

Abschlussbericht zum Vorhaben

„Verifizierung von neuen molekularen Markern für die Frühdiagnostik von Lungentumoren zum Nutzen der nachgehenden Vorsorge“ (FB-0259)

Laufzeit

01.01.2017 – 30.04.2021

Bericht vom 05.07.2021

Autoren

Yon-Dschun Ko¹, Reinhard Büttner², Jana Fassunke²,
Barbara Sitek³, Klaus Gerwert^{4,6}, Georg Johnen⁵,
Thomas Brüning^{5,6}, Thomas Behrens⁶

¹ Lungenkrebszentrum Bonn/Rhein-Sieg, Johanniter-Krankenhaus Bonn

² Institut für Pathologie, Universitätsklinikum Köln

³ Medizinisches Proteom-Center, Ruhr-Universität Bochum

⁴ Lehrstuhl für Biophysik, Ruhr-Universität Bochum

⁵ Institut für Prävention und Arbeitsmedizin der Deutschen Gesetzlichen Unfallversicherung, Institut der Ruhr-Universität Bochum (IPA)

⁶ Protein Research Unit Ruhr within Europe (PURE), Ruhr-Universität Bochum

Inhaltsverzeichnis

Kurzfassung deutsch

Kurzfassung englisch

1. Problemstellung

2. Forschungszweck/-ziel

3. Methodik

4. Ergebnisse des Gesamtvorhabens

5. Auflistung der für das Vorhaben relevanten Veröffentlichungen, Schutzrechtsanmeldungen und erteilten Schutzrechte von nicht am Vorhaben beteiligten Forschungsstellen

6. Bewertung der Ergebnisse hinsichtlich des Forschungszwecks/-ziels, Schlussfolgerungen

7. Aktueller Umsetzungs- und Verwertungsplan

8. Anhang/Anhänge

Unterschriftenseite verpflichtend für Kooperationsprojekte

Kurzfassung deutsch

Ziele:

Im Vorgängerprojekt (FP-339) wurden zahlreiche vielversprechende Markerkandidaten zur Früherkennung von Lungentumoren identifiziert. Im aktuellen Projekt sollten diese Kandidaten in einem unabhängigen Patientenkollektiv aus der Region Bonn/Rhein-Sieg verifiziert werden. Dabei sollte das bisherige Markerspektrum von epigenetischen Markern und Proteinen um genomische und RNA-basierte Marker erweitert werden. Die bereits im Vorgängerprojekt etablierte spektrale Histopathologie sollte methodisch verbessert werden, um eine genauere Gewebeklassifizierung zu erzielen, sowie per Lasermikrodissektion Gewebeproben für die weitere molekulare Charakterisierung bereitstellen. Zusätzlich zur Rekrutierung von 400 Patientinnen und Patienten mit Verdacht auf Lungentumor sollten in einer nachträglichen Fragestellung zudem eine weitere Vergleichsgruppe von 100 Patientinnen und Patienten mit anderen, gutartigen Erkrankungen der Atemwege („COPD-Kontrollen“) eingeschlossen werden.

Aktivitäten/Methoden:

Insgesamt 423 Patientinnen und Patienten, davon 333 mit Verdacht auf Lungentumor und 90 COPD-Kontrollen, wurden an zwei Standorten des Lungenkrebszentrums Bonn/Rhein-Sieg in die Studie eingeschlossen. Von den Teilnehmenden wurden Blut- und Speichelproben gesammelt. Gewebeproben wurden im Rahmen von Operationen sowie Bronchoskopien gewonnen und im Rahmen der Studie histopathologisch befundet.

DNA-Mutationen wurden per Next-Generation-Sequencing (NGS) im Gewebe sowie im Plasma identifiziert. Die im Plasma zirkulierende DNA wurde zusätzlich mit einer sensitiveren Methode (digital droplet PCR - ddPCR) detektiert. Die epigenetischen Analysen von DNA-Methylierung, RNA und Copy Number Variations (CNV) wurden anhand von Gewebeproben (OP-Resektate sowie EBUS/PE-Material aus Bronchoskopien), Vollblut-, Plasma-, und Speichelproben durchgeführt. Die Proteomanalysen erfolgten an Gewebe- und Plasmaproben. Die identifizierten Proteine wurden anschließend mittels verschiedener bioinformatischer Methoden ausgewertet. Weiterhin wurden die biospektroskopischen Gewebeanalysen um eine molekulare Erkennung des Mutationsstatus erweitert. Mittels Lasermikrodissektion wurden Tumorareale und gesunde Areale isoliert und für Proteom- und Mutationsanalysen zur Verfügung gestellt.

Ergebnisse:

Pathologie: Die Analyse der ctDNA aus den Plasmaproben zeigte keine ausreichende Detektion der DNA-Mutationen (>90% Wildtyp) im Vergleich zu den gewebebasierten Analysen (<10% Wildtyp), auch nicht nach methodischen Umstellungen (Blutentnahmesystem, sensitivere ddPCR). Dies ist möglicherweise auf eine zu geringe Tumorlast bei frühen Tumorstadien (wenig ctDNA shedding) zurückzuführen.

Epigenetik: Verschiedene Marker auf der Ebene von DNA-Methylierungen sowie die Auswertung von CNV ermöglichten eine hervorragende Differenzierung zwischen Tumor- und Nichttumorgewebe. Diese schloss auch kleinste Gewebemengen aus endobronchialen Ultraschall (EBUS) gestützten Biopsien mit ein. Eine Unterscheidung von Tumorerkrankten und Kontrollen anhand von Körperflüssigkeiten (Plasma, Speichel) mittels neu entwickelter Methoden zur DNA-Methylierung war grundsätzlich möglich, jedoch konnte keine Differenzierung zwischen Tumor- und COPD-Patientinnen und Patienten erzielt werden. Letztere Limitation galt auch für RNA-Marker. Jedoch konnten für bestimmte Subtypen (kleinzellige Lungentumoren) sowie bestimmte Markerkombinationen (miRNA + Protein) eine verbesserte Performance gegenüber den Einzelmarkern erzielt werden.

Biophysik: Die biospektroskopische Gewebeklassifikation aus dem Vorgängerprojekt (FFP-339) konnte verifiziert und die benötigte Messzeit durch den Einsatz von Quantenkaskadenlasern (QCL) auf wenige Minuten reduziert

werden. Hiermit konnte nun eine Label-freie molekulare Erkennung der Treibermutationen (TP53, KRAS, EGFR) etabliert werden, die auch auf molekularer Ebene die Label-freie, automatisierte pathologische Subklassifizierung von Treibermutationen des Adenokarzinoms der Lunge mit hoher Sensitivität und Spezifität erlaubt.

Proteomics: Beim gewebebasierten Vergleich von Adenokarzinomen und Kontrollen mit benignen Lungenerkrankungen konnten Proteine identifiziert werden, die auch in den Plasmaproben differenziell nachweisbar waren. Durch Kombination aus Proteomics-Analytik und bioinformatischen Ansätzen wurde anhand eines Teilprobensatzes ein Biomarkerpanel identifiziert, das mit einer Sensitivität von 97 % und einer Spezifität von 89 % Adenokarzinome und COPD voneinander trennte. Die Verifizierung dieses Markerpanels wurde bisher noch nicht durchgeführt.

Schlussfolgerungen:

Insgesamt konnten bisher noch keine Marker bzw. Markerkombinationen in Körperflüssigkeiten verifiziert werden, die für eine abschließende Validierung in einer Langzeitstudie zur Krebsfrüherkennung geeignet sind. Weitere Analysen sollen insbesondere neue Markerkombinationen sowie die Proteinmarker verifizieren, um die bisher nicht ausreichende Abgrenzung zu chronisch entzündlichen Lungenerkrankungen zu überprüfen.

Kurzfassung englisch

Objectives:

Numerous promising marker candidates for early detection of lung cancers were identified in a previous project (FP-339). The current project was designed to verify these candidates in an independent patient collective recruited in the Bonn/Rhein-Sieg area. Additionally, the previous marker spectrum of epigenetic markers and proteins should be expanded to include genomic and RNA-based markers. The spectral histopathology already established in the previous project was to be methodically improved in order to achieve a more precise tissue classification, as well as to provide tissue samples by means of laser microdissection for further molecular characterisation. In addition to the recruitment of 400 patients with suspected lung tumours, a group of 100 patients with benign respiratory diseases ("COPD controls") was included subsequently as control group.

Activities/Methods:

A total of 423 patients, including 333 with suspected lung tumour and 90 COPD controls, were included in the study at two clinics of the Bonn/Rhein-Sieg Lung Cancer Centre. Blood and saliva samples were collected from patients. Tissue samples were obtained during operations as well as bronchoscopies and histopathologically confirmed within the framework of the study.

DNA mutations were identified by next-generation sequencing (NGS) in tissue and plasma. Circulating DNA in plasma was additionally detected with a more sensitive method (digital droplet PCR - ddPCR). Epigenetic analyses of DNA methylation, RNA, and copy number variations (CNV) were performed in tissue samples (surgical resections and EBUS/PE material from bronchoscopies), whole blood, plasma and saliva samples. Proteome analyses were performed in tissue and plasma samples. Identified proteins were then evaluated using various bioinformatics methods. Further, biospectroscopic tissue analyses were extended to include molecular detection of the mutation status. By means of laser microdissection, tumour areas and healthy areas were isolated and made available for proteome and mutation analyses.

Results:

Pathology: The analysis of ctDNA from plasma samples did not show sufficient detection of DNA mutations (<10% wild type) compared to tissue-based analyses (<10% wild type), even after methodological modifications (blood collection system, more sensitive ddPCR). This may be due to an insufficient tumour burden at early tumour stages (low amounts of ctDNA shedding).

Epigenetics: Various markers at the DNA methylation level and evaluation of CNV enabled excellent differentiation between tumour and non-tumour tissue. This was even possible in the smallest amounts of tissue from endobronchial ultrasound-based (EBUS) biopsies. Differentiation of tumour patients and controls in body fluids (plasma, saliva) using newly developed methods for DNA methylation was also technically possible, however, no differentiation between tumour and COPD patients was achieved. This limitation also applied to RNA markers. However, marker performance was improved for small cell lung tumour subtypes as well as certain marker combinations (miRNA + protein) compared to single markers.

Biophysics: The biospectroscopic tissue classification from the preceding project (FP-339) could be verified and the required measurement time was reduced to a few minutes by using quantum cascade lasers (QCL). Hereby, a label-free molecular detection of driver mutations (TP53, KRAS, EGFR) has been established, which also allows the label-free, automated pathological subclassification of driver mutations of adenocarcinoma of the lung with high sensitivity and specificity on the molecular level.

Proteomics: In the tissue-based comparison of adenocarcinomas and controls with benign lung diseases, proteins could be identified that were also differentially detectable in the plasma samples. By combining proteomics analysis and bioinformatics approaches, a biomarker panel was identified in a study subsample that separated adenocarcinomas and COPD with a sensitivity of 97% and a specificity of 89%. Verification of this marker panel has not yet been carried out.

Conclusions:

Overall, markers or marker combinations have not been verified in body fluids yet that are suitable for final validation in a longitudinal study for early cancer detection. Further analyses should in particular verify new marker combinations and protein markers in order to improve the insufficient differentiation between cancer patients and patients with chronic inflammatory lung diseases.

1. Problemstellung

Die erfolgreiche Behandlung von Krebserkrankungen stellt weiterhin eine große Herausforderung dar. Dies liegt nicht nur an einem Mangel passender Therapieoptionen für eine große Zahl der vielen Varianten dieser Erkrankung, sondern auch an fehlenden diagnostischen Methoden für eine erfolgreiche Früherkennung. Eine deutliche Verbesserung der Prognosen wäre zu erzielen, wenn bestehende Therapien rechtzeitig, d.h. bereits in frühen Entwicklungsstadien von Tumoren, eingesetzt werden könnten. Daher ist eine Früherkennung von Krebserkrankungen, möglichst mit keiner oder minimaler Belastung für die Patientinnen und Patienten, von enormer Bedeutung. Hierbei stellen Biomarker, ggf. in Kombination mit bildgebenden Verfahren, einen vielversprechenden Ansatz dar.

Lungenkrebs ist mit ungefähr 57.500 Neuerkrankungen im Jahr 2016 die dritthäufigste Krebsart in Deutschland (Robert Koch-Institut 2019). Wegen der relativ späten Diagnosestellung überwiegen in den klinischen Stadien IIIb und IV zählt Lungenkrebs in Deutschland und weltweit zu den Erkrankungen mit der höchsten Mortalität, für den Maßnahmen zur Früherkennung, trotz Verbesserungen in der bildgebenden Diagnostik (National Lung Screening Trial Research Team 2011), weiterhin eine große Herausforderung darstellen. Ein erhöhtes Risiko, an einem Tumor der Lunge zu erkranken, ist v.a. durch das Zigarettenrauchen, aber auch durch zahlreiche berufsbedingte Expositionen bedingt. Hier können v.a. Tumoren der Lunge in Zusammenhang mit einer beruflichen Exposition gegenüber Asbest als Berufskrankheit anerkannt werden (BK-Nrn. 4104 und 4114). So wurde im Jahr 2019 bei 622 Personen eine Berufskrankheit im Zusammenhang mit einem asbestbedingten Lungentumor neu anerkannt (Deutsche Gesetzliche Unfallversicherung 2020).

Der Organisationsdienst für nachgehende Untersuchungen (ODIN) der gesetzlichen Unfallversicherungsträger sowie die Gesundheitsvorsorge der DGUV (GVS) betreuen ehemals asbestexponierte Beschäftigte durch das Angebot von nachgehenden Untersuchungen sowie dem erweiterten Vorsorgeangebot (EVA) zur Lungenkrebsfrüherkennung. Letzteres erhalten auch Versicherte mit anerkannter, durch Asbest verursachter Berufskrankheit (BK Nr. 4103). Mit diesen Konzepten zur Sekundärprävention wird eine frühzeitige Diagnose von Tumorerkrankungen angestrebt, die eine wichtige Voraussetzung für einen erfolgreichen Therapieansatz darstellt. Da aber mit den derzeit verwendeten Untersuchungsmethoden eine hinreichend frühzeitige Diagnose eines Lungentumors nur mit Einschränkung möglich ist, war es grundsätzliches Ziel dieses Vorhabens, die Sekundärprävention mittels neuer molekularer Marker zur Früherkennung von Lungentumoren zu verbessern.

Die Entwicklung von Biomarkern zur Krebsfrüherkennung durchläuft mehrere Phasen (siehe Abbildung 1). In der ersten Stufe des Vorhabens (FP339) lag der Schwerpunkt auf der Entdeckung neuer Marker (Discovery, Phase 1). Der Schwerpunkt des Folgeprojektes lag nun auf der Phase 2 der Biomarker-Entwicklung, also der Verifizierung der neu identifizierten Biomarker. Diese umfasst v.a. die Verifizierung in einem unabhängigen Kollektiv (aus dem Raum Köln/Bonn). Bei der Verifizierung kann auch eine unabhängige Messmethode als Grundlage für die Entwicklung feldtauglicher Assays dienen. Bei letzterer soll vor allem berücksichtigt werden, dass es für eine spätere Anwendung zur Früherkennung in der Praxis essentiell ist, die Biomarker auch in nicht- und minimalinvasiv gewinnbaren Körperflüssigkeiten wie z.B. Blut und Speichel nachweisen zu können.

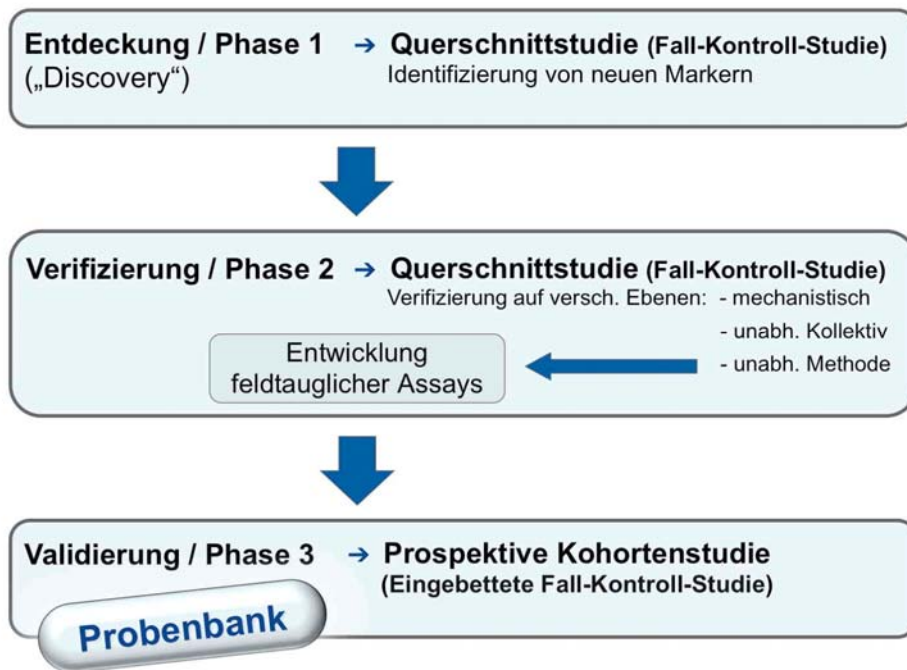


Abbildung 1: Phasen der Biomarker-Entwicklung.

Nach erfolgreichem Abschluss der Verifizierung könnte eine abschließende Validierung der Biomarker (Phase 3) für den Einsatz in der Krebsfrüherkennung dann in einem Folgeprojekt erfolgen. Hierzu wäre eine prospektive Kohorte mit angeschlossener Biobank erforderlich, wie sie beispielsweise das Projekt MoMar bereitstellen könnten.

Literatur

Deutsche Gesetzliche Unfallversicherung (2020) DGUV-Statistiken für die Praxis 2019, Berlin

National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD (2011). Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*;365(5):395-409.

Robert Koch-Institut (2019) Krebs in Deutschland für 2015/2016. 12. Ausgabe. Gemeinsame Publikation des Zentrums für Krebsregisterdaten und der Gesellschaft der Epidemiologischen Krebsregister In Deutschland e.V., Berlin

2. Forschungszweck/-ziel

Nach Abschluss der Discovery-Phase der Biomarker-Entwicklung (FP339) sollte nun die Verifizierungsphase angeschlossen werden. Ziel der Folgestudie war die Verifizierung (Phase 2) der bisherigen Ergebnisse zu Lungentumoren, um die gefundenen Markerkandidaten und neuentwickelten Methoden einen Schritt näher an die Anwendung in der arbeitsmedizinischen Praxis zu bringen. Konkret sollten die neuen Marker in einem unabhängigen klinischen Querschnittskollektiv bestätigt werden. Das unabhängige Kollektiv sollte neben Geweben vor allem auch die leicht zugänglichen Körpermaterialien Vollblut, Plasma und Speichel umfassen. Um das bisherige Markerspektrum zu erweitern und auch mechanistisch zu bestätigen, sollten neben Proteinen und epigenetischen Markern auch genomische Marker (krebsspezifische Mutationen u.a. Aberrationen) sowie RNA-basierte Marker im Plasma analysiert werden.

