure 2. Morphological features of adenocarcinoma include glandular formations (A, indicated by the dashed circle) or mucin inclusions (B, indicated by the arrow).
Photos kindly provided by Dr. Hans Brunnström.

Adenocarcinoma *in situ* refers to a tumor without invasion, while minimally invasive adenocarcinoma has an area of invasion ≤ 5 mm. Non-mucinous invasive adenocarcinomas are further subclassified into subtypes based on distinct growth patterns, and the five most common patterns are lepidic, acinar, papillary, micropapillary, and solid. Most invasive adenocarcinomas consist of a mix of these subtypes, and the tumors are classified according to the predominant growth pattern and the additional subpatterns are noted. These subtypes have been found to correlate with prognosis, and a grading system has been proposed with the lepidic predominant being of low grade (conferring better prognosis), acinar or papillary predominant intermediate, and solid or micropapillary predominant high grade (conferring poor prognosis)[35].

Squamous Cell Carcinoma

Until some decades ago, SqCC was the most common histological subtype until it was surpassed by AC[36]. This shift in incidence may partly be explained by changes in smoking habits and cigarette design, e.g., the introduction of cigarette filters[37]. However, SqCC still remains the most common type in some countries[38].

Morphological features of SqCC include keratinization and intercellular bridges (connections between cells), as illustrated in Figure 3. Squamous cell carcinomas are commonly positive for the immunohistochemical markers cytokeratin 5 (CK5), cytokeratin 5 and 6 (CK5/6), and p40. Mutations in targetable driver oncogenes are rare, and SqCC often display a high overall mutation rate due to its connection to smoking[39].

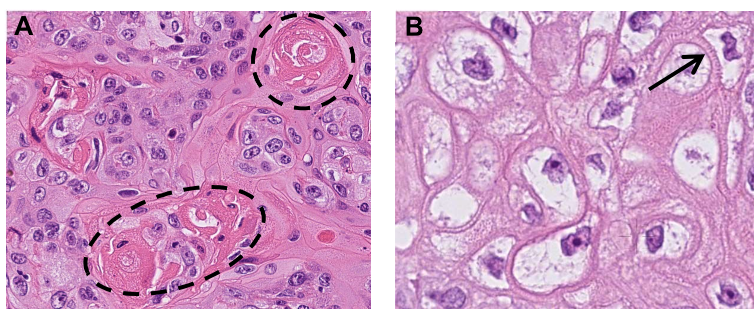


Figure 3. Morphological features of squamous cell carcinoma include keratinization (A, indicated by the dashed circles) and intercellular bridges (B, indicated by the arrow).
Photos kindly provided by Dr. Hans Brunnström.

Large Cell Carcinoma

Large cell carcinoma is defined as an NSCLC tumor that lacks the features of the other histological types[40]. This diagnosis can only be made on surgical specimens, and it has become rare since the implementation of the 2015 WHO classification, which entailed a restriction of the use of LCC. However, for biopsies and cytology specimens, the diagnostic counterpart is NSCLC not otherwise specified (NOS), which is a more commonly used diagnosis. Large cell carcinoma is suspected to represent poorly differentiated AC or SqCC where all morphological and immunohistochemical characteristics have been lost, but this remains controversial[41].

Neuroendocrine Tumors

Neuroendocrine tumors include SCLC, large cell neuroendocrine lung carcinoma (LCNEC) and carcinoids. Small cell lung carcinoma is an aggressive form of cancer with a very poor 5-year survival rate[27]. It is strongly connected to smoking and is molecularly characterized by a high mutational burden[42]. Like SCLC, LCNEC is strongly associated with smoking and is often clinically aggressive[43]. Carcinoid tumors are generally slow-growing and often have a good prognosis[44].

Molecular Alterations

In a normal cell, there is a complex balance between a myriad of functions that maintain essential cellular control mechanisms and keep the cell within its limits. For instance, a normal cell will only divide a certain number of times, and DNA damage will result in counter mechanisms, which either repair the damage or initiate apoptosis. However, a cancer cell has acquired certain capabilities that interfere with normal cell regulations. These capabilities were summarized by Hanahan and Weinberg in 2000 and organized into six hallmarks of cancer[45]. The hallmarks were further extended in 2011 with two additional hallmarks and two enabling characteristics, and recently four additional emerging hallmarks and enabling characteristics were proposed, as summarized in Figure 4[46, 47].

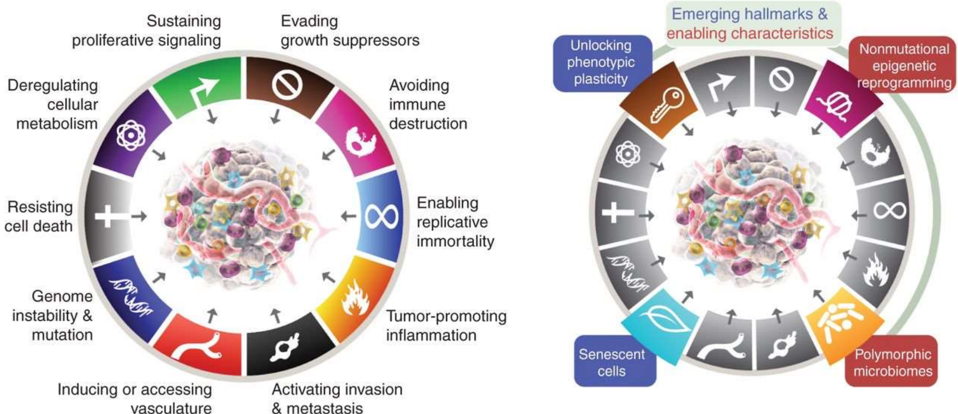


Figure 4. The hallmarks of cancer currently include eight hallmark capabilities and two enabling characteristics (left). In addition, four emerging hallmarks and enabling characteristics have recently been proposed (right).
Reprinted from Hallmarks of Cancer: New Dimensions[46], © 2021, with permission from AACR.

Many of these capabilities are acquired through various genomic changes in the future cancer cells. Some of the different types of such tumorigenic genomic alterations are described below.

A single nucleotide variant occurs if a single nucleotide in the DNA is substituted for another nucleotide. If this occurs in a protein coding region that changes the amino acid sequence of the protein, the result could be, e.g., a constantly activated protein that promotes cell proliferation.

Short insertions or deletions of nucleotides are common genetic alterations in tumor cells. Deletions that involve a number of bases that can be divided by three will result in an in-frame deletion of the amino acids encoded by the deleted bases. Insertions or deletions that involve a number of bases that cannot be divided by three will result in a frameshift, causing the aftercoming codons to code for different amino acids or a premature stop codon.

In Study II, alterations such as single nucleotide variants, small insertions, and small deletions in targetable or potentially targetable genes were studied in never smokers.

Larger deletions and duplications can result in gene copy number variations, which can affect tumorigenesis by altering gene expression levels[48]. In Study I, gene copy number alterations of *KIT* and *KITLG* were studied. Furthermore, the correlation between *KIT* gene copy numbers and expression was investigated.

Translocations are induced by double-strand breaks in the DNA, where a part of a chromosome fuses to another chromosome. Such events can promote tumorigenesis by different mechanisms, e.g., by overexpression of translocated genes, or by the formation of an oncogenic chimeric protein[49]. In most receptor tyrosine kinases (RTK), the tyrosine kinase domain is situated in the 3′ end of a gene. An oncogenic

fusion gene can occur by an in-frame translocation, where the 3′ end of the RTK is fused to the 5′ end of a partner gene. The result is a chimeric protein with a C-terminus containing the tyrosine kinase domain, and a N-terminus derived from the partner gene, typically encoding a domain capable of dimerization and activation in the absence of a ligand.

Oncogenic chimeric proteins can also occur by inversions or deletions. A common genomic alteration in lung cancer involves an inversion of a portion of chromosome 2, where the 5′ portion of the echinoderm microtubule-associated protein-like 4 (*EML4*) gene is fused to the 3′ portion of the anaplastic lymphoma kinase (*ALK*) gene, resulting in an oncogenic gene fusion, as illustrated in Figure 5. In adult lung tissue, *ALK* is not normally expressed. However, the expression of the chimeric *ALK-EML4* gene is regulated by the promoter region of *EML4* and therefore transcribed. Also, the dimerization motif of EML4 results in unregulated activation of the tyrosine kinase domains of ALK[50].

In Study II, the prevalence of targetable or potentially targetable gene fusions in never smokers was investigated.

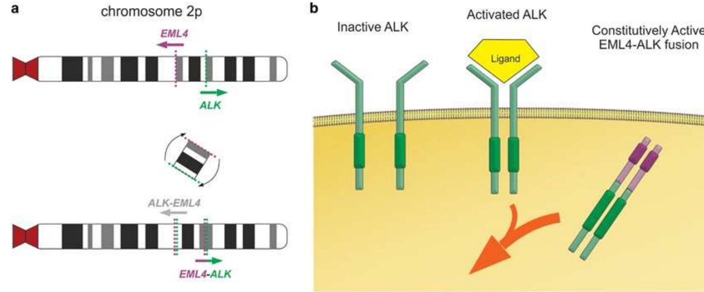


Figure 5. Inversion of a portion of chromosome 2 results in the formation of a gene fusion involving *ALK* and *EML4* (a). In adult lung tissue, *ALK* is not normally expressed. However, the expression of the chimeric *ALK-EML4* gene is regulated by the promoter region of *EML4* and therefore transcribed. Wild-type ALK (green) requires the presence of its ligand for dimerization and activation, but the EML4-ALK fusion protein is capable of dimerization in the absence of a ligand by the EML4 dimerization domain (purple), resulting in a constitutively active ALK tyrosine kinase domain (green) (b).

Reprinted from Rosenbaum et al.[50], © 2018, with permission from Springer Nature.

Epigenetic changes, such as DNA methylation and histone modification, can modify the expression of genes without altering the DNA sequence. For example, DNA methylation occurs when a methyl group is added to a cytosine (C) in the DNA that is followed by a guanine (G), a so called CpG site. Regions in the DNA that are enriched in CpG sites are called CpG islands and these are mainly found in the promoter region of a gene. Hypermethylation of CpGs in the promoter region results in gene silencing[51]. In tumors, global hypomethylation combined with hypermethylation of CpG islands are commonly observed[52].

Not only the DNA is of importance in tumorigenesis, but also the transcribed ribonucleic acids (RNA) are key mediators. There are several proteins that bind to RNA, so called RNA-binding proteins (RBPs). These proteins regulate several steps in messenger ribonucleic acid (mRNA) processing and can alter gene expression levels[53]. In Study III, the prognostic ability of RNA-binding motif protein 3 (RBM3), an RBP, was studied.

Furthermore, genetic alterations can also affect the interactions between the tumor cells and their microenvironment, e.g., by enabling the tumor cells to evade destruction by the host immune system[54].

In summary, the molecular changes in a tumor are diverse, complex, and affect different levels of the genetic regulation process.

During tumor development, multiple alterations are often stepwise accumulated over many years. A beneficial mutation in a tumor cell will confer a survival benefit, and that will facilitate the expansion of that sub-clone and generate a genetic heterogeneity within the tumor[55]. The number of alterations needed to convert a normal cell into a malignant one, and the time it takes to acquire these changes, differ. As tobacco smoke contains multiple carcinogenic compounds, lung tumors in smokers often harbor many mutations[56]. In contrast, lung cancer among never smokers are typically characterized by alterations in oncogene drivers[57]. Oncogene driver alterations are alterations in genes that are essential for tumor development.

Oncogene Drivers in Non-Small Cell Lung Cancer

There are several oncogene drivers that are commonly altered in NSCLC, presented below. These are generally known to be mutually exclusive, and, as the tumor cells have become oncogene addicted (i.e., rely on the altered signaling for their survival), these oncogene drivers constitute excellent therapeutic targets.

The EGFR protein is a transmembrane receptor with an extracellular ligand-binding domain and an intracellular tyrosine kinase domain. When the receptor is activated by its ligand, it forms homo- or heterodimers with other activated receptors and the tyrosine kinase domain is activated, leading to ATP-mediated phosphorylation of tyrosine residues. Intracellular signaling proteins can then bind to the receptor, which activates signaling cascades through the mitogen-activated protein kinase (MAPK) pathway and the phosphoinositide 3-kinase (PI3K) pathway.

Figure 6 presents a simplified description of the key mediators in these pathways. In the MAPK pathway, activation of the receptor results in activation of RAS proteins such as KRAS, further presented below. Active RAS signals through RAF protein kinases such as BRAF, also further presented below. RAF phosphorylates a second protein kinase, MEK (also known as mitogen-activated protein kinase kinase). MEK activates a third kinase, the extracellular signal-regulated kinase (ERK/MAPK), which then translocates to the nucleus and activates transcription factors resulting in several tumorigenic properties such as proliferation and survival[58]. The PI3K pathway involves the activation of two additional kinases, AKT (also known as protein kinase B) and mammalian target of rapamycin (mTOR), and activation of this pathway results in several cellular responses, e.g., proliferation and cell survival[59].

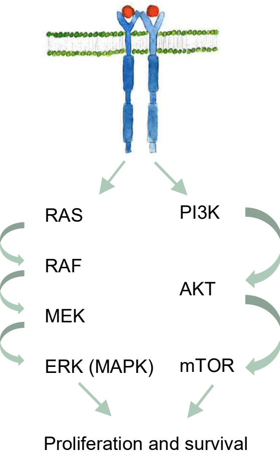


Figure 6. A simplified illustration of the key mediators in the MAPK and PI3K pathways. Through RAS – RAF – MEK – ERK signaling in the MAPK pathway, and PI3K – AKT – mTOR signaling in the PI3K pathway, several cellular responses are activated, e.g., proliferation and cell survival.

Activating mutations in the *EGFR* gene are primarily located in exons 18-21, which encode the tyrosine kinase domain, and lead to EGFR signaling in the absence of a ligand. The two most common types of activating *EGFR* mutations found in lung cancer are point mutations in exon 21 (L858R) and small (most often 15 base pair) deletions in exon 19[60]. The L858R point mutation results in a substitution of the leucine amino acid at position 858 with arginine, which has a larger charged side chain. This substitution results in a structural alteration that locks the kinase in an active state by preventing the segment from adopting its inactive conformation[61]. Similarly, deletions in exon 19 result in configurational changes that stabilize the kinase domain in an active conformation, resulting in enhanced receptor signaling[62]. Mutations in *EGFR* are more common among never smokers, ACs, and in Asian

populations, and the mutation prevalence varies from about 50% among Asian never smokers to about 10% among Caucasian patients with a smoking history[63].

The Kirsten rat sarcoma viral oncogene homolog (*KRAS*) gene encodes a protein that acts downstream of EGFR. Mutations in *KRAS* are typically substitutions in exon 2 (codons 12 or 13), or exon 3 (codon 61), which convert the protein into a constitutively active state. *KRAS* mutations are typically found among smokers, ACs and in Caucasian populations. Among patients with these three characteristics the prevalence can be about 30%[64]. However, *KRAS* mutations are also found in never smokers, and *KRAS* displays distinct mutational patterns in smokers and never smokers. For instance, the *KRAS* G12D mutation is typically found among never smokers, while the G12C mutation is more commonly found in smokers[65].

The rapidly accelerated fibrosarcoma kinase B (*BRAF*) gene encode a non-receptor serine/threonine protein kinase that, through adenosine triphosphate (ATP) mediated phosphorylation, acts downstream of KRAS and activates further signaling in the MAPK pathway[58]. Mutations in *BRAF* typically involve substitutions, such as the V600E mutation, which conforms the BRAF protein into a constitutively active state. Mutations in *BRAF* are found in about 5 % of NSCLC cases[66]. Targetable *BRAF* V600 mutations are more common among never smokers, while *BRAF* non-V600 mutations are more common among smokers[66, 67].

The human epidermal growth factor 2 receptor (HER2) belongs to a family of RTKs, which consists of four members, including EGFR. Dimerization with other activated receptors results in activation of pathways common to those induced by EGFR activation. Alterations in *HER2* include gene amplifications, mutations and overexpression[68].

The phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (*PIK3CA*) gene encodes the p110α subunit of the phosphoinositide 3-kinase (PI3K). Mutations in *PIK3CA* have been detected in about 2% of NSCLC cases, and often co-occur with other oncogene drivers[69].

The mesenchymal epithelial transition factor proto-oncogene (*MET*) encodes a receptor tyrosine kinase which signals through pathways such as the MAPK pathway and the PI3K pathway[70]. MET signaling can be dysregulated in several ways in cancer, such as gene amplification and exon 14 skipping events, the latter found among 2-4% of NSCLC cases[71]. Exon 14 encodes a domain containing a binding site for ubiquitin which, upon binding, marks the receptor for degradation, and receptors lacking exon 14 thus have a decreased degradation and increased signaling[72]. Around 130 different mutations, such as substitutions and deletions, leading to aberrant splicing in which a receptor lacking exon 14 is generated, have been described[73].

The anaplastic lymphoma kinase (*ALK*) gene encodes a receptor tyrosine kinase, which, upon activation and phosphorylation, activates signaling cascades including

the MAPK pathway and the PI3K pathway[74]. Gene fusions involving *ALK* are found in about 5% of NSCLC cases and, as with *EGFR* mutations, are more frequent in never smokers, ACs, and in Asian populations[75]. Rearrangements involving *ALK* are heterogeneous, with many possible fusion partners. Studies suggest that the fusion partner influences the properties of the ALK fusion protein[74, 76]. The most common fusion partner is *EML4*, but different breakpoints in *EML4* generate different *EML4-ALK* fusion variants. Studies have suggested that particular *EML4-ALK* variants may influence the degree of response to ALK inhibitors and the propensity to develop specific secondary *ALK* resistance mutations during treatment[77, 78].

Other oncogene driver gene fusions in NSCLC involve the regions which encode the tyrosine kinases of the ROS proto-oncogene 1 (*ROS1*), the rearranged during transfection proto-oncogene (*RET*), and the neurotrophic tropomyosin receptor kinase 1-3 (*NTRK*) genes. Upon activation, these receptor tyrosine kinases activate signaling cascades through the MAPK pathway and the PI3K pathway[79]. The neuregulin 1 (*NRG1*) gene encodes a protein that serves as a ligand for receptors in the epidermal growth factor receptor family[80]. Gene fusions involving *NRG1* result in aberrant expression of the ligand and increased signaling through the receptors. Translocations involving *ROS1*, *RET*, *NRG1*, and the *NTRK* genes occur in approximately 1-2% of NSCLC cases and are more common among never smokers[81].

Diagnostics

Lung cancer symptoms include symptoms from the lungs such as cough and shortness of breath, systemic symptoms such as fatigue and weight loss, or symptoms from a metastasis such as pain or neurological symptoms. However, symptoms usually occur at later stages when the disease has already advanced, which ultimately leads to high mortality rates. Since the symptoms in most cases are nonspecific and could be caused by benign conditions, there can be both a patient's delay and a doctor's delay in diagnosing the disease[82]. In some cases, lung cancer is detected before the onset of symptoms due to incidental findings in a radiology exam performed for other reasons. These patients often have earlier-stage disease and therefore may have a better chance of survival[83].

The diagnostic work-up should not only establish the lung cancer diagnosis but also define information that will guide the choice of treatment. For this purpose, tumor stage, histological subtype, molecular diagnostics and the general condition of the patient, are essential.

A contrast-enhanced computed tomography (CT) of the thorax and upper abdomen is performed for examination of the suspected primary lesion, lymph nodes and

potential metastases. For patients who might become candidates for curatively intended treatment, the CT is often performed in combination with a fluorodeoxyglucose positron emission tomography (FDG PET) for a more accurate staging.

Through a flexible bronchoscope, the suspected lesion can be sampled using biopsy forceps and bronchial brushes. Peripheral lesions can be sampled through transthoracic needle biopsies (fine-needle aspiration or core-needle biopsy), usually performed using imaging guidance such as CT or ultrasound.

Assessment of lymph node status is a key component in the staging procedure and is used for decision making regarding treatment strategy. For patients that are candidates for curatively intended treatment, the lymph nodes can be further examined by bronchoscopy combined with endobronchial ultrasound (EBUS), which enables the lymph nodes to be sampled with a transbronchial needle aspiration (TBNA). Since not all lymph nodes are accessible through EBUS, it can be complemented by endoscopic ultrasound (EUS) through the esophagus. However, if the tumor is PET positive, ≤ 3 cm, and located in the peripheral third of the lung, and the lymph nodes are not enlarged and are PET negative, EBUS is not necessarily required since such patients are considered clear candidates for surgery.

Whereas radiology of the brain is performed on generous grounds, also in asymptomatic lung cancer patients, examination of other potential metastatic sites is governed by the patient’s symptoms. Common metastatic locations include (besides the pulmonary lymph nodes and pleura) the brain, skeleton, liver and adrenal glands.

Molecular diagnostics of advanced lung cancer disease should include, as a minimum, alterations in targetable genes and immunohistochemical analysis of programmed death-ligand 1 (PD-L1) expression[1]. As the field of targeted therapies is constantly expanding with the emergence of new targets and drugs, the minimum set of genes that must be analyzed is constantly being revised. In Sweden, single gene assays have now been replaced by multiplexed next generation sequencing (NGS), where a large set of genes are analyzed for detection of mutations and gene fusions in DNA and RNA.

Both tumor-related factors and patient-related factors influence the choice of treatment. The patient's operability, performance status (PS) and other comorbidities must be assessed before decision-making regarding surgery, radiotherapy, and systemic treatment. Performance status is a score used to estimate the patient's ability to perform the activities of daily living. The scale used in the studies included in this thesis refers to the Eastern Cooperative Oncology Group (ECOG) scale[84], and is further presented in Table 1. For patients with a poor performance status, it can be difficult to tolerate demanding treatments, and they generally have a worse prognosis compared to more fit patients[85]. Pre-operative assessment of the patient's operability also includes pulmonary- and cardiac-function tests.

Table 1. Description of the Eastern Cooperative Oncology Group (ECOG) performance status scale[84].

Grade	Description
0	Fully active, able to carry on all pre-disease performance without restriction
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work
2	Ambulatory and capable of all selfcare but unable to carry out any work activities. Up and about more than 50% of waking hours
3	Capable of only limited selfcare, confined to bed or chair more than 50% of waking hours
4	Completely disabled. Cannot carry on any selfcare. Totally confined to bed or chair
5	Dead

Staging

The tumor, node, and metastasis (TNM) classification is used for classifying the extent of the spread of a tumor. Three main parameters are taken into account: the size and local invasion of the primary tumor (T), lymph node involvement (N), and distant metastases (M), as further described in Table 2. The different TNM combinations dictate what stage the tumor is in, as further described in Table 3. Tumor stage is used for decision making regarding treatment and is a well-established prognostic factor[85, 86]. The TNM classification system is periodically updated and currently, for lung cancer, the 8th edition is in clinical use[87]. All tumors in the studies included in this thesis were staged according to the 7th edition, published in 2009 and clinically used in Sweden until 2018[88]. A major difference between the 7th and 8th editions is a further subdivision of the T and M components in the 8th edition which aims to achieve a more precise TNM classification and thereby more accurate predictions[89].

Table 2. T, N, and M descriptors in the 7th edition of the TNM classification for lung cancer[88].

Primary Tumor (T)	Definitions
TX	Primary tumor cannot be assessed, or tumor proven by the presence of malignant cells in sputum or bronchial washings but not visualized by imaging or bronchoscopy
T0	No evidence of primary tumor
Tis	Carcinoma *in situ*
T1	Tumor 3 cm or less in greatest dimension, surrounded by lung or visceral pleura, without bronchoscopic evidence of invasion more proximal than the lobar bronchus • T1a: Tumor 2 cm or less in greatest dimension • T1b: Tumor more than 2 cm but 3 cm or less in greatest dimension
T2	Tumor more than 3 cm but 7 cm or less or tumor with any of the following features: - involves main bronchus 2 cm or more distal to the carina; - invades visceral pleura; - associated with atelectasis or obstructive pneumonitis that extends to the hilar region but does not involve the entire lung • T2a: Tumor more than 3 cm but 5 cm or less in greatest dimension • T2b: Tumor more than 5 cm but 7 cm or less in greatest dimension
T3	Tumor more than 7 cm or - one that directly invades parietal pleural, chest wall, diaphragm, phrenic nerve, mediastinal pleura or parietal pericardium - tumor in the main bronchus less than 2 cm distal to the carina but without involvement of the carina; - associated atelectasis or obstructive pneumonitis of the entire lung or separate tumor nodule(s) in the same lobe
T4	Tumor of any size that invades any of the following: mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body, carina, or separate tumor nodule(s) in a different ipsilateral lobe
Regional Lymph Nodes (N)	**Definitions**
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastases
N1	Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes, including involvement by direct extension
N2	Metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s)
N3	Metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph node(s)
Distant Metastasis (M)	**Definitions**
M0	No distant metastasis
M1	Distant metastasis • M1a: Separate tumor nodule(s) in a contralateral lobe, tumor with pleural nodules or malignant pleural or pericardial effusion • M1b: Distant metastasis in extrathoracic organs

Table 3. Stage grouping according to the 7th edition of the TNM classification for lung cancer[88].

Stage	T Component	N Component	M Component
Occult carcinoma	TX	N0	M0
Stage 0	Tis	N0	M0
Stage IA	T1a, T1b	N0	M0
Stage IB	T2a	N0	M0
Stage IIA	T2b	N0	M0
	T1a, T1b	N1	M0
	T2a	N1	M0
Stage IIB	T2b	N1	M0
	T3	N0	M0
Stage IIIA	T1a, T1b, T2a, T2b	N2	M0
	T3	N1, N2	M0
	T4	N0, N1	M0
Stage IIIB	T4	N2	M0
	Any T	N3	M0
Stage IV	Any T	Any N	M1a, M1b

Screening

Almost half of the patients with lung cancer are diagnosed with advanced disease[27], and since the survival rates are heavily dependent on stage, an important factor for improving 5-year survival rates is earlier detection. Two large screening trials for lung cancer have been conducted. The National Lung Cancer Screening Trial (NLST) was an American study that demonstrated a reduction in lung cancer mortality by screening with low-dose CT compared to chest radiography[90]. In the more recently completed Dutch-Belgian NELSON trial, it was demonstrated that low-dose CT screening reduced lung cancer mortality compared to no screening[91]. Based on the results from these trials, implementation programs for high-risk groups (i.e. current or former smokers above a certain age) are under development in several countries. However, there are several issues to overcome before a screening program can be launched, such as how to reach participants, guidelines for nodule management and other incidental findings, and general organizational issues.

Treatment

Lung cancer treatment is very much dictated by tumor stage. For patients with early stage disease, curatively intended treatment is an option. Unfortunately, a majority of the patients are diagnosed when the tumor has already advanced, and for these patients the treatment is given with a palliative intent to relieve pain, improve quality of life and, if possible, to prolong survival. Besides stage, there are several other factors that determine treatment strategy. These include both tumor- and patient-

related factors such as histology, molecular profile, comorbidities, performance status, and patient preferences. Due to this complexity, treatment strategies are generally discussed at multidisciplinary conferences where pulmonologists, oncologists, thoracic surgeons, pathologists, and radiologists attend.

An overview of lung cancer treatment for newly diagnosed patients is presented in Figure 7. However, lung cancer research is a rapidly evolving field, and the treatment options and rationale are constantly changing. Therefore, guidelines and treatment recommendations are constantly being revised as new drugs, combinations, and indications are established.

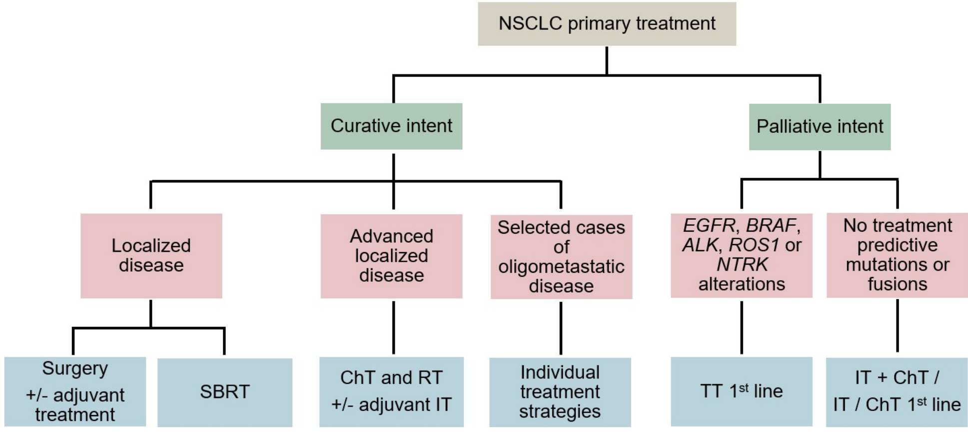


Figure 7. An overview of lung cancer treatment for newly diagnosed patients.
Abbreviations: NSCLC = non small cell lung cancer, SBRT = stereotactic body radiotherapy, ChT = chemotherapy, RT = radiotherapy, IT = immunotherapy, TT = targeted treatment.

Surgery

All tumor tissue used in the studies in this thesis have been surgically obtained. Surgery is an option for patients with lung cancer in a potentially curable stage. For patients with lung cancer in stage I and II, and for selected cases in stage IIIA, an assessment of tumor resectability and patient operability, including a risk evaluation of cardiopulmonary function and comorbidities, is performed to determine if surgery is feasible and to predict post-surgical status.

Lung cancer surgery can be performed as an open thoracotomy or by video-assisted thoracoscopic surgery (VATS). Studies have demonstrated advantages with VATS such as less intra-operative blood loss, less pain, and shorter hospital stay[92, 93]. Regarding long-term survival, retrospective studies have suggested at least equivalent results between VATS and open thoracotomy, but results from randomized studies are awaited[94, 95].

The most common surgical procedure is lobectomy, where the tumor-containing lobe is removed[1]. If all lobes of a lung are affected, or if the tumor has a central growth, a pulmectomy can be considered, but this demands a higher pre-surgical pulmonary function and is associated with a higher risk of post-operative complications compared to a lobectomy. If the patient has a limited cardiopulmonary function that does not allow for a lobectomy, a sub-lobar resection can sometimes be performed. In the mid-nineties, a randomized study demonstrated a higher rate of local recurrence after sub-lobar resections compared to lobectomies, and therefore lobectomies are recommended as first choice also for small tumors[96]. Newer studies have been published, and meta-analyses indicate that sub-lobar resections might be considered as an alternative to lobectomies for small tumors[97]. Also, long-term results from new randomized trials are awaited[98, 99].

During surgery, a systematic sampling of lymph nodes (removal of one or more representative nodes from specified lymph node stations) or dissection (complete removal of lymph nodes and surrounding tissue) is performed to more accurately determine the N-status of the TNM-classification, which is needed for decision-making regarding post-operative treatment. Although results have differed between studies, dissections are reported to confer more accurate nodal staging and improved survival, but a higher incidence of postoperative complications[100, 101].

Radiotherapy

Thanks to substantial technical development during the past decades, radiotherapy has become an important curative treatment modality for local or locally advanced lung cancer. Also, radiotherapy have an important role in the palliative setting. The effects, and side-effects, of radiation therapy are dose-dependent. Side-effects include inflammatory pneumonitis, esophagitis, lung fibrosis, and negative effects on the spinal cord and heart. Hence, the offered doses and the fractioning of the doses depend on the localization of the target, as well as on the treatment intent[1].

For patients with early-stage disease, where curative surgery is not an option due to patient-related factors, stereotactic body radiotherapy (SBRT) can be an option. In SBRT, the radiation is given with high precision to the tumor, thus sparing the surrounding healthy tissue. Tumors suitable for SBRT are typically small, peripherally situated, and not close to any risk organ such as the heart. Randomized trials comparing SBRT and surgery are scarce due to challenges in recruiting patients[102]. However, a recent systematic review and meta-analysis suggested better outcomes after lobar resections than SBRT[103]. For patients with locally advanced disease, curative radiotherapy is the primary option and should ideally be combined with chemotherapy, since chemoradiotherapy improves survival compared to radiotherapy alone[104]. Concomitant treatment with a platinum-based combination has shown to confer the best survival benefits in this group of lung cancer patients[104, 105].

Radiotherapy can also be offered post-surgically to patients with non-radical excisions or pre-surgically, combined with chemotherapy, in the neoadjuvant setting[1].

Palliative radiotherapy can be used for relieving distressing symptoms, caused by either the primary tumor or metastases, such as airway obstruction, pain, or neurological symptoms[106].

Chemotherapy

Chemotherapy can be given in the curative (as part of chemoradiotherapy, see above), adjuvant, or palliative setting.

Postoperative adjuvant chemotherapy can be offered to patients with NSCLC in stage IB or higher, and has been demonstrated to increase 5-year survival by about 5%[1, 107, 108]. For optimal effect, adjuvant chemotherapy should start within 8 weeks after surgery and be given as a combination of platinum and vinorelbine in four 21-day cycles, which is the most well-documented adjuvant regimen[107]. Chemotherapy can also be offered in the neoadjuvant setting, alone or in combination with radiotherapy, but this is less well documented than post-operative adjuvant chemotherapy[1, 109].

As first-line treatment in the palliative setting, chemotherapy can be offered to patients who are not candidates for targeted therapy. A combination of immunotherapy and chemotherapy is then preferred over only chemotherapy and has become the first choice for the majority of patients, although single immunotherapy may be an alternative for patients where PD-L1 is expressed in more than 50% of the tumor cells. For patients where immunotherapy is contraindicated, conventional chemotherapy is still an option.

The standard chemotherapy regimen in first-line is typically four cycles of a platinum-based doublet. For patients with weaker general condition or comorbidities, single agent chemotherapy is sometimes offered. The regimen also depends on histology as, e.g., platinum-based combinations with pemetrexed are more beneficial in tumors with non-squamous histology[110].

Common side effects of chemotherapy include bone marrow suppression leading to anemia, neutropenia, lymphopenia, and thrombocytopenia, gastro-intestinal side effects such as diarrhea and nausea, and fatigue.

Immunotherapy

Tumor cells use several mechanisms to evade the immune system. Programmed cell death protein 1 (PD-1) is a transmembrane receptor expressed by T cells. Activation by its ligand, PD-L1, results in inhibition of T cell responses. PD-1 can be expressed on tumor cells, resulting in immune evasion. Drugs targeting this pathway (either PD-1 or PD-L1) are known as immune checkpoint inhibitors, and by blocking these inhibitory signals the immune system can destroy cancer cells. Immunohistochemical expression of PD-L1 in tumors is currently being used as a predictive marker for response to immunotherapy.

Immunotherapy can be given in the curative or palliative setting. In the curative setting, immunotherapy can be offered as adjuvant treatment to patients who have been treated with chemoradiotherapy, who have not progressed during treatment, and if > 1% of the tumor cells are positive for PD-L1 expression.

In the palliative setting, immunotherapy in combination with chemotherapy is currently offered as first-line treatment to patients with a good performance status and without contraindications such as certain autoimmune conditions. For patients where > 50% of the tumor cells express PD-L1, a single immunotherapy agent might be considered in first line.

Side effects from immunotherapy are immune-related and include, e.g., pneumonitis, colitis, thyroiditis, and hepatitis. Although often manageable with corticosteroids, these side effects are potentially severe, and these treatment regimens thus require thorough toxicity monitoring.

Targeted Therapies

Targeted therapies are either antibodies or so-called small molecules. These drugs exert their anticancer effect thanks to the tumor's dependence of the targeted signaling pathway within the tumor cells. Whereas the monoclonal antibodies may target either the ligand or the extracellular domain of a transmembrane receptor, small molecule kinase inhibitors act intracellular and inhibit the kinase function of signaling proteins.

At the turn of the millennium, the first TKI became approved, namely imatinib for the treatment of patients with chronic myeloid leukemia. Imatinib targets cells that harbor the Philadelphia translocation, where the gene breakpoint cluster region (*BCR*) is fused to the v-abl Abelson murine leukemia viral oncogene homolog 1 (*ABL1*) gene. This fusion gene encodes a chimeric protein with a constitutively active tyrosine kinase domain, which is essential in the pathogenesis of the disease. Imatinib also blocks the activity of the receptor tyrosine kinase KIT, which was studied in Study I. Imatinib is used in the treatment of gastrointestinal stromal

tumors, which commonly express KIT protein and frequently harbor activating mutations in *KIT*.

There are several oncogene drivers that are commonly altered in lung cancer. For some of the alterations, targeted therapies are now well established in the clinical management of lung cancer patients. In addition, there are other driver alterations where drugs are currently entering the clinical setting or are being explored in clinical trials. The overview presented below tries to summarize the current status of targeted therapies in NSCLC, but, given the rapid advances in this field, the landscape of targeted therapies in lung cancer is constantly evolving.

EGFR

Drugs targeting EGFR were the first targeted therapies approved in lung cancer. First-generation EGFR TKIs like erlotinib and gefitinib, reversibly bind to the ATP binding site in the tyrosine kinase domain, thereby blocking ATP from binding and hindering the ATP-mediated phosphorylation and downstream signal transduction. Second-generation EGFR TKIs like afatinib and dacomitinib, irreversibly bind to the ATP binding site.

In 2004, it was discovered that patients harboring certain activating mutations in *EGFR* responded to treatment with EGFR TKI[111]. These typically include point mutations in exon 21 (L858R) and small deletions in exon 19. Subsequent trials demonstrated a longer progression-free survival (PFS) for patients treated with EGFR TKIs than patients treated with chemotherapy in the first-line setting for patients with advanced *EGFR* mutation-positive disease[112-114]. The drugs are generally well tolerated but common side-effects include adverse skin reactions and gastrointestinal side effects.

As with all oncological drugs, resistance to EGFR TKIs almost inevitably occurs at some point during treatment. A broad variety of resistance mechanisms has been described[115]. The most common of these is a secondary mutation in the *EGFR* gene, the T790M mutation, which is found in about 50% of the resistant cases during treatment with first- or second-generation EGFR TKIs. Other resistance mechanisms include activation of bypass pathways through amplification of *HER2* and *MET*, or mutations in genes that encode downstream mediators such as *KRAS*, *BRAF*, and *PIK3CA*[116]. Furthermore, histologic transformation by conversion to small-cell lung cancer or epithelial to mesenchymal transition has been observed[116].

To overcome T790M-mediated resistance, a third-generation TKI, osimertinib, was developed. Osimertinib irreversibly and with high affinity binds to certain mutant forms of the EGFR protein, including T790M-mutant EGFR[117]. Trials have demonstrated an improved PFS with osimertinib compared to first-generation TKIs[118]. However, various mechanisms of resistance to osimertinib have been reported. Many of these are similar to those seen among patients treated with first- and second-generation TKIs, such as activation of bypass pathways, *MET*

amplification and histologic transformation, but also loss of T790M and secondary *EGFR* mutations are observed[119].

Today, targeted therapies have an important role in the treatment of patients with advanced lung cancer. Second- or third-generation EGFR TKIs are recommended as first-line treatment for patients with tumors harboring the classical sensitizing mutations. In addition, the role of targeted therapies is currently being expanded to surgically treated patients, as osimertinib has demonstrated improved disease-free survival in the adjuvant setting and received approval as part of the adjuvant treatment of surgically resected NSCLC[120, 121]. Furthermore, there are also trials of EGFR-TKI therapy in the neoadjuvant setting, although larger studies are needed[122, 123]. In conclusion, since the introduction of EGFR TKIs into the clinical management of lung cancer patients, there has been a constant evolution with new and improved drugs and new settings have been explored. Therefore, it is highly probable that the role of EGFR TKIs will continue to evolve in the forthcoming years.

ALK

Patients whose tumors harbor *ALK* rearrangements are offered ALK TKIs, which inhibit aberrant ALK signaling by interacting with the ATP binding site. Crizotinib was the first ALK TKI to demonstrate a longer PFS compared to chemotherapy as first-line treatment for advanced NSCLC with *ALK* rearrangements[124]. Ceritinib, alectinib, and brigatinib are more recent ALK TKIs that have demonstrated an improved PFS when compared either to chemotherapy or to crizotinib. For alectinib and brigatinib, a much-improved intracranial activity has been observed and therefore these drugs are now recommended for first-line treatment. As previously mentioned, studies suggest that the *ALK* fusion partner influences drug sensitivity, but currently there are no clinical guidelines on how to address this issue[76]. Common side effects from treatment with ALK TKIs include gastrointestinal symptoms such as nausea, vomiting and diarrhea, as well as fatigue, increased liver enzymes, and myalgia[125].

As with EGFR TKIs, there are several mechanisms of resistance that can develop during treatment. The resistance mechanisms are complex and involve secondary *ALK* mutations, *ALK* amplifications and activation of bypass signaling pathways[126].

Other alterations that are targetable in the first-line setting

Other molecular alterations with approved drugs for the first-line setting include *BRAF* V600 mutations and rearrangements involving *ROS1* or *NTRK*.

Patents harboring *BRAF* V600 mutations can be offered a combination of the BRAF inhibitor dabrafenib and the MEK inhibitor trametinib based on results from two phase II trials[127, 128]. Because BRAF kinases activates MEK kinases, this

combination acts on two levels in the MAPK pathway, thereby avoiding resistance through bypass MEK signaling.

As ALK and ROS1 are related tyrosine kinases, some ALK inhibitors also inhibit ROS1. Currently, crizotinib and entrectinib are approved in the first-line setting for patients displaying *ROS1* gene fusions[1]. Entrectinib also inhibits NTRK, and is currently recommended for patients with tumors harboring *NTRK* rearrangements[1].

Other targetable or potentially targetable alterations

Insertions in exon 20 of *EGFR* are heterogenous, and typically confer resistance to EGFR TKIs[129, 130]. Currently, there are two drugs that target EGFR with exon 20 insertions that have received approval from the Food and Drug Administration (FDA). Mobocertinib is a TKI specifically designed for tumors harboring *EGFR* exon 20 insertions and has demonstrated promising results in a phase I/II trial[131]. Amivantanab is an antibody targeting EGFR and MET, and has demonstrated promising results in a phase I trial[132]. Mobocertinib or amivantanab as first-line therapy are currently being investigated in phase III trials

For *RET* rearrangements, the specific RET TKIs, selpercatinib and pralsetinib, have demonstrated promising results, and both drugs have received FDA approval[133, 134].

Crizotinib, which also inhibits ALK and ROS1, have demonstrated efficacy for treatment of tumors harboring *MET* exon 14 skipping events[135]. Capmatinib and tepotinib are selective MET TKIs that have demonstrated promising results in clinical trials and have received FDA approval[136, 137].

Since *NRG1* encode a ligand for the HER3 receptor, which upon activation forms homodimers or heterodimers with other activated receptors in the epidermal growth factor receptor family, gene fusions involving *NRG1* can be targeted by inhibiting HER3 through the use of anti-HER3 antibodies or afatinib, a pan-HER inhibitor[138].

It has proven to be difficult to target KRAS, mainly because the protein has a high affinity for guanosine triphosphate (GTP) and a lack of deep pockets where small molecule inhibitors can bind. However, the G12C mutation creates a mutation-specific site on the KRAS protein where inhibitors can bind, and therefore several specific G12C inhibitors have been developed and tested in clinical trials[139]. Sotorasib, a selective G12C inhibitor, has demonstrated promising results and was the first to receive FDA approval[140]. However, more than 50% of the *KRAS* mutations in NSCLC are other than the G12C, but so far, no direct inhibitors of non-G12C *KRAS* mutation-positive tumors are available[65, 141, 142]. Selective KRAS G12D inhibitors have been developed but not yet reached clinical trials. Another approach under current investigation is to combine several therapies that target pathways downstream of KRAS[142].

Approximately half of *BRAF*-mutated lung tumors harbor *BRAF* mutations other than the now targetable V600 mutations, described above[66, 67, 143]. Pre-clinical

models suggest that non-V600 *BRAF*-mutated tumors are primary resistant to BRAF inhibitors[144]. So far, there are no targeted treatments available for tumors harboring non-V600 mutations. However, the effect of pan-RAF-inhibitors is explored in *BRAF*-mutated NSCLC and new generations of BRAF inhibitors are under development[145].

Alterations in *HER2* include amplification, overexpression, and mutations. Several different strategies to target HER2 have been explored, such as tyrosine kinase inhibitors, antibodies, and antibody drug conjugates[146]. However, results have been conflicting, and tumors with *HER2* alterations are a heterogenous group, and there is a lack of consensus regarding subgroup definitions, which could further complicate results from studies exploring the potential role of *HER2* alterations in lung cancer.

The importance of *PIK3CA* mutations have also been explored in lung cancer. Previous studies have demonstrated a high frequency of co-existing alterations, particularly *EGFR* and *KRAS* mutations, among cases harboring *PIK3CA* mutations[69, 147, 148]. PI3K-inhibitors have not yet performed optimally in the clinical setting, and the role of *PIK3CA* mutations as a potential target for therapies is debated[149].

Other targeted treatments

Angiogenesis inhibitors are another type of targeted treatment used in lung cancer. For patients with non-squamous histology who are treated with chemotherapy (primarily carboplatin and paclitaxel), addition of a monoclonal antibody blocking the vascular endothelial growth factor (VEGF) can be offered.

Follow-up and Prognosis

There are few studies on how the follow-up and surveillance of patients who have undergone curative-intent therapy should optimally be organized[150, 151]. The frequency, duration and content of follow-up visits are based on guidelines, and aim to identify side effects from the treatment, recurrences, second primary tumors, and to rehabilitate the patient[1]. Most of the recurrences occur within the first two years, but they can develop as late as 10 years or more after treatment[152, 153]. Many studies aim to identify patients with a high risk of relapse, but currently there are no markers in clinical use other than tumor stage to help identify these individuals. In Sweden, follow-up visits (including regular CT-scans and clinical examinations) are recommended for at least three years after resected stage IA, while patients with later-stage disease are followed for five years after surgery.

The prognosis in lung cancer is heavily dependent on tumor stage according to the TNM classification, described above, which is a well-established prognostic factor

and clinically used to guide choice of therapy. The 5-year survival rate is around 60% for patients diagnosed in stage I, while it is below 5% for stage IV patients[27], as illustrated in Figure 8.

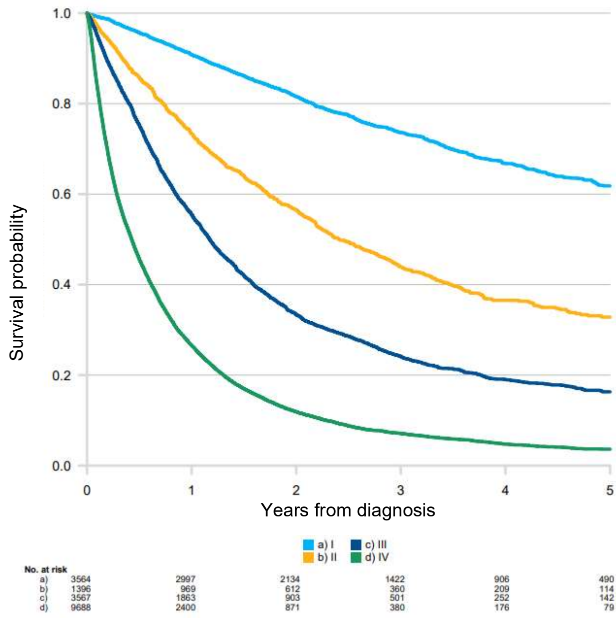


Figure 8. The 5-year survival rates in NSCLC for stages I, II, III and IV.
Reprinted with permission from "Nationell lungcancerrapport för diagnosår 2013-2017"[27].

Even though tumor stage reflects the degree of tumor burden on the patient, it does not fully reflect the general condition of the patient. Lung cancer patients, especially those with a history of heavy smoking, often have cardiovascular or pulmonary comorbidities that affect the patients' ability to cope with demanding treatments. Hence, patient performance status is an important factor in the decision-making regarding the treatment strategy. It is also an important prognostic factor that has been evaluated in several studies[85].

Tumor stage and patient performance status are the most well-established and clinically used prognostic factors in lung cancer, and these well reflect the prognosis. However, there is still heterogeneity in the outcome of patients, and more prognostic markers are needed for additional stratification to guide choice of treatment and follow-up. Many studies have aimed to identify such markers, which can be patient- or tumor-related factors. According to a systematic overview from 2002, 887 articles on prognostic factors in NSCLC were published between 1990 and 2001, and those papers identified 169 prognostic factors[86]. Two decades have

now passed since that publication, and the number of potential prognostic factors has, with all certainty, been further expanded.

Examples of patient-related prognostic factors that have frequently been explored are gender, weight loss, smoking history and age[85]. Female sex, less weight loss at baseline and absence of a history of heavy smoking have been associated with a better outcome in some studies. For age, results have been inconsistent, and few studies have reported significant findings.

Similarly, examples of tumor-related prognostic factors that have frequently been explored are histology, growth pattern, expression of protein markers, and genetic markers, e.g., gene expression patterns, methylation status, and mutations.

Of the different histological subtypes, AC have been associated with the best outcome[85]. As previously mentioned, adenocarcinomas are further subclassified based on growth patterns into subtypes, and these subtypes have been found to correlate with prognosis, where a predominant lepidic growth pattern has been associated with better outcome and micropapillary or solid growth pattern with worse outcome[35, 154, 155].

There are many studies exploring the prognostic value of single protein biomarkers assessed by IHC. Despite a plethora of such studies, there are no IHC-based prognostic markers in clinical use today[156]. As discussed in the methods section, there are several advantages with IHC, but the method also suffers from drawbacks. For many potential prognostic markers assessed by IHC, results have been inconsistent across studies, often due to a lack of standardization across studies[157]. Furthermore, the independent prognostic value and clinical benefit of the potential marker is rarely validated in independent studies.

Several different types of genetic prognostic markers have been studied. Gene expression profiling can be used for risk stratification by dividing tumors into different gene expression phenotypes. In breast cancer, four main intrinsic subtypes have been identified (luminal A, luminal B, HER2-enriched and basal-like), but no gene expression phenotypes have been established in lung cancer[158, 159]. Other proposed genetic prognostic markers are methylation patterns, where different epigenetic subgroups have been found to be associated with outcome[160], and gene mutations, e.g., mutations in *TP53*[161].

While a prognostic factor has an impact on patient outcome regardless of the treatment, a predictive factor predicts response to a specific treatment. For example, alterations in targetable genes such as *EGFR* mutations and *ALK* fusions are used for predicting response to treatments that target these alterations. A factor can be both prognostic and predictive. In lung cancer, *EGFR* mutations and PD-L1 status has been suggested to be both prognostic and predictive, although the prognostic role of both markers remains inconclusive[162, 163].

The prognosis and treatment of lung cancer patients has improved in the past decades, but further advancement is needed. A deeper understanding of tumor biology, earlier detection, refined tools for prognostication, and additional improvements in treatment will hopefully lead to further improvements in the survival rates in the coming decades.

Lung Cancer in Never Smokers

A never smoker is generally defined as someone who has smoked less than 100 cigarettes in his or her lifetime. If regarded as a separate entity, lung cancer in never smokers is estimated to be the seventh leading cause of cancer-related mortality worldwide[9]. Indeed, studies have demonstrated major differences in the epidemiology, histology and tumor biology of lung cancer between smokers and never smokers. The proportion of never smokers among lung cancer patients vary in different populations, ranging from over 50% in Southeast Asian women to only around 5% in men in Western studies[9, 164]. However, the prevalence of smokers in a population affect the proportion of never smokers, which is, in part, reflected in these numbers. Furthermore, studies suggest that among smokers there is a subset of patients with smoking-independent lung cancer, and in a population where the smoking prevalence is high, patients are more likely to be biased by the smoking status[165, 166].