Zielgruppen

Die Verifizierung der Markerkandidaten aus der Identifizierungsphase sollte in einem allgemeinen klinischen Kollektiv durchgeführt werden, um an einer umfangreichen Zahl von Patientinnen und Patienten statistisch signifikant zwischen an Krebs erkrankten und anderen Lungenerkrankungen sowie gesunden Probanden unterscheiden zu können. Mit dem Konzept verfolgten die Antragsteller das Ziel, wirksame Maßnahmen zur Sekundärprävention zu entwickeln, die es erlauben, eine frühzeitige Lungentumordiagnose zu stellen, von der auch ehemals beruflich gegenüber Kanzerogenen exponierte Beschäftigte profitieren können, die im Rahmen der nachgehenden Vorsorge untersucht werden.

2.1 Wissenschaftlich-epidemiologisches Studienzentrum (WESZ)

Aufgabe des WESZ war, zu Beginn der Studie die ethisch-rechtlichen Rahmenbedingungen der Studie sicherzustellen und die Etablierung der Studienabläufe zu koordinieren. Dazu gehörten die Erstellung eines Datenschutzkonzeptes, das Einholen der Zustimmung der beteiligten Ethikkommissionen sowie die Abstimmung aller benötigten Standard Operating Procedures (SOPs). Weiterhin wurden studienspezifische Instrumente zur Datenerfassung erstellt, Datenbanken aufgebaut und die Software zur Probenverwaltung (Starlims) für eine qualitätsgesicherte Probensammlung angepasst. Fortlaufende Aufgaben des WESZ waren Logistik und Verwaltung von gesammelten Proben und die Prüfung, Weiterverarbeitung und Auswertung der gesammelten Daten.

2.2 Kliniken Bonn

Das Lungenkrebszentrum der Region Bonn/Rhein-Sieg mit den Standorten Malteser-Krankenhaus Seliger Gerhard Bonn/Rhein-Sieg und Johanniter-Krankenhaus Bonn sollte Gewebe von 150 operablen Lungenkarzinomen pro Jahr kryoasservieren. Es sollten von denselben Patientinnen und Patienten Plasma-, Blutkuchen- sowie Speichelproben gewonnen werden. Zusätzlich sollten pro Jahr von 50 weiteren Teilnehmenden Lymphknoten-Biopsien gewonnen werden. Ein weiteres Ziel war der Einschluss von 100 Patientinnen und Patienten mit einer entzündlichen Atemwegserkrankung, wie z.B. der chronisch-obstruktiven Lungenerkrankung (COPD). Die Datenerhebung (Interviews, Dokumentation) erfolgte in Kooperation mit dem Wissenschaftlich-Epidemiologischen Studienzentrum.

2.3 Pathologie Uniklinik Köln

Die Aufgabe der Uniklinik Köln war, das zuvor in Bonn gewonnene Gewebe und Plasma weiter zu verarbeiten. Hierbei sollte Gewebe von 150 operablen Lungentumoren formalinfixiert und in Paraffin eingebettet werden (FFPE; „formalin-fixed paraffin-embedded“). Das FFPE-Gewebe wurde für die teilnehmenden Institute zugeschnitten und zur Feststellung der Tumorareale und des Tumorgehalts begutachtet. Zusätzlich sollte eine molekularbiologische Analyse der Tumoren über das jeweilige FFPE Material und über das korrespondierende Plasma durch das „Next Generation Sequencing“ (NGS) erfolgen.

2.4 Biophysik

In der Biophysik (Teilprojekt C) sollten die Ergebnisse aus der Discovery-Phase zur histologischen Subklassifizierung von Tumoren verifiziert werden. Weitere Herausforderungen waren eine Optimierung der CARS-Technologie für die Untersuchung kleinster Probenmengen (z.B. EBUS-Proben) unter Berücksichtigung begrenzter Probenmengen und feststehender Analysen sowie allgemein die Heterogenität von Tumoren/Metastasen innerhalb von Patientinnen und Patienten durch das IR-Imaging. Durch neue technische Entwicklungen hat sich der Fokus auf die Etablierung des schnellen Laser-basierten QCL-IR-Imagings für die Gewebesanalyse verschoben. Eine Erweiterung der Gewebeklassifizierung in Richtung automatisierter molekularer Erkennung zur Subklassifizierung der Treibermutationen des Adenokarzinoms der Lunge sollte entwickelt werden. Basierend hierauf sollten Label-frei Gewebeproben für die Proteomik (Prof. Sitek) und – nach Absprache im Verlauf der Studie – die Genomik (Prof. Büttner) gesammelt werden.

2.5 Medizinisches Proteom-Center (MPC)

Mit Hilfe der Proteomicsmethoden sollten potenzielle Biomarker im Gewebe sowie im Serum und Speichel identifiziert und quantifiziert werden. Die gewebebasierten Studien wurden in Zusammenarbeit mit der Biophysik (Prof. Gerwert) durchgeführt werden.

Das primäre Vorhabensziel der MPC-Bioinformatik bestand darin, gemeinsam mit den anderen Partnern die Validierungsstudie der Marker-Kandidaten aus dem Vorgänger-Projekt durchzuführen und die innerhalb dieses Projektes generierten Daten (neben massenspektrometrischen Daten beinhaltete dies v.a. die im Rahmen dieses Projektes durchgeführten Methylierungsexperimente) bioinformatisch und statistisch auszuwerten. Dafür war die MPC-Bioinformatik an allen Arbeitspaketen des MPC (Arbeitspaket 1-4) beteiligt, wo sie als sekundäre Vorhabensziele die Optimierung des Studiendesigns (Arbeitspaket 1), die Optimierung des MRM (multi reaction monitoring)-Datenanalyse-Workflows (Arbeitspaket 2), die Qualitätsanalyse der gemessenen Daten (Arbeitspaket 3), die eigentliche bioinformatische und statistische Datenauswertung (Arbeitspaket 3) sowie die Pathway- und Enrichment-Analyse zur Einordnung der Ergebnisse (Arbeitspaket 4) verfolgte.

2.6 Epigenetik

Die Analysen auf der Protein-Ebene sollten durch Analysen auf anderen molekularen Ebenen ergänzt werden. So können weitere neue Biomarker entdeckt und mit den Proteinmarkern kombiniert werden, um möglicherweise die Sensitivität eines Marker-Panels zu erhöhen. Da über die Epigenetik die Gene reguliert und somit die Expression von Proteinen gesteuert wird, liefert sie einen wichtigen Beitrag zum Verständnis pathologisch veränderter Proteinmuster, was die Aufdeckung möglicher Kausalketten bei der Kanzerogenese ermöglicht. Ähnliches gilt für Veränderungen auf der Ebene der Gene selbst (z.B. copy number variations, CNV) oder auf der RNA-Ebene. Dadurch kann die Plausibilität neu entdeckter Proteinbiomarker untermauert und die Zahl von Screening-Artefakten reduziert werden. Aus dem Vorprojekt FP 339 sowie Voruntersuchungen in anderen Projekten des IPA lagen zahlreiche Marker-Kandidaten auf der Ebene der DNA-Methylierung, der RNA und der CNV vor. Ziel des Teilprojektes Epigenetik war die Verifizierung dieser Marker-Kandidaten im Gewebe und insbesondere in Körperflüssigkeiten von Probanden an einem unabhängigen klinischen Kollektiv.

Nachweis von DNA-Methylierungen

Neuentdeckte epigenetische Veränderungen auf der Ebene der DNA-Methylierung wurden im Vorprojekt bereits teilweise mittels einer unabhängigen Methode, dem Pyrosequencing, im Gewebe und in Speichelproben bestätigt. Diese waren nun auch in einem unabhängigen Kollektiv von Patientinnen und Patienten zu verifizieren. Da zudem ein Nachweis in Plasmaproben erfolgen sollte und das Pyrosequencing sich als nicht ausreichend sensitive Methode

erwiesen hat, war es Ziel, eine geeignete Deep-Sequencing Methode (Bisulfite Next-Generation Sequencing, „Bisulfite-Seq“) zu etablieren und damit die Kandidatenmarker aus dem Vorprojekt zu bestätigen.

Nachweis von tumorspezifischer RNA

Die RNA-Ebene ist neben den molekularen Ebenen DNA (Gen-Kodierung), DNA-Methylierung (Gen-Steuerung) und Proteine (Gen-Produkte) ein wichtiges Bindeglied zur mechanistischen Verifizierung und Korrelation der Biomarker-Ergebnisse. Das Ziel dieses Teilprojekts war daher, potentielle RNA-Marker wie *long-non-coding* RNA (lncRNA), die vermutlich eine wichtige Rolle in der Lungentumorgenese spielen, sowie *messenger* RNA (mRNA) gezielt in Plasma- und Speichelproben mittels der quantitativen Real-Time PCR (qPCR) zu untersuchen, um zu klären, ob sie als minimal-invasive Marker zur Detektion von Lungenkrebs geeignet sind.

Nachweis von CNV in Gewebeproben

Zum Nachweis von Tumor-spezifischen Mutationen sollten chromosomale Defekte untersucht werden, die sich in einer variierenden Kopienzahl bestimmter Genbereiche, sogenannten CNV, äußern. Die Verwendung des hochempfindlichen Verfahrens der digitalen PCR (dPCR) ermöglicht eine quantitative Bestimmung von Gewinnen und Verlusten ausgewählter Gene.

3. Methodik

Der Start der Rekrutierung der Studienteilnehmerinnen und Studienteilnehmer erfolgte verzögert ab 08/2017, insbesondere aufgrund der Abstimmung der teils unterschiedlichen Anforderungen der beteiligten Ethik-Kommissionen (Ruhr-Universität Bochum (RUB) (Leitethik), Universität zu Köln, Ärztekammer Nordrhein). Erste Gewebeschnitte standen ab Anfang 2018 für Analysen zur Verfügung. Zu Beginn der Studie erfolgte keine Befundung der Hämatoxylin-Eosin-(HE)gefärbten Schnitte aus Nativgewebe in der Klinik, so dass die Befundung nach Köln verlegt wurde. Die bis zur Umstellung fehlenden Begutachtungen wurden jedoch nachträglich digital erstellt. Des Weiteren wurde zu Beginn der Laufzeit beschlossen, zu den im Antrag genannten 400 Patientinnen und Patienten mit Verdacht auf Lungenkrebs weitere Kontrollen mit chronischen, nicht-malignen Lungenerkrankungen („COPD-Kontrollen“) einzuschließen.

Aufgrund der anfänglichen Verzögerungen des Rekrutierungsbeginns und bei der Gewebeverarbeitung, der Erhöhung der Zahl der Teilnehmer und Teilnehmerinnen durch Einbezug der COPD-Kontrollen sowie personeller Engpässe bei der Rekrutierung in 2019 wurde zweimal eine kostenneutrale Verlängerung der Studienlaufzeit um jeweils ein Jahr bis Ende 2019 bzw. Ende 2020 beantragt und von der DGUV bewilligt. Bei den Treffen des Forschungsbegleitkreises sowie in den Zwischenberichten zum Vorhaben wurde die Notwendigkeit zur Verlängerung jeweils dargestellt. Aufgrund der Corona-Pandemie wurde die Studie letztmals um weitere vier Monate verlängert.

Die im Antrag geplanten sowie die tatsächlichen Abläufe sind in Zeitplänen schematisch dargestellt (s. Anhang).

3.1 WESZ

Der ethischen Voraussetzungen der Studie wurden durch die zuständigen Ethikkommissionen (Ruhr-Universität Bochum, Universität zu Köln, Ärztekammer Nordrhein) geprüft. Das Datenschutzkonzept wurde in Form einer Vorabkontrolle mit dem Datenschützer der RUB abgestimmt.

Vorbereitend wurden Materialien für die Proben- und Datensammlung über Barcodes digital verknüpft und zu Sets zusammengestellt. Neue Sets wurden regelmäßig nach Bonn transportiert und im Gegenzug Sets mit den gesammelten Proben und ausgefüllten Dokumenten zurück nach Bochum transportiert. Dort wurden Proben und Daten auf Vollständigkeit und Plausibilität geprüft und in die Probenverwaltungssoftware Starlims integriert.

3.2 Kliniken Bonn

Insgesamt wurden 423 Patientinnen und Patienten in 441 Untersuchungen in die Studie eingeschlossen. Dreizehn Personen nahmen zweimal teil, 5 Teilnehmende wurden aufgrund einer Tuberkulose (n=4) oder sonstiger Nichterfüllung der Ein- bzw. Ausschlusskriterien (n=1) nachträglich ausgeschlossen.

Zwischenzeitliche personelle Engpässe beim rekrutierenden Personal sorgten dafür, dass die avisierten Fallzahlen nicht vollständig erreicht wurden: Bis 2019 wurden durch eine der beiden Study Nurses des Johanniter-Krankenhauses regelmäßig auch Patientinnen und Patienten im Malteser-Krankenhaus vor Ort im Rahmen der Thorax-Sprechstunde rekrutiert, seit 2017 unterstützt durch eine Fachpflegekraft vor Ort. In 2019 wechselte eine der beiden Study Nurses ihren Tätigkeitsbereich innerhalb der Klinik, so dass die Rekrutierung im Malteserkrankenhaus von April bis August 2019 nur unregelmäßig stattfinden konnte. Von September 2019 bis Juni 2020 wurden dann wieder regelmäßig Patientinnen und Patienten im Malteserkrankenhaus rekrutiert.

3.3 Pathologie Köln

Die in Kooperation mit dem Johanniter-Krankenhaus in Bonn gesammelten Biopsien wurden in Köln formalinfixiert und anschließend in Paraffin eingebettet. Das FFPE-Gewebe wurde anschließend in Köln für die teilnehmenden Institute der Studie zugeschnitten.

Bei dem Zuschnitt des nativen und des FFPE-Gewebes wurden aus Gründen der Qualitätssicherung pro Studienteilnehmer zwölf mit Hämatoxylin-Eosin (HE) gefärbte Schnitte in unterschiedlichen Schnittebenen hergestellt. Diese Schnitte wurden durch einen Pathologen in Köln zur Feststellung der Tumorareale und des Tumorgehalts begutachtet.

Zusätzlich sollte die DNA aus dem FFPE-Gewebe der Studienteilnehmenden mit der korrespondierenden zirkulierenden Tumor-DNA (ctDNA) aus Plasma molekularbiologisch durch das NGS Verfahren untersucht und verglichen werden. Da der Vergleich bis zum Zeitpunkt des ersten Zwischenberichtes zu keinem zufriedenstellenden Ergebnis gelangte, wurden ausgewählte Proben, bei denen eine *KRAS*-Mutationen in der Tumor-DNA festgestellt werden konnte, mit einer sensitiveren Methode validiert. Hierfür wurde die „digital droplet PCR“ (ddPCR) verwendet. Zusätzlich wurde das Blutentnahmesystem auf BD/PaxGene-Röhrchen umgestellt, bei dem geringere Scherkräfte erwartet wurden und so die Zerstörung von Leukozyten reduziert werden sollte.

Für den Vergleich der Sequenzierung von DNA und ctDNA der Studienteilnehmenden wurde im Vorfeld DNA durch das Maxwell® 16 FFPE Plus LEV DNA Purification Kit (Promega) aus FFPE-Gewebe nach Herstellerangaben isoliert. Dabei wurden ausschließlich Tumorareale des Materials verwendet, die zuvor von einem Pathologen auf einem HE-gefärbten Schnitt eingezeichnet wurden. Die Isolation von ctDNA aus Plasma erfolgte mit dem QIAamp MinElute ccfDNA Kit (Qiagen) nach Herstellerangaben. Anschließend erfolgte eine Bibliothekerstellung mit dem Lun5-Panel für stark ($>1\text{ng}/\mu\text{l}$) und mit dem Lun3-Panel für schwach ($<1\text{ng}/\mu\text{l}$) konzentrierte DNA-Proben. Für die Sequenzierung wurde das MiSeq-System (Illumina) verwendet.

Insgesamt konnte das Material von 194 Studienteilnehmenden ausgewertet werden, von denen für 160 Proben ein NGS Ergebnis für das FFPE-Material vorlag. Für 98 Proben existierte ein korrespondierendes NGS-Ergebnis der ctDNA aus Plasma. Eine Analyse durch die ddPCR-Methode für Proben, die eine *KRAS* c.35G>T-Mutation trugen, wurde für 20 Proben realisiert.

In Zusammenarbeit mit dem Institut der Biophysik der Ruhr-Universität Bochum wurde DNA mittels NGS untersucht, die mittels „quantum cascade laser based infrared imaging“ (QCL-IR) und lasergesteuerter Mikrodissektion aus FFPE-Gewebe von Adenokarzinomen gewonnen wurde. Die Resultate wurden mit den korrespondierenden Ergebnissen der DNA aus der ursprünglichen klinischen Diagnostik verglichen. Hierbei wurde untersucht, ob die Mutationsprofile der, mit der QCL-IR-Imaging-basierten Mikrodissektion gewonnenen, mit denen der konventionell isolierten Proben vergleichbar sind. Jeweils 20 *KRAS*, *EGFR* und *TP53* mutierte Proben wurden für diese Untersuchung bereitgestellt und diskrepante Fälle mit dem NGS-Verfahren überprüft.

3.4 Biophysik

FTIR-/IR-Imaging

Die im ersten Förderzeitraum auf Basis des FTIR-Imaging identifizierten spektralen Marker zur Klassifizierung von histologischen Subtypen der Lungentumore konnten erfolgreich verifiziert und auf ein 160-fach schnelleres System mit Quantenkaskadenlasern übertragen werden. So konnte die differentielle Diagnostik der Lungentumoren, insbesondere der Plattenepithel- und Adenokarzinome, bis hin zu der klassischen morphologischen Betrachtung folgenden Subtypisierung für das schnelle IR-Imaging verifiziert werden.

Durch Abgleich mit in Köln/Bonn etablierten Gen-Panels konnte ein weiterer Klassifizierer etabliert werden, der auch auf molekularer Ebene die Label-freie, automatisierte Subklassifizierung von Treibermutationen des Adenokarzinoms räumlich aufgelöst erlaubt (TP53, EGFR, KRAS). Basierend auf dieser Arbeit wurden Adenokarzinome mit den genannten Mutationen für die Proteomik gesammelt. Hierdurch sollten neue Proteinbiomarker identifiziert werden, die eine Erkennung in z. B. Blut ermöglichen.

EBUS-Proben mit Tumor von 10 Teilnehmenden wurden mit IR-Imaging untersucht. Aufgrund fehlender zusammenhängender Gewebestruktur konnten die Schnitt-basierten Klassifizierer nicht übertragen werden. Nur bei vereinzelt Proben mit Gewebestrukturen konnte eine Tumorerkennung durchgeführt werden. Allerdings müsste ein neuer Klassifizierer etabliert werden, um verlässliche Ergebnisse für EBUS-Proben zu liefern. Insgesamt ist hierfür die Probenmenge zu klein und zu divers. Zukünftig ist dieser Ansatz bei größerer Menge an EBUS-Proben für die spektrale Analyse sehr vielversprechend.

Raman/CARS Imaging

Um eine höhere räumliche Auflösung gerade für EBUS-Proben zu erreichen, war angedacht, auch Raman- und SRS-Imaging (*stimulated Raman scattering*) einzusetzen. Dies wurde aufgrund der sehr kleinen Probenmenge nicht begonnen. Zukünftig könnte dieser Ansatz für EBUS-Proben aber noch interessant sein.

Arbeitspakete Spektrale Histopathologie/Biophysik

Arbeitspaket 1 (abgeschlossen):

Die Optimierung und Anpassung der SOPs wurde vorgenommen. Proben aus den Kliniken wurden gewonnen und bearbeitet.

Arbeitspaket 2 (abgeschlossen):

Die Verifizierung der spektralen Marker des ersten Förderzeitraums mit FTIR-Imaging wurde erfolgreich durchgeführt und bestätigt die veröffentlichten Ergebnisse. Der Transfer auf das viel schnellere QCL-IR-Imaging konnte erfolgreich durchgeführt werden. Die automatisierte Erkennung von Lungengewebe und Subtypisierung durch das QCL-basierte IR-Imaging wurde durch eine erweiterte Sammlung von Proben etabliert und verifiziert. Mit dem IR-Imaging konnte außerdem die räumlich aufgelöste Erkennung von Treibermutationen für Adenokarzinome etabliert werden. Dies erlaubt die Label-freie Erstellung einer molekularen Karte des Gewebes.

Erste Tests an EBUS-Gewebeproben zeigen, dass eine Tumorerkennung möglich ist.

Arbeitspaket 3 (abgeschlossen):

Der in AP2 etablierte Label-freie Ansatz wurde zur Gewinnung homogener Proben für die Genomik und Proteomik genutzt. Von 30 Teilnehmenden wurden mit QCL-basiertem IR-Imaging Proben von Adenokarzinomen mit verschiedenen Mutationen (KRAS, TP53, EGFR) sowie Proben von 29 Adenokarzinomen und 29 korrespondierenden Nicht-Tumorproben für die Proteomik gesammelt. Von diesen Proben wurde das Proteom analysiert, um neue Proteinbiomarker für die Diagnostik in z. B. Blut zu identifizieren.

Arbeitspaket 4 (abgeschlossen):

Es wurden alle in der Studie gewonnenen und der Biophysik zur Verfügung gestellten Gewebeproben vermessen und mit IR-Imaging analysiert. Durch die Korrelation der Ergebnisse der Genomanalysen mit dem IR-Imaging konnten wir eine molekulare Darstellung der Treibermutationen ermöglichen, um beispielsweise Label-frei Patientinnen und Patienten mit EGFR-Mutationen zu selektieren. Weiterhin konnte für Adenokarzinome versus

Tumor-freies Lungengewebe und unterschiedlich mutierte Adenokarzinome eine Proteomanalyse durchgeführt werden. Hierbei konnten interessante Proteinbiomarker identifiziert werden. Aufgrund des Ausscheidens der auf Seiten der Proteomanalyse beteiligten Post-Doktorandin konnten diese Daten bis zum Abschluss des Projekts noch nicht in die IR-Imaging-Klassifizierer einfließen. Dies wird aktuell durch eine beteiligte Doktorandin aus Hausmitteln der Biophysik bearbeitet.

3.5 MPC

MPC-Proteomics

Zusätzlich zu dem gut etablierten Workflow der gewebebasierten Studien, die zusammen mit der Biophysik durchgeführt wurden, wurde ein neuer Workflow zu Plasmaanalytik etabliert und erfolgreich angewendet.

Bei den Gewebestudien ist es uns gelungen, zwei Fragestellungen zu adressieren. Zum einen wurden Proteine identifiziert, die zwischen Adenokarzinomen und Kontrollen unterscheiden. Einige der differentiellen Proteine konnten in den Plasmaproben verifiziert werden. Zum anderen konnten Proteine identifiziert werden, die Tumore mit verschiedenen Mutationen voneinander unterscheiden.

Bei dem Workflow zu Plasmaanalytik ist es nun möglich, global die Proteine in einer Probe zu identifizieren (mehr als 200 Proteine) und zu quantifizieren, so dass wir von unserem ursprünglichen Plan, nur einzelne Proteine im Blut mit der MRM-Methode zu analysieren, abgewichen sind. Zusätzlich wurde der Workflow mit den etablierten Auswerte-Methoden aus der Bioinformatik ergänzt mit dem finalen Ergebnis der Identifikation von einem Biomarkerpanel (Abbildung 2). Insgesamt konnten mit dem Verfahren in einer finalen Studie mehr als 200 Proben analysiert werden (Adenokarzinome, COPD, Kontrollen).

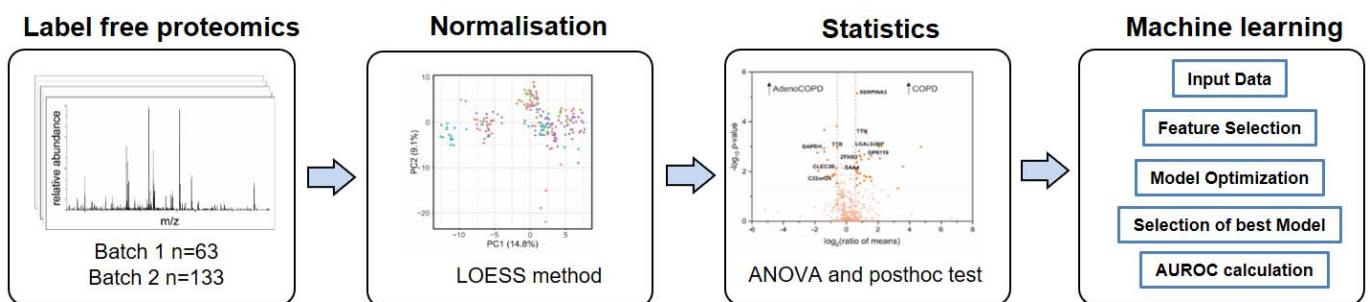


Abbildung 2: Neu etablierter Workflow für die Analytik und Auswertung von Plasmaproben.

MPC-Bioinformatik

Klassifikationsmodelle mit Proteomics-Messungen im Blutplasma (Arbeitspaket Proteomics)

Aufgrund der neuen Fragestellung zur Unterscheidung von Adenokarzinomen und COPD-Kontrollen in Plasmaproben wurden neue Verfahren in der Bioinformatik etabliert, um diese 2 ähnlichen Entitäten basierend auf einem Biomarkerpanel voneinander zu trennen. Es wurden drei Klassifikationsmodelle entwickelt: Modell 1 trennte 64 Adenokarzinom-Erkrankte von 76 COPD-Erkrankten ohne Adenokarzinom, Modell 2 trennte 21 Adenokarzinom-Erkrankte mit COPD von den 76 COPD-Erkrankten und Modell 3 trennte die 64 Adenokarzinom-Erkrankte von 35 Kontrollen.

Die Modelle nutzen die normalisierten Protein-Quantitäten als Einflussvariablen und wurde in mehreren Schritten optimiert. Zunächst wurden redundante Proteine entfernt. Hierzu wurde bei einer paarweisen absoluten Spearman-Korrelation $\geq 0,7$ jeweils nur eines der beiden Proteine für die Analyse ausgewählt und das andere entfernt. Im nächsten Schritt wurden die in Bezug auf die Fragestellung nicht relevanten Proteine entfernt. Diese wurden über den Kruskal-Wallis-Tests auf Lageverschiebung und den Levene-Test auf unterschiedliche Varianzen zwischen den

beiden Gruppen der Zielvariable (z.B. Adenokarzinom- Erkrankte vs. COPD) identifiziert. Diejenigen Proteine, für welche sowohl der p-Wert des Kruskal-Wallis-Tests als auch derjenige des Levene-Tests unter 0,05 lagen, wurden entfernt. Dabei wurden die p-Werte nach der Methode von Benjamini und Hochberg angepasst. Die Thresholds für die p-Werte der statistischen Tests und die Korrelation wurden über je vier ausgewählte Werte optimiert. Die getesteten Alternativen sind 0,01, 0,05, 0,1 und 0,2 für die p-Werte sowie 0,3, 0,5, 0,7 und 0,8 für die Korrelationen. Die verbliebenen Proteine wurden genutzt, um zu ermitteln, welches Klassifikationsverfahren die besten Ergebnisse liefert. Hierzu wurde die 10mal wiederholte 10-fold Crossvalidierung und als Optimierungskriterium die AUC der Precision-Recall-Kurve verwendet. Es wurden die Machine-Learning-Verfahren Random Forest, Support Vector Machine und künstliche Neuronale Netze, logistische Regression und LDA (linear discriminant analysis) verglichen. Im finalen Schritt wurde die Teilmenge der Proteine, auf welcher das ausgewählte Klassifikationsverfahren die besten Vorhersagen trifft, bestimmt und die Modellparameter final optimiert. Die Precision der Modelle wurde anhand von AUROC (Area Under the curve of the Receiver Operating Characteristic), Accuracy, Sensitivität und Spezifität beurteilt. Anschließend wurde die Permutation Feature Importance berechnet, um die Wichtigkeit der Proteine für Vorhersagen der Modelle zu bestimmen.

Klassifikator und Biomarker-Ensemble für Bisulfit DNA-Methylierungsmessungen (Arbeitspaket IPA):

DNA-Bisulfit-Methylierungsdaten von 394 Teilnehmenden, die 10 verbleibende Assays umfassen, wurden mithilfe von Methoden des maschinellen Lernens (Gradient Boosted Trees und Random Forests) ausgewertet. Dieser Datensatz umfasst auch Proben aus dem Vorgängerprojekt. Um zu verhindern, dass Assays mit einer geringen Zahl an Counts (methyliert + unmethyliert) die Ergebnisse verfälschen, wurden die Prozentwerte der Methylierung auf 0 gesetzt, sobald der Count unterhalb von 1250 lag. Die Proben wurden in zwei Gruppen eingeteilt, Teilnehmende mit Lungenkrebs (n=171) und Teilnehmende mit COPD bzw. Kontrollen (n=223).

Die Daten wurden in einen Trainingsdatensatz, auf dem die Algorithmen eine Regel zur Klassifizierung von Proben in eine der beiden Gruppen lernen, und einen Testdatensatz zur anschließenden Evaluierung aufgeteilt. Dabei wurden Untergruppen, wie z.B. Adenokarzinome oder COPD-Kontrollen berücksichtigt, um die Ähnlichkeit der beiden Datensätze zu gewährleisten. Es wurde außerdem eine Variablenselektion durchgeführt, die die wichtigsten Assays für eine gute Klassifizierung auswählt, einmal nach optimaler Accuracy und einmal nach optimaler Spezifität.

3.6 Epigenetik

Nachweis von DNA-Methylierungen

In diesem Projekt wurde DNA aus insgesamt 316 Nativgewebeproben (162 Teilnehmende mit Tumor- und/oder Nicht-Tumorgewebe), 70 EBUS-TBNA-/PE-Proben, 354 Speichelproben und 439 Plasmaproben extrahiert. Auf Basis der in der ersten Förderstufe mittels MBD-Sequenzierung (Screening) entdeckten ca. 800 potenziellen Methylierungsmarker wurden 40 Einzelassays (Pyrosequencing) entwickelt und getestet. Die dabei gefundenen 20 vielversprechendsten Assays wurden erfolgreich mit 68 Teilnehmenden (Tumor- gegen Nicht-Tumorgewebe) aus dem unabhängigen Kollektiv verifiziert. Aufgrund der hoch spezifischen Assays hat die Anzahl an Proben für die Verifizierung gereicht. Das ausgewählte Kollektiv beinhaltete die häufigsten Tumorsubtypen mit einem hohen Tumoranteil im Gewebeschnitt. Für die minimalinvasiven und nichtinvasiven Probeabnahmen, wie die Biomaterialien EBUS-TBNA-/PE-, Plasma- und Speichelproben, wurde eine spezifische Auswahl an Assays zusammengestellt und analysiert. Dabei konnten in 61 EBUS-TBNA-/PE-Proben 8 Assays getestet werden und in 354 Speichelproben weitere fünf Assays. Ebenso wurde die *deep sequencing* Methode für den Nachweis von DNA-Methylierungsmustern (Bisulfit-Seq) im Plasma auf Basis von 10 Assays etabliert. Mit dieser BS-Seq Methode wurden 423 Proben sequenziert. Zusätzlich wurde die Pyrosequencing Methode für die Analyse von Plasmaproben optimiert. Des Weiteren wurden dreizehn Marker zur Unterscheidung von COPD gegen Lungenkrebs auf Basis einer Literaturrecherche und MBD-Sequenzierung überprüft. Außerdem wurde ein Marker-Panel auf Methylierungs- und

Protein-Ebene für kleinzellige Lungen-Karzinome in Plasma identifiziert. Für alle Proben wurden Patientinnen und Patienten mit Metastasen und mit nicht eindeutigen Befunden ausgeschlossen. Für die Analysen in Plasma ergeben sich höhere Probenzahlen, weil viele Teilnehmende mit unterschiedlichen Methoden und teilweise doppelt untersucht wurden.

Nachweis von tumorspezifischer RNA

Aus allen im Studienzeitraum von Pure-Lunge 2 gesammelten Plasma- und Speichel-Proben wurde die RNA erfolgreich isoliert. Somit lag von insgesamt 362 Plasma- und 310 Speichel-Proben die isolierte RNA zur Analyse vor. Die Auswahl möglicher lncRNAs erfolgte fortlaufend anhand aktueller Literatur. Da in diesen Studien die lncRNAs in der Regel im Gewebe nachgewiesen wurden, wurde zunächst in einem kleinen Subkollektiv (N=20) geprüft, ob die ausgewählten lncRNAs überhaupt im Plasma und im Speichel nachweisbar sind. So wurden im Studienverlauf insgesamt rund 50 lncRNAs im Plasma und 40 lncRNAs im Speichel auf ihre generelle Nachweisbarkeit in diesen Körperflüssigkeiten mittels qPCR getestet. Die Mehrzahl der analysierten lncRNAs ließen sich allerdings nicht oder nur sporadisch im Plasma bzw. im Speichel nachweisen. In allen zur Verfügung stehenden Proben wurden schließlich aufgrund ihrer generell guten Nachweisbarkeit die lncRNAs *MALAT1*, *GAS5*, *RMRP*, *HNRPA1* und *FPR1* im Plasma und *MALAT1*, *GAS5*, *RMRP* und *BZYRN1* im Speichel als Marker-Kandidaten analysiert. Als Referenzen wurden die mRNAs *HPRT1*, *GUSB*, *B2M*, *RPLPO* und *PPIA* in den Plasma- und Speichel-Proben gemessen und die jeweils geeignetste Kombinationen aus zwei mRNAs als Referenz zur Normalisierung verwendet. Dies waren *B2M* und *PPIA* im Plasma sowie *B2M* und *RPLPO* im Speichel.

Nachweis von CNV in Gewebeproben

Im Studienzeitraum von Pure-Lunge 2 standen aufgrund der notwendigen Qualitätskriterien (Primärtumoren, Verfügbarkeit sowohl von Tumor- als auch Nicht-Tumorgewebe, Nicht-Tumoranteil sollte keine Tumorzellen aufweisen und frei von Entzündungen sein) Material von insgesamt 114 Probanden für die Analyse der CNV zur Verfügung. Aus allen geeigneten Tumor- und Nicht-Tumorgewebeproben der Probanden wurde die DNA erfolgreich isoliert und die Kopienzahl der Gene *NKX2-1*, *CCND1*, *MYC*, *ERBB2*, *MET* und *TERT* anschließend mittels digitaler PCR (dPCR) bestimmt, um Unterschiede der potentiellen CNV-Marker zwischen Tumor- und Nicht-Tumor Gewebe zu untersuchen.

4. Ergebnisse des Gesamtvorhabens

4.1 Rekrutierung und Probensammlung

Insgesamt wurden im Studienzeitraum 423 Patientinnen und Patienten rekrutiert (n=500 geplant), davon 333 Patientinnen und Patienten mit Verdacht auf Lungentumor (n=400 lt. Antrag geplant) sowie 90 COPD-Kontrollen (n=100 im Anschluss an Antrag geplant). Die Rekrutierung blieb somit leicht unter den geplanten Fallzahlen. Nicht für alle Teilnehmenden, die mit Verdacht auf Lungentumor eingeschlossen wurden, stand wie ursprünglich geplant zusätzliches Gewebematerial für Studienzwecke zur Verfügung, dies traf auch auf EBUS/PE-Gewebematerial zu (n=71 von n=100 geplanten Proben gesammelt, s. Tabelle 2).

Tabelle 1: Rekrutierte Patientinnen und Patienten in PURE Lunge II.

	Gesamt	Maligne Primärtumoren der Lunge	Benigne Lungenerkrankung ^a	Lungenmetastasen	COPD-Kontrollen
N	423 ^b	204	91	27	90
Alter [Jahre] (Median (IQR^c))	68 (61-75)	68 (62-73)	64 (58-74)	67 (59-75)	70 (62-78)
Geschlecht (N %)					
männlich	219 (52%)	108 (53%)	46 (51%)	14 (48%)	43 (48%)
weiblich	204 (48%)	96 (47%)	45 (49%)	13 (52%)	47 (52%)
Rauchstatus (N %)					
Nie-Raucher	72 (17%)	33 (16%)	19 (21%)	9 (33%)	9 (10%)
Ex-Raucher	145 (34%)	68 (33%)	31 (34%)	9 (33%)	37 (41%)
aktiver Raucher	198 (47%)	101 (49%)	40 (44%)	8 (30%)	41 (46%)
unbekannt	8 (2%)	2 (1%)	1 (1%)	1 (4%)	3 (3%)

^a Patientinnen und Patienten mit Verdacht auf Lungenkrebs, bei denen eine Krebsdiagnose nicht bestätigt wurde

^b enthalten sind 11 Teilnehmende mit sonstigen Erkrankungen, die nicht in einer eigenen Spalte dargestellt werden.

^c IQR: Interquartilsrange

Tabelle 2: Gesammelte Daten und Materialien der Teilnehmer.

Material	Total (N)	Total (%)	Maligner		Lungenmetastase	COPD-Kontrolle
			Primärtumor der Lunge	Benigner Tumor/Lungenerkrankung		
Teilnehmer insgesamt	423 ^a	100.0	204	91	27	90
Interview	400	94.6	195	88	26	82
Akte	423	100,0	204	91	27	90
Blut	405 ^b	95,7	193	87	27	87
davon Plasma	405 ^b	95,7	193	87	27	87
davon Blutkuchen	405 ^b	95,7	193	87	27	87
Speichel	404 ^c	95,5	195	89	27	83
Gewebe	255	76,6 ^d	175	55	21	0
davon EBUS/PE	71	21,3 ^d	39	25	4	0
davon Nativ	193	58,0 ^d	144	31	17	0
davon Paraffin	181	54,4 ^d	139	24	17	0
Histologischer Befund	324	97,3 ^d	204	91	27	0

^a enthalten sind 11 Teilnehmende mit sonstigen Erkrankungen, die nicht in einer eigenen Spalte dargestellt werden.

^b davon zwei Teilnehmende mit zwei Blutproben

^c davon zwei Teilnehmende mit zwei Speichelproben

^d Prozentzahl bezogen auf insgesamt n=333 Teilnehmende, bei denen Gewebe gesammelt werden sollte (d.h. ohne COPD-Kontrollen)

Tabelle 3: Subtypen maligner Primärtumoren der Lunge und Art der gewonnenen Gewebeproben.