Both in Sweden and worldwide, more never-smoking women than men are diagnosed with lung cancer[9, 24, 27]. This could in part be explained by differences in the smoking behavior between men and women, also across different regions, and globally there are more never-smoking women than men, which could in part account for this finding. Nevertheless, it is unclear if there are gender-dependent, or regional, differences in the susceptibility or exposure to risk factors for lung cancer other than smoking.

Age is one of the most studied risk factors for cancer, and the incidence of cancer increases with age[167]. Reginal differences in age at diagnosis for never smokers and smokers have been reported. Studies from Asian countries have reported a younger age at diagnosis in never smokers, while studies from the United States and Europe have demonstrated the opposite, i.e., an older age at diagnosis in never smokers than in smokers[9]. In a previous review of the National Lung Cancer Register in Sweden, never smokers were older at the time of diagnosis compared to former and current smokers[168]. At the same time, in that study, a higher proportion of never smokers were younger than 50 years compared to former and current smokers. The latter finding is in line with previous studies demonstrating that among younger lung cancer patients, the proportion of never smokers is greater than in older lung cancer

patients[169, 170]. Previous studies have suggested a hereditary component in the etiology of smoking-independent lung cancer among young patients, as family history of lung cancer is associated with increased risk for early-onset lung cancer[171].

Adenocarcinoma is the histological subtype that is the least associated with smoking and is the far most common type among never smokers, although never smokers can be diagnosed with any histological subtype[172, 173].

Lung tumors in never smokers often harbor genomic and epigenomic alterations that differ from those in smokers. Due to the many carcinogenic substances in cigarette smoke, tumors from smokers harbor more mutations than never smokers[56]. Furthermore, tumors in never smokers more often harbor oncogene driver alterations, some of which are today targetable with specific drugs. For instance, activating mutations in the *EGFR* or fusions involving *ALK* are frequently found among never smokers[174, 175].

Also, there are differences in the epigenetic and transcriptional profile between tumors arising in smokers and never smokers. Smokers more often display promoter hypermethylation in some specific tumor suppressor genes than never smokers[176]. Furthermore, gene expression profiles in never smokers differ from that in smokers[165, 177]. In a previous study exploring different lung cancer subgroups based on DNA methylation patterns, one of the subgroups was enriched for never smokers and also contained a subset of the smokers[160]. Similar findings have been observed based on gene expression patterns, with subgroups enriched for tumors from never smokers, together with a subset of tumors from smokers[165, 166]. Taken together, these findings imply that lung tumors in never smokers, together with a subset of tumors from smokers, have a similar tumor biology with distinct molecular features, and may represent a distinct entity.

Aims

Overall Aim

The overall aim of this thesis was to deepen the understanding of molecular alterations in lung cancer. In the long term, the studies aim to improve and individualize treatment prediction and disease prognostication.

Specific Aims

Studies I and II focus on alterations in potential targets for therapy.

- Study I aims to explore alterations in the receptor tyrosine kinase KIT and its ligand KITLG on the DNA, RNA and protein levels.
- Study II aims to investigate the prevalence of alterations in targetable or potentially targetable genes among never smokers in a nationwide and population-based cohort.

Studies III and IV focus on prognostic markers.

- Study III aims to examine the prognostic ability of RBM3 in lung cancer.
- Study IV aims to test a gene expression-based, multi-cohort strategy to identify genes with prognostic potential.

Materials

Patient Cohorts

Tumor specimens and clinical data from four patient cohorts were analyzed in the four studies included in this thesis. The cohorts are based on surgically obtained tumor specimens of primary lung carcinoma, and are summarized in Table 4.

Table 4. Characteristics of patient cohorts included in Studies I-IV.

	Study I	**Study II**	**Study III**	**Study IV**
Cohort	The ”Lund Lung Cancer Cohort” (a retrospective cohort from Lund University Hospital)	The “SMIL” cohort (a nationwide and population-based study consisting of tumors from never smokers)	- Cohort I: The ”Southern Swedish Lung Cancer Study” - Cohort II: The “Uppsala NSCLC II cohort”	- Cohort I: The ”Southern Swedish Lung Cancer Study” - Cohort II: The “Uppsala NSCLC II cohort”
Number of samples included in the study	72	431	Cohort I: 213 Cohort II: 306	Cohort I: 213 Cohort II: 194 (AC)
Period of surgical treatment for patients included in the study	1989 – 2007	2005 – 2014	Cohort I: 2005 – 2011 Cohort II: 2006 –2010	Cohort I: 2005 – 2011 Cohort II: 2006 –2010
Histology	- 47 AC - 13 SqCC - 9 LCNEC - 3 SCLC	- 391 AC - 17 SqCC - 7 AdSq - 3 LCC - 7 SC - 3 LCNEC - 3 combined SCLC	Cohort I: - 131 AC - 69 SqCC - 2 AdSq - 1 LCC - 2 SC - 8 LCNEC Cohort II: - 194 AC - 91 SqCC - 5 AdSq - 5 LCC - 2 SC - 9 LCNEC	Cohort I: - 131 AC - 69 SqCC - 2 AdSq - 1 LCC - 2 SC - 8 LCNEC Cohort II: - 194 AC
Tumor materials used in the study	Surgically resected tumor specimens - FFPE - Fresh frozen	Surgically resected tumor specimens - FFPE	Surgically resected tumor specimens - FFPE - Fresh frozen (cohort II)	Surgically resected tumor specimens - FFPE - Fresh frozen

Abbreviations: SMIL = Swedish Molecular Initiative agains Lung cancer, AC = adenocarcinoma, SqCC = squamous cell carcinoma, AdSq = adenosquamous carcinoma, LCC = large cell carcinoma, SC = sarcomatoid carcinoma, NSCLC = non-small cell lung cancer, LCNEC = large cell neuroendocrine carcinoma, SCLC = small cell lung cancer, FFPE = formalin-fixed paraffin embedded.

Study I

Study I was conducted on 72 surgically obtained specimens of primary NSCLC of different histological subtypes (47 AC, 13 SqCC, nine LCNEC, and three SCLC). The patients were surgically treated between 1989 and 2007 and retrospectively included in the study, based on the availability of fresh frozen tissue. Histopathological slides for the cases were reviewed by a thoracic pathologist, and the diagnoses were updated in accordance with guidelines current at the time of the study (WHO 2004 and TNM 7th edition)[88, 178]. Both formalin-fixed paraffin embedded (FFPE) samples and fresh frozen tissue were used. Data collected from patients' medical charts and from registers included age, gender, smoking status, histology, stage, overall survival (OS), and cancer specific survival (CSS). Patient characteristics and other clinicopathological data are further presented in Study I. The study was approved by the Regional Ethical Review Board in Lund, Sweden (Registration numbers 2004/762 and 2008/702).

Study II

Study II was based on a nationwide, population-based cohort of lung tumors from never smokers. The cohort was collected through the "Swedish Molecular Initiative against Lung cancer", SMIL, which is a national lung cancer research network including experimental researchers, pulmonologists, oncologists, surgeons, and pathologists.

Through the National Lung Cancer Registry (NLCR), all patients in Sweden who were registered as never smokers and had undergone lung cancer surgery between 2005 and 2014 were identified. In total, 545 individuals were identified, and tumor tissue and clinical data from these patients were collected through the SMIL network.

FFPE tumor tissue was obtained from regional clinical biobanks. Histopathological slides for identified cases were reviewed by a thoracic pathologist and the diagnoses were updated in accordance with the 2015 WHO classification and TNM 7th edition[40, 88].

Clinical data were collected from patients' medical charts and included variables such as smoking history, baseline patient characteristics, lung cancer treatments, time to recurrence and subsequent treatments. Data on OS were obtained from the Swedish Cancer Registry. The study was conducted in adherence with the Declaration of Helsinki and approved by the Regional Ethical Review Boards in Lund (Dnr 2014/546).

After reviewing the histopathological slides and patients' medical charts, 116 patients were excluded, as schematically illustrated in Figure 9. In total, 36 patients were found to have a history of smoking. Two patients were excluded due to having developed lung cancer in transplanted lungs, and smoking status of the donors was unknown. Furthermore, 17 patients never underwent surgery, and an additional 41 cases were missing in the archives. After histopathological review, 20 patients were excluded due to a revised diagnosis of either pulmonary carcinoid (n=4), salivary gland tumor (n=9), or a malignancy with other origin than the lungs (n=7). Hence, a total of 429 patients were included, whereof two patients had synchronous tumors, thus resulting in 431 tumors included in the analyses.

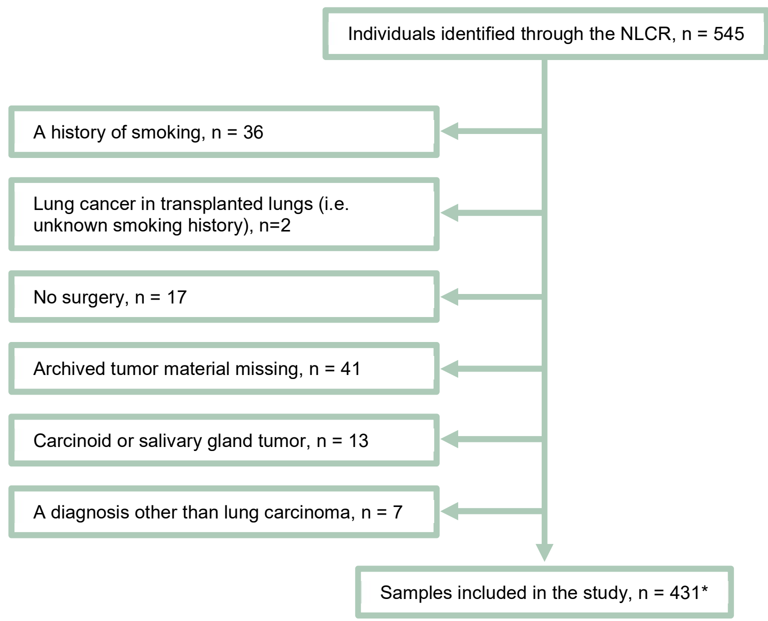


Figure 9. Scematical illustration of patients excluded in Study II.
*Two patients had two synchronous tumors. Abbreviations: NLCR = National Lung Cancer Registry.

Studies III and IV

Studies III and IV were conducted on the same two cohorts, referred to as cohorts I and II in the studies.

Cohort I was based on our in-house biobank, the "Southern Swedish Lung Cancer Study"[179]. This biobank prospectively included patients with primary lung cancer who underwent surgical treatment at Skåne University Hospital, Lund, Sweden,

between 2005 and 2011. Cohort II was based on a biobank from Uppsala, Sweden; the "Uppsala NSCLC II cohort"[180, 181]. This biobank retrospectively included consecutive samples of primary NSCLC from patients that had been surgically treated at the University Hospital in Uppsala, Sweden, between 2006 and 2010. The histopathological slides for these cases were previously reviewed, and the diagnoses were updated in accordance with the 2015 WHO classification and TNM 7th edition.

In Studies III and IV, tissue microarrays (TMA) based on FFPE tissue were used for the IHC staining. In addition, for subsets of patients in both cohorts, fresh frozen tissue was available, enabling use of previously generated RNA expression data in both studies.

Baseline patient and disease characteristics, as well as data on treatment and follow up, were collected through patients' medical charts. Clinical data available included age, gender, smoking status, disease stage, adjuvant treatment, performance status (available for cohort II only), and follow-up data with regard to recurrences. For both cohorts, survival data were obtained from the Swedish Cancer Registry, consulted on June 26, 2018, for cohort I, and on March 29, 2019, for cohort II.

In Studies III and IV, cohort I was used as an IHC discovery cohort for identification of optimal cut-offs for classifying samples as having a low or a high expression of the studied markers. These cut-offs were then applied to cohort II, which served as the validation cohort.

External Data Sets

Several external microarray-based data sets were used in the studies included in this thesis to complement our own patient material. In Studies I and III, external data sets were used for validation of our findings. In Study IV, we used external gene expression data sets to generate potential prognostic markers. The external data sets used in the four studies are summarized in Table 5 and further presented in Studies I, III, and IV.

Table 5. Overview of external data sets used in the thesis.

	Study I	**Study III**	**Study IV**
External data sets	Gene copy number alterations: - Pooled data (in total 1,142 AC and 458 SqCC) Gene expression levels: - 2 data sets (in total 502 AC and 53 SqCC) - 6 data sets (in total 360 normal airway epithelial specimens and 107 normal lung specimens)	Gene expression levels: - 13 data sets (in total 2,087 AC) - 8 data sets (in total 899 SqCC)	Gene expression levels: - 6 data sets (in total 1,167 AC)

Abbreviations: AC = adenocarcinoma, SqCC = squamous cell carcinoma.

Methods

The main methods used in the four studies included in this thesis are summarized in Table 6.

Table 6. Main methods used in the four studies included in this thesis.

Method	Analysis	Study I	Study II	Study III	Study IV
Immunohistochemistry	Protein expression	X		X	X
Sanger sequencing	Mutational status	X			
Next generation sequencing	Mutational status		X		
Real-time qRT-PCR	RNA expression levels	X			
NanoString technology	Gene fusions		X		
Analysis of data generated from previous microarray-based studies	-Gene copy numbers -RNA expression levels	X X		 X	 X

Tumor Tissue Handling

Formalin-fixed paraffin embedded tissues were used in all four studies included in this thesis. In addition, data generated from previous studies based on fresh frozen tumor tissue were used in Studies I, II and IV.

In clinical routine, tumors are fixed in formalin and embedded in paraffin blocks to preserve the tissue and enable microscopic evaluation of thin tissue sections. First, the tumor is placed in formalin which will penetrate the tissue and stabilize it. Since the wax used in the final paraffin embedding is hydrophobic, the specimen must be dehydrated. This is achieved by immersing the specimen in a series of alcohol solutions, which will replace the water. Since alcohol and wax will not fully mix, the alcohol must be replaced by xylene, a process called clearing. Then, the specimen can be embedded in paraffin-wax.

Although this procedure will ensure preservation and enable the cutting of thin sections with preserved tissue morphology, it can damage the DNA, RNA, and proteins in the tissue, through cross-linking, degradation and modification. Thus, studies based on nucleic acids retrieved from FFPE tissue can be technically challenging. In contrast, nucleic acids extracted from fresh frozen tissue are generally of high quality. However, fresh frozen tissue is usually not collected in the clinical routine, and all fresh frozen tissue used in this thesis work was collected through research biobanks.

Tissue Microarrays

Tissue microarrays (TMA) consist of multiple tissue cores that are assembled in a paraffin block in a grid pattern, as illustrated in Figure 10. The TMAs are constructed by first identifying areas of interest in the FFPE blocks of the tumors that are to be included in the TMA. Next, small tissue cores (typically 1 mm in diameter) are removed from the tumor donor blocks and inserted in a receiver paraffin block, which can hold tissue cores from many patients. Usually, several cores are included from a case to address, to some extent, potential intra-tumor heterogeneity.

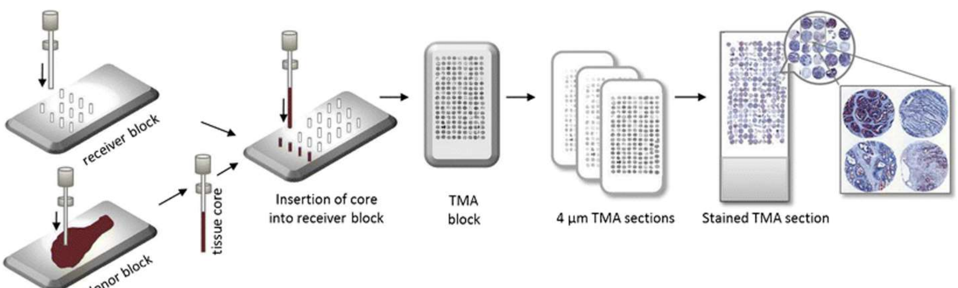


Figure 10. A tissue microarray (TMA) is constructed by removing small tissue cores from a donor block and inserting it into a receiver block, which can hold tissue cores from many patients. Multiple specimens can be stained and evaluated simultaneously in a TMA slide.
Reprinted with permission from Oberländer et al.[182], © 2014, with permission from Springer Nature

Multiple specimens can be stained and evaluated simultaneously in a TMA slide, thereby reducing both reagent costs and labor. Furthermore, because many cases are assembled in a few slides, the experimental conditions can be standardized. Also, the evaluation of the IHC staining is simplified because many cases can be evaluated in a relatively short time in a TMA slide.

A disadvantage of TMAs is that only a small portion of the tumor is analyzed, which increases the risk of not detecting intra-tumor heterogeneity which can be a potential source of error. In our studies, we sought to address this issue by including two or three cores from each tumor, and the cores were taken from tumor areas selected by an experienced lung pathologist.

Immunohistochemistry

For many decades, IHC has been a valuable tool in health care and research. Immunohistochemical staining of FFPE tissue was used in Studies I, III, and IV.

Immunohistochemical techniques use antibodies for detection of antigens in a tissue section. First, the FFPE specimen must be deparaffinized, typically by using xylene, and then rehydrated in graded washes of ethanol to water. As mentioned, the FFPE procedure can cause cross-linking of proteins, which might interfere with the antibodies' ability to recognize its epitopes. Therefore, a process called antigen retrieval is performed to unmask epitopes that are tied up in cross-links, typically with heat or digestive enzymes.

There are several IHC staining techniques. The EnVision™ System (DAKO, Hamburg, Germany), schematically presented in Figure 11, was used in Studies I, III, and IV. In this assay, the sections are first stained with a primary antibody that binds to specific epitopes on the antigen. Antibodies can be either monoclonal (binding to only one epitope) or polyclonal (binding to different epitopes on the same antigen). A polymer conjugate is added, which consists of multiple secondary antibodies bound to a dextran backbone, labelled with enzyme molecules, horseradish peroxidase (HRP). The secondary antibodies bind to the primary antibodies. Then, a chemical substrate, 3,3´-diaminobenzidine (DAB), is added that reacts with the HRP enzyme, which creates a brown precipitate that can be studied in a microscope. Finally, a counterstaining with hematoxylin stain can visualize cellular structures. If desirable, stained sections can be digitalized using a scanner, which enables evaluation by a computer and indefinite storage in electronic files.

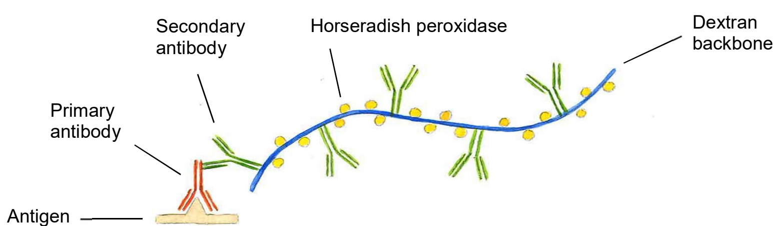


Figure 11. A schematic overview of the EnVision™ System. A primary antibody binds to specific epitopes on the antigen. A polymer conjugate is added, which consists of multiple secondary antibodies bound to a dextran backbone, labelled with enzyme molecules (horseradish peroxidase). The secondary antibodies bind to the primary antibodies. A chemical substrate that reacts with the HRP enzyme is added, which creates a brown precipitate that can be studied in a microscope.

Several factors can influence the results from an IHC experiment. These include the different steps during the tissue handling, the staining procedure and how assessment of the final staining is performed. For example, the tissue must be properly handled since both under- and over-fixation can affect the intensity of the staining[183]. Furthermore, since the sensitivity and specificity of the antibody will affect the results, proper validation of the antibody is important. In addition, the stained slide or scanned image is assessed by the human eye, which can result in potential differences in the interpretation. A cut-off for defining samples as having a positive or negative expression is often applied for statistical purposes, but there is usually no standardized method for setting such cut-offs for a given marker, and often there are discrepancies between studies. A lack of standardization across studies can cause inconsistent results and difficulties to compare studies.

Despite the potential disadvantages with IHC, there are many advantages. Formalin-fixed paraffin embedded tissue sections are used in the clinical routine and are therefore readily available. Immunohistochemistry is a fast, economic and well-known method, and many research studies are therefore conducted with IHC. Immunohistochemistry provides the ability to visualize the studied antigen, which enables quantification and *in situ* assessment of the cellular localization of the marker/markers. Furthermore, the expression of the antigen in different types of cells in the tissue section can be assessed.

Sequencing

In Study I, we used Sanger sequencing for studying mutations in the *KIT* gene in exons 9, 11, 13, and 17. In this method, the DNA region of interest is first amplified in a polymerase chain reaction (PCR). Then, a second type of PCR is performed where a mix of normal deoxynucleoside triphosphates (dNTPs) and fluorescently labelled dideoxynucleoside triphosphates (ddNTPs) are used. When DNA polymerase adds a fluorescently labelled ddNTP the extension is terminated, producing DNA strands (fragments of DNA) terminated at random lengths. These are separated by size through capillary electrophoresis. Finally, the sequence is determined by detecting the terminal ddNTP in each band using a laser to excite the florescent labels. A chromatogram visualizes the results, and double peaks indicate a deviation from the reference DNA, i.e., a mutation.

The Sanger method can only sequence short DNA fragments and one sample at a time. However, if the intention is to study only a short sequence on a limited number of samples, the Sanger method can be advantageous. Also, due to the limited data output, the data analysis is relatively straight-forward compared to more advanced sequencing methods.

In Study II, sequencing was performed using next generation sequencing (NGS). In NGS, multiple samples and many regions of the genome can be sequenced simultaneously. In short, the NGS assay can be divided into four steps as illustrated in Figure 12. First, a library preparation is performed. If using targeted sequencing, the regions of interest are selected. Sequencing adapters and index sequences (for sample identification) are added, and the library is amplified by PCR before being pooled with other samples treated equally. The next step is cluster amplification, where the pooled library is loaded onto a flow cell. The flow cell is covered with oligonuleotides complementary to the adapters and will therefore bind the fragments. Each individual fragment is then amplified, creating a cluster of identical fragments. Next is the sequencing step, in which fluorescently labelled nucleotides are added, and the bases are detected as they are incorporated. The final step is to align the generated data to a reference genome for identification of aberrations, before the data can be further analyzed.

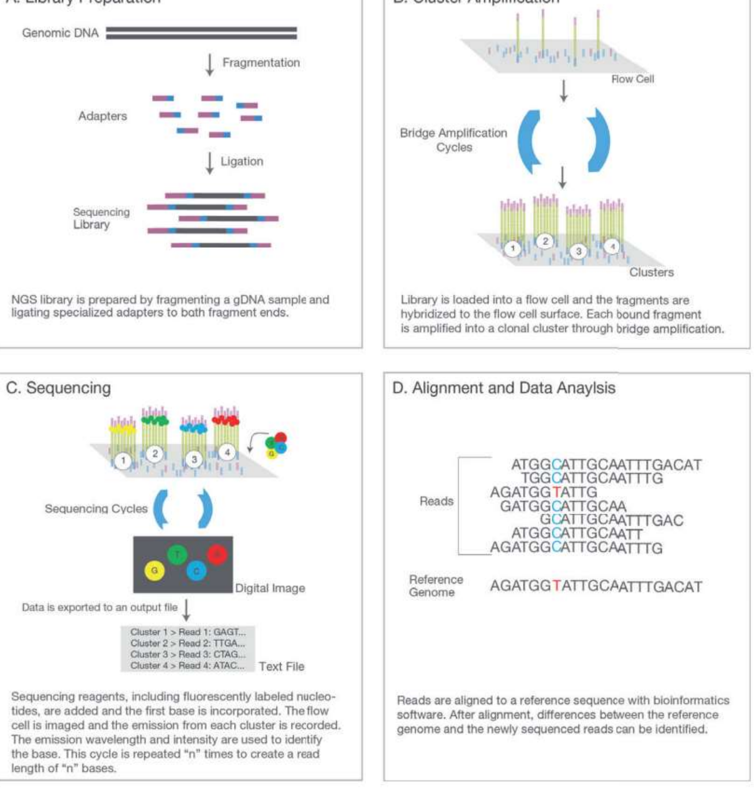


Figure 12. The next generation sequencing assay (Illumina) can be divided into four steps. The DNA is ligated with adapters (A). These adapters can bind to the surface of a flow cell, where each fragment is amplified into a cluster (B). Fluorescently labelled nucleotides are added, and the bases are detected as they are incorporated (C). The generated data is aligned to a reference genome for identification of aberrations (D).
Courtesy of Illumina, Inc.

In Study II, we used a commercially available exon-focused 26- and 15-gene panel. These panels are designed to detect mutations in clinically relevant genes and are optimized for FFPE samples in which the DNA can be heavily fragmented and modified from the FFPE-process.

Real-time Quantitative Reverse Transcription PCR

Real-time quantitative reverse transcription PCR (real-time qRT-PCR) combines reverse transcription (converting RNA into complementary DNA) and real-time monitoring of amplification by PCR, and can be used for quantifying gene expression levels.

In the real-time qRT-PCR assay, as used in Study I, RNA from the tumor is first converted into complementary DNA (cDNA) by a reverse transcriptase. An aliquot of the cDNA reaction is then used as a template in a real-time PCR, in which a fluorescent dye called SYBR Green is added. The fluorescent dye binds to double-stranded (ds) DNA and, once bound, fluoresces a green light. The fluorescent signal intensifies gradually as the amount of dsDNA increases with the number of PCR cycles. The cycle value when the fluorescent signal crosses a threshold is denoted as the cycle threshold (Ct) value. The Ct-value will depend on the amount of input cDNA and can thus be used for relative quantification of RNA in a sample. The PCR amplification curve can be visualized in an amplification plot, as illustrated in Figure 13.