Hauptgruppe primärer Lungentumor	Anzahl Gewebeprobe				
	Patienten	erhalten	Nativgewebe	Paraffingewebe	EBUS/PE
Plattenepithelkarzinom	47	42	35	35	8
Kleinzelliges Karzinom	7	4	0	0	4
Adenokarzinom	124	108	90	85	22
Nicht-kleinzelliges Karzinom NOS*	5	4	4	4	1
Karzinoid	8	5	5	5	0
Sonstiges	13	12	10	10	4

*NOS – not otherwise specified

4.2 Pathologie

Im Universitätsklinikum Köln wurden Gewebeproben von 194 Patientinnen und Patienten formalinfixiert, geschnitten und begutachtet. Diese teilten sich auf in primäre Lungentumoren (n=137, ca. 70 %), Metastasen anderer Entitäten (n=19, ca. 10 %) und nicht-maligne Tumoren bzw. Teilnehmende, für die das UK-Köln kein

Tumorgewebe erhalten hat (n=39, ca. 20 %). Von den insgesamt 194 eingegangenen Teilnehmenden konnte von 160 Teilnehmenden DNA aus FFPE-Gewebe gewonnen und sequenziert werden (Abbildung 3). Für 98 Teilnehmende wurde aus dem korrespondierenden Plasma ctDNA gewonnen, diese sequenziert und mit den Ergebnissen der Sequenzierung der DNA aus FFPE Material verglichen (Abbildung 4). Für 9 Teilnehmende, bei denen in der DNA eine *KRAS* c.35G>T Mutation im Gewebe nachgewiesen werden konnte, wurde die ddPCR-Methode Methode für korrespondierende ctDNA Proben verwendet (Tabelle 4).

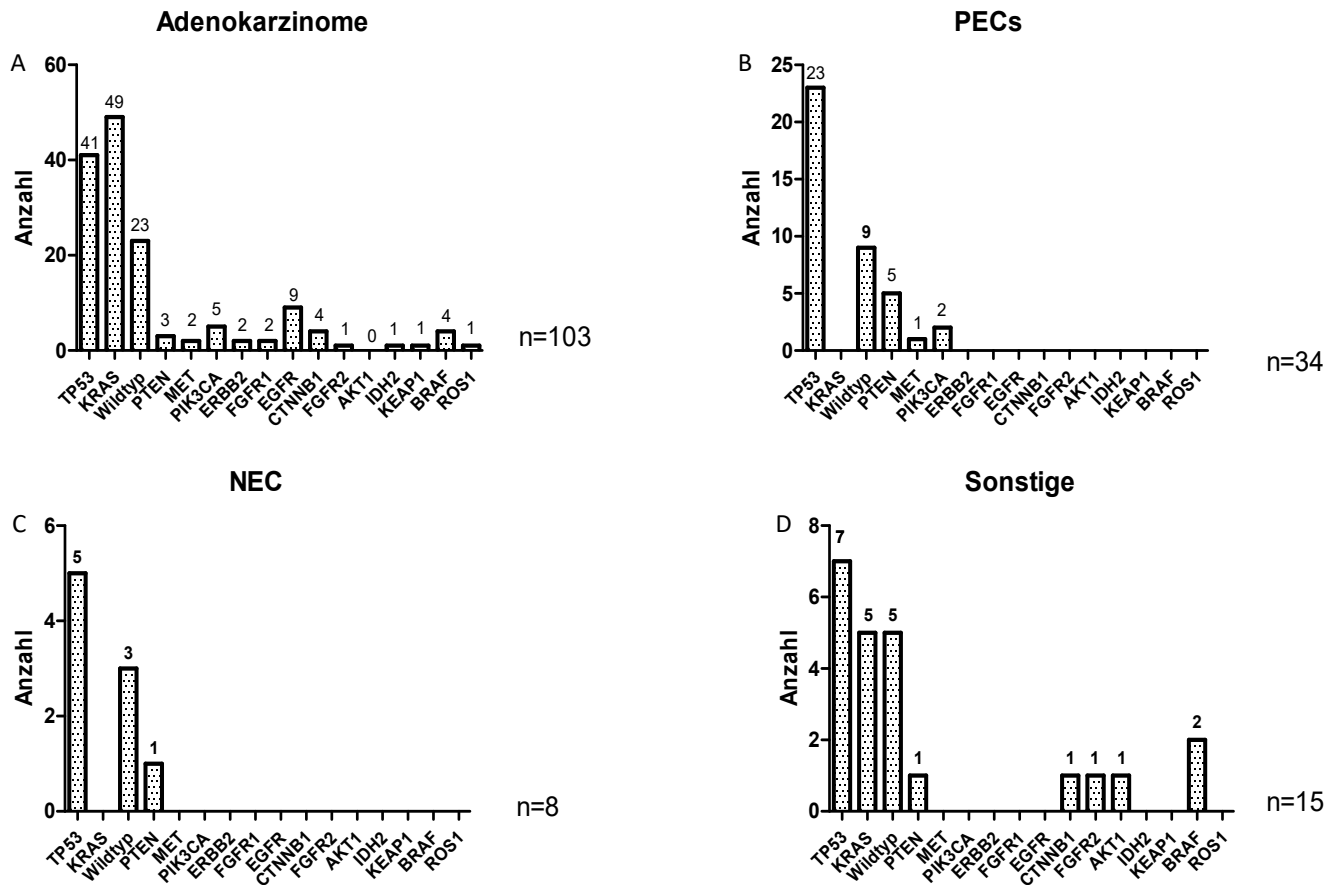


Abbildung 3: Prozentuale Aufteilung der in der DNA aus FFPE Material gefundenen Mutationen in (A) Adenokarzinomen, (B) Plattenepithelkarzinomen (PEC), (C) neuroendokrinen Tumoren (NEC) und (D) sonstigen Tumoren.

In DNA aus FFPE-Material konnten in den verschiedenen Entitäten unterschiedliche Mutationsprofile erfasst werden. In Adenokarzinomen wurden in den Genen *KRAS* und *TP53* am häufigsten Aberrationen gefunden. Bei Plattenepithelkarzinomen ist das *TP53*-Gen am häufigsten von Veränderungen betroffen. Ebenso verhielt es sich in neuroendokrinen Tumoren (Abbildung 3).

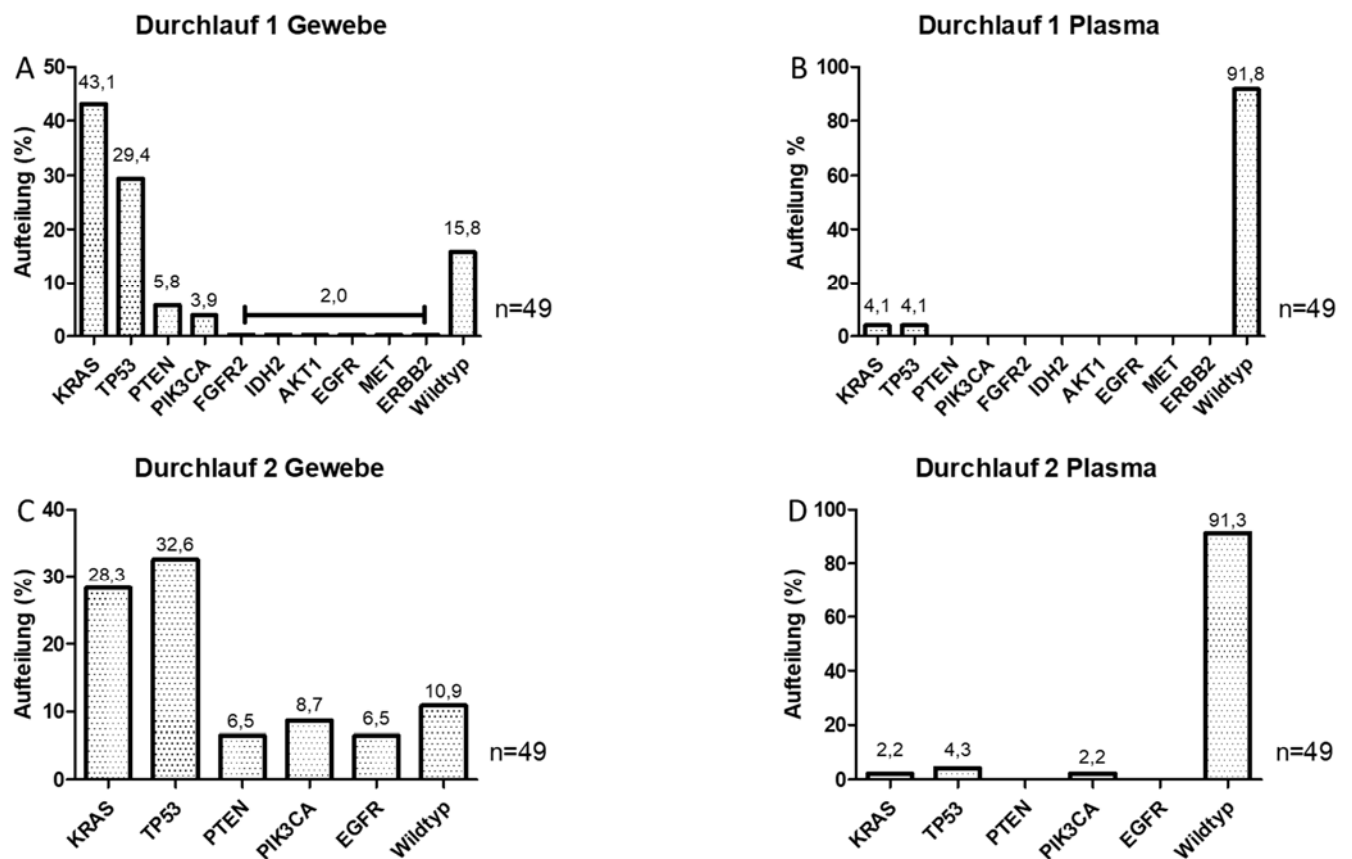


Abbildung 4: Prozentuale Aufteilung der gefundenen Mutationen in 98 Fällen mit Material aus FFPE-Gewebe und des korrespondierenden Plasmas. Gefundene Mutationen in DNA aus FFPE-Material im Durchlauf 1(A) und 2(C) und in ctDNA der korrespondierenden Plasmaproben im Durchlauf 1 (B) und 2 (D).

In 98 Fällen lag sowohl DNA aus FFPE Material als auch ctDNA aus den korrespondierenden Plasmaproben vor. Im Vergleich konnten in Durchlauf 1 8,2 % der Mutationen bzw. Deletionen im NGS von beiden DNA-Quellen gefunden werden. Diese waren gleichverteilt zu je 4,1 % im *KRAS*- und *TP53*-Gen (Abb. 4A, 4B). In Durchlauf 2 zeigte sich ein ähnliches Verhältnis. Hier konnten 8,7 % der Mutationen in beiden DNA-Quellen gefunden werden. Diese Verteilten sich zu jeweils 2,2 % auf Mutationen im *KRAS*- und *PIK3CA*-Gen und 4,3 % auf Mutationen im *TP53*-Gen (Abb. 4C, 4D).

Tabelle 4: ddPCR-Ergebnisse der Proben mit der *KRAS* c.35G>T Mutation.

Probe	Detektion c.35G>T	Tumor- Kopien/Gesamtreaktion	Konzentration ctDNA ng/μl
1	Ja	105	0,075
2	Nein	2	0,001
3	Nein	0	0
4	Nein	0	0
5	Nein	0	0
6	Nein	0	0
7	Nein	0	0
8	Ja	31	0,022
9	Nein	0	0

Tabelle 5: ddPCR-Ergebnisse der Proben mit der KRAS c.34G>T Mutation.

Probe	Detektion c.34G>T	Tumor- Kopien/Gesamtreaktion	Konzentration ctDNA ng/μl
1	Nein	0	0
2	Ja	2,4	0,002
3	Nein	0	0
4	Nein	0	0
5	Nein	0	0
6	Nein	0	0
7	Ja	2,2	0,002

Tabelle 6: ddPCR-Ergebnisse der Proben mit der KRAS c.35G>A Mutation

Probe	Detektion c.35G>A	Tumor- Kopien/Gesamtreaktion	Konzentration ctDNA ng/μl
1	Ja	219	0,16
2	Nein	0	0
3	Nein	0	0
4	Nein	0	0

Von den 20 Proben, die mit ddPCR untersucht wurden konnte nur in 5 Fällen (25 %) die jeweilige Mutation nachgewiesen werden (Tabellen 4-6). Die Konzentrationen der ctDNA betrug in den Proben zwischen 0,001 ng/μl und 0,16 ng/μl. Bei den restlichen Proben konnte nur extrem wenig (Tabelle 4, Probe 2) oder keine ctDNA nachgewiesen werden.

Tabelle 7: Vergleich der Sequenzierung mit DNA aus FFPE und FITR (Fourier-Transformations-Infrarotspektrometer).

#	Erste Sequenzierung (Lun3/ Lun5)	Allelfrequenz	Zweite Sequenzierung (Lun5) FTIR	Allelfrequenz
1	ERBB2: c.2313_2324dup p.Y772_A775dup; PTEN: c.219_222delAAGA p.E73Dfs*25; TP53: c.658T>C p.Y220H	21,9/63,3/66	ERBB2: c.2313_2324dup p.Y772_A775dup; PTEN: c.219_222delAAGA p.E73Dfs*25; TP53: c.658T>C p.Y220H, FGFR2: c.935T>G	29,9/73,8/83,8/5,2
2	EGFR: c.2153T>A p.L718Q; EGFR: c.2573T>G p.L858R	9,6/11,7	EGFR: c.2153T>A p.L718Q; EGFR: c.2573T>G p.L858R; ROS1: c.5704G>A p.E1902K	15,5/29,6/24,6
3	EX8: c.886delC p.H296Tfs*55	20,7	EX8: c.886delC p.H296Tfs*55	13,1
4	KRAS: c.34G>T p.G12C	15,3	KRAS: c.34G>T p.G12C	21,7
5	KRAS: c.34G>T p.G12C	18	KRAS: c.34G>T p.G12C	24,2
6	KRAS: c.34G>T p.G12C	34,1	KRAS: c.34G>T p.G12C	8,1
7	KRAS: c.37G>T p.G13C; TP53: c.659A>C p.Y220S	11,3/7,3	KRAS: c.37G>T p.G13C; TP53: c.659A>C p.Y220S	16,4/15,6
8	KRAS: c.35G>T p.G12V	30,1	KRAS: c.35G>T p.G12V	20,5
9	TP53: c.329G>T p.R110L	28,5	wt	-
10	KRAS: c.35G>T p.G12V; TP53: c.102_103insC p.P36Afs*7	51,1/45,6	KRAS: c.35G>T p.G12V	55,1
11	KRAS: c.34G>T p.G12C; TP53: c.376T>G p.Y126D	25/28,6	KRAS: c.34G>T p.G12C; TP53: c.376T>G p.Y126D	23,2/30,6
12	KRAS: c.35G>T p.G12V	32,7	KRAS: c.35G>T p.G12V (Sehr schlechte Coverage)	41,5
13	KRAS: c.34G>T p.G12C	77,7	KRAS: c.34G>T p.G12C (tw. Schlechte Cov.)	86
14	TP53: c.653T>G p.V218G	40,3	TP53: c.653T>G p.V218G	45,5
15	KRAS: c.35G>T p.G12V	36,4	KRAS: c.35G>T p.G12V	33,1
16	KRAS: c.37G>T p.G13C	24	KRAS: c.37G>T p.G13C	34,7
17	TP53: c.473_476delGCGC p.R158Pfs*11; TP53: c.838A>G p.R280G	16,7/23,6	TP53: c.473_476delGCGC p.R158Pfs*11; TP53: c.838A>G p.R280G	19,7/31,6

Der Vergleich der Sequenzierung mit DNA aus FFPE- und FITR-Material erfolgte mit 17 Proben. Hierfür wurden ausschließlich Adenokarzinome verwendet. Die Proben 1 und 2 aus FITR-Material zeigten nach der DNA-Isolation eine DNA-Konzentration von $\geq 1\text{ng}/\mu\text{l}$. Dadurch konnte das NGS mit dem Lun5-Panel (vgl. 3.3) durchgeführt werden. Da dieses Panel generell mehr Gene abdeckt, konnten zwei zusätzliche Mutationen entdeckt werden (

Tabelle 7, fett geschrieben). Probe 9 zeigte nach der Sequenzierung der durch FITR isolierten DNA keine Mutationen. Die restlichen Proben zeigten keinerlei Abweichungen.

Des Weiteren wurden jeweils 20 10µm dicke FFPE-Schnitte von Adenokarzinomen der Lunge, die eine *KRAS*, *TP53* oder *EGFR*-Mutation tragen bereitgestellt, um die Klassifizierung des „RF Classifiers“ zu trainieren und diese Mutationen in FFPE-Material zu erkennen.

4.3 Biophysik

Die Verifizierung der FTIR-Imaging-Ergebnisse der ersten Förderperiode (FP-339, Ruhrlandklinik, Essen) mit den Proben aus dem Lungenkrebszentrum Bonn/Rhein-Sieg konnte erfolgreich abgeschlossen werden. Die hohen Sensitivitäten und Spezifitäten der Methodik (über 93%) zur Tumorerkennung konnten anhand von 60 Tumorgewebeproben bestätigt werden. Die Tumorerkennung konnte erfolgreich auf das im vorherigen Berichtszeitraum eingeführte IR-Imaging-System mit Quantenkaskadenlasern (QCL) übertragen werden. Die Erkennung pathologischer und gesunder Gewebebereiche, die Differenzierung zwischen Tumor und Entzündung sowie die nachfolgende Tumortypisierung konnten erfolgreich etabliert werden. Die Tumorerkennung konnte mit einer Spezifität von 97% und Sensitivität von 95% anhand der Proben von 200 Patientinnen und Patienten (alle eingeschlossenen Patienten, von denen der Biophysik Gewebe zur Verfügung stand) gezeigt werden. Die erreichte Spezifität von 97% und Sensitivität von 94% für die Erkennung von Adeno- und Plattenepithelkarzinomen und jeweils 100 % für kleinzellige, neuroendokrine Karzinome und Karzinoide ist vergleichbar mit den Ergebnissen des FTIR-Imagings. Im Vergleich zum FTIR-Imaging konnte die Messzeit jedoch um das 160-fache beschleunigt und damit von Tagen auf wenige Minuten reduziert werden. Durch die große Anzahl von Studienteilnehmenden konnte nun neben dem klassischen maschinellen Lernen mit Ansätzen des Deep Learning's begonnen werden. Hierdurch versprechen wir uns eine noch größere Robustheit der Methodik.

Ein wichtiges neues Ergebnis dieser Förderperiode ist die Etablierung von molekularen Karten, die die direkte molekulare Identifizierung eines Adenokarzinoms Label-frei und automatisch anhand eines Gewebeschnitts erlauben. Hier konnte eine Erkennung für Tumoren mit einer K-Ras-, TP53- oder EGFR-Treibermutation etabliert und an jeweils 20 Proben validiert werden (Abbildung 5). Dabei werden eine Sensitivität und Spezifität von 95% erreicht. Eine Veröffentlichung der Ergebnisse konnte erfolgreich beim American Journal of Pathology eingereicht werden und befindet derzeit im Proof-Reading-Prozess (Goertzen et al. 2021).

Basierend auf der Erkennung der Treibermutationen wurden für die Proteomik Proben von 30 Teilnehmenden mit Adenokarzinomen (jeweils n=10 KAS, TP53 und EGFR) gesammelt. Zudem wurden für die Proteomik Proben von 29 Teilnehmenden mit Adenokarzinomen und korrespondierende Gesundproben gesammelt und der Proteomik zur Verfügung gestellt. Die gefundenen Gewebeproteinmarker wurden dann mit den im Blut gefundenen Markern abgeglichen. Dieser erfolgreich etablierte Ansatz wird aktuell auf weitere klinisch interessante Fragestellungen in dem gesammelten Kollektiv übertragen.