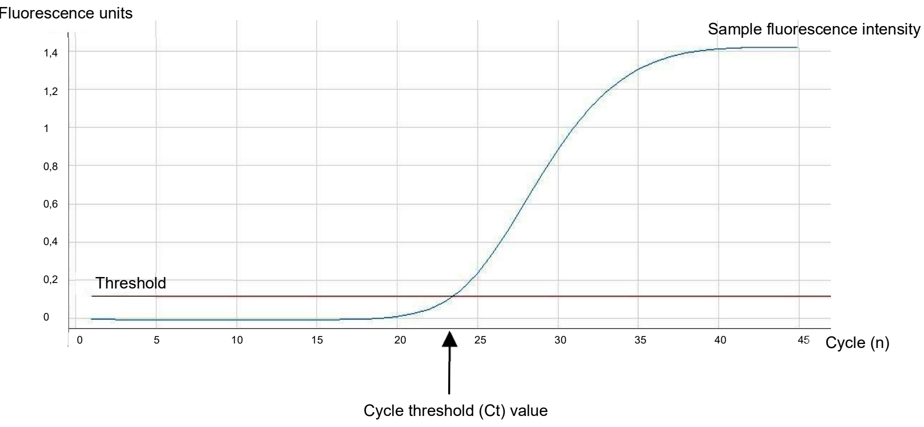


Figure 13. An amplification plot from a real-time qRT-PCR experiment. The cycle threshold (Ct) value represents the value (cycle) when the flourescent signal generated from the sample crosses the threshold.

In Study I, the expression level of the studied genes (*KIT* and *KITLG*) were for each sample reported as a relative ratio, calculated according to the Pfaffl method, summarized in Figure 14[184]. A no-template control (omitting RNA template) was added in each run and served as a control for contamination with exogenous nucleic acids. The relative ratio, calculated by the Pfaffl method, reports the expression level in a sample relative to the expression level in a calibrator. In Study I, we used a pool of RNA purified from normal lung tissue from six lung cancer patients as a calibrator. Also, to adjust for variations in PCR efficiency between different experiments, a serial dilution of this pool of RNA was used for determining PCR efficiency (E) in each run. All samples were run in duplicates and an average was calculated from the Ct values of the duplicate samples (ΔCt value).

The input amount of cDNA depends on the general RNA concentration in the sample and on the expression level of the studied gene. The RNA concentration in a sample is measured, and the input amount is adjusted in accordance. However, to adjust for errors and insufficient precision in sample quantification, endogenous reference genes are included in the assay. These are genes with expression levels that are assumed to be constant over time and different conditions across the sample set. In Study I, we used RNA polymerase II subunit A (*POLR2A*) and ribosomal RNA 18S (*rRNA 18S*). However, the assumption that the expression of these genes is constant can be a source of error[185]. Ideally, many housekeeping genes should be included in the assay to minimize the effect from variability in these, albeit the number of housekeeping genes included in the assay must be weighed against time and cost efficacy.

Advantages of real-time qRT-PCR includes time- and cost efficiency and sensitivity. Furthermore, only small amounts of tumor material are required. Disadvantages include the low multiplex possibility of different genes for the same sample. Thus, if expression of many genes is to be investigated the method may become cumbersome, expensive and time consuming.

$$\text{Ratio} = \frac{(E_{\text{target gene}})^{\Delta\text{Ct value target (calibrator – sample)}}}{(E_{\text{reference gene}})^{\Delta\text{Ct value reference gene (calibrator – sample)}}}$$

Figure 14. The Pfaffl method [184]. The ratio reports the expression level of the studied target gene in a sample relative to the expression level in a calibrator, also considering the PCR efficency (E). All samples were run in duplicates, and the ΔCt value represents an average of the cycle threshold (Ct) values of these duplicate samples.

NanoString

The NanoString technology[186] can be used as an assay for multiplexed gene expression analysis. A main advantage of NanoString is that no enzymatic steps, such as cDNA synthesis or PCR amplification is required. When synthesizing or amplifying DNA/RNA, editing errors or an imbalance in amplification efficiency can introduce experimental errors. Other advantages of NanoString are that it is a multiplexed analysis allowing hundreds of targets (genes) to be analyzed simultaneously, it has a short experimental turn-around time, and the technology can overcome the challenges with degraded RNA through specific probe design.

In Study II, NanoString was primarily used to detect gene fusions and *MET* exon 14 skipping events by analysis of tumor RNA. A gene fusion causes an imbalance in the expression of the 3′ end, where the tyrosine kinase domain is located, and the 5′ end of the gene. This imbalance in expression can be utilized for detection of gene fusions by counting and comparing the number of transcripts from both ends, as first reported by Lira et al.[187].

The NanoString assay, as used in Study II, had multiple probes per gene. As illustrated in Figure 15, imbalance probes that measure the expression of exons located 3' and 5' of the breakpoint of known gene fusions were included in the assay. Also, junction probes that span the junction of known fusions, e.g., the junction of exon 13 of *EML4* and exon 20 of *ALK*, were included to identify the exact fusion if present.

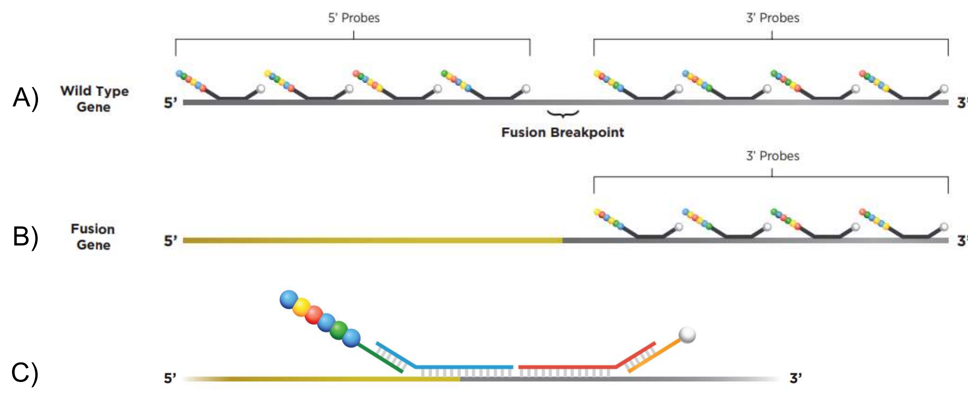


Figure 15. Illustration of the 5' and 3' imbalance design, with four probes to measure the expression of exons located 3' and four probes to measure the expression of exons located 5' of the breakpoint of the studied gene (A). An imbalance in the ratio of counts from 5' *versus* 3' probes is indicative of a gene fusion (B). Junction probes that span the junction of known fusions were included in the assay to identify the exact fusion if present (C).
Courtesy of NanoString Technologies, Inc.

Briefly, in the NanoString assay, the probes first hybridize to target RNA. Next, excess probes are washed away, and hybridized probes are then immobilized to a surface and subsequently counted by the NanoString instrument. In the gene fusion analysis, more counts from probes that bind to the 3′ end compared to the 5′ end indicates a fusion. If junction probes that span the junction of the fusion are included in the assay, these probes will also generate counts and indicate the exact fusion.

Imbalance between the 3' probes and the 5' probes of a gene, but no counts from any of the junction probes included in the assay indicate the presence of a fusion partner or a breakpoint other than the ones included in the assay design. In such cases, other methods, such as RNA sequencing, are needed to further elucidate the fusion partner. Based on such analyses, the NanoString assay can then be updated with a new junction probe for future analyses. Figure 16 illustrates the NanoString results from two samples harboring *RET* fusions.

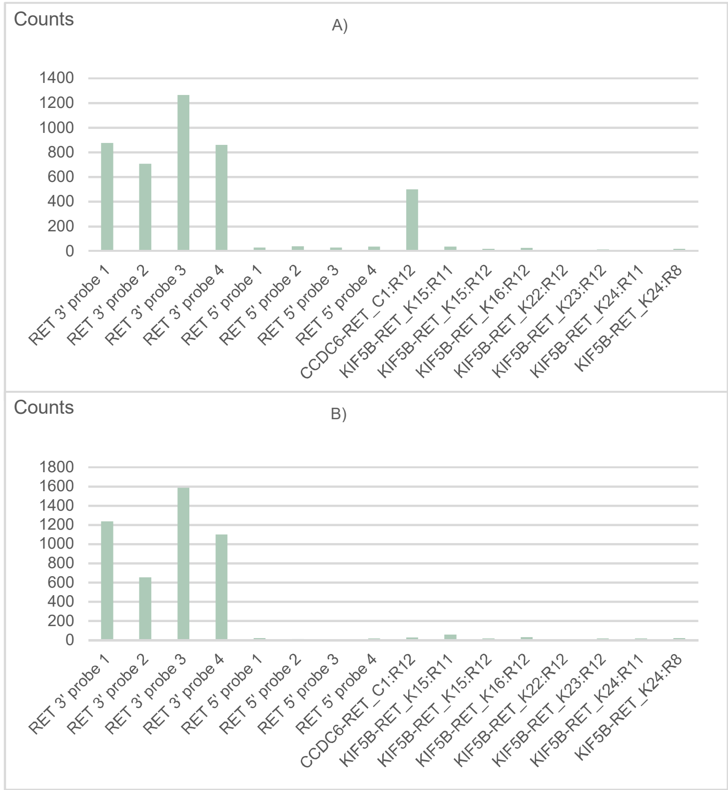


Figure 16. Results from the NanoString analysis demonstrating two sampes harboring *RET* fusions. A) More counts generated from 3' probes than 5' probes indicate the presence of a fusion. Also, counts generated from junction probes indicate the exact breakpoint and fusion partner. B) Imbalance between the 3' probes compared to the 5' probes, but no counts from any of the junction probes included in the assay, indicate the presence of a fusion with a partner or a breakpoint other than the ones included in the assay design.

Microarrays

Results from previously reported microarray-based studies were used in Studies I, III and IV. Briefly, a microarray is a chip on to which thousands of microscopic spots are attached in a grid pattern. Each spot contains many copies of a unique oligonucleotide, which are called probes, corresponding to a short fragment of, e.g., a gene. Microarrays can be designed to analyze both RNA levels (gene expression) and DNA alterations, e.g., gene copy number alterations.

For gene expression analysis, fluorescently labeled sample RNA, converted into cDNA, is flooded over the chip and complementary strands hybridize to matching probes. Unhybridized cDNA is washed away. The chip is scanned, during which a laser excites the labelled sample and the intensity of the fluorescent signal is measured. The data generated are then pre-processed to correct for, e.g., background noise due to unspecific binding of sample cDNA and technical variability such as differences across samples in staining efficiency. The intensity signal is typically log2-transformed to better present proportional changes, where a doubling of the intensity corresponds to one unit (+1) on the log2 scale.

Protocols for DNA analysis include similar steps but starts from genomic DNA. Array-based comparative genomic hybridization (aCGH) was used in Study I to identify gene copy number alterations of *KIT* and *KITLG*. In this assay, copy number alterations in a sample is compared with a normal sample by labeling each with a different fluorescent dye. The tumor sample and control sample competitively hybridize to the probes, allowing the samples to be compared by measuring the fluorescent signals. Gene copy number alterations in a sample can be visualized by a copy number plot, as illustrated in Figure 17.

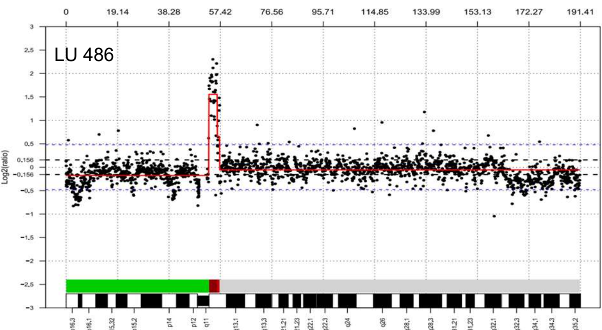


Figure 17. Results from an array-based comparative genomic hybridization experiment can be visualized by a copy number plot. The sample, LU 486, displays amplification of the 4q12 region, where the *KIT* gene is located. Each black dot represents individual probes. Positive log2 values indicate more DNA in the tumor sample than in the control sample, and negative values indicate less DNA in the tumor sample than in the control sample. By a segmentation algorithm, adjacent probes are connected into coherent segments, and the red line demonstrates copy number alterations in these segments. In the annotation track above the cytoband, green color corresponds to segments with gene copy number loss, red color corresponds to gains or amplifications, and gray color corresponds to unchanged gene copy numbers.

Statistical Methods

The main statistical methods used in this thesis work are listed in Table 7.

Table 7. Main statistical methods used in this thesis work.

Statistical Method	Purpose	Study I	Study III	Study IV
Fisher's exact test	Determine the association between categorical variables	X	X	X
Mann-Whitney U test/Wilcoxon rank sum test	Compare the distribution of a continuous variable between two groups	X	X	X
Kruskal-Wallis test	Compare the distribution of a continuous variable between more than two groups	X	X	
Log-rank test	Comparison of survival curves		X	X
Univariable Cox proportional hazards regression models	Estimate the effects of the studied factor on survival with a hazard ratio	X	X	X
Multivariable Cox proportional hazards regression models	Estimate the effects of the studied factor on survival with a hazard ratio, adjusted for other factors in the model		X	X

A study is often set out to investigate a hypothesis, such as finding a difference between two groups. However, there is always a probability that an observed difference occurs by chance alone and that there is, in fact, no difference between the groups. The assumption of no difference between the groups is called the null hypothesis. The probability value (*P*-value, or simply *P*) represents the probability of finding a difference between the groups when the null hypothesis is true. A *P*-value of 0.05 means that this probability is 5%. A threshold of when the probability is low enough for rejecting the null hypothesis is defined, and this value is usually, albeit arbitrary, set at 0.05. Therefore, in the studies included in this thesis, results with a *P*-value below 0.05 were considered to be statistically significant.

However, the magnitude or importance of the difference cannot be interpreted from the *P*-value, and the *P*-value is also affected by the sample size. A very small difference between two large groups can be statistically significant but may not be clinically meaningful. Also, a large and clinically relevant difference can be difficult to prove to be statistically significant if the number of observations is too low. A power calculation can be performed to estimate the number of patients that must be included in a study to obtain statistically significant results. In the studies included in this thesis, no power calculations were performed. This is mainly due to the cohorts in each study being set and limited to the material that was available through research biobanks.

The patient cohorts used in Studies I, III and IV were relatively small and only non-parametric tests were used. The advantages of non-parametric tests include less assumptions about the distribution of data, e.g., normal distribution, compared to parametrical tests. Drawbacks of non-parametrical tests are that they are less powerful than parametric tests and therefore have a higher risk of wrongly accepting

the null hypothesis, and that they use ranking instead of actual data which can be more difficult to interpret.

Overall survival was used as an endpoint in survival analyses in all four studies included in this thesis. Advantages of using OS include that it is easily collected through registers, rarely misinterpreted, and of clear benefit for the patient. A disadvantage is that it does not consider the cause of death, which is of importance when studying a population of high age, low performance status and comorbidities other than the studied disease.

Cancer specific survival was used in Study I, an endpoint that is dependent on the accuracy of the medical death records, which are not always filled out correctly. To address the issues with OS or CSS as clinical endpoints, we also used recurrence-free interval (RFI) as an endpoint in the two prognostic studies (Studies III and IV). The main drawback of RFI is that it is more complicated and demanding to determine, as it requires detailed reviewing of patients' medical charts by a trained person to avoid misinterpretations.

Study I

Fisher's exact test was used to analyze associations between categorical variables. The Mann-Whitney U test was used to compare the distribution of continuous variables between two groups. If more than two groups were included in the analysis, the Kruskal-Wallis test was used. For analyses of correlations between gene copy number alterations, RNA levels, and protein expression, Fisher's exact test (categorical variables) and Mann-Whitney U test (continuous variables) were used. Survival analyses were performed using Univariate Cox proportional hazards regression models.

Study II

No statistical analyses were performed in Study II.

Study III and IV

Fisher's exact test was used to analyze associations between categorical variables. The Mann-Whitney U test or the Kruskal-Wallis test were used to compare the distribution of continuous variables between groups. Kaplan-Meier plots with log-rank test were used to identify optimal cut-offs for classifying samples as having a low or a high expression of the markers. Furthermore, Kaplan-Meier plots with log-rank test were used for OS and RFI analyses. Cox proportional hazards regression models were used to further compare groups. Multivariable models were adjusted

for gender, age, stage (I, II, III, and IV), smoking (current, past, and never), adjuvant therapy, growth pattern (AC only), and patients' performance status (cohort II only).

Results and Discussion

The four studies included in this thesis jointly aimed to characterize lung cancer on the molecular level. By investigating alterations in oncogene drivers, Studies I and II focused on established or potential targets for therapy. Studies III and IV focused on prognostication by exploring potential prognostic markers. In addition, in Studies I, III, and IV, correlations between RNA levels and protein expression were explored.

Targets for Therapy

Therapies that target alterations in oncogene drivers have an important role in the clinical management of lung cancer patients. Given the rapid progression in this field, mapping of oncogene drivers implicated in the tumorigenesis of lung cancer is essential to our understanding of lung cancer biology and to the development of novel targeted therapies. In Study I, the receptor tyrosine kinase KIT and its ligand (KITLG) were studied on the DNA, RNA and protein levels in different subtypes of NSCLC. In Study II, alterations in targetable and potentially targetable genes were investigated in never smokers.

Study I

Upon stimulation by its ligand, KITLG, the receptor tyrosine kinase KIT activates signaling cascades, which involve the MAPK and PI3K pathways, leading to cell survival and proliferation[188]. In gastrointestinal stromal tumors, KIT is frequently expressed and is used as a diagnostic marker[189]. Furthermore, activating mutations in *KIT* are common in this disease and imatinib, a TKI which targets KIT, is the standard treatment for patients with advanced disease[189]. In lung cancer, *KIT* has been reported to be altered by different mechanisms such as gene copy number alterations, overexpression, and autocrine/paracrine stimulation by its ligand[190-192]. In this context, we sought to explore alterations in *KIT* and *KITLG* in lung cancer by mapping alterations at the DNA, RNA, and protein levels, and performed mutational analyses in a subset of the tumors.

The tumor material in Study I consisted of 72 surgical specimens of primary NSCLC (47 AC, 13 SqCC, nine LCNEC, and three SCLC). Gene copy numbers were investigated using aCGH, mRNA expression levels were assessed by real-time qRT-PCR, and protein expression was evaluated with IHC. Mutation status of *KIT* was investigated in the LCNEC samples using Sanger sequencing. For validation of our findings, external microarray data sets were used, where gene copy number alterations and RNA expression levels were investigated in 1,600 and 555 primary lung tumors, respectively.

Since NSCLC is a heterogeneous disease, we examined the role of *KIT* and *KITLG* in different histological subtypes. As illustrated in Figure 18, different types of alterations in *KIT* were identified in the histological subtypes, with SqCCs more commonly harboring gene copy number gains and high-level amplifications, while gene copy number loss was most frequently found among the LCNEC samples. However, the SqCC samples displayed the lowest expression levels of *KIT* mRNA, and the LCNECs displayed the highest levels (Kruskal-Wallis test, $P < 0.0001$). A similar pattern was noted for protein expression, as all the LCNEC samples were positive for KIT expression but none of the SqCC samples were positive (Fisher's exact test, $P < 0.0001$).

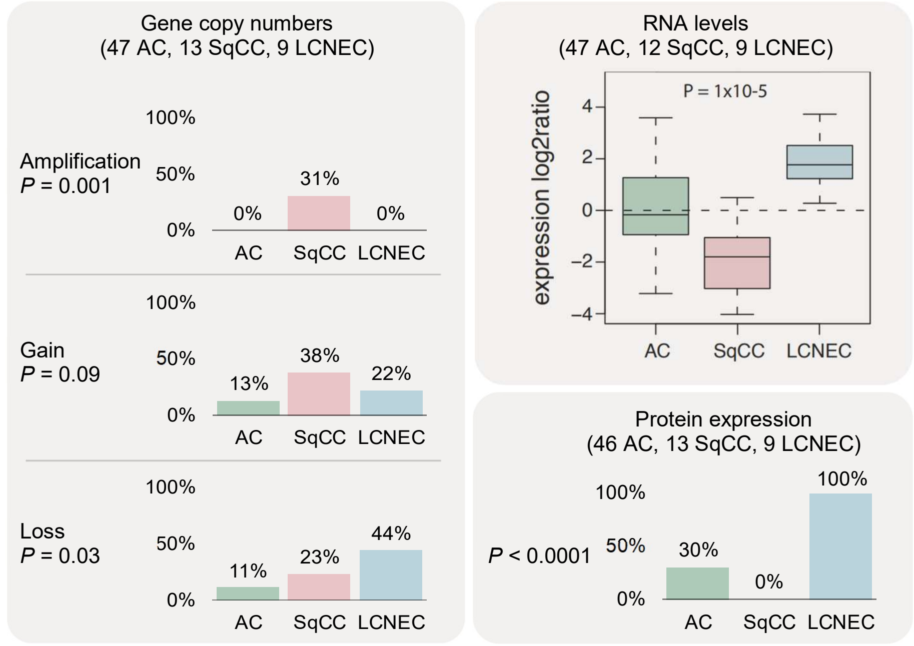


Figure 18. Amplifications and gains of KIT were more commonly found in SqCC, while gene copy number loss was most frequently found in LCNEC. The SqCC samples displayed the lowest RNA levels of KIT, and none were positive for KIT protein expression. The LCNECs displayed the highest RNA levels of KIT, and all were positive for KIT protein expression.

Abbreviations: AC = adenocarcinoma, SqCC = squamous cell carcinoma, LCNEC = large cell neuroendocrine carcinoma.

Thus, we found a poor correlation between *KIT* gene copy number alterations and expression of mRNA or protein (Mann-Whitney U test, Fisher's exact test, $P > 0.05$ all comparisons), but a significant correlation between mRNA levels and protein expression (Mann-Whitney U test, $P < 0.0001$). Previous studies have found similar results, with a poor correlation between *KIT* gene copy numbers and protein expression[190, 193, 194].

The reason for this poor correlation between *KIT* gene copy numbers and expression remains unclear, but the general correlations between gene copy number alterations, RNA levels and protein expression are complex and in many aspects unexplored[195, 196]. Regulation of gene expression occurs through several mechanisms and at multiple levels. Results from this study indicate that the expression of *KIT* mRNA and protein are well correlated but not driven by gene copy numbers, thus being possibly regulated at the epigenetic or transcriptional level. Indeed, promoter hypermethylation of *KIT* has been observed in breast tumors, which supports the role of an epigenetic regulation[197].

The 4q12 region, where the *KIT* gene is located, also harbors genes that encode other receptor tyrosine kinases, such as the platelet-derived growth factor receptor α (*PDGFRA*) and the vascular endothelial growth factor receptor-2 (*VEGFR2/KDR*), and amplification of these genes have been observed in several types of cancer[190, 193, 198]. Due to the poor correlation between gene copy numbers and expression, it is possible that other genes might be the targets of the increased gene copy numbers of the 4q12 region in SqCC and thus of oncogenic importance. However, further studies are needed to elucidate the importance and implications of 4q12 amplifications in lung cancer.

Consistent with results from other studies, the LCNEC samples displayed high expression levels of *KIT*[199-202]. To further explore potential *KIT* alterations in this subtype, mutation status of *KIT* (exons 9, 11, 13 and 17) in these tumors was investigated, but no mutations were found. This finding is consistent with previous studies which demonstrate a low frequency of activating *KIT* mutations in lung cancer[203, 204].

Similar to LCNEC, SCLC is classified as a neuroendocrine carcinoma and SCLC often display KIT protein expression[199]. The efficacy of the KIT inhibitor imatinib in SCLC has been evaluated in several phase II trials. However, imatinib alone or in combination with chemotherapy failed to demonstrate any activity on SCLC tumors that express KIT[205-209]. Thus, in lung cancer, the oncogenic importance of KIT and the value of targeting it have hitherto been difficult to prove. Possibly, protein expression is not an optimal biomarker for predicting response to KIT inhibitors, and perhaps other biomarkers, such as activating mutations, should be considered.

Since alterations in *KIT*, such as mutations and amplifications, have been described as a mechanism of acquired resistance to targeted therapies[210, 211], KIT inhibitors

may have a potential role for treatment of lung cancer patients in this setting. Furthermore, there are multi-target tyrosine kinase inhibitors, that inhibit KIT among others, that are currently approved in China as a third-line treatment for patients with advanced NSCLC[212].

We also studied the ligand of KIT, KITLG. No correlations or clear differences between the histological subtypes were observed for *KITLG* regarding gene copy numbers, mRNA levels or protein expression. Protein expression of KITLG was found in about 30% of all samples, regardless of histology. In total, only five samples demonstrated positive immunohistochemical staining for both KIT and KITLG, three of which were LCNECs. Since all LCNECs were positive for KIT protein expression, the LCNEC samples that were positive for KITLG were positive for both markers. Hence, we could not find much evidence for the importance of an autocrine loop involving KIT and KITLG, but further investigation among LCNEC might be warranted.

Among the AC cases, current smokers had the highest *KIT* mRNA levels and never smokers the lowest, both in our data set and in the external data set (Kruskal-Wallis test, $P = 0.05$ and $P = 0.01$, respectively). In a previous study, two main molecular subgroups of AC were defined by gene expression analyses in smokers and never smokers[165]. One of the subgroups, which contained only smokers, more often displayed KIT protein expression than the other subgroup, which contained all never smokers together with a subset of the smokers. These results suggest that, possibly, KIT expression is connected to certain molecular subtypes of AC, rather than directly related to the effects of smoking.

In a recent review, overexpression of KIT was suggested to be of prognostic importance in lung carcinoma[213]. However, we could not demonstrate any such associations in our own sample set or in the two external data sets used in the study. Furthermore, *KIT* alterations have been associated with secondary resistance to targeted therapies as discussed above, and therefore *KIT* alterations may implicate a worse prognosis within these subgroups.

The tumor material was a limitation in Study I. A relatively small number of samples of different histological subtypes were included, based on the availability of fresh frozen tumor tissue. The patients were surgically treated over a wide time span, ranging from 1989 to 2007. However, the diagnoses were updated in accordance with guidelines current at the time of the study. The limited number of samples, especially when performing subgroup analyses, could affect the results by reducing the power of the study. To compensate for this limitation, in part, large external datasets were used to validate some of the findings.