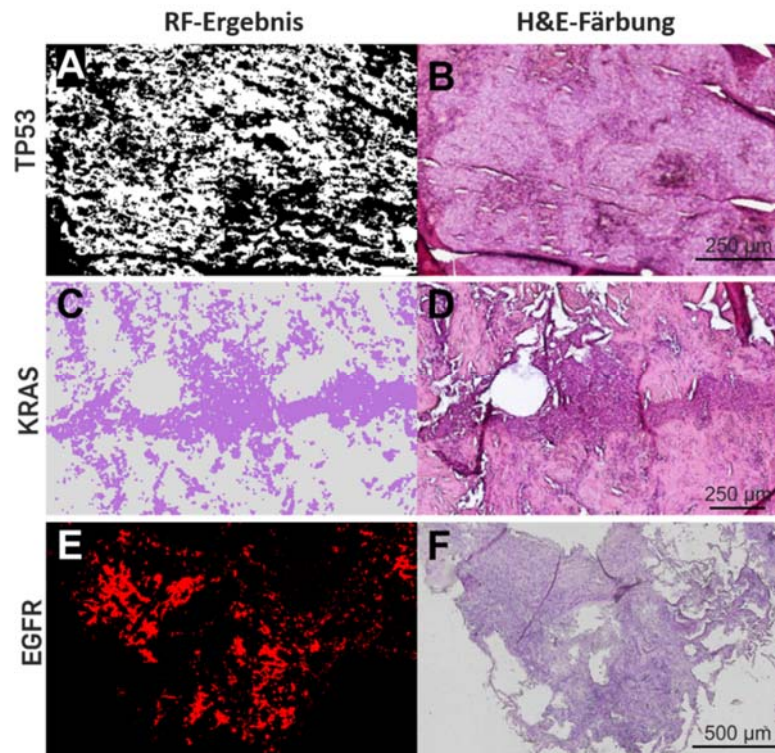


Abbildung 5: Random forest(RF)-Erkennung (A, C, E) der Treiber Mutationen im Vergleich zur H&E-Färbung (B, D, F) des QCL-IR-vermessenen Gewebedünnschnittes von Teilnehmenden mit Adenokarzinomen mit TP53- (A, weiß), K-Ras- (C, magenta) und EGFR-Mutation (E, rot)

4.4 Epigenetik

Nachweis von DNA-Methylierungen

Gewebe (Resektate)

Aus 40 erfolgreich etablierten Assays wurden die 20 vielversprechendsten DNA-Methylierungsmarker in Nativgewebe von 68 Patientinnen und Patienten (jeweils Tumor- und Nichttumorgewebe vom selben Lungenerkrankten) in diesem unabhängigen Kollektiv mittels Pyrosequencing verifiziert. Das Kollektiv beinhaltete 23 Adenokarzinome, 24 Plattenepithelkarzinome, vier Karzinoide, vier großzellige neuroendokrine Karzinome, drei nichtkleinzellige Karzinome ohne weitere Spezifizierung, vier sonstige und sechs benigne Lungentumoren. Für alle Assays wurde ein signifikanter Methylierungsunterschied mit einem p-Wert von $<0,000004$ zwischen malignem Tumorgewebe und Nicht-Tumorgewebe festgestellt (Beispiele s. Abbildung 6). Bei einer hohen Spezifität von 100% lag die Sensitivität für 19 Assays zwischen 87-40%. Für eine Sensitivität von 97% bei 100% Spezifität reichten drei Biomarker (DLMS002, DLMS004 und DLMS030) aus, um alle Lungentumoren (maligne und benigne) im Gewebe zu identifizieren. Für maligne Lungentumoren betrug die Sensitivität 95% (Spezifität: 100%) bei Nutzung eines erweiterten Markerpanels mit DLMS002, DLMS030, DLMS480, DLMS482 und DLMS003. Das Tumor Staging zeigt die Tendenz für alle Assays, dass Tumoren im Stadium T1 (N=11) meistens niedriger methyliert sind als in T2 (N=31), T3 (N=9) und T4 (N=8). Die Assays DLMS492 und DLMS025 weisen ähnlich hohe Methylierungswerte in allen T-Stadien auf und könnten somit besser für die Früherkennung geeignet sein.

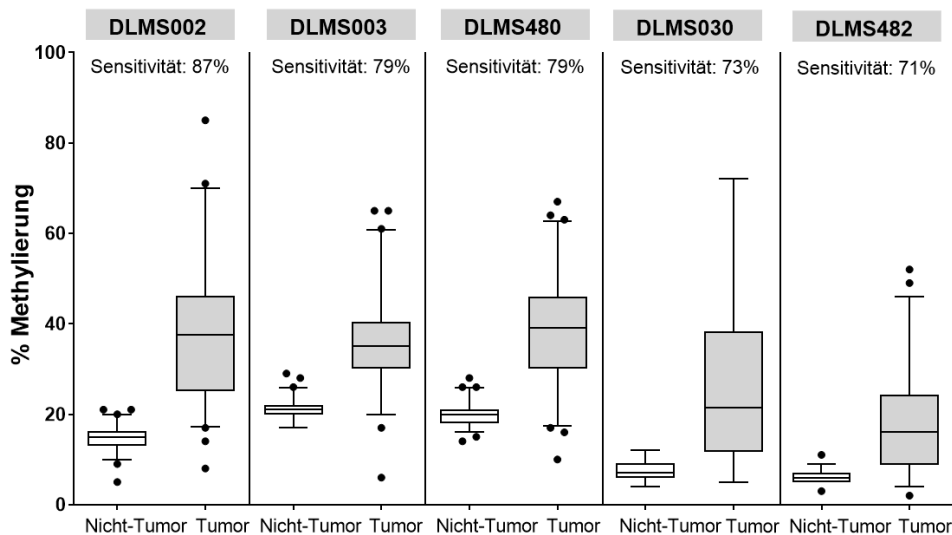


Abbildung 6: Boxplots der DNA-Methylierung von potenziellen Biomarkern in Nativgewebe (jeweils Paare von Nicht-Tumor- und Tumorgewebe) von 62 Lungenkrebskranken ($p < 0,000004$). Angegeben ist die Sensitivität bei jeweils 100% Spezifität. Die waagerechten Linien zeigen den Median und die 25% bzw. 75% Perzentile.

Gewebe (Biopsien)

Für die Untersuchung der DNA-Methylierung in Bronchialbiopsien standen 61 analysierbare EBUS-TBNA-/PE-Proben (31 Lymphknoten und 30 Lungengewebe, 100 geplant) zur Verfügung. Es konnten bei allen 8 potenziellen Biomarkern signifikante Methylierungsunterschiede zwischen 31 tumorfreien und 30 tumorinfiltrierten Lymphknoten/Lungengewebe gefunden werden ($p \leq 0,0007$) (Abbildung 7). Die EBUS-TBNA- (N=17) und PE-Proben (N=14) wurden zusammengefasst, weil die Unterschiede zwischen den beiden Gruppen meist klein waren und die Zuordnung häufig unklar war. Nur im Assay DLMS008 zeigten sich signifikante Unterschiede in dem tumorfreien Gewebe zwischen EBUS-TBNA und PE-Proben. Der vielversprechendste Assay DLMS492 zeigte im Median 37% Methylierung. Mit 49,5% Methylierung lag der Median bei den kleinzelligen Karzinomen (SCLC, N=6) am höchsten, gefolgt von Plattenepithelkarzinomen (SQCC, N=7) mit 42% und Adenokarzinomen (ADC, N=17) mit 31% Methylierung, wobei das tumorfreie Gewebe bei einem medianen Wert von 5% Methylierung lag. Bei einer Spezifität von 100% lagen die individuellen Sensitivitäten für die 8 Assays zwischen 20% und 73%. Bei einer Spezifität von 100% konnten 77% der Lungentumore mit der Assay-Kombination DLMS006 und DLMS492 erkannt werden (Abbildung 8). Die Sensitivität ist im Vergleich zum Nativgewebe niedriger, da der Anteil an reinen Tumorzellen in Biopsien oft geringer ist, außerdem ist eine Streuung des Tumors in die Lymphknoten nicht immer gegeben. Die Verifizierung war damit auch an den kleinen EBUS-Biopsieproben erfolgreich.

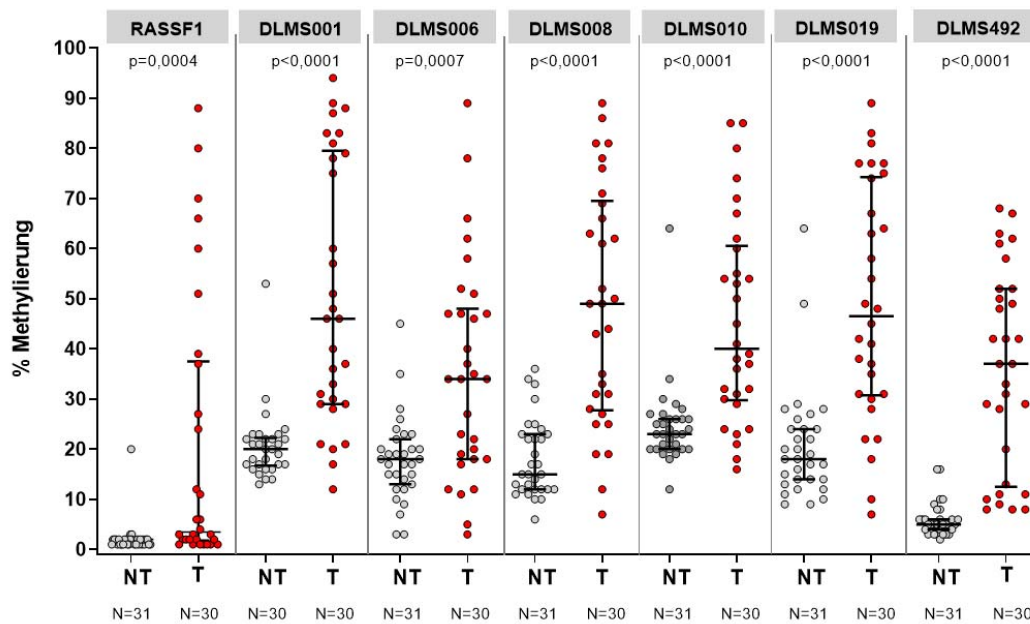


Abbildung 7: Boxplots der DNA-Methylierung von potenziellen Biomarkern in EBUS-TBNA-/PE-Proben (Nicht-Tumor- und Tumorgewebe) von Lungenkrebserkrankten ($p < 0,0007$). Die waagerechten Linien zeigen den Median und die 25% und 75% Perzentile an.

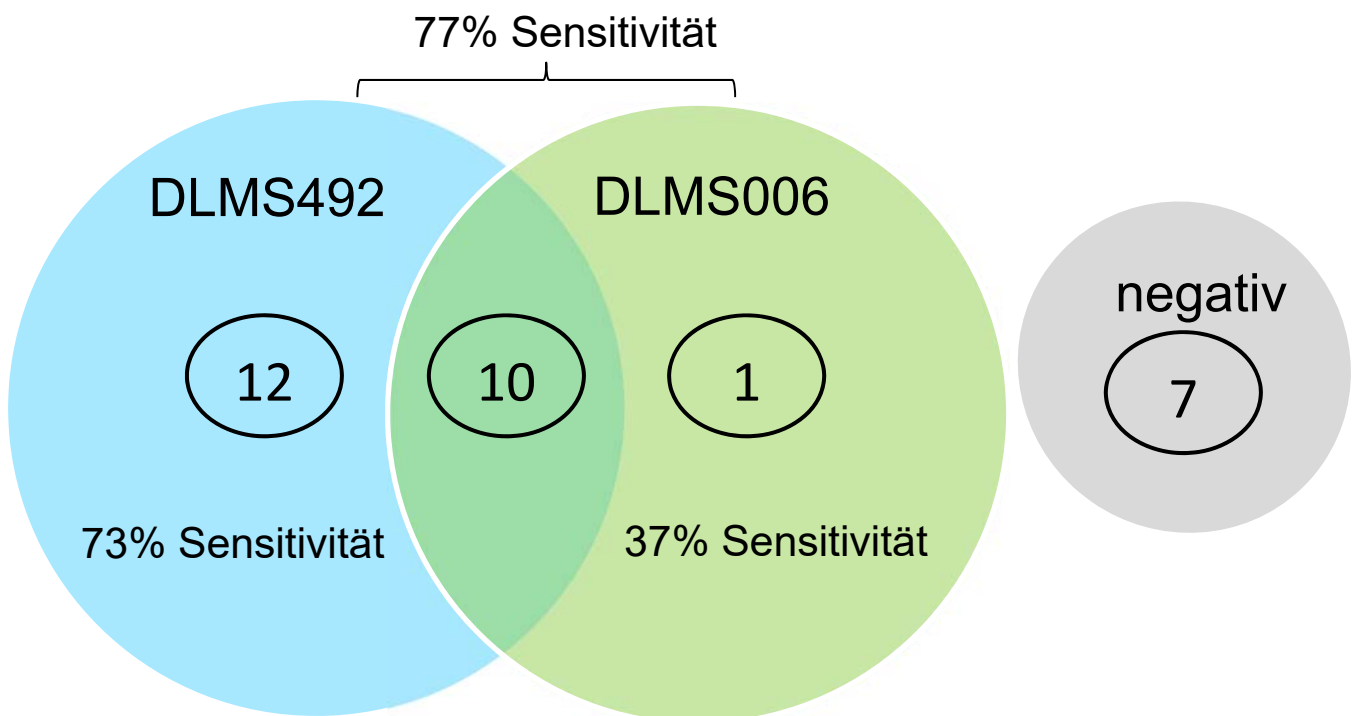


Abbildung 8: Venn-Diagramm aus dem Markerpanel DLMS492 (blauer Kreis) und DLMS006 (grüner Kreis) in EBUS-TBNA-/PE-Proben von Lungenkrebserkrankten. Die Zahlen im Kreis zeigen die Anzahl an positiv erkannten Lungentumoren. Der graue Kreis steht für die Anzahl an nicht erkannten Tumoren. Die Kombination der beiden Assays führt zu einer Erhöhung der Sensitivität auf 77%.

Plasma

Für die DNA-Methylierungsanalyse in Plasma wurde die Bisulfit-Seq erfolgreich etabliert und schrittweise weiter optimiert. Die aufwendige Auswertung wurde über die Bioinformatik des MPC entwickelt (siehe Abschnitt 3.5). Dazu wurden in einem ersten Schritt 31 Marker Kandidaten aus der Vorstudie (FP 339) in 48 Plasmaproben mittels Bisulfit-Seq analysiert. Diese Ergebnisse lieferten die Grundlage zur Etablierung eines Marker-Panels aus 10 Assays für die Methylierungsanalyse in Plasma. Bei der Methodenoptimierung wurde von 1 ml auf 3 ml Plasma und auf ein neues DNA-Plasma-Isolierungskit (Apostle MiniMax™ High Efficiency cfDNA Isolation Kit von Beckman Coulter) umgestellt, um die sehr geringe zirkulierende zellfreie DNA (cfDNA)-Menge in Plasma zu kompensieren. Außerdem wurde das Kit für die Bisulfit-Konvertierung gegen ein enzymatisches Bisulfit-Konvertierungskit ausgetauscht, weil damit nur noch minimal 10 ng cfDNA benötigt werden und auch weniger DNA bei der Konvertierung zerstört wird. Diese Optimierung führte dazu, dass in den meisten Fällen alle 10 Assays ein PCR Produkt lieferten und somit fast alle Proben sequenziert werden konnten. In 170 malignen Lungentumoren gegenüber 78 Bevölkerungskontrollen wurden signifikante Methylierungsunterschiede ($p < 0,01$) in den Assays DLMS019, DLMS025, DLMS034 und DLMS482 beobachtet (Abbildung 9). In den ROC-Kurven waren aber nur geringe Sensitivitäten (maximal 18%) bei 98% Spezifität der einzelnen Marker festzustellen. Ein Panel aus acht Biomarkern (DLMS001, DLMS002, DLMS006, DLMS019, DLMS025, DLMS034, DLMS492 und DLMS493) erreichte eine Performance von 32% Sensitivität bei einer Spezifität von 90%. Bei den kleinzelligen Lungenkarzinomen (N=20) erreichte die Bisulfit-Seq-Methode eine Sensitivität von 70% bei einer Spezifität von 90%, wobei hier oft mehrere Marker gleichzeitig informativ und die Methylierungen insgesamt höher waren. Die Plattenepithelkarzinome (N=37) erreichten nur 38% und die Adenokarzinome (N=85) 22 % Sensitivität, bei jeweils 90% Spezifität.

In der Vorstudie wurden keine COPD-Kontrollen eingeschlossen, in der aktuellen Studie wurden die COPD-Erkrankten jedoch als zusätzliche Gruppe bei Auswertung der Bisulfit-Seq-Ergebnisse berücksichtigt. Beim Vergleich von 170 Patientinnen und Patienten (maligne Tumoren) mit 145 Krankenhauskontrollen (benigne Erkrankungen/benigne Tumoren, einschließlich 84 Fälle (58%) mit COPD) konnten im Plasma keine Unterschiede beobachtet werden (Beispiel s. Abbildung 10). Der Hauptteil der erhöhten Methylierung bei den Krankenhauskontrollen wurde unter den COPD-Kontrollen beobachtet. Setzt man den Cutoff für eine positive Testentscheidung so, dass eine Spezifität von 98% im Vergleich zu den Bevölkerungskontrollen erzielt wird, würden 22% der COPD-Kontrollen irrtümlich als Lungenkrebs klassifiziert. Für die Subgruppe der kleinzelligen Tumoren konnte ein signifikanter Methylierungsunterschied gegenüber den Krankenhauskontrollen (COPD und benigne Lungenerkrankungen) ($p = 0,0113$) ermittelt werden. Hierbei wurde eine Sensitivität von 65% bei einer Spezifität von 90% erreicht. Weitere Ergebnisse der Analyse von Plasma-DNA Proben mittels Bisulfit-Seq sind dem Abschnitt 3.5 der Bioinformatik des MPC zu entnehmen.

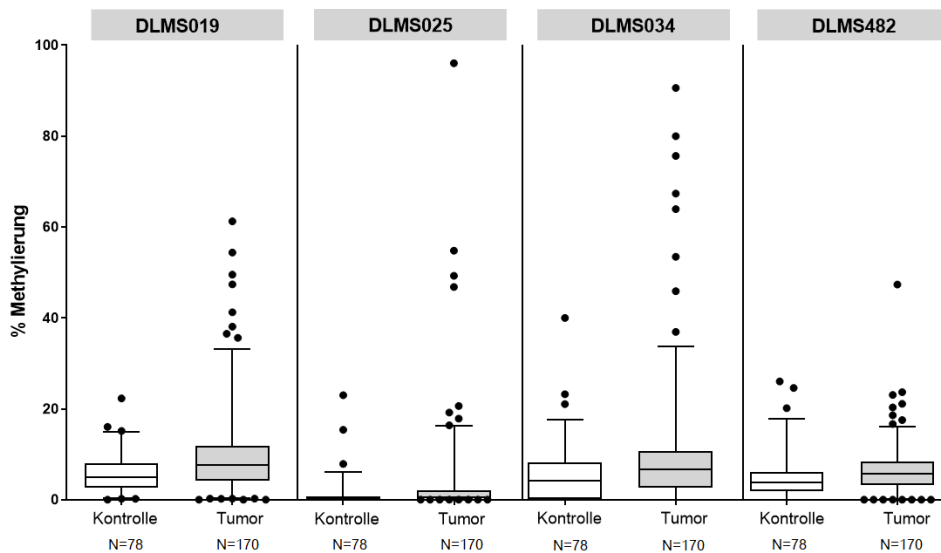


Abbildung 9: Boxplots der DNA-Methylierung von potenziellen Biomarkern in Plasma von Bevölkerungskontrollen aus FP-339 gegen maligne Lungentumoren aus beiden Projekten ($p < 0,01$). Die waagerechten Linien zeigen den Median und die 25% und 75% Perzentile.

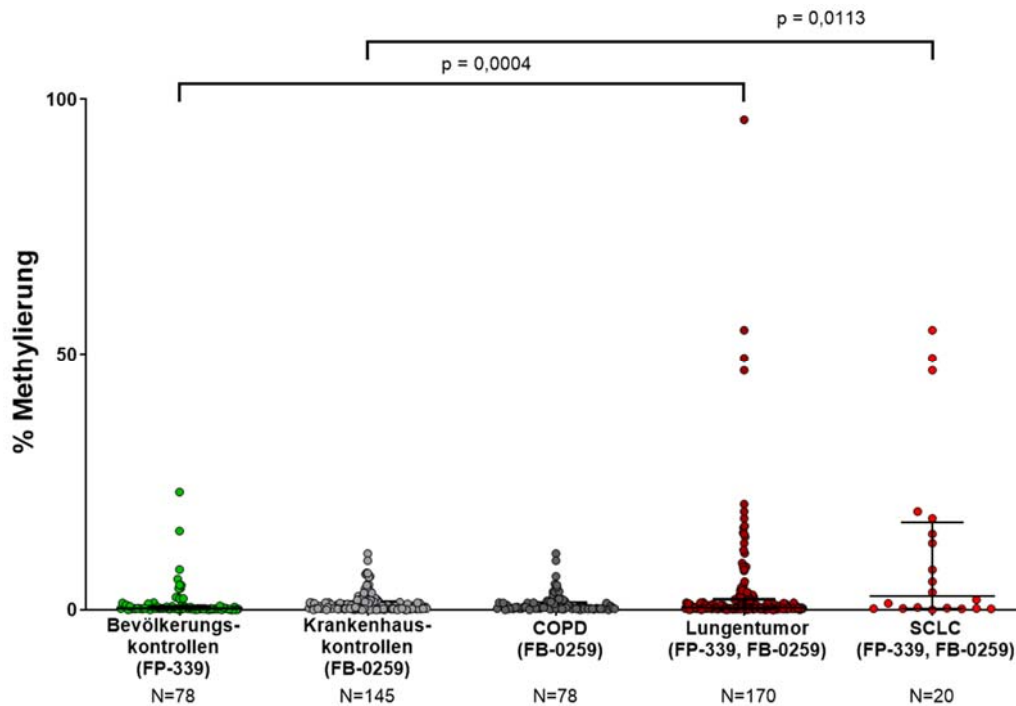


Abbildung 10: Messwerte der DNA-Methylierung in Dotplots des potenziellen Biomarkers DLMS025 in Plasma von Bevölkerungskontrollen (aus PURE-Lunge 1, FP-339), von Krankenhauskontrollen aus FB-0259 (benigne Lungenerkrankungen/Tumoren, darunter 58% COPD-Kontrollen), von COPD-Kontrollen separat (FB-0259), von Lungentumoren und kleinzelligen Karzinomen (SCLC) aus FP-339 und FB-0259. Ein signifikanter Unterschied war zwischen Bevölkerungskontrollen und malignen Lungentumoren festzustellen ($p = 0,0004$) und Krankenhauskontrollen und SCLC ($p = 0,0113$). Die waagerechten Linien zeigen den Median an und die 25% und 75% Perzentile.

Weitere Analysen mit Plasma-cfDNA zeigten, dass Plasmaproben von Tumor- und COPD-Kontrollen höhere DNA-Mengen im Vergleich zu Krankenhauskontrollen aufweisen. Vor allem in Plasmaproben von kleinzelligen Lungentumoren wurden deutlich größere DNA-Mengen isoliert. Die Pyrosequencing-Methode mit cfDNA aus Plasma wurde optimiert, indem eine Mindestmenge von 30ng bei der Bisulfit-Konvertierung eingesetzt und

grundsätzlich Dreifachbestimmungen durchgeführt wurden. Für manche Assays wurden höhere Eingangsmengen benötigt, sodass die minimale DNA-Menge bei der Etablierung für jeden neuen Assay überprüft werden sollte. Im Projekt FP-339 wurde schon gezeigt, dass ein signifikanter Unterschied zwischen kleinzelligen Lungenkarzinomen und Kontrollen für die Pyro-Seq Einzelassays DLMS001 und DLMS002 nachweisbar waren. In diesem Projekt mit einem unabhängigen Kollektiv zeigten kleinzellige Tumoren ebenfalls einen signifikanten Unterschied zu Kontrollen und COPD-Fällen, mit einem p-Wert von 0,010 für den Assay DLMS002 (Abbildung 11). Durch die Kombination der Methylierungsassays DLMS001 (12% Sensitivität) und DLMS002 (5% Sensitivität) mit dem Proteinassay Calretinin (34% Sensitivität) konnten kleinzellige Karzinome (N=41) aus FP-339 und FB-0259 mit einer Sensitivität von 38,5% bei einer Spezifität von 99% erkannt werden. Unter zusätzlicher Berücksichtigung der empfindlicheren Bisulfit-Seq Assays stieg die Sensitivität auf 51,2%, jedoch sank die Spezifität auf 90%. Parallel zum Bisulfit-Seq Ansatz wird auch eine Optimierung der Pyro-Seq Methode für Plasma verfolgt, da sie in der Praxis einfacher anzuwenden ist. Die DNA-Isolierungsmethode und die enzymatische Konvertierung der Bisulfit-Seq Methode wurde auf die Pyro-Seq Methode übertragen. Dadurch wurde die Methode sensitiver und meistens war nur noch eine Doppelbestimmung nötig. Die Anwendung im SCLC Kollektiv war noch nicht möglich, da nicht genügend Plasma von Patientinnen und Patienten mit SCLC vorlag. In einem anderen Kollektiv mit COPD-Markern sah die neue Methode sehr vielversprechend aus. Außerdem wurde eine neue DNA-Methylierungsanalyse auf Basis der digitalen PCR (dPCR) etabliert, die ebenfalls einfacher als die Bisulfit-Seq Methode durchführbar ist. Die Umsetzung auf cfDNA in Plasma ist noch in der Etablierungsphase und könnte ein weiterer Baustein zur Früherkennung von Lungenkrebs mittels des MoMar-Kollektivs sein.

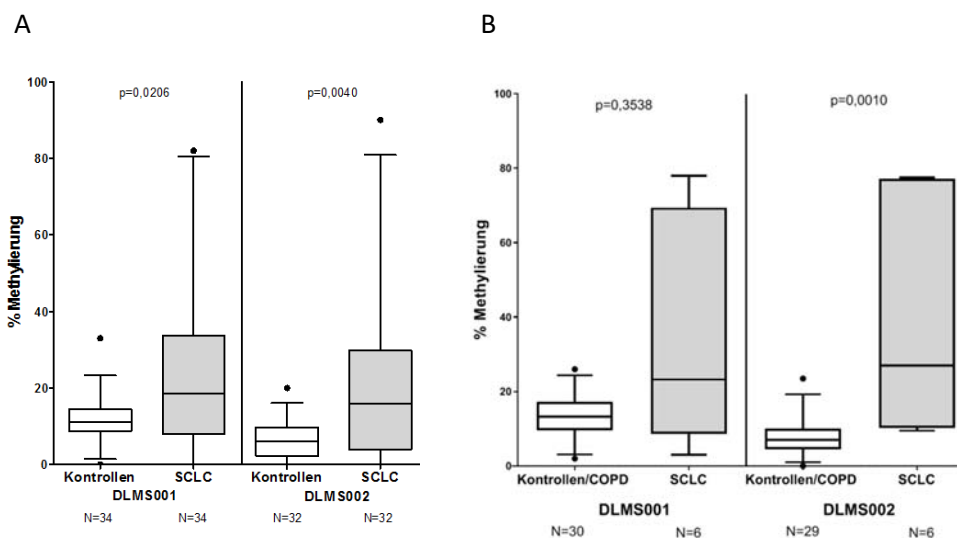


Abbildung 11: Boxplots der DNA-Methylierung von DLMS001 und DLMS002 in Plasma von Teilnehmenden mit kleinzelligen Lungenkarzinomen im Vergleich zu gesunden und Krankenhauskontrollen mit benignen Lungenerkrankungen. Die waagerechten Linien zeigen den Median und die 25% und 75% Perzentile. A) Kollektiv FP-0339, B) Kollektiv FB-0259.

Speichel

Aus insgesamt 367 Speichelproben wurde DNA extrahiert, da Proben von Teilnehmenden mit Metastasen ausgeschlossen wurden (Tabelle 2). Von diesen konnten insgesamt 354 Proben mit ausreichendem DNA-Material für die weitere Analyse verwendet werden: In Speichelproben wurden von 188 Patientinnen und Patienten mit malignen Lungentumoren und 166 Krankenhauskontrollen (benigne Lungenerkrankungen/Tumoren mit und ohne COPD) aus FB 259 im Vergleich zu 203 Bevölkerungskontrollen aus FP-339 fünf ausgewählte Methylierungsmarker

(DLMS001, 025, 482, 492 und 495) mittels Pyrosequencing analysiert. Die im vorherigen Projekt gefundenen signifikanten Unterschiede ($p < 0,001$) zwischen den Bevölkerungskontrollen und den Tumoren wurden verifiziert (Beispiel s. Abbildung 12). Der Methylierungsgrad der Tumorproben bewegte sich bei diesen Assays im gleichen Rahmen wie in den Proben der Vorstudie. Im Vergleich zum Vorprojekt sank der Median der Tumorproben etwas, jedoch lediglich für die Assays DLMS001 und DLMS025. Bei einer Spezifität von 98% lag die Sensitivität jedoch nur zwischen 4,4-23,1% (bei 95% Spezifität: Sensitivität 8,7-37,8%) für diese fünf Biomarker. Mit einem Markerpanel, bestehend aus DLMS001, DLMS482, DLMS495 und DLMS25, konnten 28% der malignen Tumoren bei einer Spezifität von 98% richtig erkannt werden. Für das Plattenepithelkarzinom lag der Wert bei 32% und für kleinzellige Karzinome bei 43% Sensitivität bei gleichbleibender Spezifität von 98%.

Eine Unterscheidung zwischen Krankenhauskontrollen (benigne Lungenerkrankungen/Tumoren, darunter 56% COPD-Kontrollen) und Lungenkrebskranken in Speichelproben war mit diesen fünf Markern nicht möglich. Der Anteil an falsch als Lungenkrebs diagnostizierten COPD-Kontrollen machte 30% in der COPD-Gruppe aus, wenn der Cutoff auf 98% Spezifität auf Basis der Bevölkerungskontrollen festgelegt wird. Krankenhauskontrollen ohne COPD gegen Kontrollen mit COPD zeigen signifikante Methylierungsunterschiede in den Assays DLMS001, 025 und 482 ($p < 0,03$). COPD-Fälle sind der Hauptgrund für die erhöhte Methylierung in der Gruppe der Krankenhauskontrollen. Außerdem sind, wie bei den Tumoren, pro Fall mindestens zwei oder mehr Assays gleichzeitig positiv. Hier stellt sich die Frage, ob diese positiven Assays bei den COPD-Erkrankten einen ersten Hinweis auf einen bevorstehenden Lungenkrebs liefern. Da von den 90 COPD-Kontrollen ein Tumor bei ca. 14 Personen innerhalb von 2-3 Jahren erwartet werden kann, von denen nur ein Teil das entsprechende Markerspektrum aufweisen wird, ist fraglich, ob dieses die fehlende Differenzierung zwischen Lungenkrebs und COPD erklären kann.

Es wurden 13 neue Methylierungsmarker auf Basis von Literaturangaben und Ergebnissen der MBD-Seq aus FP-339 und mit 11 weiteren Nicht-Tumorgewebeproben von Lungenkrebskranken mit einer zusätzlichen COPD aus FB-0259 für die Pyrosequencer-Plattform etabliert, um COPD-Kontrollen von Tumorerkrankten zu unterscheiden. Im Gewebe wurden Unterschiede zwischen Nicht-Tumor mit COPD gegen Tumor in 12 Assays gefunden. Die Biomarker PAX9, HOXD10, IRX4 und DLMS815 sahen vielversprechend im Nativgewebe aus. Im Speichel wurden anschließend 10 Assays in einem kleinen Kollektiv untersucht. Die vielversprechendsten Biomarker IRX4 und DLMS815 wurden anschließend in einem größeren Kollektiv analysiert, jedoch zeigten sie im Speichel nicht den erwünschten Erfolg (Beispiel s. Abbildung 13).

Eine Verifizierung von Methylierungsmarkern für Gewebe (Resektate und Biopsien), Speichel sowie Plasma war somit im Vergleich mit gesundem Gewebe bzw. gesunden Probanden möglich. Für die Krankenhauskontrollen mit benignen Lungenerkrankungen (FB-0259) konnte auf Basis der DNA-Methylierungsanalyse im Vergleich zu den Lungenkrebskranken bisher kein signifikanter Unterschied gefunden werden, der insbesondere auf die höhere DNA-Methylierung in den Assays von COPD-Kontrollen zurückzuführen ist (Beispiel siehe Abbildung 12).

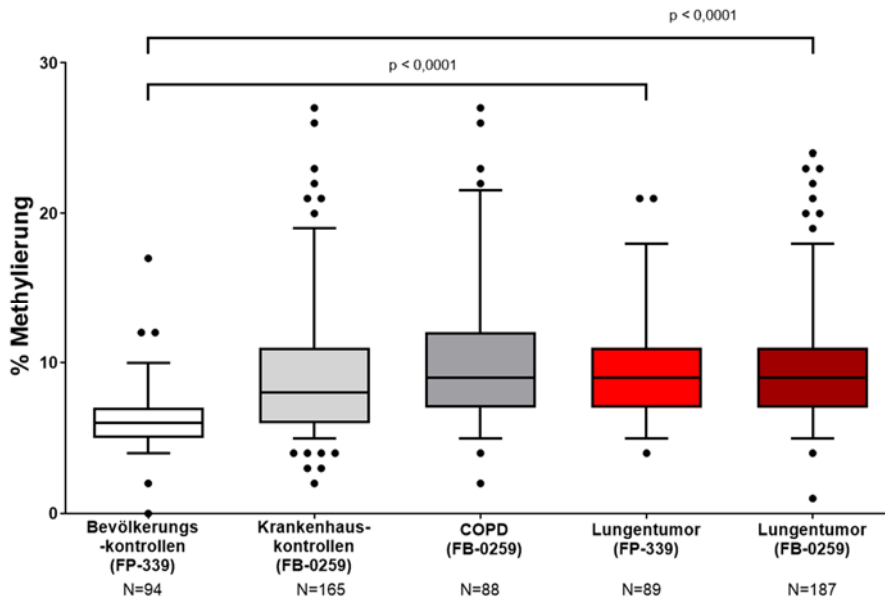


Abbildung 12: Boxplots der DNA-Methylierung des potenziellen Biomarkers DLMS482 in Speichel von Bevölkerungskontrollen (FP-339), von Krankenhauskontrollen aus FB-0259 (benigne Lungenerkrankungen/Tumoren, darunter 56% COPD-Kontrollen), COPD-Kontrollen in FB-0259 und von Lungenkrebs erkrankten aus PURE FP-339 und FB-0259. Ein signifikanter Unterschied war zwischen Bevölkerungskontrollen und malignen Lungentumoren festzustellen ($p < 0,0001$). Die waagerechten Linien zeigen den Median und die 25% und 75% Perzentile.

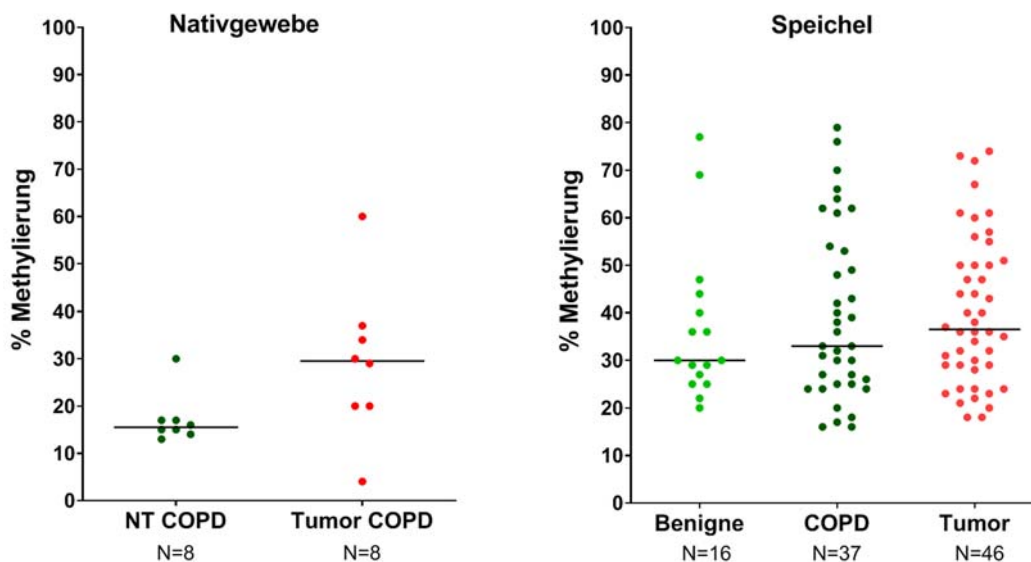


Abbildung 13: Dotplots der DNA-Methylierung des Biomarker-Kandidaten IRX4 in Nativgewebe- und Speichelproben (PURE Lunge 2). In Nativgewebe wurden Nicht-Tumor- und Tumorgewebe von Lungenkrebs erkrankten analysiert. Im Speichel wurden Kontrollen mit benignen Erkrankungen, COPD-Kontrollen und Lungenkrebs erkrankten untersucht. Die waagerechten Linien zeigen den Median an.

Nachweis von tumorspezifischer RNA

Mit allen aus Plasma isolierten RNA-Proben wurde eine Analyse der lncRNA-Kandidaten durchgeführt. Dabei konnten RNA-Proben von insgesamt 190 Lungenkrebskranken und 172 Kontrollen verwendet werden. Es konnten mehr Plasma-Proben als ursprünglich geplant gesammelt und ohne Mehrkosten analysiert werden (geplant: (n=150 Lungenkrebskranken, n=150 Kontrollen)). Im Plasma zeigten die vier Marker-Kandidaten *MALAT1*, *GAS5*, *HNRPA1* und *FPR1* einen statistisch signifikanten Gruppenunterschied ($p < 0,05$) zwischen den Lungenkrebskranken und den Kontrollen, *RMRP* hingegen nicht (Abbildung 14).