In summary, the results from Study I demonstrated distinct *KIT* alterations in different histological subtypes, with SqCC more often harboring gene copy number gains and LCNEC displaying higher expression levels. It also points out a poor

correlation between *KIT* gene copy number alterations and expression levels. KIT inhibitors have hitherto performed poorly in lung cancer patients, and there is a paucity of studies. However, considering the possible role of KIT in the development of resistance to other targeted therapies used in the first-line setting, there might still be a role for treatments targeting KIT in lung cancer.

Study II

In Study II, we performed a mapping of alterations in targetable or potentially targetable genes in a nationwide and population-based cohort consisting of over 400 lung tumors from never smokers.

Studies have demonstrated clinical, pathological, and molecular differences between lung tumors in smokers and never smokers, suggesting that lung cancer in never smokers represents a different entity[9]. Tumors from never smokers often harbor specific oncogene driver alterations, which generally are mutually exclusive, illustrating their functional importance in the carcinogenesis. Today, some of these oncogene drivers are targetable in the clinical praxis, and there are several drugs under development, targeting additional alterations. Given the increased awareness of lung cancer in never smokers, and the increasing opportunity to treat these patients with targeted therapies, we performed a mapping of oncogene driver alterations among never smokers, with the aim of obtaining an increased understanding of smoking-independent lung cancer biology and with implications for therapy response prediction.

The National Lung Cancer Registry (NLCR) is a Swedish quality register where information regarding patient characteristics, diagnostic work-up and patient management is collected[214]. It is compulsory to report newly detected cancer cases to the Swedish Cancer Register, and the NLCR includes about 97% of the patients reported to the Swedish Cancer Register, and thus the NLCR has a good coverage rate. Through the NLCR, we identified all patients that were registered as never smokers and had undergone surgery for lung cancer in Sweden during a ten-year period (2005 to 2014). As further detailed in the Materials section, 545 patients were identified in the NLCR, and after histopathological review and review of patients' medical charts, 431 tumors could be further analyzed in the study.

Patient characteristics and clinicopathological data are presented in Table 8. As expected, the most frequent histological subtype was AC (n = 391, 91%), although other histological subtypes were also present among the cases. The median age at surgery was 71 years, ranging from 25 to 86 years. In Sweden, the median age at diagnosis for all lung cancer patients is 69 years[1], and thus well in line with the median age of the patients in this study. In total, 32 (7%) of the patients in our study were younger than 50 years. The majority of patients had undergone curatively intended surgery, hence most tumors were in stage I (n = 274, 64%) or stage II (n =

76, 18%). The 5-year overall survival rate for all patients included in the analyses was 65%, which is in good agreement with the 5-year survival data after surgical treatment of lung cancer in Sweden[1].

Table 8. Patient characteristics and clinicopathological data of the SMIL cohort.

Patient age	
Median age at surgery	71 years
Min – max age at surgery	25-86 years
Patient gender	
Female	302 (70%)
Male	129 (30%)
Tumor histology	
Adenocarcinoma	391 (90.7%)
Squamous cell carcinoma	17 (3.9%)
Adenosquamous carcinoma	7 (1.6%)
Large cell carcinoma	3 (0.7%)
Pulmonary sarcomatoid carcinoma	7 (1.6%)
Large cell neuroendocrine carcinoma	3 (0.7%)
Combined small cell lung carcinoma	3 (0.7%)
Tumor stage	
0, carcinoma *in situ*	5 (1.2%)
IA	173 (40.1%)
IB	101 (23.4%)
IIA	54 (12.5%)
IIB	22 (5.1%)
IIIA	58 (13.5%)
IIIB	2 (0.5%)
IV	16 (3.7%)

Abbreviations: SMIL = Swedish Molecular Initiative agains Lung cancer

In Study II, we defined *MET* exon 14 skipping events, gene fusions involving *ALK*, *RET*, *ROS1*, *NRG1*, and *NTRK1*, and mutations in *EGFR* (exons 18-21), *BRAF* (exon 15), *KRAS* (exons 2 and 3), *HER2* (exon 20), and *PIK3CA* (exons 9 and 20) as targetable or potentially targetable oncogene driver alterations.

We used the NanoString nCounter (NanoString Technologies, Seattle, WA, USA) for analysis of gene fusions and *MET* exon 14 skipping events. Mutation analysis was performed by Next Generation Sequencing (NGS) using a 26- or 15-gene exon-focused panel (Illumina TruSight Tumor 26 or 15, Illumina, San Diego, CA, US). For the first 161 samples, NanoString analysis and targeted sequencing were performed in parallel. As oncogene driver alterations often are mutually exclusive, only samples without any findings in the NanoString analysis proceeded to NGS analysis for the remaining samples.

When combining the results from NanoString analysis and targeted sequencing, we were able to identify a targetable or potentially targetable driver alteration in 307 out of 428 samples (72%). Figure 19 illustrates the frequency of targetable or potentially targetable findings identified by Nanostring analysis and targeted sequencing among the samples in Study II.

The most frequently altered gene was *EGFR*, where alterations were detected among 150 patients, harboring altogether 155 alterations. Exon 19 deletions were detected in 77 tumors, and the L858R point mutation was detected in 47 patients. In addition, *EGFR* exon 20 insertions were identified in 16 tumors. Uncommon sensitizing *EGFR* mutations were detected in 12 patients, out of which five had co-occurring mutations in *EGFR*. Cases with co-occurring mutations and gene fusions are further presented in Study II, Table 2.

Furthermore, *MET* exon 14 skipping events were detected in 55 patients, and *ALK* fusions were identified in 34 tumors. Mutations in *KRAS* were detected in 34 patients, where G12V and G12D were the most commonly detected variants. Other alterations identified among our samples included *RET* fusions (n = 11), *NRG1* fusions (n = 4), *NTRK1* fusions (n = 2), *ROS1* fusions (n = 2), *BRAF* mutations (n = 4), *HER2* mutations (n = 10), and *PIK3CA* mutations (n = 10).

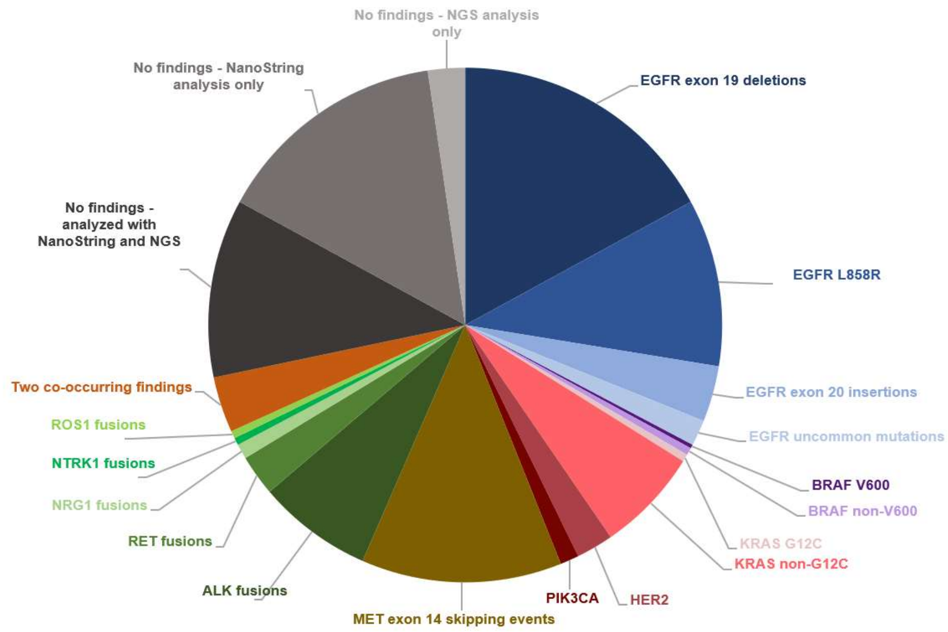


Figure 19. A pie chart illustrating the frequency of targetable or potentially targetable alterations identified by Nanostring analysis and targeted sequencing in the samples in Study II.

A more detailed description of therapies targeting the alterations herein defined as targetable or potentially targetable can be found in the Introduction section. However, as PI3K-inhibitors have not yet performed optimally in the clinical setting, and given the frequent co-occurrence of *PIK3CA* mutations with other oncogene drivers (particularly in *EGFR* and *KRAS*), the role of *PIK3CA* as a potential target for therapies can be debated. Indeed, of the 10 tumors herein that

harbored *PIK3CA* mutations, five displayed co-occurring alterations: *BRAF* V600E mutation (n = 1), *EGFR* exon 19 deletions (n = 3), and *EGFR* exon 20 insertion (n = 1).

In total, there were 121 samples without findings in the NanoString or NGS analyses. Due to poor DNA/RNA quality among some of these samples, a proportion of the cases could only be analyzed with either NanoString or NGS. Indeed, only 48 out of the 121 samples with no findings had been successfully analyzed with both methods. For the remaining samples with no identified oncogene driver, 63 samples had only been analyzed with NanoString, and 10 had only been analyzed with NGS. For an additional three samples, the DNA and RNA quality was too poor for both NanoString analysis and targeted sequencing. Therefore, it is likely that our results, with 72% of the samples harboring a targetable or potentially targetable alteration, is an underestimation.

The cohort in Study II will serve as a platform for future studies on smoking-independent lung cancer. The study was performed on tumor tissue from surgically treated patients, as there for patients with advanced disease generally are only small biopsies or cytology specimens available. In a previous review of the NLCR, 57% of never smokers in Sweden had stage IV disease at diagnosis[168], which is the relevant category for targeted therapies. The applicability of our results to patients with advanced lung cancer therefore deserves interest. In a study by Couraud et al., where most patients had advanced disease, a targetable molecular alteration was found in 73% of all never smoking patients (including alterations in *EGFR*, *ALK*, *KRAS*, *HER2*, *BRAF*, and *PIK3CA*)[215]. In the present study, 150 patients out of 428 (35%) harbored *EGFR* mutations, and as mutational data was missing for a proportion of the samples as mentioned above, the frequency of *EGFR* mutated tumors in this cohort was most likely slightly underestimated. Indeed, in a previous study where results from clinical NGS-testing in southern Sweden was reviewed, *EGFR* mutations were identified in 44% of the never smokers[216].

After patients' medical charts had been reviewed, 36 patients (with a total of 37 tumors) were excluded from the study as it was likely that their smoking status had been incorrectly registered in the NLCR. A commonly used definition of a never smoker is an individual who has smoked less than 100 cigarettes over his/her lifetime. However, patients' medical charts rarely state the exact number of cigarettes the patient has smoked. Hence, the application of such a definition on a retrospective material is not feasible. Some of the patients reported very light smoking (mainly social smoking) during a short and limited time period, a long time before diagnosis (>30 years). These patients were kept in the study. The smoking history of the 36 patients that were excluded from the study varied widely. For example, one patient reported having smoked for 80 pack years, whereas more debatable cases included patients who had stopped smoking >30 years before diagnosis, but where there was no additional information regarding the smoking

behavior. The latter category of patients was excluded from the study. Examples of the smoking history for some of the excluded patients are listed in Table 9.

Table 9. Example of patients that were excluded from the study due to smoking history.

Sample Number	Smoking History	Year of Surgery	Age at Surgery
S_0031	Former smoker since the 1970s, smoked very little.	2012	77
S_0115	Former smoker since the 1960s, no further information available	2006	74
S_0152	Former smoker since 1976. Smoked very little.	2010	66
S_0338	Smoked 10 cigarettes per day during one year, 1979 to 1980. Also sparse social smoking.	2006	65
S_0417	Habitually smoked marijuana as a teenager. Started smoking cigarettes in 2013, smokes 4-5 cigarettes per day at the time of diagnosis	2014	35

Some of the cases that were excluded due to smoking were analyzed by NanoString or NGS, as the clinical information in some cases were obtained after the analyses had been performed. Of the 37 tumors (from 36 patients) that were excluded due to smoking history, 28 tumors were analyzed by Nanostring. Out of these, four tumors displayed the following alterations: one *ALK* fusion, one *RET* fusion, one uncertain *ALK* fusion, and one uncertain *NTRK1* fusion. For one sample, the NanoString data was not further analyzed due to poor data quality. The remaining 23 samples had no findings in the NanoString analysis.

In total, 12 of the tumors that were excluded due to smoking were analyzed by NGS. For one sample, the sequencing data could not be interpreted due to insufficient data quality. For the remaining samples, 8 tumors displayed the following mutations: *KRAS* G12C (n = 4), *KRAS* G12D (n = 2), *KRAS* G12V (n = 1), and *EGFR* exon 19 deletion (n = 1).

Based on the results from the NanoString and NGS analyses from the tumors that were excluded due to smoking, we identified slightly more confirmed or suspected oncogene driver alterations than would have been expected among NSCLC from smokers. This may be indicative of a smoking-independent tumorigenesis for some of the cases within the excluded subgroup, illustrating the difficulties of assessing smoking status retrospectively on the basis of medical records. However, the four cases with *KRAS* G12C mutation, indicative of a history of smoking, suggest that it was correct to regard those cases as incorrectly registered in the NLCR.

In summary, we detected a targetable or potentially targetable alteration among 72% of the samples in this nationwide and population-based study, which includes over 400 surgically treated never smokers with lung cancer. This underlines the importance of oncogene driver alterations among these patients. For the smoking-independent lung tumors with no driver alteration identified by the NanoString or NGS assays, further investigation of the landscape of oncogenic alterations is warranted.

Prognostic Markers

The poor prognosis in lung cancer is partly due to that many patients are diagnosed at a late stage. However, also among surgically treated cases there is a substantial cancer-related mortality since about a third of these patients will relapse, and the 5-year overall survival rate for surgically treated patients in Sweden is around 60%[1, 217]. Postoperative adjuvant chemotherapy can reduce the risk of recurrence, and has been demonstrated to increase the 5-year survival by about 5%[107, 108]. Today, tumor stage and patient performance status are the main parameters used for treatment decision concerning adjuvant chemotherapy. However, heterogenous outcomes also within the same disease stage highlight the need for additional tools to stratify patients into low- or high-risk groups. Patients with tumors in stage IA are typically not given adjuvant treatment and are monitored for at least three years to detect any recurrences. However, as some of these patients will relapse, it is possible that a subgroup of patients with small tumors would benefit from intensified treatment strategies or increased surveillance. Patients who have been surgically treated for tumors in stage IB or higher are offered adjuvant treatment and are typically followed for five years to detect any recurrences. Some of these patients would never have relapsed even without adjuvant treatment, but, so far, there are no reliable markers to identify those individual patients for whom the risks of adjuvant treatment might exceed the benefits.

In Studies III and IV, we used two different approaches to explore prognostic markers in lung cancer. The traditional method to study prognostic markers is by IHC, which is a well-established method and is easily applicable in the clinical setting. In Study III, we studied the prognostic ability of RBM3 protein expression by IHC. In Study IV, instead, we tested a multi-cohort, gene expression-based strategy as a proof of concept for identification of genes with prognostic potential and used IHC to validate some of these candidate markers.

Study III

The RNA binding motif protein 3 (*RBM3*) gene is located at the p11.23 region of the X-chromosome[218]. The expression of *RBM3* is increased by cellular stress such as hypothermia and hypoxia[219]. RBM3 contains an RNA recognition motif domain that is evolutionarily conserved across species, and can bind to RNA and thereby modulate the translation[219]. By interacting with ribosomes, initiation factors, and micro-RNAs, RBM3 modifies the translational process in several ways, generally leading to an enhanced global protein synthesis[220, 221]. By regulating the translational process, RBM3 is involved in many different and complex cellular mechanisms such as cell cycle progression and regulation of apoptosis[222, 223]. In part due to this complexity, the exact role of RBM3 in normal cells is not fully understood, but it is

thought to protect the cells from apoptosis by facilitating translation during stressful conditions[224, 225].

Similarly, the role of RBM3 in cancer remains to be fully elucidated. Several studies on different types of cancer have demonstrated RBM3 protein expression to be a prognostic factor, with high levels conferring a better prognosis[226-236]. This is somewhat contradictory to the above-mentioned role of RBM3 in normal cells, where RBM3 is thought to protect the cell from apoptosis. In a study by Ehlén et al., a relationship between RBM3 and processes involved in DNA damage responses was demonstrated[237]. Also, a link between RBM3 protein expression and sensitivity to chemotherapy has been proposed[233, 238]. However, additional studies are needed to further explore these hypotheses and the functional role of RBM3 in cancer.

Despite an incomplete understanding of the biological functions of RBM3 in normal cells and in cancer cells, RBM3 has emerged as a promising prognostic marker. In Study III, the prognostic ability of RBM3 in lung cancer was investigated. Protein expression of RBM3 was assessed with IHC in two independent surgically treated lung cancer patient populations consisting of 213 (cohort I) and 306 (cohort II) cases. Results were correlated to clinicopathological parameters, overall survival, and recurrence-free interval (RFI). Furthermore, the gene expression levels of *RBM3* were investigated in cohort II and in public gene expression data sets consisting of 2,087 ACs and 899 SqCCs.

For assessment of RBM3 protein expression, both the fraction of viable tumor cells expressing RBM3 in the nucleus (nuclear fraction, NF) and the nuclear staining intensity (NI) were scored, and a combined nuclear score (NS) was denoted by multiplying NF and NI. A combination of NF and NI has also been used in other cancer cohorts for assessing RBM3 protein expression[226-234, 238]. As the expression levels may vary between different types of cancer, thereby affecting the optimal cut-off, the exact way of classifying samples as having a high or a low expression has differed between previous studies. In Study III, it was noted that the protein expression of RBM3 differed between histological subtypes, as ACs had a higher NS than SqCCs (Mann-Whitney test, $P = 0.07$ and $P = 0.001$ in cohorts I and II, respectively), as illustrated in Figure 20. Hence, different cut-offs for identifying patients with a high or low expression were applied to AC and SqCC in Study III.

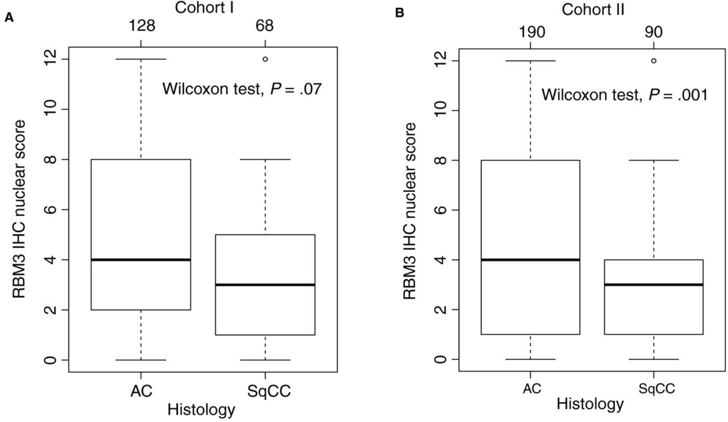


Figure 20. RBM3 immunohistochemical (IHC) protein expression among adenocarcinomas (AC) and squamous cell carcinomas (SqCC) for cohort I (A) and cohort II (B).
Abbreviations: AC = adenocarcinoma, SqCC = squamous cell carcinoma, IHC = immunohistochemistry.
Reprinted from Salomonsson et al.[239]

As illustrated in Figure 21, high RBM3 protein expression was found to be a favorable prognostic factor among AC (although not statistically significant in the multivariable analysis for OS in cohort II). Among the SqCC cases, a trend for the opposite relationship between RBM3 expression and prognosis was observed, although not statistically significant. Due to a limited number of SqCC cases, results must be considered with care. Similar results were demonstrated in a previous study by Melling et al., where high protein expression of RBM3 was found to be a favorable prognostic marker in lung AC but not in SqCC[240]. However, in the study by Melling et al., no multivariable analysis was performed, different antibodies were used, and criteria for classifying samples as having a low or a high RBM3 expression were not clarified, thus making direct comparison between that study and Study III difficult.

As for many prognostic markers, there is no established cut-off for identifying patients with a low or high expression of RBM3. In such cases, it is important to have a strategy for validation of the generated cut-off. In Study III, we used cohort I as a discovery cohort for identification of an optimal cut-off among these patients. This cut-off was then applied to cohort II, which was used as a validation cohort to test the performance of the selected cut-off.

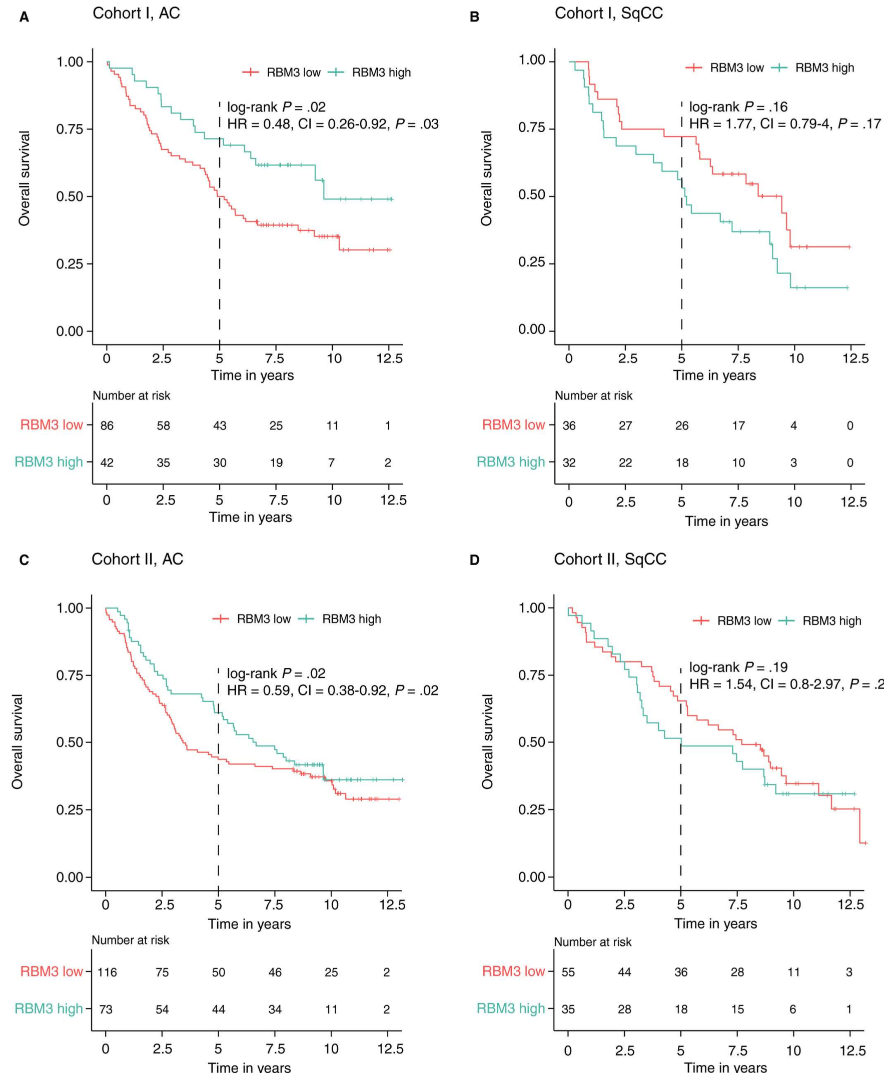


Figure 21. Prognostic value of RBM3 protein expression on overall survival in adenocarcinomas (AC) in cohort I (A), squamous cell carcinomas (SqCC) in cohort I (B), AC in cohort II (C) and SqCC in cohort II (D).
Abbreviations: AC = adenocarcinoma, SqCC = squamous cell carcinoma, HR = hazard ratio, CI = confidence interval.
Reprinted from Salomonsson et al.[239]

In cohort I, the protein expression of marker of proliferation Kiel 67 (Ki67) was assessed. Here, AC cases had a lower expression of Ki67 than the SqCC cases (Mann-Whitney test, $P < 0.0001$), as illustrated in Figure 22, which is in line with results from previous studies[241-243]. For SqCC, no association between Ki67 and RBM3 protein expression could be demonstrated. In AC, we found a negative correlation between Ki67 and RBM3 protein expression (Mann-Whitney test, $P =$

0.02), as illustrated in Figure 22, suggesting that the prognostic ability of RBM3 could be linked to proliferation. As further discussed in Study IV, genes linked to proliferation have been demonstrated to be of value for assessment of prognosis. Furthermore, previous studies have demonstrated a limited prognostic ability of Ki67 among SqCC compared to AC, similar to the prognostic ability of RBM3 demonstrated in our study and in the study by Melling et al.[240, 242-244]. These observations, with a correlation between Ki67 and RBM3 protein expression, and a similar prognostic ability of both markers among AC and SqCC, are intriguing and suggests a connection between the two markers that merits further investigation.

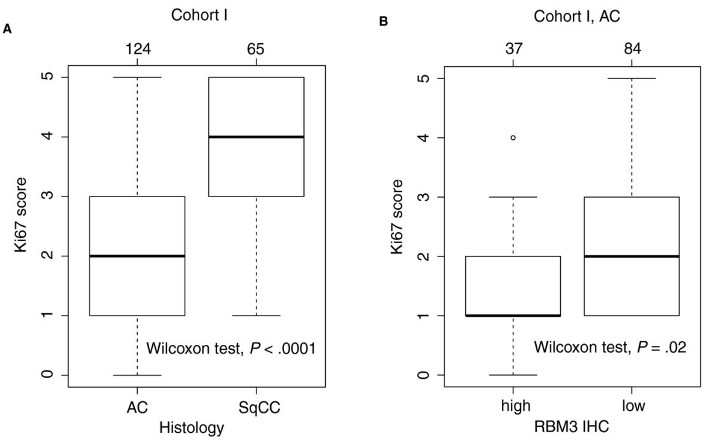


Figure 22. Protein expression of Ki67 among adenocarcinomas (AC) and squamous cell carcinomas (SqCC) in cohort I (A). Among AC in cohort I, a negative correlation between Ki67 and RBM3 protein expression could be demonstrated (B).
Abbreviations: AC = adenocarcinoma, SqCC = squamous cell carcinoma, IHC = immunohistochemistry.
Reprinted from Salomonsson et al.[239]

A weak correlation between *RBM3* mRNA levels and RBM3 protein expression was demonstrated. As previously discussed in Study I, where we explored the association between *KIT* gene copy numbers, RNA levels and protein expression, these correlations are complex and are not always linear. However, in Study I, a similar correlation between *KIT* mRNA levels and KIT protein expression was detected. Despite the correlation between *RBM3* mRNA levels and RBM3 protein expression observed in Study III, gene expression levels of *RBM3* were not found to be prognostic, suggesting that the correlation is not absolute and perhaps also that the mRNA expression is more tightly regulated.

In summary, in Study III, we found that RBM3 protein expression was a favorable prognostic factor for patients with lung AC. Its potential future clinical usefulness merits further research, ideally in a prospective study.

Study IV

Traditionally, potential prognostic markers have often been studied by using IHC to evaluate the expression of individual proteins. However, the markers explored in such studies must be pre-defined and only a limited number of markers can be evaluated simultaneously. Despite hundreds of prognostic markers that have been explored in such studies, there are no IHC-based prognostic markers for lung cancer in clinical use today.

Given the complexity of the lung cancer genome and the heterogeneity of lung tumors, a gene expression-based approach, where gene signatures integrate the expression of multiple genes, could potentially better predict prognosis. However, the reproducibility and clinical applicability of such gene expression signatures remain to be ascertained[157].

In Study IV, a gene expression-based multi-cohort discovery and validation strategy was tested as a proof of concept, with subsequent validation of generated possible markers with IHC. Strengths of this approach include a broad, unbiased initial search among numerous potential prognosticators, a selection of candidates based on a robust association with OS in multiple gene expression data sets, and a final assessment of the clinical applicability of the selected markers by evaluating these with IHC.

In the gene expression-based discovery and validation strategy, publicly available gene expression data sets from six microarray-based studies were used to identify genes with correlation to OS. Thereby, the prognostic ability of numerous genes among many samples was assessed. As further described in Study IV, the data sets were divided into four discovery data sets and two validation data sets. In the discovery step, genes associated with OS in all four discovery data sets were identified. In the validation step, the prognostic potential of each gene generated from the discovery step was assessed in the two validation data sets. Genes associated with OS in both validation data sets were identified, which generated a list of 19 candidate prognostic markers, specified in Table 10.

Table 10. Genes with prognostic potential identified by the gene expression-based discovery and validation strategy.

Gene Symbol	Gene Name
KI67	Marker of proliferation Kiel 67
MCM4	Minichromosome maintenance complex component 4
TYMS	Thymidylate synthetase
CCNA2	Cyclin A2
CCNE1	Cyclin E1
BUB1B	Budding uninhibited by benzimidazoles 1 homolog beta
DLGAP5	Discs large homolog associated protein 5
KIF14	Kinesin family member 14
NUSAP1	Nucleolar and spindle-associated protein 1
RACGAP1	Rac GTPase activating protein 1
ECT2	Epithelial cell transforming sequence 2 oncogene
ASPM	Abnormal spindle-like microcephaly-associated protein
PRC1	Protein regulator of cytokinesis 1
BTG2	B-cell translocation gene 2
HLF	Hepatic leukemia factor
GDF10	Growth differentiation factor 10
CTTN	Cortactin
COL4A3	Collagen, type IV, alpha 3
CIRBP	Cold inducible RNA binding protein

Many of the genes identified by our gene expression-based discovery and validation strategy are genes linked to proliferation. Proliferation is, as one of the hallmarks of cancer, a central ability of cancer cells that contributes to tumor progression, and the prognostic impact of proliferation has long been recognized in many types of cancer[47]. A short overview of the 19 genes identified by the gene expression-based strategy is presented below.

The exact function of Ki67 is unknown, but as it is present during all active phases of the cell cycle, and absent from resting cells, Ki67 is widely used as a marker of proliferation in different types of cancer[245]. Many studies have assessed the prognostic ability of Ki67 in lung cancer, but consensus is lacking[246].

The mini-chromosome maintenance (MCM) proteins form a hexameric complex that constitutes the core of the replicative helicase that separates the double-stranded DNA into single strands during DNA replication[247]. The association between expression of the MCM proteins and prognosis have previously been explored in lung cancer[248-251].

Thymidylate synthase (TYMS) is an enzyme essential for the metabolism of thymidine monophosphate, a nucleotide used in DNA synthesis[252]. As the main

target of some chemotherapeutic agents, the predictive value of TYMS expression has been explored in many studies[253, 254]. Also, the prognostic ability of TYMS has been investigated in lung cancer, but further studies are needed[255].

Two cyclins, cyclin A2 (CCNA2) and cyclin E1 (CCNE1), were identified by the gene expression-based strategy. Cyclins are key components in the regulation of the cell cycle and essential for cell proliferation[256].

During cell division, the segregation of chromosomes is dependent on the mitotic spindle to separate the chromosomes. This complex process involves many proteins, and several of the genes identified as potential prognostic markers in Study IV are involved in this process. Budding uninhibited by benzimidazoles 1 homolog beta, BUB1B, is involved in the spindle assembly checkpoint, a cell cycle checkpoint that ensure correct separation of chromosomes during cell division[257]. Discs Large Homolog Associated Protein 5, DLGAP5, is involved in the formation of the mitotic spindle[258]. Kinesin family member 14 (KIF14) is a microtubule motor protein which associates with the mitotic spindle[259]. Nucleolar and spindle-associated protein 1 (NUSAP1) has an essential role in spindle formation[260]. Rac GTPase activating protein 1 (RACGAP1) is part of the centralspindlin complex, important for organization of the mitotic spindle[261], and it interacts with epithelial cell transforming sequence 2 oncogene (ECT2), which localizes to the central spindle during cell division[262, 263]. Abnormal spindle-like microcephaly-associated protein (ASPM) is essential for orientation of the mitotic spindle[264]. Protein regulator of cytokinesis 1 (PRC1) binds to spindle microtubules[265].

B-cell translocation gene 2 (BTG2) is involved in the regulation of the cell-cycle and cellular stress response pathways, and its expression is induced by TP53[266].

Hepatic leukemia factor (HLF) is a transcription factor, and gene fusions involving this gene and the transcription factor E2-alpha (E2A) gene are implicated in the carcinogenesis of some types of leukemias[267].

Growth differentiation factor 10 (GDF10) is a cytokine implicated in signaling pathways that regulates proliferation and epithelial to mesenchymal transition[268].

Cortactin (CTTN) regulates the structure of the actin cytoskeleton and has also been linked to processes involving epithelial to mesenchymal transition[269, 270].

Collagen, type IV, alpha 3 (COL4A3), also known as Goodpasture antigen, is one of the subunits that constitutes type IV collagen, a structural component of the basement membranes[271]. It is involved in the pathogenesis of Goodpasture syndrome where antibodies directed towards COL4A3 cause glomerulonephritis and alveolitis.

Cold inducible RNA binding protein (CIRBP) has many similarities to RBM3 that was studied in Study III. Both CIRBP and RBM3 are RNA-binding proteins and share a conserved RNA-recognition motif domain[219]. These two proteins are

induced in response to cellular stresses such as hypothermia and hypoxia. As with RBM3, CIRBP is involved in many biological functions, one of which is proliferation[272].

To ensure a clinical applicability of the results, three of the 19 genes identified by the gene expression-based strategy were selected for further evaluation with IHC: Ki67, MCM4 and TYMS. These three genes were chosen based on the availability of reliable antibodies and on reports from previous lung cancer studies. However, also other genes of the 19 candidate prognostic markers have previously been suggested to have prognostic potential in lung cancer as individual markers or as part of gene expression signatures[273-276].

IHC was performed in two independent and clinically well-characterized lung cancer cohorts, the same cohorts as used in Study III. As in Study III, neither of the markers had an established cut-off for classifying samples as having a high or a low expression. Therefore, an IHC discovery and validation strategy was employed, as in Study III. Cohort I was used as an IHC discovery cohort, and different cut-offs were tested by using Kaplan-Meier plots with log-rank test. The cut-offs which most clearly identified prognostic groups were chosen.

For Ki67, a cut-off of >10% positive tumor cells was chosen for the AC cases. By applying this cut-off, 74 ACs (56%) were classified as having a high Ki67 protein expression. For MCM4, a cut-off of >75% positive tumor cells was chosen, resulting in 15 cases (12%) identified as having a high MCM4 expression. For TYMS, a score (obtained by multiplying fraction and intensity) of >2 p was chosen among the AC cases, which resulted in 19 cases (16%) classified as having a high expression of TYMS.

In cohort I, high expression of Ki67 and MCM4 was associated with poor prognosis in the 5-year OS analysis (log rank-test), and in the univariable and multivariable Cox proportional hazards regression models. For TYMS, a trend was observed with a high expression conferring a worse prognosis in the 5-year OS analysis (log-rank test), although the results were not statistically significant. The prognostic value of the three markers in cohort I is illustrated in Figure 23.

For SqCC, no relevant subgroups could be identified by any of the markers and therefore only AC cases were evaluated in the IHC validation cohort.

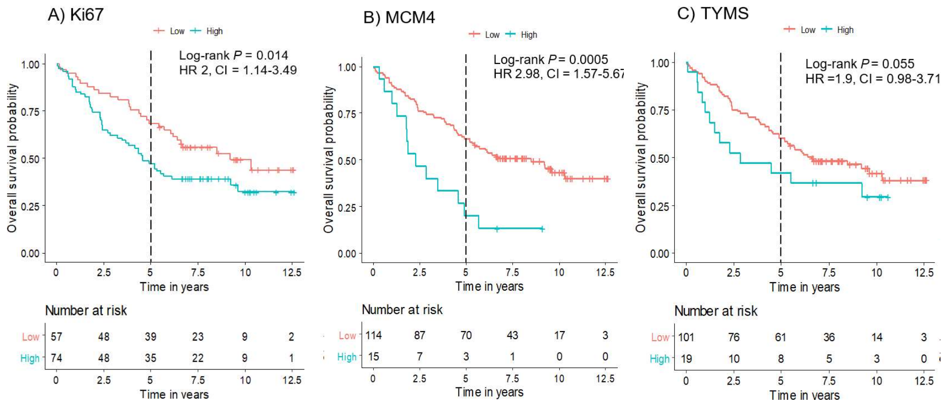


Figure 23. The prognostic value of Ki67 (A), MCM4 (B), and TYMS (C) protein expression in cohort I. Abbreviations: HR = hazard ratio, CI = confidence interval.

In the IHC validation phase, the AC cases in cohort II were simply scored as having a high or a low expression of the markers, according to the cut-offs generated in the IHC discovery cohort. By applying the selected cut-off to the IHC validation cohort, 91 cases (57%) demonstrated a high Ki67 protein expression. A prognostic ability of this marker was demonstrated in the log rank test and in the univariable Cox proportional hazards regression model, but did not remain statistically significant in the multivariable model. A high protein expression of MCM4 and TYMS was demonstrated in 17 cases (10%) and 17 cases (12%), respectively. However, the prognostic ability of the markers could not be confirmed, neither in the log-rank test nor in the univariable and multivariable Cox proportional hazards regression models. The prognostic value of the markers is illustrated in Figure 24.

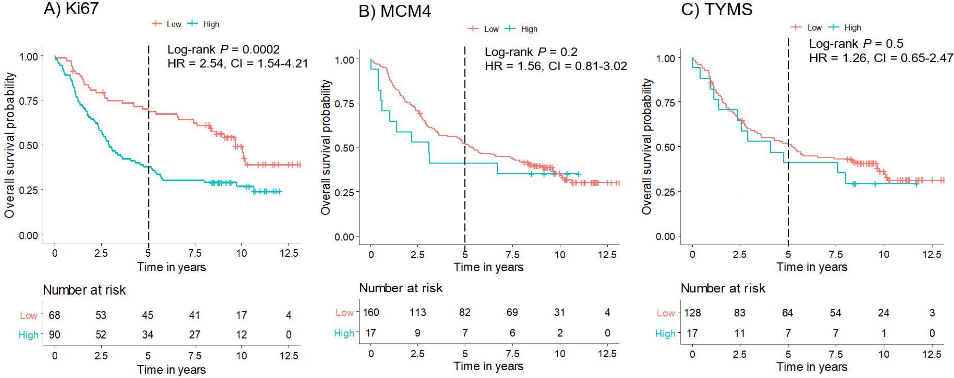


Figure 24. The prognostic value of Ki67 (A), MCM4 (B), and TYMS (C) protein expression in cohort II. Abbreviations: HR = hazard ratio, CI = confidence interval.

We sought to explore if a signature of multiple prognostic IHC markers would more accurately stratify the patients into low- and high-risk groups. Therefore, the combined prognostic ability of the three markers was examined by each case receiving one point per positive maker, thus denoting a combined score ranging from 0 to 3 points. However, in both cohorts, the combined score was dependent on the prognostic ability of Ki67 alone. In both cohorts, there were no patients that had a high expression of MCM4 alone. Also, only two patients in cohort I and one patient in cohort II had a high expression of TYMS alone, as illustrated in Figure 25. Perhaps this overlap between the markers reflects that the cut-offs for MCM4 and TYMS classified only a small proportion of the patients (10-16% of the patients) as having a high expression. Cut-offs more resembling that of Ki67 (identifying around half of the patients) would perhaps have identified additional patients. Also, the overlap between the markers might be due to all three being linked to proliferation, thus measuring the same biological processes. It cannot be ruled out that selecting markers from the 19 generated candidates in a way that instead measures several biological processes would have led to a more informative combination of potential markers. However, in a study by Grinberg et al., five protein markers, representing diverse biological processes, were included in a prognostic biomarker panel[277]. Each marker was also associated with prognosis in gene expression data sets. However, the markers did not add prognostic information beyond clinical parameters alone, underlining the difficulties of identifying reliable prognostic markers that add prognostic information to the already existing tools.

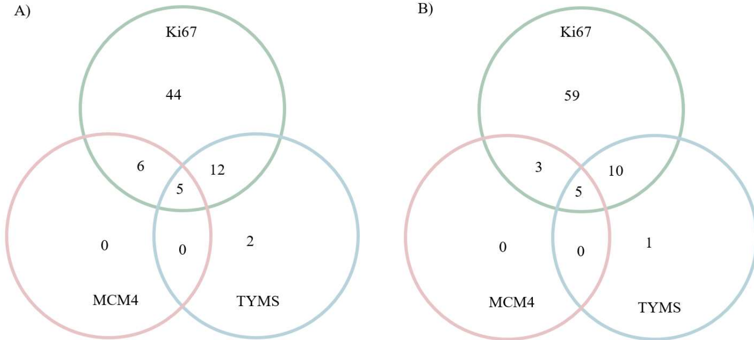


Figure 25. The overlap between cases that were positive for the three markers. Only two patients in cohort I (A) and one patient in cohort II (B) had a high expression of TYMS alone, and there were no patients that had a high expression of MCM4 alone.

For a subset of the patients in cohorts I and II, gene expression data of *Ki67*, *MCM4* and *TYMS* were available. For all three markers, a correlation between gene expression levels and IHC classification could be demonstrated. In cohort II, the prognostic value of *Ki67*, *MCM4*, and *TYMS* gene expression levels was assessed. For all three markers, potentially prognostic subgroups could be visualized in the Kaplan-Meier plots, although not statistically significant for *Ki67* and *TYMS*.

As previously mentioned, the correlations between gene copy numbers, gene expression levels and protein expression are complex. Many factors influence the relationship between RNA levels and protein expression[278]. It has been demonstrated that the correlation between RNA levels and protein expression varies between different genes and conditions[278, 279]. In Studies I, III, and IV, we demonstrated a correlation between RNA levels and protein expression of the studied genes (*KIT*, *RBM3*, *Ki67*, *MCM4*, and *TYMS*). Despite a prognostic potential of RBM3 protein expression, and the observed correlation between RNA levels and protein expression, we could not demonstrate any prognostic potential of *RBM3* RNA levels. Similarly, for Ki67, MCM4, and TYMS, the prognostic potential of these genes identified by our gene expression-based strategy could not be confirmed by IHC, despite a correlation between RNA levels and protein expression.

The difference in *RBM3* RNA levels between samples classified as having a high or a low RBM3 protein expression was small. Possibly, subtle differences in RNA levels between samples, and insufficient precision in the assays, can weaken the correlations in an OS analysis. For the markers in Study IV, insufficient precision in the IHC classification and suboptimal cut-offs could potentially reduce the prognostic associations on the protein level. Moreover, a binary cut-off might not accurately reflect the continuous distribution of proliferation between samples. In breast cancer, it has been demonstrated that assessment of Ki67 staining is reliable for tumors displaying an expression below 5% and above 30%, but there is a lack of consistency in the intermediate range, reflecting the difficulties to accurately separate groups by using a binary cut-off[280]. Also, it is possible that different cut-offs for the markers would have performed better in the validation cohort. For MCM4 and TYMS, the cut-offs selected in the IHC discovery cohort identified a small subgroup of the patients with a worse prognosis. Such cut-offs would ideally require larger patient materials for validation. Furthermore, the results from Studies III and IV emphasize the necessity to employ a validation cohort when there are no generally accepted cut-offs for a marker.

Generally, there is often a poor reproducibility among IHC-based prognostic studies. As mentioned in the Methods section, many factors influence the results from an IHC experiment, such as the handling and processing of tissues, the staining procedure and the analytical steps. These different steps have proven to be difficult to standardize across different laboratories, as demonstrated by the marker Ki67. Despite great efforts to implement a standardized assessment of Ki67 in breast cancer, the clinical utility of Ki67 staining remains controversial[280]. Therefore, as

with gene expression signatures, the reproducibility and clinical applicability of a universal IHC-based prognostic marker remain to be ascertained.

In summary, Study IV demonstrates that a gene expression-based discovery and validation strategy is feasible for exploring potential prognostic markers in lung cancer. We identified 19 genes robustly associated with OS on the RNA level and selected three for subsequent immunohistochemical validation to be applicable in the clinical practice. The value of this concept for exploring prognostic markers that can add information to the already existing tools merit further investigation.

Conclusions

The four studies in this thesis characterize lung cancer on the molecular level by investigating alterations in oncogene drivers and potential prognostic markers.

Studies I and II

- Histological subgroups of lung cancer display different types of *KIT* alterations, and there is a poor correlation between *KIT* gene copy number alterations and expression.
- Among never smokers, alterations in targetable or potentially targetable genes are common, which highlight the necessity of extensive molecular testing in this subgroup to optimize patient therapy.

Studies III and IV

- High RBM3 protein expression is a marker of better prognosis in lung adenocarcinoma.
- A gene expression-based discovery and validation strategy, with subsequent immunohistochemical validation, is a feasible strategy for exploring potential prognostic markers in lung cancer.

Future Perspectives

High incidence rates, in combination with a high mortality, makes lung cancer the leading cause of cancer-related death, and a major health problem worldwide. At present, the majority of lung cancer cases are attributable to smoking. Therefore, smoking prevention and cessation are key factors to decrease the high lung cancer incidence rates worldwide.

A large proportion of lung cancer patients are diagnosed at a late stage, when curatively intended treatment is no longer an option. For these patients, great advancements in the clinical management have been achieved in the past decades. The rapid progress within the research area of targeted therapies will, with utmost probability, continue in the forthcoming decades, with the emergence of improved drugs, new targets, and new settings. Even now there is a need for an increased understanding of the therapies that are already available today, regarding, e.g., how drug sensitivity is influenced by both the exact character of the targeted alteration and by co-occurring alterations. Furthermore, as these tumors inevitably acquire resistance to the drug, more knowledge of different resistance mechanisms, and how to detect and overcome them, is needed.

The tumor biology in never smokers differs from that in smokers. Targetable alterations are more commonly detected among never smokers, whose tumors are also more often less proliferative and harbor fewer mutations and copy number alterations. The large cohort of tumors from never smokers that was collected in Study II is a unique nationwide, population-based cohort that will be used as a platform for future studies. Potential projects include further applications of the data generated in Study II (NanoString and NGS data), IHC-based studies on TMAs, and studies with a focus on etiology and hereditary factors.

Lung cancer screening programs, which are currently being evaluated or will soon be implemented in many countries, could aid in reducing the high mortality rates by earlier detection of lung cancer. However, also among surgically treated cases there is a substantial mortality. The high frequency of relapse among surgically treated patients reinforces the need of risk stratification beyond stage and performance status.

Great efforts have been made to explore potential prognostic markers in lung cancer. However, no such marker is in clinical use today, and the clinical usefulness of a single prognostic marker has proven to be difficult to demonstrate. Despite the

discouraging results of previous studies and efforts, ongoing work to develop prognostic markers is still of great importance. Additional advances, e.g., improved image analysis tools by artificial intelligence, will possibly facilitate coming discoveries and validations of molecular prognosticators. Advancements in gene expression-based prognostic tools may result in predictions that are more accurate and enable a broader examination of multiple prognosticators simultaneously. However, as with single protein prognostic markers, the reproducibility and clinical value of such tests remains to be demonstrated.

In summary, within the rapidly evolving field of lung cancer research, continuous improvements in treatment strategies and refined predictions of treatment response and prognosis may further reduce the burden of this disease in the future.

Acknowledgements

The work presented in this thesis would not have been possible without the support from many people and several collaborations. I would like to express my gratitude to all of you!

In particular, I would like to thank:

Maria Planck, my main supervisor. Whether you are meeting a patient, supervising a PhD-student, hosting a group meeting, or organizing a large research gathering, you have the most extraordinary way of making people feel at home. With your extensive knowledge, kind heart, and ability to focus on things that really matter, you are a source of inspiration.

Johan Staaf, my co-supervisor. Thank you for introducing bioinformatics to my studies, always with a careful balance between guidance and independent work. Thanks for sharing your skills and for always having a new perspective on research projects.

Mats Jönsson, my co-supervisor. From the Pfaffl method a decade ago to next generation sequencing today (via beer brewing in Flyinge) -- thank you for patiently and cheerfully sharing your wide knowledge with me throughout the years!

Hans Brunnström, my co-supervisor. Thank you for always, and with great enthusiasm, answering all my questions. You truly are an encyclopedia of pathology.

Karin Jirström, my co-supervisor. Thank you for welcoming me into the RBM3 family, and for sharing one of your many inspiring projects with me.

Åke Borg, my co-supervisor. Thank you for creating a thriving research environment at the department.

All former and present members of the Breast and Lung Cancer Research Group: *Anna Karlsson, Sofi Isaksson, Bassam Hazem, Daniel Ehinger, Frida Rosengren, Christel Reuterswärd, Srinivas Veerla, Helena Liljedahl, Elsa Arbajian, Gudrun Oskarsdottir, Suze Roostee, Deborah Figueiredo Nacer de Oliveira, Mattias Aine, Christel Estberg, Lennart Hohmann, Kamila Kaminska, Jari Häkkinen,* and *Marija Kotevska*. A special thanks to Anna for your invaluable support as a colleague and friend.

All *co-authors* for their contribution and collaboration.

Former and present *laboratory and health care personnel* at the Departments of Thoracic Surgery, Respiratory Medicine, and Pathology for engagement in the Southern Lung Cancer Study.

All *lung cancer patients*, who generously contributed to the studies.

Pär-Ola Bendahl, for all the help with statistical issues.

Anne Vähäniemi, Björn Frostner and *Susanne André* for excellent administrative work and support.

All *other colleagues* at the department. A special thanks to *Quinci Romero* for your friendship and support throughout the years.

All funders of this work: *The Swedish Cancer Society, The Mrs Berta Kamprad Foundation, The Sjöberg Foundation, Governmental funding of clinical research within the national health services (ALF), The Crafoord foundation, Maja and Hjalmar Leander foundation, Märta Winkler Foundation, Anna-Lisa and Sven-Eric Lundgren foundation, The King Gustaf V Jubilee Fund, John and Augusta Persson Foundation, The Gunnar Nilsson Cancer Foundation, Inga and Johan Hains Foundation.*

Last but not least: my family. *Kristina, Eva, Anna Katarina, Nils Olof, Per Olof, David* and *Valter*. I love you all ♡

References

1. *Nationellt vårdprogram Lungcancer 2021. Available at:* https://www.cancercentrum.se.
2. Howlader, N., et al., *The Effect of Advances in Lung-Cancer Treatment on Population Mortality.* N Engl J Med, 2020. 383(7): p. 640-649.
3. Sung, H., et al., *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.* CA Cancer J Clin, 2021. 71(3): p. 209-249.
4. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2018.* CA Cancer J Clin, 2018. 68(1): p. 7-30.
5. Cao, M. and W. Chen, *Epidemiology of lung cancer in China.* Thorac Cancer, 2019. 10(1): p. 3-7.
6. Ferlay, J., et al., *Cancer statistics for the year 2020: An overview.* Int J Cancer, 2021.
7. Parkin, D.M., et al., *Global cancer statistics, 2002.* CA Cancer J Clin, 2005. 55(2): p. 74-108.
8. Hecht, S.S., *Lung carcinogenesis by tobacco smoke.* Int J Cancer, 2012. 131(12): p. 2724-32.
9. Sun, S., J.H. Schiller, and A.F. Gazdar, *Lung cancer in never smokers--a different disease.* Nat Rev Cancer, 2007. 7(10): p. 778-90.
10. Sasco, A.J., M.B. Secretan, and K. Straif, *Tobacco smoking and cancer: a brief review of recent epidemiological evidence.* Lung Cancer, 2004. 45 Suppl 2: p. S3-9.
11. Pinsky, P.F., C.S. Zhu, and B.S. Kramer, *Lung cancer risk by years since quitting in 30+ pack year smokers.* J Med Screen, 2015. 22(3): p. 151-7.
12. Darby, S., et al., *Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies.* Bmj, 2005. 330(7485): p. 223.
13. *Global and regional burden of cancer in 2016 arising from occupational exposure to selected carcinogens: a systematic analysis for the Global Burden of Disease Study 2016.* Occup Environ Med, 2020. 77(3): p. 151-159.
14. Raaschou-Nielsen, O., et al., *Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE).* Lancet Oncol, 2013. 14(9): p. 813-22.