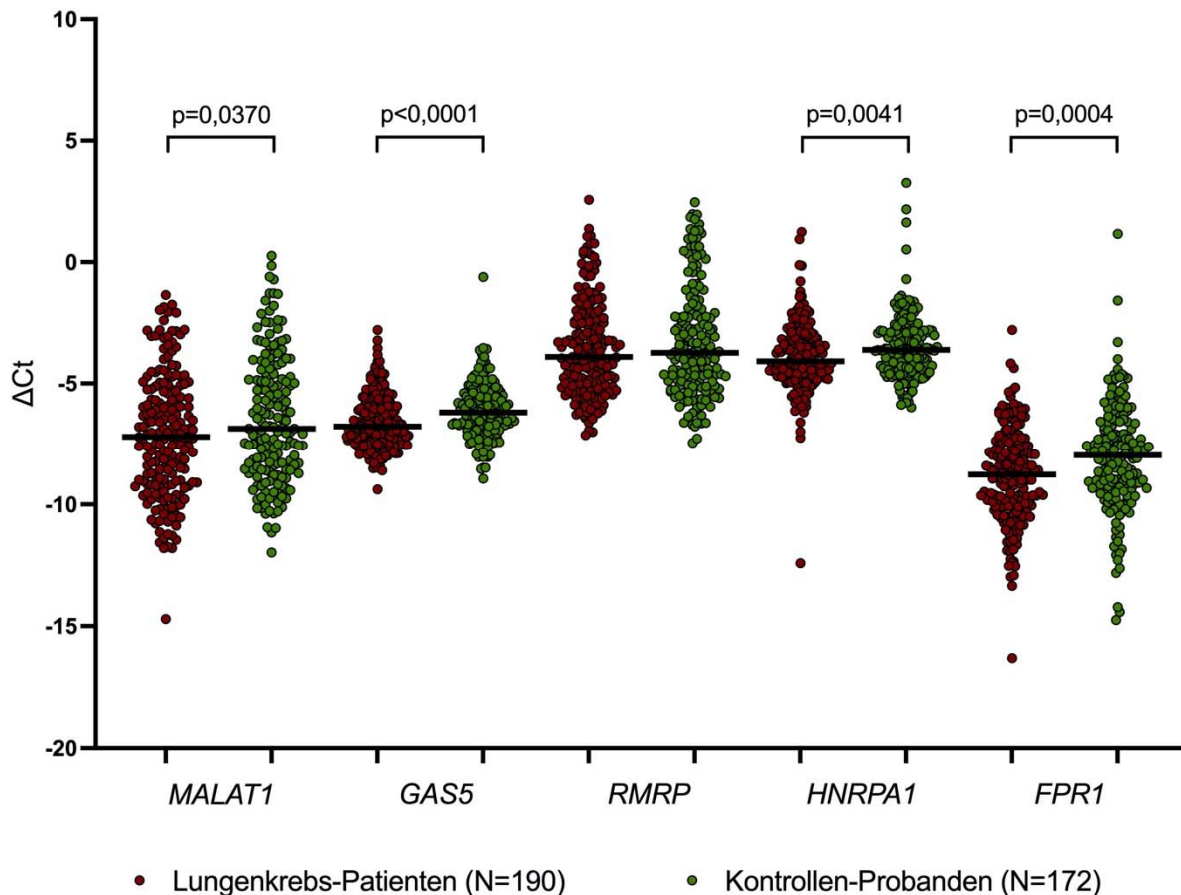


Abbildung 14: Verteilung von *MALAT1*, *GAS5*, *HNRPA1*, *FPR1* und *RMRP* im Plasma von Lungenkrebskranken und Kontrollen. Die horizontale Linie zeigt den Median.

Für die Analyse der lncRNAs im Speichel konnten RNA-Proben von insgesamt 173 Lungenkrebskranken und 137 Kontrollen verwendet werden. Analog zu den Plasma-Proben konnten mehr gesammelten Speichelproben gesammelt und ohne Mehrkosten analysiert werden. Im Speichel zeigte nur der Marker-Kandidat *BZYRN1* einen statistisch signifikanten Gruppenunterschied ($p < 0,05$) zwischen den Lungenkrebskranken und den Kontrollen, die übrigen drei lncRNAs *MALAT1*, *GAS5* und *RMRP* hingegen nicht (Abbildung 15).

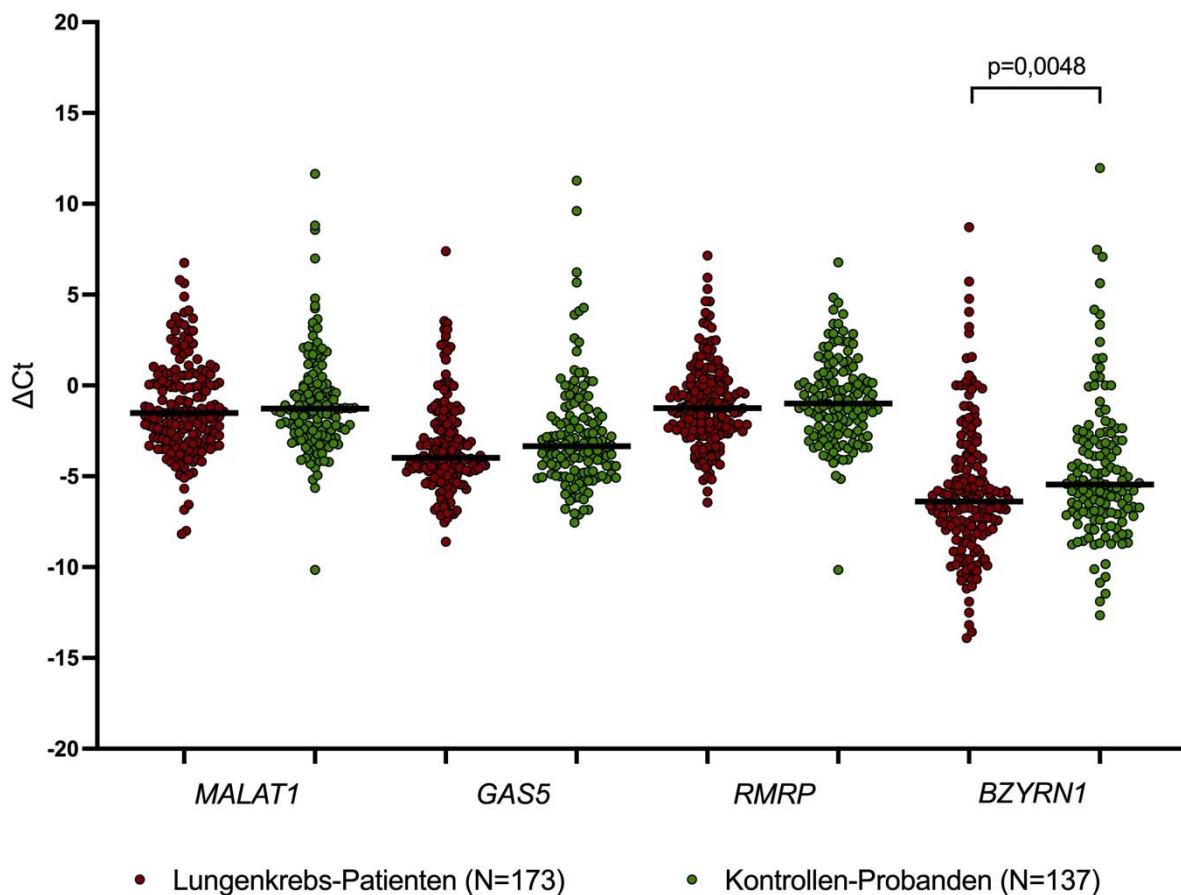


Abbildung 15: Verteilung von MALAT1, GAS5, RMRP und BZYRN1 im Speichel von Lungenkrebskranken und Kontrollen. Die horizontale Linie zeigt den Median.

Seit Beginn der Studie ist es auffällig gewesen, dass die unterschiedlichen lncRNAs relativ große und vor allem zwischen den beiden untersuchten Gruppen vergleichbare Streuungen aufweisen. Diese Streuungen könnten auf bisher unbekannte Faktoren zurückzuführen sein, denen biologische Variationen zugrunde liegen könnten bzw. die im Bereich der Analytik bzw. Präanalytik zu suchen sind und die einen Einfluss auf die Messwerte haben könnten. So ergibt sich allerdings dementsprechend auch keine zufriedenstellende Performance der untersuchten Marker-Kandidaten, trotz der beobachteten signifikanten Gruppenunterschiede. Bei einer festgesetzten hohen Spezifität von mindestens 98% sind durchgehend sehr niedrige Sensitivitäten <10% für die analysierten lncRNAs im Plasma und im Speichel festzustellen. Eine Kombination von unterschiedlichen Markern verdoppelt zwar die Sensitivitäten bei gleichbleibend hoher Spezifität auf 12,6% im Plasma und 5,2% im Speichel, allerdings sind die erreichten Sensitivitäten weiterhin nicht ausreichend für den Einsatz der lncRNAs als Marker zur Detektion von Lungenkrebs (

Tabelle 8).

Tabelle 8: Sensitivitäten der Marker-Kandidaten und der geeignetsten Kombination von Markern bei einer festgesetzten Spezifität von mindestens 98% zur Diskriminierung von Lungenkrebskranken und Kontrollen.

	MALAT1	GAS5	RMRP	HNRPA1	FPR1	BZYRN1	Kombination*
Plasma	5,3%	2,1%	1,9%	6,8%	2,1%	n.b.	12,6%
Speichel	2,9%	2,9%	2,9%	n.b.	n.b.	2,9%	5,2%

*Plasma: MALAT1, RMRP, HNRPA1, FPR1; Speichel: MALAT1, GAS5, RMRP, BZYRN1, n.b.: nicht bestimmt

Auch bei der Analyse bezüglich des histologischen Subtyps (Adenokarzinom, Plattenepithelkarzinom und kleinzelliges Karzinom) zeigte sich keine relevante Verbesserung der Performance der analysierten lncRNAs und so blieben die individuellen Sensitivitäten bei der festgesetzten hohen Spezifität von mindestens 98% sowohl in Plasma als auch Speichel durchgehend bei <10%.

Im Lauf der Studie zeigte sich, dass eine Unterscheidung von Lungenkrebskranken und COPD-Kontrollen mittels der untersuchten RNA-Marker nicht eindeutig möglich ist. COPD ist allerdings innerhalb der Studiengruppe in größerer Anzahl vertreten. So wiesen aus dem Plasma-Kollektiv 54 Teilnehmende mit Lungenkrebs zusätzlich auch eine COPD auf und 136 Teilnehmende keine COPD. Von den Kontrollen hatten 96 Teilnehmende eine COPD und 76 Teilnehmende nicht. Im Speichel-Kollektiv wiesen 52 der Lungenkrebskranken auch eine COPD auf, 121 Teilnehmende hingegen nicht. Von den Kontrollen hatten 76 eine COPD und 61 Teilnehmende keine.

Daher wurden im Verlauf des Projekts zusätzlich unterschiedliche in der Literatur beschriebene molekulare Marker sowohl auf RNA-Ebene (miR-212, miR-145 und miR-132-3p) als auch auf Protein-Ebene (ALCAM, Calretinin, OPG, Osteopontin und Mesothelin) analysiert, um zu klären, ob sie zur Diskriminierung von Lungenkrebskranken und Kontrollen mit COPD beitragen können. Bei der Analyse in einem kleineren Subkollektiv (N=78), bestehend aus 39 Teilnehmenden mit Lungenkrebs mit/ohne COPD und 39 Kontrollen mit diagnostizierter COPD zeigte sich, dass die beste Kombination aus miR-212 und Calretinin bei einer festgesetzten Spezifität von mindestens 98% eine Sensitivität von 20,5% erreicht.

Nachweis von Copy Number Variations (CNV) in Gewebeproben

Für die Analyse der CNV wurde die DNA aus den Gewebeproben (jeweils Tumor- und Nicht-Tumor Material) von 114 Lungenkrebskranken isoliert (geplant n=150). Die Anzahl der untersuchten Gewebeproben basierte auf den zugrunde liegenden Auswahlkriterien. So sollte pro Fall immer ein Paar bestehend aus Tumor und Nicht-Tumor Material zur Verfügung stehen, das Nicht-Tumor-Gewebe sollte tumor- und entzündungsfrei sein und nach der Isolation musste eine ausreichende Menge an isolierter DNA vorliegen. Diese Bedingungen konnten in 114 Probenpaaren realisiert werden. Die CNV-Analyse erfolgte mittels digitaler PCR unter Einsatz von kommerziellen Assays für die Marker *NKX2-1*, *CCND1*, *MYC*, *TERT*, *ERBB2* und *MET*. Als Referenz wurde *RNaseP* eingesetzt.

Während *NKX2-1* und *CCND1* keine Unterschiede zwischen Tumor- und Nicht-Tumor Gewebe aufwiesen, zeigten sich signifikante Gruppenunterschiede für *ERBB2* (p=0,0065), *MET* (p=0,0019), *TERT* (p<0,0001) und *MYC* (p<0,0001) zwischen den Tumor- und Nicht-Tumorproben. Ausgehend von einer festgesetzten Spezifität von 99% ließen sich mittels ROC-Analyse geeignete Cutoffs für die vier Marker bestimmen. Darauf basierend war die Kopienzahl von *ERBB2* bei 27, von *MET* bei sieben, von *TERT* bei 47 und von *MYC* bei 58 der 114 Lungenkrebskranken im Tumorgewebe erhöht (Abbildung 16). Für keinen der untersuchten Marker konnte ein signifikanter Gruppenunterschied bezüglich Alter, Rauchstatus, Tumorart, Grading oder T-Stage zwischen Tumorgewebe und Nicht-Tumorgewebe gezeigt werden. Beim Geschlecht zeigte sich nur bei *MYC* ein statistisch signifikanter

Unterschied ($p=0,0217$), während bei den anderen Markern der Unterschied zwischen Männern und Frauen nicht signifikant war.

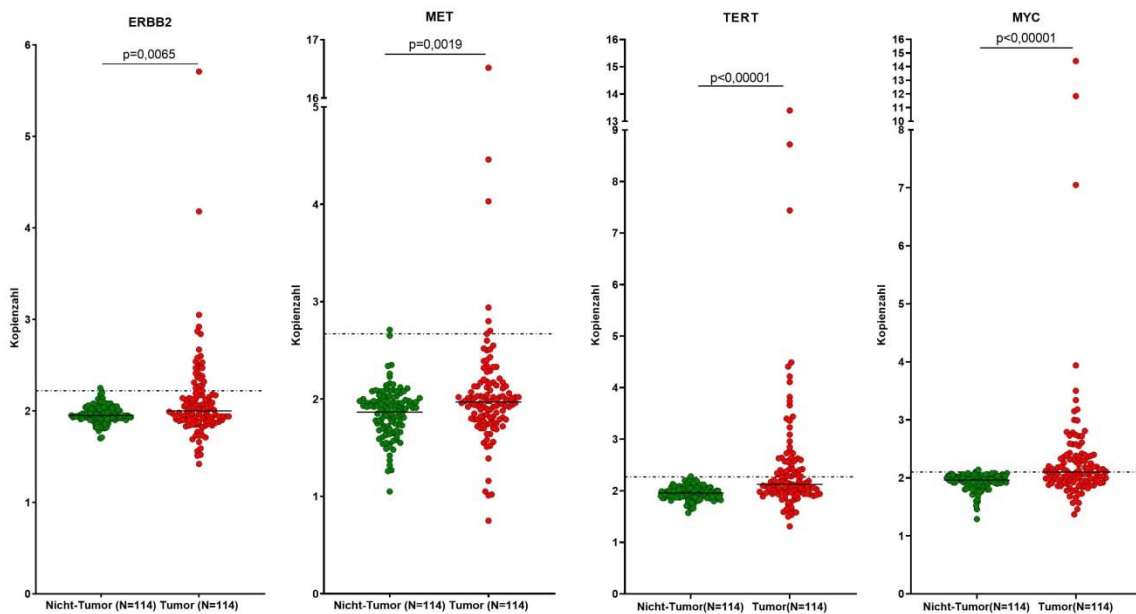


Abbildung 16: Verteilung der vier CNV-Assays in Tumor- (rot) und Nicht-Tumor- (grün) Gewebeproben von Lungenkrebserkrankten. Die gepunkteten Linien zeigen den jeweiligen Cutoff (Spezifität von mindestens 99%) und die horizontalen Linien den Median an.

Während *ERBB2* mit 24% und *MET* mit 6% niedrige Sensitivitäten aufwiesen, zeigten sich für *TERT* mit 41% und *MYC* mit 50% ausreichend hohe Sensitivitäten. Durch die ODER-Kombination von *MYC* und *TERT* kann die Sensitivität bei gleichbleibend hoher Spezifität (99%) von 50% auf 60% erhöht werden. Mit der angewendeten ODER-Kombination von *MYC* und *TERT* können so insgesamt 68 Lungenkrebserkrankte richtig positiv erkannt werden, 25 durch *MYC*, 16 durch *TERT* und 27 durch beide Marker (Abbildung 17). CNV-Ergebnisse für *MYC* wurden inzwischen publiziert (Brik et al. 2020, siehe Kapitel 5).

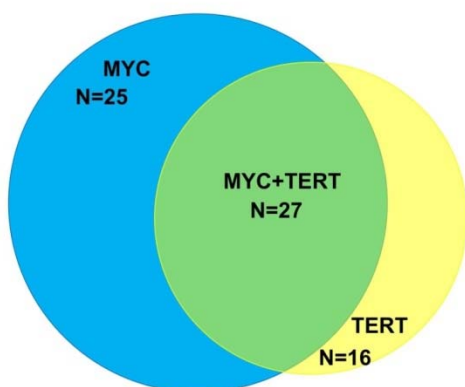


Abbildung 17: Venn-Diagramm zur Anzahl richtig positiver Tests von Lungenkrebserkrankten mittels *MYC* (blau) und *TERT* (gelb).

Kombination von Markern aus unterschiedlichen molekularen Ebenen

Basierend auf der Annahme, dass eine Kombination von einzelnen Markern aus unterschiedlichen molekularen Ebenen sowohl zu einer verbesserten Sensitivität als auch zu einer Reduzierung der notwendigen Einzelmarker innerhalb des Panels führen kann, wurden die geeigneten CNV- und DNA-Methylierungsmarker in Panels kombiniert und ihre Performance bezüglich der Diskriminierung von malignen Lungentumoren und Kontrollen (angrenzendes Nichttumorgewebe) analysiert.

Es zeigte sich, dass zur Diskriminierung der Gewebe von malignen Tumoren und Kontrollen die Kombination aus den zwei Methylierungsmarkern DLMS002 und DLMS030 sowie dem CNV-Marker *TERT* bei einer festgesetzten Spezifität von 100% eine Sensitivität von 95% erreicht und damit die identische Performance wie das Panel bestehend aus fünf Methylierungsmarkern (DLMS002, DLMS030, DLMS480, DLMS482 und DLMS003) besitzt. Bei der Zusammenstellung der Marker wurde die Performance im Gegensatz zu den bisher verwendeten ODER-Kombinationen innerhalb einer molekularen Ebene mittels der logistischen Regression ermittelt.