15. Matakidou, A., T. Eisen, and R.S. Houlston, *Systematic review of the relationship between family history and lung cancer risk.* Br J Cancer, 2005. 93(7): p. 825-33.
16. Schwartz, A.G., et al., *The molecular epidemiology of lung cancer.* Carcinogenesis, 2007. 28(3): p. 507-18.
17. Gazdar, A., et al., *Hereditary lung cancer syndrome targets never smokers with germline EGFR gene T790M mutations.* J Thorac Oncol, 2014. 9(4): p. 456-63.
18. Wang, Y., et al., *Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer.* Nat Genet, 2014. 46(7): p. 736-41.
19. Hwang, S.J., et al., *Lung cancer risk in germline p53 mutation carriers: association between an inherited cancer predisposition, cigarette smoking, and cancer risk.* Hum Genet, 2003. 113(3): p. 238-43.
20. Araldi, R.P., et al., *The human papillomavirus (HPV)-related cancer biology: An overview.* Biomed Pharmacother, 2018. 106: p. 1537-1556.
21. Ragin, C., et al., *HPV-associated lung cancers: an international pooled analysis.* Carcinogenesis, 2014. 35(6): p. 1267-75.
22. Ramqvist, T., et al., *Analysis of human papillomaviruses and human polyomaviruses in lung cancer from Swedish never-smokers.* Acta Oncol, 2020. 59(1): p. 28-32.
23. Fidler-Benaoudia, M.M., et al., *Lung cancer incidence in young women vs. young men: A systematic analysis in 40 countries.* Int J Cancer, 2020. 147(3): p. 811-819.
24. Wakelee, H.A., et al., *Lung cancer incidence in never smokers.* J Clin Oncol, 2007. 25(5): p. 472-8.
25. Hsu, L.H., N.M. Chu, and S.H. Kao, *Estrogen, Estrogen Receptor and Lung Cancer.* Int J Mol Sci, 2017. 18(8).
26. Brenner, D.R., et al., *Previous lung diseases and lung cancer risk: a pooled analysis from the International Lung Cancer Consortium.* Am J Epidemiol, 2012. 176(7): p. 573-85.
27. *Nationell lungcancerrapport för diagnosår 2013-2017. Available at* https://cancercentrum.se/samverkan/cancerdiagnoser/lunga-och-lungsack/kvalitetsregister/rapporter/.
28. Navada S, L.P., Schwartz AG, Kalemkerian GP. Temporal trends in small cell lung cancer: analysis of the national Surveillance Epidemiology and End-Results (SEER) database [abstract 7082] J Clin Oncol. 2006;24(18S) suppl:384S.
29. Travis WD, B.E., Müller-Hermelink HK, Harris CC. Pathology and Genetics of Tumours of the Lung, Pleura, Thymus and Heart: World Health Organization Classification of Tumours. Lyon, France: IARC Press; 2004.
30. Travis WD, W.H.O., International Agency for Research on Cancer, International Association for the Study of Lung Cancer, International Academy of Pathology. WHO Classification of Tumours of the Lung, Pleura, Thymus and Heart. IARC Press, Oxford University Press (distributor), Lyon and Oxford (2015).

31. Travis, W.D., et al., *The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification.* J Thorac Oncol, 2015. 10(9): p. 1243-1260.
32. Nicholson, A.G., et al., *The 2021 WHO Classification of Lung Tumors: Impact of advances since 2015.* J Thorac Oncol, 2021.
33. Khuder, S.A., *Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis.* Lung Cancer, 2001. 31(2-3): p. 139-48.
34. Greulich, H., *The genomics of lung adenocarcinoma: opportunities for targeted therapies.* Genes Cancer, 2010. 1(12): p. 1200-10.
35. Moreira, A.L., et al., *A Grading System for Invasive Pulmonary Adenocarcinoma: A Proposal From the International Association for the Study of Lung Cancer Pathology Committee.* J Thorac Oncol, 2020. 15(10): p. 1599-1610.
36. Lortet-Tieulent, J., et al., *International trends in lung cancer incidence by histological subtype: adenocarcinoma stabilizing in men but still increasing in women.* Lung Cancer, 2014. 84(1): p. 13-22.
37. Thun, M.J., et al., *Cigarette smoking and changes in the histopathology of lung cancer.* J Natl Cancer Inst, 1997. 89(21): p. 1580-6.
38. Cheng, T.Y., et al., *The International Epidemiology of Lung Cancer: Latest Trends, Disparities, and Tumor Characteristics.* J Thorac Oncol, 2016. 11(10): p. 1653-71.
39. *Comprehensive genomic characterization of squamous cell lung cancers.* Nature, 2012. 489(7417): p. 519-25.
40. Travis WD, B.E., Burke AP, Marx A, Nicholson AG (ed). *WHO Classification of Tumours of the Lung, Pleura, Thymus and Heart.* 4th ed. Lyon, France: IARC Press; 2015.
41. Rekhtman, N., et al., *Distinct profile of driver mutations and clinical features in immunomarker-defined subsets of pulmonary large-cell carcinoma.* Mod Pathol, 2013. 26(4): p. 511-22.
42. Rudin, C.M., et al., *Small-cell lung cancer.* Nat Rev Dis Primers, 2021. 7(1): p. 3.
43. Fasano, M., et al., *Pulmonary Large-Cell Neuroendocrine Carcinoma: From Epidemiology to Therapy.* J Thorac Oncol, 2015. 10(8): p. 1133-41.
44. Bertino, E.M., et al., *Pulmonary neuroendocrine/carcinoid tumors: a review article.* Cancer, 2009. 115(19): p. 4434-41.
45. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. 100(1): p. 57-70.
46. Hanahan, D., *Hallmarks of Cancer: New Dimensions.* Cancer Discov, 2022. 12(1): p. 31-46.
47. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. 144(5): p. 646-74.
48. Shao, X., et al., *Copy number variation is highly correlated with differential gene expression: a pan-cancer study.* BMC Med Genet, 2019. 20(1): p. 175.

49. Mitelman, F., B. Johansson, and F. Mertens, *The impact of translocations and gene fusions on cancer causation.* Nat Rev Cancer, 2007. 7(4): p. 233-45.
50. Rosenbaum, J.N., et al., *Genomic heterogeneity of ALK fusion breakpoints in non-small-cell lung cancer.* Mod Pathol, 2018. 31(5): p. 791-808.
51. Esteller, M., *Cancer epigenomics: DNA methylomes and histone-modification maps.* Nat Rev Genet, 2007. 8(4): p. 286-98.
52. Pfeifer, G.P. and T.A. Rauch, *DNA methylation patterns in lung carcinomas.* Semin Cancer Biol, 2009. 19(3): p. 181-7.
53. Wurth, L., *Versatility of RNA-Binding Proteins in Cancer.* Comp Funct Genomics, 2012. 2012: p. 178525.
54. Lamberti, G., et al., *The Mechanisms of PD-L1 Regulation in Non-Small-Cell Lung Cancer (NSCLC): Which Are the Involved Players?* Cancers (Basel), 2020. 12(11).
55. Greaves, M. and C.C. Maley, *Clonal evolution in cancer.* Nature, 2012. 481(7381): p. 306-13.
56. Govindan, R., et al., *Genomic landscape of non-small cell lung cancer in smokers and never-smokers.* Cell, 2012. 150(6): p. 1121-34.
57. Subramanian, J. and R. Govindan, *Molecular profile of lung cancer in never smokers.* EJC Suppl, 2013. 11(2): p. 248-53.
58. McCubrey, J.A., et al., *Roles of the Raf/MEK/ERK pathway in cell growth, malignant transformation and drug resistance.* Biochim Biophys Acta, 2007. 1773(8): p. 1263-84.
59. Vivanco, I. and C.L. Sawyers, *The phosphatidylinositol 3-Kinase AKT pathway in human cancer.* Nat Rev Cancer, 2002. 2(7): p. 489-501.
60. Castellanos, E., E. Feld, and L. Horn, *Driven by Mutations: The Predictive Value of Mutation Subtype in EGFR-Mutated Non-Small Cell Lung Cancer.* J Thorac Oncol, 2017. 12(4): p. 612-623.
61. 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. 11(3): p. 217-27.
62. Tamirat, M.Z., et al., *Structural characterization of EGFR exon 19 deletion mutation using molecular dynamics simulation.* PLoS One, 2019. 14(9): p. e0222814.
63. 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, 2016. 7(48): p. 78985-78993.
64. Dogan, S., et al., *Molecular epidemiology of EGFR and KRAS mutations in 3,026 lung adenocarcinomas: higher susceptibility of women to smoking-related KRAS-mutant cancers.* Clin Cancer Res, 2012. 18(22): p. 6169-77.
65. Riely, G.J., et al., *Frequency and distinctive spectrum of KRAS mutations in never smokers with lung adenocarcinoma.* Clin Cancer Res, 2008. 14(18): p. 5731-4.

66. Marchetti, A., et al., *Clinical features and outcome of patients with non-small-cell lung cancer harboring BRAF mutations.* J Clin Oncol, 2011. 29(26): p. 3574-9.
67. Litvak, A.M., et al., *Clinical characteristics and course of 63 patients with BRAF mutant lung cancers.* J Thorac Oncol, 2014. 9(11): p. 1669-74.
68. Riudavets, M., et al., *Targeting HER2 in non-small-cell lung cancer (NSCLC): a glimpse of hope? An updated review on therapeutic strategies in NSCLC harbouring HER2 alterations.* ESMO Open, 2021. 6(5): p. 100260.
69. Zhu, F., et al., *Concealed driver in lung adenocarcinoma with single PIK3CA mutation: a case report and single-center genotyping review.* Ann Transl Med, 2021. 9(3): p. 271.
70. Organ, S.L. and M.S. Tsao, *An overview of the c-MET signaling pathway.* Ther Adv Med Oncol, 2011. 3(1 Suppl): p. S7-s19.
71. Bylicki, O., et al., *Targeting the MET-Signaling Pathway in Non-Small-Cell Lung Cancer: Evidence to Date.* Onco Targets Ther, 2020. 13: p. 5691-5706.
72. Kong-Beltran, M., et al., *Somatic mutations lead to an oncogenic deletion of met in lung cancer.* Cancer Res, 2006. 66(1): p. 283-9.
73. Frampton, G.M., et al., *Activation of MET via diverse exon 14 splicing alterations occurs in multiple tumor types and confers clinical sensitivity to MET inhibitors.* Cancer Discov, 2015. 5(8): p. 850-9.
74. Hallberg, B. and R.H. Palmer, *Mechanistic insight into ALK receptor tyrosine kinase in human cancer biology.* Nat Rev Cancer, 2013. 13(10): p. 685-700.
75. Chapman, A.M., et al., *Lung cancer mutation profile of EGFR, ALK, and KRAS: Meta-analysis and comparison of never and ever smokers.* Lung Cancer, 2016. 102: p. 122-134.
76. Childress, M.A., et al., *ALK Fusion Partners Impact Response to ALK Inhibition: Differential Effects on Sensitivity, Cellular Phenotypes, and Biochemical Properties.* Mol Cancer Res, 2018. 16(11): p. 1724-1736.
77. Yoshida, T., et al., *Differential Crizotinib Response Duration Among ALK Fusion Variants in ALK-Positive Non-Small-Cell Lung Cancer.* J Clin Oncol, 2016. 34(28): p. 3383-9.
78. Lin, J.J., et al., *Impact of EML4-ALK Variant on Resistance Mechanisms and Clinical Outcomes in ALK-Positive Lung Cancer.* J Clin Oncol, 2018. 36(12): p. 1199-1206.
79. Chu, Q.S., *Targeting non-small cell lung cancer: driver mutation beyond epidermal growth factor mutation and anaplastic lymphoma kinase fusion.* Ther Adv Med Oncol, 2020. 12: p. 1758835919895756.
80. Stein, R.A. and J.V. Staros, *Insights into the evolution of the ErbB receptor family and their ligands from sequence analysis.* BMC Evol Biol, 2006. 6: p. 79.

81. Guo, Y., et al., *Recent Progress in Rare Oncogenic Drivers and Targeted Therapy For Non-Small Cell Lung Cancer.* Onco Targets Ther, 2019. 12: p. 10343-10360.
82. Sulu, E., et al., *Delays in the diagnosis and treatment of non-small-cell lung cancer.* Tumori, 2011. 97(6): p. 693-7.
83. Quadrelli, S., et al., *Clinical characteristics and prognosis of incidentally detected lung cancers.* Int J Surg Oncol, 2015. 2015: p. 287604.
84. Oken, M.M., et al., *Toxicity and response criteria of the Eastern Cooperative Oncology Group.* Am J Clin Oncol, 1982. 5(6): p. 649-55.
85. Cuyún Carter, G., et al., *A comprehensive review of nongenetic prognostic and predictive factors influencing the heterogeneity of outcomes in advanced non-small-cell lung cancer.* Cancer Manag Res, 2014. 6: p. 437-49.
86. Brundage, M.D., D. Davies, and W.J. Mackillop, *Prognostic factors in non-small cell lung cancer: a decade of progress.* Chest, 2002. 122(3): p. 1037-57.
87. Brierley, J.D., Gospodarowicz, M.K., Wittekind, C., Eds. (2016). International Union against Cancer (UICC): TNM Classification of Malignant Tumours. 8th Edition, Wiley-Blackwell, Hoboken.
88. Sobin, L.H., Gospodarowicz, M.K. and Wittekind, C. (2009) International Union against Cancer (UICC): TNM Classification of Malignant Tumours. 7th Edition, Wiley-Blackwell, Chicester.
89. 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, 2016. 11(1): p. 39-51.
90. Aberle, D.R., et al., *Reduced lung-cancer mortality with low-dose computed tomographic screening.* N Engl J Med, 2011. 365(5): p. 395-409.
91. de Koning, H.J., et al., *Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial.* N Engl J Med, 2020. 382(6): p. 503-513.
92. Chen, F.F., et al., *Video-assisted thoracoscopic surgery lobectomy versus open lobectomy in patients with clinical stage Ⅰ non-small cell lung cancer: a meta-analysis.* Eur J Surg Oncol, 2013. 39(9): p. 957-63.
93. Bendixen, M., et al., *Postoperative pain and quality of life after lobectomy via video-assisted thoracoscopic surgery or anterolateral thoracotomy for early stage lung cancer: a randomised controlled trial.* Lancet Oncol, 2016. 17(6): p. 836-844.
94. Lim, E., et al., *Study protocol for VIdeo assisted thoracoscopic lobectomy versus conventional Open LobEcTomy for lung cancer, a UK multicentre randomised controlled trial with an internal pilot (the VIOLET study).* BMJ Open, 2019. 9(10): p. e029507.
95. Shagabayeva, L., et al., *Open, Video- and Robot-Assisted Thoracoscopic Lobectomy for Stage II-IIIA Non-Small Cell Lung Cancer.* Ann Thorac Surg, 2022.

96. Ginsberg, R.J. and L.V. Rubinstein, *Randomized trial of lobectomy versus limited resection for T1 N0 non-small cell lung cancer. Lung Cancer Study Group.* Ann Thorac Surg, 1995. 60(3): p. 615-22; discussion 622-3.
97. Winckelmans, T., et al., *Segmentectomy or lobectomy for early-stage non-small-cell lung cancer: a systematic review and meta-analysis.* Eur J Cardiothorac Surg, 2020. 57(6): p. 1051-1060.
98. Altorki, N.K., et al., *Perioperative mortality and morbidity after sublobar versus lobar resection for early-stage non-small-cell lung cancer: post-hoc analysis of an international, randomised, phase 3 trial (CALGB/Alliance 140503).* Lancet Respir Med, 2018. 6(12): p. 915-924.
99. Nakamura, K., et al., *A phase III randomized trial of lobectomy versus limited resection for small-sized peripheral non-small cell lung cancer (JCOG0802/WJOG4607L).* Jpn J Clin Oncol, 2010. 40(3): p. 271-4.
100. Mokhles, S., et al., *Systematic lymphadenectomy versus sampling of ipsilateral mediastinal lymph-nodes during lobectomy for non-small-cell lung cancer: a systematic review of randomized trials and a meta-analysis.* Eur J Cardiothorac Surg, 2017. 51(6): p. 1149-1156.
101. Darling, G.E., et al., *Randomized trial of mediastinal lymph node sampling versus complete lymphadenectomy during pulmonary resection in the patient with N0 or N1 (less than hilar) non-small cell carcinoma: results of the American College of Surgery Oncology Group Z0030 Trial.* J Thorac Cardiovasc Surg, 2011. 141(3): p. 662-70.
102. Chang, J.Y., et al., *Stereotactic ablative radiotherapy versus lobectomy for operable stage I non-small-cell lung cancer: a pooled analysis of two randomised trials.* Lancet Oncol, 2015. 16(6): p. 630-7.
103. Ijsseldijk, M.A., et al., *Oncologic Outcomes of Surgery Versus SBRT for Non-Small-Cell Lung Carcinoma: A Systematic Review and Meta-analysis.* Clin Lung Cancer, 2020.
104. O'Rourke, N., et al., *Concurrent chemoradiotherapy in non-small cell lung cancer.* Cochrane Database Syst Rev, 2010(6): p. Cd002140.
105. Aupérin, A., et al., *Meta-analysis of concomitant versus sequential radiochemotherapy in locally advanced non-small-cell lung cancer.* J Clin Oncol, 2010. 28(13): p. 2181-90.
106. Fairchild, A., et al., *Palliative thoracic radiotherapy for lung cancer: a systematic review.* J Clin Oncol, 2008. 26(24): p. 4001-11.
107. Burdett, S., et al., *Adjuvant chemotherapy for resected early-stage non-small cell lung cancer.* Cochrane Database Syst Rev, 2015(3): p. Cd011430.
108. Pignon, J.P., et al., *Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE Collaborative Group.* J Clin Oncol, 2008. 26(21): p. 3552-9.
109. Lim, E., et al., *Preoperative versus postoperative chemotherapy in patients with resectable non-small cell lung cancer: systematic review and indirect comparison meta-analysis of randomized trials.* J Thorac Oncol, 2009. 4(11): p. 1380-8.
110. Scagliotti, G.V., et al., *Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients*

with advanced-stage non-small-cell lung cancer. J Clin Oncol, 2008. 26(21): p. 3543-51.
111. 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. 350(21): p. 2129-39.
112. Mok, T.S., et al., *Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma.* N Engl J Med, 2009. 361(10): p. 947-57.
113. Zhou, C., et al., *Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study.* Lancet Oncol, 2011. 12(8): p. 735-42.
114. Sequist, L.V., et al., *Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations.* J Clin Oncol, 2013. 31(27): p. 3327-34.
115. Huang, L. and L. Fu, *Mechanisms of resistance to EGFR tyrosine kinase inhibitors.* Acta Pharm Sin B, 2015. 5(5): p. 390-401.
116. Westover, D., et al., *Mechanisms of acquired resistance to first- and second-generation EGFR tyrosine kinase inhibitors.* Ann Oncol, 2018. 29(suppl_1): p. i10-i19.
117. Cross, D.A., et al., *AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer.* Cancer Discov, 2014. 4(9): p. 1046-61.
118. Soria, J.C., et al., *Osimertinib in Untreated EGFR-Mutated Advanced Non-Small-Cell Lung Cancer.* N Engl J Med, 2018. 378(2): p. 113-125.
119. Mu, Y., et al., *Acquired resistance to osimertinib in patients with non-small-cell lung cancer: mechanisms and clinical outcomes.* J Cancer Res Clin Oncol, 2020. 146(9): p. 2427-2433.
120. Wu, Y.L., et al., *Osimertinib in Resected EGFR-Mutated Non-Small-Cell Lung Cancer.* N Engl J Med, 2020. 383(18): p. 1711-1723.
121. Koch, A.L., et al., *FDA Approval Summary: Osimertinib for adjuvant treatment of surgically resected non-small cell lung cancer, a collaborative Project Orbis review.* Clin Cancer Res, 2021.
122. Sun, L., et al., *Neoadjuvant EGFR-TKI Therapy for EGFR-Mutant NSCLC: A Systematic Review and Pooled Analysis of Five Prospective Clinical Trials.* Front Oncol, 2020. 10: p. 586596.
123. Tsuboi, M., et al., *Neoadjuvant osimertinib with/without chemotherapy versus chemotherapy alone for EGFR-mutated resectable non-small-cell lung cancer: NeoADAURA.* Future Oncol, 2021. 17(31): p. 4045-4055.
124. Solomon, B.J., et al., *First-line crizotinib versus chemotherapy in ALK-positive lung cancer.* N Engl J Med, 2014. 371(23): p. 2167-77.
125. Wang, L. and W. Wang, *Safety and efficacy of anaplastic lymphoma kinase tyrosine kinase inhibitors in non-small cell lung cancer (Review).* Oncol Rep, 2021. 45(1): p. 13-28.
126. Lin, J.J., G.J. Riely, and A.T. Shaw, *Targeting ALK: Precision Medicine Takes on Drug Resistance.* Cancer Discov, 2017. 7(2): p. 137-155.

127. Planchard, D., et al., *Dabrafenib plus trametinib in patients with previously untreated BRAF(V600E)-mutant metastatic non-small-cell lung cancer: an open-label, phase 2 trial.* Lancet Oncol, 2017. 18(10): p. 1307-1316.
128. Planchard, D., et al., *Dabrafenib plus trametinib in patients with previously treated BRAF(V600E)-mutant metastatic non-small cell lung cancer: an open-label, multicentre phase 2 trial.* Lancet Oncol, 2016. 17(7): p. 984-993.
129. van Veggel, B., et al., *Osimertinib treatment for patients with EGFR exon 20 mutation positive non-small cell lung cancer.* Lung Cancer, 2020. 141: p. 9-13.
130. Vyse, S. and P.H. Huang, *Targeting EGFR exon 20 insertion mutations in non-small cell lung cancer.* Signal Transduct Target Ther, 2019. 4: p. 5.
131. Zhou, C., et al., *Treatment Outcomes and Safety of Mobocertinib in Platinum-Pretreated Patients With EGFR Exon 20 Insertion-Positive Metastatic Non-Small Cell Lung Cancer: A Phase 1/2 Open-label Nonrandomized Clinical Trial.* JAMA Oncol, 2021. 7(12): p. e214761.
132. Park, K., et al., *Amivantamab in EGFR Exon 20 Insertion-Mutated Non-Small-Cell Lung Cancer Progressing on Platinum Chemotherapy: Initial Results From the CHRYSALIS Phase I Study.* J Clin Oncol, 2021. 39(30): p. 3391-3402.
133. Drilon, A., et al., *Efficacy of Selpercatinib in RET Fusion-Positive Non-Small-Cell Lung Cancer.* N Engl J Med, 2020. 383(9): p. 813-824.
134. Gainor, J.F., et al., *Pralsetinib for RET fusion-positive non-small-cell lung cancer (ARROW): a multi-cohort, open-label, phase 1/2 study.* Lancet Oncol, 2021. 22(7): p. 959-969.
135. Drilon, A., et al., *Antitumor activity of crizotinib in lung cancers harboring a MET exon 14 alteration.* Nat Med, 2020. 26(1): p. 47-51.
136. Paik, P.K., et al., *Tepotinib in Non-Small-Cell Lung Cancer with MET Exon 14 Skipping Mutations.* N Engl J Med, 2020. 383(10): p. 931-943.
137. Wolf, J., et al., *Capmatinib in MET Exon 14-Mutated or MET-Amplified Non-Small-Cell Lung Cancer.* N Engl J Med, 2020. 383(10): p. 944-957.
138. Laskin, J., et al., *NRG1 fusion-driven tumors: biology, detection, and the therapeutic role of afatinib and other ErbB-targeting agents.* Ann Oncol, 2020. 31(12): p. 1693-1703.
139. Ostrem, J.M., et al., *K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions.* Nature, 2013. 503(7477): p. 548-51.
140. Li BT, S.F., Falchook G, et al: CodeBreak 100: Registrational phase 2 trial of sotorasib in KRAS p.G12C mutated non-small cell lung cancer. 2020 World Conference on Lung Cancer. Abstract PS01.07. Presented January 30, 2021.
141. Passiglia, F., et al., *KRAS inhibition in non-small cell lung cancer: Past failures, new findings and upcoming challenges.* Eur J Cancer, 2020. 137: p. 57-68.
142. Jacobs, F., et al., *Targeting KRAS in NSCLC: Old Failures and New Options for "Non-G12c" Patients.* Cancers (Basel), 2021. 13(24).

143. Cardarella, S., et al., *Clinical, pathologic, and biologic features associated with BRAF mutations in non-small cell lung cancer.* Clin Cancer Res, 2013. 19(16): p. 4532-40.
144. Lin, L., et al., *Mapping the molecular determinants of BRAF oncogene dependence in human lung cancer.* Proc Natl Acad Sci U S A, 2014. 111(7): p. E748-57.
145. Dankner, M., et al., *Classifying BRAF alterations in cancer: new rational therapeutic strategies for actionable mutations.* Oncogene, 2018. 37(24): p. 3183-3199.
146. Jebbink, M., et al., *The force of HER2 - A druggable target in NSCLC?* Cancer Treat Rev, 2020. 86: p. 101996.
147. Chaft, J.E., et al., *Coexistence of PIK3CA and other oncogene mutations in lung adenocarcinoma-rationale for comprehensive mutation profiling.* Mol Cancer Ther, 2012. 11(2): p. 485-91.
148. Scheffler, M., et al., *PIK3CA mutations in non-small cell lung cancer (NSCLC): genetic heterogeneity, prognostic impact and incidence of prior malignancies.* Oncotarget, 2015. 6(2): p. 1315-26.
149. Tan, A.C., *Targeting the PI3K/Akt/mTOR pathway in non-small cell lung cancer (NSCLC).* Thorac Cancer, 2020. 11(3): p. 511-518.
150. Colt, H.G., et al., *Follow-up and surveillance of the patient with lung cancer after curative-intent therapy: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines.* Chest, 2013. 143(5 Suppl): p. e437S-e454S.
151. Schmidt-Hansen, M., D.R. Baldwin, and E. Hasler, *What is the most effective follow-up model for lung cancer patients? A systematic review.* J Thorac Oncol, 2012. 7(5): p. 821-4.
152. Sonoda, D., et al., *Ultra-late recurrence of non-small cell lung cancer over 10 years after curative resection.* Cancer Manag Res, 2019. 11: p. 6765-6774.
153. Maeda, R., et al., *Late recurrence of non-small cell lung cancer more than 5 years after complete resection: incidence and clinical implications in patient follow-up.* Chest, 2010. 138(1): p. 145-50.
154. Borczuk, A.C., *Prognostic considerations of the new World Health Organization classification of lung adenocarcinoma.* Eur Respir Rev, 2016. 25(142): p. 364-371.
155. Russell, P.A., et al., *Does lung adenocarcinoma subtype predict patient survival?: A clinicopathologic study based on the new International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary lung adenocarcinoma classification.* J Thorac Oncol, 2011. 6(9): p. 1496-504.
156. Zhu, C.Q., et al., *Immunohistochemical markers of prognosis in non-small cell lung cancer: a review and proposal for a multiphase approach to marker evaluation.* J Clin Pathol, 2006. 59(8): p. 790-800.
157. Zhu, C.Q. and M.S. Tsao, *Prognostic markers in lung cancer: is it ready for prime time?* Transl Lung Cancer Res, 2014. 3(3): p. 149-58.

158. Sørlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proc Natl Acad Sci U S A, 2001. 98(19): p. 10869-74.
159. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. 406(6797): p. 747-52.
160. Karlsson, A., et al., *Genome-wide DNA methylation analysis of lung carcinoma reveals one neuroendocrine and four adenocarcinoma epitypes associated with patient outcome.* Clin Cancer Res, 2014. 20(23): p. 6127-40.
161. Steels, E., et al., *Role of p53 as a prognostic factor for survival in lung cancer: a systematic review of the literature with a meta-analysis.* Eur Respir J, 2001. 18(4): p. 705-19.
162. Zhang, M., et al., *PD-L1 expression in lung cancer and its correlation with driver mutations: a meta-analysis.* Sci Rep, 2017. 7(1): p. 10255.
163. Suda, K. and T. Mitsudomi, *Role of EGFR mutations in lung cancers: prognosis and tumor chemosensitivity.* Arch Toxicol, 2015. 89(8): p. 1227-40.
164. Couraud, S., et al., *Lung cancer in never smokers--a review.* Eur J Cancer, 2012. 48(9): p. 1299-311.
165. Staaf, J., et al., *Relation between smoking history and gene expression profiles in lung adenocarcinomas.* BMC Med Genomics, 2012. 5: p. 22.
166. Karlsson, A., et al., *Genomic and transcriptional alterations in lung adenocarcinoma in relation to smoking history.* Clin Cancer Res, 2014. 20(18): p. 4912-24.
167. White, M.C., et al., *Age and cancer risk: a potentially modifiable relationship.* Am J Prev Med, 2014. 46(3 Suppl 1): p. S7-15.
168. Löfling, L., et al., *Clinical characteristics and survival in non-small cell lung cancer patients by smoking history: a population-based cohort study.* Acta Oncol, 2019. 58(11): p. 1618-1627.
169. Liu, N.S., et al., *Adenocarcinoma of the lung in young patients: the M. D. Anderson experience.* Cancer, 2000. 88(8): p. 1837-41.
170. Liam, C.K., K.H. Lim, and C.M. Wong, *Lung cancer in patients younger than 40 years in a multiracial Asian country.* Respirology, 2000. 5(4): p. 355-61.
171. Cassidy, A., et al., *Family history and risk of lung cancer: age-at-diagnosis in cases and first-degree relatives.* Br J Cancer, 2006. 95(9): p. 1288-90.
172. Dias, M., et al., *Lung cancer in never-smokers - what are the differences?* Acta Oncol, 2017. 56(7): p. 931-935.
173. Toh, C.K., et al., *Never-smokers with lung cancer: epidemiologic evidence of a distinct disease entity.* J Clin Oncol, 2006. 24(15): p. 2245-51.
174. Kosaka, T., et al., *Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications.* Cancer Res, 2004. 64(24): p. 8919-23.

175. 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. 27(26): p. 4247-53.
176. Toyooka, S., et al., *Smoke exposure, histologic type and geography-related differences in the methylation profiles of non-small cell lung cancer.* Int J Cancer, 2003. 103(2): p. 153-60.
177. Takeuchi, T., et al., *Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors.* J Clin Oncol, 2006. 24(11): p. 1679-88.
178. WD Travis, E.B., HK Müller-Hermelink, CC Harris. Pathology and Genetics: Tumours of the Lung, Pleura, Thymus and Heart, IARC, Lyon (2004).
179. Brunnström, H., et al., *Immunohistochemistry in the differential diagnostics of primary lung cancer: an investigation within the Southern Swedish Lung Cancer Study.* Am J Clin Pathol, 2013. 140(1): p. 37-46.
180. La Fleur, L., et al., *Mutation patterns in a population-based non-small cell lung cancer cohort and prognostic impact of concomitant mutations in KRAS and TP53 or STK11.* Lung Cancer, 2019. 130: p. 50-58.
181. Tran, L., et al., *Various Antibody Clones of Napsin A, Thyroid Transcription Factor 1, and p40 and Comparisons With Cytokeratin 5 and p63 in Histopathologic Diagnostics of Non-Small Cell Lung Carcinoma.* Appl Immunohistochem Mol Morphol, 2016. 24(9): p. 648-659.
182. Oberländer, M., et al., *A 'waterfall' transfer-based workflow for improved quality of tissue microarray construction and processing in breast cancer research.* Pathol Oncol Res, 2014. 20(3): p. 719-26.
183. Engel, K.B. and H.M. Moore, *Effects of preanalytical variables on the detection of proteins by immunohistochemistry in formalin-fixed, paraffin-embedded tissue.* Arch Pathol Lab Med, 2011. 135(5): p. 537-43.
184. Pfaffl, M.W., *A new mathematical model for relative quantification in real-time RT-PCR.* Nucleic Acids Res, 2001. 29(9): p. e45.
185. Vandesompele, J., et al., *Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes.* Genome Biol, 2002. 3(7): p. Research0034.
186. *Nanostring Technologies. Available at* https://www.nanostring.com.
187. Lira, M.E., et al., *A single-tube multiplexed assay for detecting ALK, ROS1, and RET fusions in lung cancer.* J Mol Diagn, 2014. 16(2): p. 229-43.
188. Lennartsson, J. and L. Rönnstrand, *The stem cell factor receptor/c-Kit as a drug target in cancer.* Curr Cancer Drug Targets, 2006. 6(1): p. 65-75.
189. Casali, P.G., et al., *Gastrointestinal stromal tumours: ESMO-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Ann Oncol, 2018. 29(Suppl 4): p. iv68-iv78.
190. Ramos, A.H., et al., *Amplification of chromosomal segment 4q12 in non-small cell lung cancer.* Cancer Biol Ther, 2009. 8(21): p. 2042-50.

191. Donnenberg, A.D., et al., *KIT (CD117) expression in a subset of non-small cell lung carcinoma (NSCLC) patients.* PLoS One, 2012. 7(12): p. e52885.
192. Pietsch, T., et al., *Expression of the c-Kit receptor and its ligand SCF in non-small-cell lung carcinomas.* Int J Cancer, 1998. 75(2): p. 171-5.
193. Jansson, S., et al., *The three receptor tyrosine kinases c-KIT, VEGFR2 and PDGFRα, closely spaced at 4q12, show increased protein expression in triple-negative breast cancer.* PLoS One, 2014. 9(7): p. e102176.
194. Holtkamp, N., et al., *Characterization of the amplicon on chromosomal segment 4q12 in glioblastoma multiforme.* Neuro Oncol, 2007. 9(3): p. 291-7.
195. Geiger, T., J. Cox, and M. Mann, *Proteomic changes resulting from gene copy number variations in cancer cells.* PLoS Genet, 2010. 6(9): p. e1001090.
196. Maier, T., M. Güell, and L. Serrano, *Correlation of mRNA and protein in complex biological samples.* FEBS Lett, 2009. 583(24): p. 3966-73.
197. Janostiak, R., et al., *Loss of c-KIT expression in breast cancer correlates with malignant transformation of breast epithelium and is mediated by KIT gene promoter DNA hypermethylation.* Exp Mol Pathol, 2018. 105(1): p. 41-49.
198. Nobusawa, S., et al., *Amplification of the PDGFRA, KIT and KDR genes in glioblastoma: a population-based study.* Neuropathology, 2011. 31(6): p. 583-8.
199. Matsumura, Y., et al., *Expression profiling of receptor tyrosine kinases in high-grade neuroendocrine carcinoma of the lung: a comparative analysis with adenocarcinoma and squamous cell carcinoma.* J Cancer Res Clin Oncol, 2015. 141(12): p. 2159-70.
200. Araki, K., et al., *Frequent overexpression of the c-kit protein in large cell neuroendocrine carcinoma of the lung.* Lung Cancer, 2003. 40(2): p. 173-80.
201. Casali, C., et al., *The prognostic role of c-kit protein expression in resected large cell neuroendocrine carcinoma of the lung.* Ann Thorac Surg, 2004. 77(1): p. 247-52; discussion 252-3.
202. Pelosi, G., et al., *CD117 immunoreactivity in high-grade neuroendocrine tumors of the lung: a comparative study of 39 large-cell neuroendocrine carcinomas and 27 surgically resected small-cell carcinomas.* Virchows Arch, 2004. 445(5): p. 449-55.
203. Sihto, H., et al., *KIT and platelet-derived growth factor receptor alpha tyrosine kinase gene mutations and KIT amplifications in human solid tumors.* J Clin Oncol, 2005. 23(1): p. 49-57.
204. Iyoda, A., et al., *Expression profiling and identification of potential molecular targets for therapy in pulmonary large-cell neuroendocrine carcinoma.* Exp Ther Med, 2011. 2(6): p. 1041-1045.
205. Spigel, D.R., et al., *Irinotecan, carboplatin, and imatinib in untreated extensive-stage small-cell lung cancer: a phase II trial of the Minnie Pearl Cancer Research Network.* J Thorac Oncol, 2007. 2(9): p. 854-61.

206. Schneider, B.J., et al., *Phase II trial of imatinib maintenance therapy after irinotecan and cisplatin in patients with c-Kit-positive, extensive-stage small-cell lung cancer.* Clin Lung Cancer, 2010. 11(4): p. 223-7.
207. Dy, G.K., et al., *A phase II trial of imatinib (ST1571) in patients with c-kit expressing relapsed small-cell lung cancer: a CALGB and NCCTG study.* Ann Oncol, 2005. 16(11): p. 1811-6.
208. Krug, L.M., et al., *Imatinib mesylate lacks activity in small cell lung carcinoma expressing c-kit protein: a phase II clinical trial.* Cancer, 2005. 103(10): p. 2128-31.
209. Soria, J.C., B.E. Johnson, and T.L. Chevalier, *Imatinib in small cell lung cancer.* Lung Cancer, 2003. 41 Suppl 1: p. S49-53.
210. Dziadziuszko, R., et al., *An Activating KIT Mutation Induces Crizotinib Resistance in ROS1-Positive Lung Cancer.* J Thorac Oncol, 2016. 11(8): p. 1273-1281.
211. Katayama, R., et al., *Mechanisms of acquired crizotinib resistance in ALK-rearranged lung Cancers.* Sci Transl Med, 2012. 4(120): p. 120ra17.
212. Shen, G., et al., *Anlotinib: a novel multi-targeting tyrosine kinase inhibitor in clinical development.* J Hematol Oncol, 2018. 11(1): p. 120.
213. Su, Y., et al., *Clinical and Prognostic Significance of CD117 in Non-Small Cell Lung Cancer: A Systemic Meta-Analysis.* Pathobiology, 2021. 88(4): p. 267-276.
214. *Available at* https://cancercentrum.se/samverkan/cancerdiagnoser/lunga-och-lungsack/kvalitetsregister/.
215. Couraud, S., et al., *BioCAST/IFCT-1002: epidemiological and molecular features of lung cancer in never-smokers.* Eur Respir J, 2015. 45(5): p. 1403-14.
216. Isaksson, S., et al., *Clinical Utility of Targeted Sequencing in Lung Cancer: Experience From an Autonomous Swedish Health Care Center.* JTO Clin Res Rep, 2020. 1(1): p. 100013.
217. Taylor, M.D., et al., *Tumor recurrence after complete resection for non-small cell lung cancer.* Ann Thorac Surg, 2012. 93(6): p. 1813-20; discussion 1820-1.
218. Derry, J.M., J.A. Kerns, and U. Francke, *RBM3, a novel human gene in Xp11.23 with a putative RNA-binding domain.* Hum Mol Genet, 1995. 4(12): p. 2307-11.
219. Zhu, X., C. Bührer, and S. Wellmann, *Cold-inducible proteins CIRP and RBM3, a unique couple with activities far beyond the cold.* Cell Mol Life Sci, 2016. 73(20): p. 3839-59.
220. Smart, F., et al., *Two isoforms of the cold-inducible mRNA-binding protein RBM3 localize to dendrites and promote translation.* J Neurochem, 2007. 101(5): p. 1367-79.
221. Dresios, J., et al., *Cold stress-induced protein Rbm3 binds 60S ribosomal subunits, alters microRNA levels, and enhances global protein synthesis.* Proc Natl Acad Sci U S A, 2005. 102(6): p. 1865-70.

222. Sureban, S.M., et al., *Translation regulatory factor RBM3 is a proto-oncogene that prevents mitotic catastrophe.* Oncogene, 2008. 27(33): p. 4544-56.
223. Yang, H.J., et al., *RNA-binding protein RBM3 prevents NO-induced apoptosis in human neuroblastoma cells by modulating p38 signaling and miR-143.* Sci Rep, 2017. 7: p. 41738.
224. Wellmann, S., et al., *The RNA-binding protein RBM3 is required for cell proliferation and protects against serum deprivation-induced cell death.* Pediatr Res, 2010. 67(1): p. 35-41.
225. Chip, S., et al., *The RNA-binding protein RBM3 is involved in hypothermia induced neuroprotection.* Neurobiol Dis, 2011. 43(2): p. 388-96.
226. Jonsson, L., et al., *High RBM3 expression in prostate cancer independently predicts a reduced risk of biochemical recurrence and disease progression.* Diagn Pathol, 2011. 6: p. 91.
227. Hjelm, B., et al., *High nuclear RBM3 expression is associated with an improved prognosis in colorectal cancer.* Proteomics Clin Appl, 2011. 5(11-12): p. 624-35.
228. Boman, K., et al., *Decreased expression of RNA-binding motif protein 3 correlates with tumour progression and poor prognosis in urothelial bladder cancer.* BMC Urol, 2013. 13: p. 17.
229. Olofsson, S.E., et al., *Low RBM3 protein expression correlates with clinical stage, prognostic classification and increased risk of treatment failure in testicular non-seminomatous germ cell cancer.* PLoS One, 2015. 10(3): p. e0121300.
230. Boman, K., et al., *Podocalyxin-like and RNA-binding motif protein 3 are prognostic biomarkers in urothelial bladder cancer: a validatory study.* Biomark Res, 2017. 5: p. 10.
231. Siesing, C., et al., *High RBM3 expression is associated with an improved survival and oxaliplatin response in patients with metastatic colorectal cancer.* PLoS One, 2017. 12(8): p. e0182512.
232. Jonsson, L., et al., *High expression of RNA-binding motif protein 3 in esophageal and gastric adenocarcinoma correlates with intestinal metaplasia-associated tumours and independently predicts a reduced risk of recurrence and death.* Biomark Res, 2014. 2: p. 11.
233. Ehlén, A., et al., *Expression of the RNA-binding protein RBM3 is associated with a favourable prognosis and cisplatin sensitivity in epithelial ovarian cancer.* J Transl Med, 2010. 8: p. 78.
234. Vidarsdottir, H., et al., *Clinical significance of RBM3 expression in surgically treated colorectal lung metastases and paired primary tumors.* J Surg Oncol, 2021. 123(4): p. 1144-1156.
235. Jögi, A., et al., *Nuclear expression of the RNA-binding protein RBM3 is associated with an improved clinical outcome in breast cancer.* Mod Pathol, 2009. 22(12): p. 1564-74.
236. Jonsson, L., et al., *Low RBM3 protein expression correlates with tumour progression and poor prognosis in malignant melanoma: an analysis of 215*

cases from the Malmö Diet and Cancer Study. J Transl Med, 2011. 9: p. 114.
237. Ehlén, Å., et al., *RBM3-regulated genes promote DNA integrity and affect clinical outcome in epithelial ovarian cancer.* Transl Oncol, 2011. 4(4): p. 212-21.
238. Karnevi, E., et al., *Translational study reveals a two-faced role of RBM3 in pancreatic cancer and suggests its potential value as a biomarker for improved patient stratification.* Oncotarget, 2018. 9(5): p. 6188-6200.
239. Salomonsson, A., et al., *Comprehensive analysis of RNA binding motif protein 3 (RBM3) in non-small cell lung cancer.* Cancer Med, 2020. 9(15): p. 5609-5619.
240. Melling, N., et al., *Prevalence and clinical significance of RBM3 immunostaining in non-small cell lung cancers.* J Cancer Res Clin Oncol, 2019. 145(4): p. 873-879.
241. Warth, A., et al., *Tumour cell proliferation (Ki-67) in non-small cell lung cancer: a critical reappraisal of its prognostic role.* Br J Cancer, 2014. 111(6): p. 1222-9.
242. Haga, Y., et al., *Ki-67 expression and prognosis for smokers with resected stage I non-small cell lung cancer.* Ann Thorac Surg, 2003. 75(6): p. 1727-32; discussion 1732-3.
243. Hommura, F., et al., *Prognostic significance of p27KIP1 protein and ki-67 growth fraction in non-small cell lung cancers.* Clin Cancer Res, 2000. 6(10): p. 4073-81.
244. Takahashi, S., et al., *Relationship between postoperative recurrence and expression of cyclin E, p27, and Ki-67 in non-small cell lung cancer without lymph node metastases.* Int J Clin Oncol, 2002. 7(6): p. 349-55.
245. Gerdes, J., et al., *Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67.* J Immunol, 1984. 133(4): p. 1710-5.
246. Jakobsen, J.N. and J.B. Sørensen, *Clinical impact of ki-67 labeling index in non-small cell lung cancer.* Lung Cancer, 2013. 79(1): p. 1-7.
247. Bochman, M.L. and A. Schwacha, *The Mcm2-7 complex has in vitro helicase activity.* Mol Cell, 2008. 31(2): p. 287-93.
248. Kikuchi, J., et al., *Minichromosome maintenance (MCM) protein 4 as a marker for proliferation and its clinical and clinicopathological significance in non-small cell lung cancer.* Lung Cancer, 2011. 72(2): p. 229-37.
249. Yu, S., et al., *MCMs in Cancer: Prognostic Potential and Mechanisms.* Anal Cell Pathol (Amst), 2020. 2020: p. 3750294.
250. Huang, C., et al., *Potential Prospective Biomarkers for Non-small Cell Lung Cancer: Mini-Chromosome Maintenance Proteins.* Front Genet, 2021. 12: p. 587017.
251. Liu, Y.Z., et al., *MCMs expression in lung cancer: implication of prognostic significance.* J Cancer, 2017. 8(18): p. 3641-3647.

252. Carreras, C.W. and D.V. Santi, *The catalytic mechanism and structure of thymidylate synthase.* Annu Rev Biochem, 1995. 64: p. 721-62.
253. Hu, H.B., et al., *Predictive value of thymidylate synthase expression in gastric cancer: a systematic review with meta-analysis.* Asian Pac J Cancer Prev, 2012. 13(1): p. 261-7.
254. Liu, Y., et al., *Expression of thymidylate synthase predicts clinical outcomes of pemetrexed-containing chemotherapy for non-small-cell lung cancer: a systemic review and meta-analysis.* Cancer Chemother Pharmacol, 2013. 72(5): p. 1125-32.
255. Liu, Q., et al., *Prognostic and predictive significance of thymidylate synthase protein expression in non-small cell lung cancer: a systematic review and meta-analysis.* Cancer Biomark, 2015. 15(1): p. 65-78.
256. Sánchez, I. and B.D. Dynlacht, *New insights into cyclins, CDKs, and cell cycle control.* Semin Cell Dev Biol, 2005. 16(3): p. 311-21.
257. Bolanos-Garcia, V.M. and T.L. Blundell, *BUB1 and BUBR1: multifaceted kinases of the cell cycle.* Trends Biochem Sci, 2011. 36(3): p. 141-50.
258. Koffa, M.D., et al., *HURP is part of a Ran-dependent complex involved in spindle formation.* Curr Biol, 2006. 16(8): p. 743-54.
259. Carleton, M., et al., *RNA interference-mediated silencing of mitotic kinesin KIF14 disrupts cell cycle progression and induces cytokinesis failure.* Mol Cell Biol, 2006. 26(10): p. 3853-63.
260. Raemaekers, T., et al., *NuSAP, a novel microtubule-associated protein involved in mitotic spindle organization.* J Cell Biol, 2003. 162(6): p. 1017-29.
261. Lekomtsev, S., et al., *Centralspindlin links the mitotic spindle to the plasma membrane during cytokinesis.* Nature, 2012. 492(7428): p. 276-9.
262. Chalamalasetty, R.B., et al., *Influence of human Ect2 depletion and overexpression on cleavage furrow formation and abscission.* J Cell Sci, 2006. 119(Pt 14): p. 3008-19.
263. Yüce, O., A. Piekny, and M. Glotzer, *An ECT2-centralspindlin complex regulates the localization and function of RhoA.* J Cell Biol, 2005. 170(4): p. 571-82.
264. Higgins, J., et al., *Human ASPM participates in spindle organisation, spindle orientation and cytokinesis.* BMC Cell Biol, 2010. 11: p. 85.
265. Mollinari, C., et al., *PRC1 is a microtubule binding and bundling protein essential to maintain the mitotic spindle midzone.* J Cell Biol, 2002. 157(7): p. 1175-86.
266. Rouault, J.P., et al., *Identification of BTG2, an antiproliferative p53-dependent component of the DNA damage cellular response pathway.* Nat Genet, 1996. 14(4): p. 482-6.
267. Inaba, T., et al., *Fusion of the leucine zipper gene HLF to the E2A gene in human acute B-lineage leukemia.* Science, 1992. 257(5069): p. 531-4.
268. Zhou, T., et al., *GDF10 inhibits proliferation and epithelial-mesenchymal transition in triple-negative breast cancer via upregulation of Smad7.* Aging (Albany NY), 2019. 11(10): p. 3298-3314.

269. Wu, H. and J.T. Parsons, *Cortactin, an 80/85-kilodalton pp60src substrate, is a filamentous actin-binding protein enriched in the cell cortex.* J Cell Biol, 1993. 120(6): p. 1417-26.
270. Ji, R., et al., *Cortactin in Epithelial-Mesenchymal Transition.* Front Cell Dev Biol, 2020. 8: p. 585619.
271. Hudson, B.G., S.T. Reeders, and K. Tryggvason, *Type IV collagen: structure, gene organization, and role in human diseases. Molecular basis of Goodpasture and Alport syndromes and diffuse leiomyomatosis.* J Biol Chem, 1993. 268(35): p. 26033-6.
272. Masuda, T., et al., *Cold-inducible RNA-binding protein (Cirp) interacts with Dyrk1b/Mirk and promotes proliferation of immature male germ cells in mice.* Proc Natl Acad Sci U S A, 2012. 109(27): p. 10885-90.
273. Zhang, L., et al., *Identification of a panel of mitotic spindle-related genes as a signature predicting survival in lung adenocarcinoma.* J Cell Physiol, 2020. 235(5): p. 4361-4375.
274. Ko, E., et al., *Synergistic effect of Bcl-2 and cyclin A2 on adverse recurrence-free survival in stage I non-small cell lung cancer.* Ann Surg Oncol, 2013. 20(3): p. 1005-12.
275. Corson, T.W., et al., *KIF14 messenger RNA expression is independently prognostic for outcome in lung cancer.* Clin Cancer Res, 2007. 13(11): p. 3229-34.
276. Zhang, L., et al., *Genome-wide investigation of the clinical significance and prospective molecular mechanisms of kinesin family member genes in patients with lung adenocarcinoma.* Oncol Rep, 2019. 42(3): p. 1017-1034.
277. Grinberg, M., et al., *Reaching the limits of prognostication in non-small cell lung cancer: an optimized biomarker panel fails to outperform clinical parameters.* Mod Pathol, 2017. 30(7): p. 964-977.
278. Liu, Y., A. Beyer, and R. Aebersold, *On the Dependency of Cellular Protein Levels on mRNA Abundance.* Cell, 2016. 165(3): p. 535-50.
279. Gry, M., et al., *Correlations between RNA and protein expression profiles in 23 human cell lines.* BMC Genomics, 2009. 10: p. 365.
280. Nielsen, T.O., et al., *Assessment of Ki67 in Breast Cancer: Updated Recommendations From the International Ki67 in Breast Cancer Working Group.* J Natl Cancer Inst, 2021. 113(7): p. 808-819.

Printed by Media-Tryck, Lund 2022 NORDIC SWAN ECOLABEL 3041 0903

Department of Clinical Sciences Lund
Division of Oncology, Lund University

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2022:47
ISBN 978-91-8021-208-3
ISSN 1652-8220