4.5 MPC

Proteomics

Für die Proteomicsarbeiten wurden sowohl Gewebe- als auch Plasmaproben verwendet. Da es uns gelungen ist, die Plasmaanalytik so zu optimieren, dass wir nicht nur einzelne Proteine wie ursprünglich geplant mit dem Multi-Reaction-Monitoring (MRM)-Verfahren, sondern mehr als 200 Proteine in einer Messung quantifizieren und identifizieren können, wurde das MRM-Verfahren nicht weiter verfolgt. Zusätzlich konnten wir in Zusammenarbeit mit der Bioinformatik die Auswertung auf die klinische Fragestellung anpassen, indem durch die Implementierung von Machine Learning Biomarkerpanel identifiziert wurden, die statistisch signifikant verschiedene Gruppen voneinander trennen können.

Gewebebasierte Studien

Im Rahmen der Gewebecharakterisierung mit anschließender Identifikation von differentiellen Proteinen wurden in Kooperation mit der Arbeitsgruppe Biophysik zwei verschiedene Fragestellungen bearbeitet. Zum einen wurde eine Studie am frischen Gewebe durchgeführt, um differentielle Proteine zu finden, die zwischen Adenokarzinomen und Kontrollen unterscheiden. Dabei wurden pro Patient jeweils Tumorbereiche und gesunde Bereiche gesammelt und mithilfe von labelfreier Proteomics eine differenzielle Studie durchgeführt (3.656 Proteine wurden identifiziert, 3.227 quantifiziert).

Wie erwartet, wurden zahlreiche Proteine identifiziert, die eine Unterscheidung zwischen Adenokarzinomen und Kontrollen ermöglichten (ca. 1.000 Proteine). Dabei wurden auch Proteine gefunden, die bereits im Zusammenhang mit Lungenkrebs bekannt sind. Zusätzlich wurden auch Proteine identifiziert, die bislang nicht als bekannte Biomarker für Adenokarzinome beschrieben wurden. Diese gefundenen differentiellen Proteine wurden mit einer Liste signifikanter Proteine aus der Plasmastudie abgeglichen. So konnten 6 Proteine bestätigt werden, die sowohl im Gewebe als auch im Plasma differentiell zwischen Adenokarzinomen und Kontrollen waren.

Zum anderen wurde eine FFPE-basierte Studie durchgeführt, bei der jeweils 10 Proben an Geweben mit verschiedenen Mutationen verglichen wurden (KRAS- vs. EGFR- vs. p53-Mutationen). Da FFPE-Gewebe eine Herausforderung in der Proteinanalytik darstellt, konnten nicht so viele Proteine identifiziert und quantifiziert werden (1.800 wurden identifiziert, 1.000 quantifiziert). Signifikante Proteine ohne FDR-Korrektur sind in den Volcano plots in Abbildung 18 dargestellt.

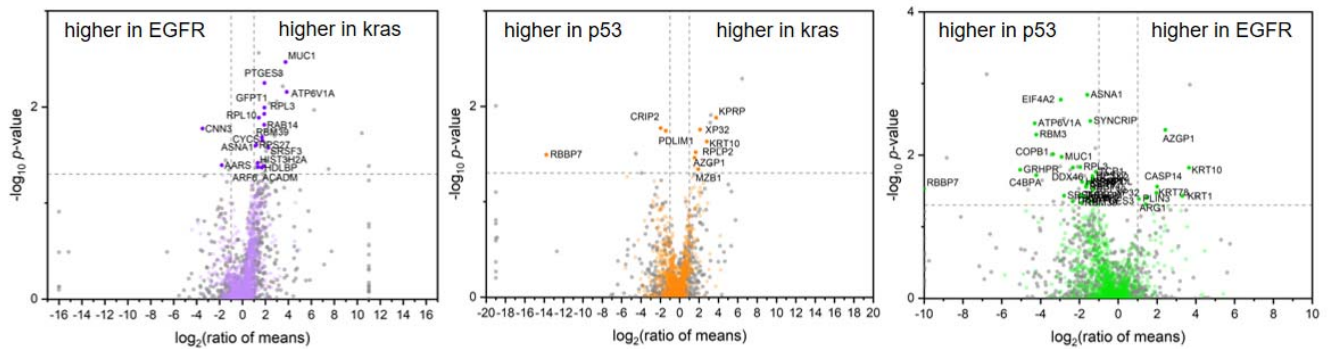


Abbildung 18: Volcano plots mit differentiellen Proteinen bei verschiedenen Mutationen.

Plasma-basierte Studien

Aufgrund der ersten Ergebnisse aus den Methylierungsanalysen (s. Epigenetik) ergab sich eine große Herausforderung, die Proteine zu identifizieren, die zwischen den Erkrankungen Adenokarzinom und COPD unterscheiden können. Deshalb wurden in den Proteomics-Analysen Plasmaproben von Patientinnen und Patienten mit Adenokarzinom mit oder ohne COPD, COPD-Kontrollen und von anderen Krankenhauskontrollen mittels Massenspektrometrie vermessen. Es wurden zwei unabhängige Studien durchgeführt, wobei im ersten Batch 63 und im zweiten Batch 133 Proben vermessen wurden. Zunächst wurden die zwei Batches einzeln mit der LOESS-Methode normalisiert und anschließend mithilfe eines linearen Modells der Batch-Effekt entfernt. Einige Proben wurden in beiden Batches gemessen und konnten als Kontrolle für eine erfolgreiche Entfernung des Batch-Effektes genutzt werden. Nach der Normalisierung beider Kohorten wurden insgesamt 487 Proteingruppen identifiziert und quantifiziert. Die Anwendung univariater statistischer Ansätze ermöglichte die Identifizierung signifikanter Proteine zwischen den einzelnen Gruppen. Für die klinisch-relevante Fragestellung zu Unterscheidung zwischen Adenokarzinomen und COPD wurde angestrebt ein Biomarkerpanel zu finden, welches eine möglichst hohe Spezifität und Sensitivität hat.

Die Auswertung der Ergebnisse in Bezug auf Spezifität und Sensitivität erfolgte in Zusammenarbeit mit der Bioinformatikgruppe.

MPC-Bioinformatik

Klassifikationsmodelle mit Proteomics-Messungen im Blutplasma (Arbeitspaket Proteomics)

Es wurden drei Vergleichsbetrachtungen (Klassifikationsmodelle) für die Unterscheidung zwischen 64 Adenokarzinom-Erkrankten von 76 COPD-Kontrollen ohne Adenokarzinom (Klassifikationsmodell 1), 21 Adenokarzinom-Erkrankten mit COPD von den 76 COPD-Kontrollen (Modell 2) und 64 Adenokarzinom-Erkrankten von 35 Kontrollen (Modell 3) entwickelt. Im Verlauf der Optimierung zeigte sich, dass der Random Forest von den üblichen Klassifikationsverfahren für alle drei Fragestellungen die besten Ergebnisse liefert. Dieser erzielte beim Gütekriterium AUROC höhere Werte als Support Vector Machine, künstliche Neuronale Netze, logistische Regression und LDA. Die drei finalen Modelle benötigen lediglich 10 Proteine. Modell 1 trennt 64 Adenokarzinom-Erkrankte von 76 COPD-Kontrollen ohne Adenokarzinom. In der Vorselektion wurden 46 Proteine für die Modellentwicklung ausgewählt. Die Vorhersagegenauigkeit des optimierten Modelles ist bei einer AUROC von 94% sowie einer Accuracy von 88%, Sensitivität von 89% und Spezifität von 87% sehr hoch.

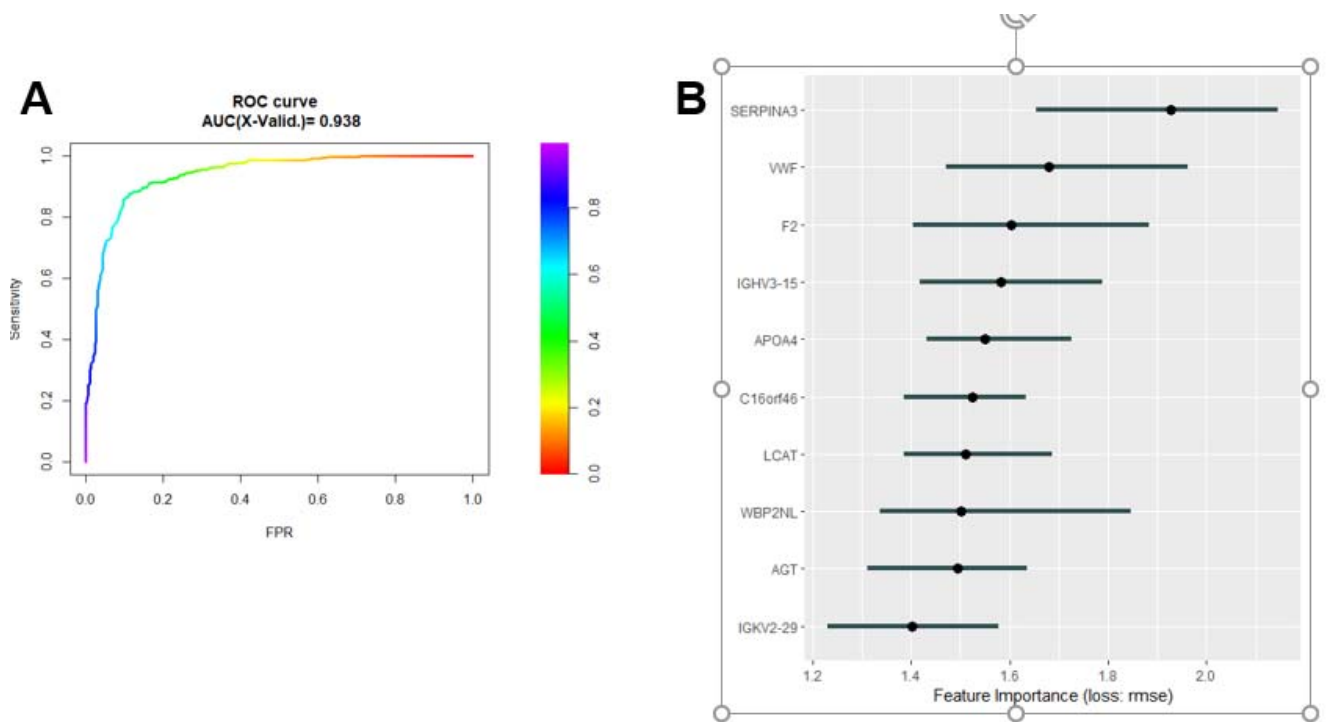


Abbildung 19: ROC-Kurve (A) und Permutation Feature Importance (B) des Klassifikationsmodelles für den Vergleich von allen Adenokarzinom-Erkrankten und COPD-Kontrollen ohne Adenokarzinom.

Modell 2 trennt 21 Adenokarzinom- Erkrankte mit COPD von den 76 COPD-Kontrollen. In der Vorselektion wurden 19 Proteine für die Modellentwicklung ausgewählt. Auch dieses Modell klassifiziert die Teilnehmenden sehr genau. Die AUROC liegt bei 92% und Accuracy und Spezifität sind mit Werten von 90% und 97% ebenfalls sehr hoch. Die Sensitivität ist mit 62% deutlich niedriger, aber dennoch in einem akzeptablen Bereich.

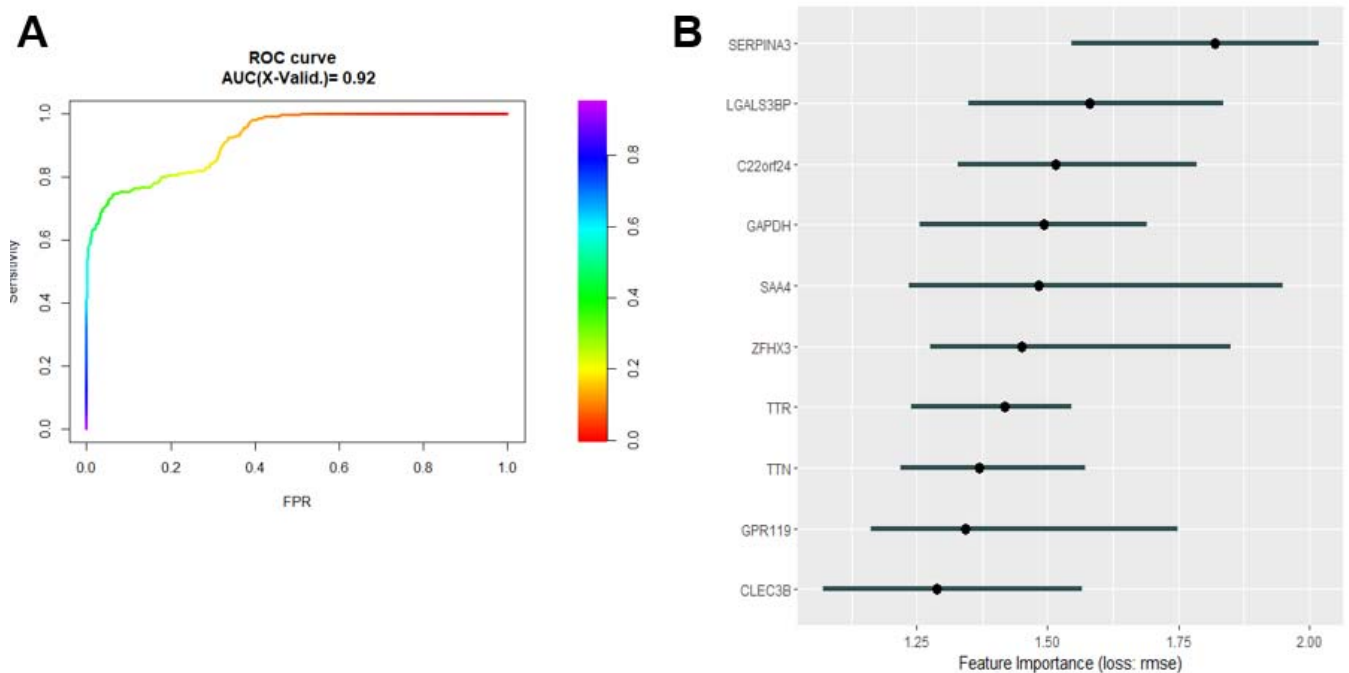


Abbildung 20: ROC-Kurven (A) und Permutation Feature Importance (B) des Klassifikationsmodelles für den Vergleich von Adenokarzinom-Erkrankten mit COPD und COPD-Kontrollen ohne Adenokarzinom.

Modell 3 trennt die 64 Adenokarzinom- Erkrankten von den 35 Kontrollen. In der Vorselektion wurden 119 Proteine für die Modellentwicklung ausgewählt. Die Vorhersagegenauigkeit des optimierten Modelles ist bei einer AUROC von 98% sowie einer Accuracy von 94%, Sensitivität von 97% und Spezifität von 89% ebenfalls sehr hoch.

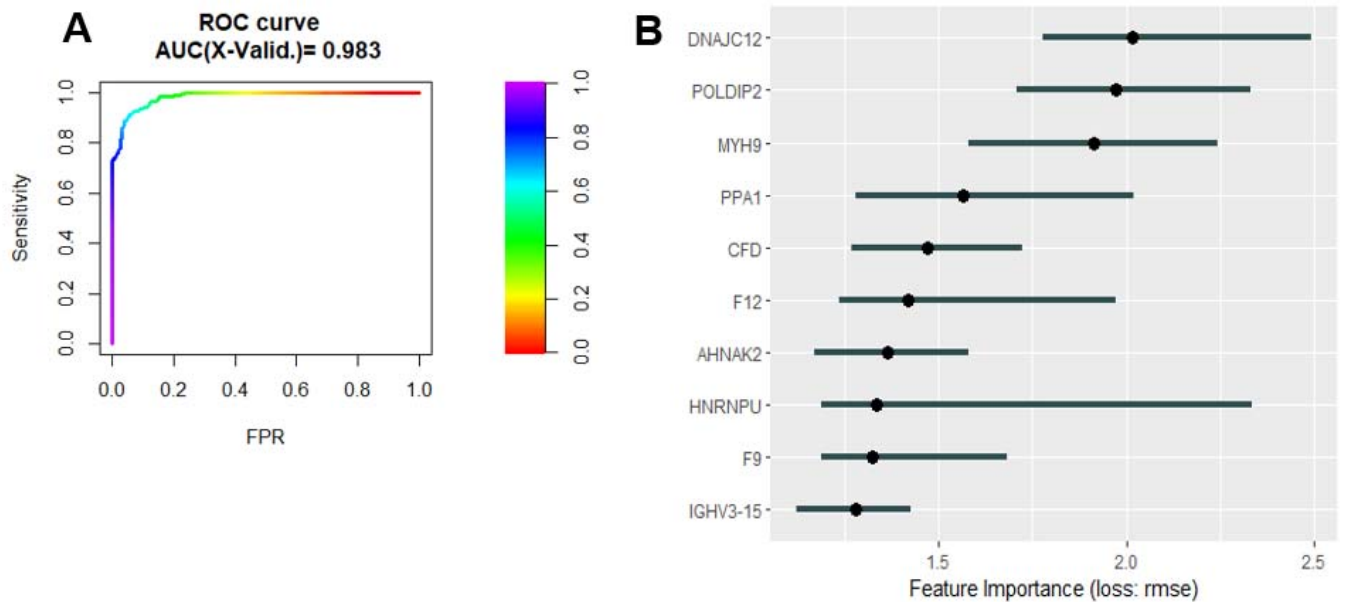


Abbildung 21: ROC-Kurve (A) und Permutation Feature Importance (B) des Klassifikationsmodelles für den Vergleich von allen Patientinnen und Patienten mit Adenokarzinom und Kontrollpersonen.

Klassifikator und Biomarker-Ensemble für Bisulfit DNA-Methylierungsmessungen (Arbeitspaket IPA):

Die vier verschiedenen Modelle zeigen, dass die Erhöhung der Spezifität einen gegenteiligen Effekt auf die Sensitivität hat, während die Gesamtgenauigkeit (Accuracy) relativ stabil bleibt. Tatsächlich zeigten die Gradient Boosted Trees (GBT) Modelle eher eine Tendenz zu einer hohen Spezifität, während bei den Random Forest (RF) Modellen Sensitivität und Spezifität gleichermaßen maximiert wurden. Bei Anwendung der Modelle im Testdatensatz konnte die hohe Spezifität des GBT Modells mit etwa 90% bestätigt werden, während die Sensitivität von 22% auf 12% sank. Bei den RF Modellen lag die Sensitivität im Testdatensatz etwas höher bei 29% bzw. 40%; dieses ging allerdings zu Lasten der Spezifität, die nur 68 bzw. 71% betrug.

Tabelle 9: Die Ergebnisse (getrennt nach Trainings- und Testdatensatz) der trainierten Klassifikatoren (GBT = Gradient Boosted Trees, RF = Random Forest) zur Unterscheidung von Lungenkrebs und der gemischten Gruppe aus Kontrollen und COPD auf Bisulfit-Methylierungs-Sequenzierungen.

Methode	Trainingsdatensatz			Testdatensatz		
	Sensitivität	Spezifität	Accuracy	Sensitivität	Spezifität	Accuracy
GBT (optimiert nach Accuracy)	0,227	0,894	0,560	0,119	0,911	0,571
GBT (optimiert nach Spezifität)	0,222	0,896	0,559	0,119	0,911	0,571
RF (optimiert nach Accuracy)	0,661	0,561	0,587	0,286	0,679	0,510
RF (optimiert nach Spezifität)	0,667	0,548	0,573	0,405	0,714	0,582

5. Auflistung der für das Vorhaben relevanten Veröffentlichungen, Schutzrechtsanmeldungen und erteilten Schutzrechte von nicht am Vorhaben beteiligten Forschungsstellen

Epigenetik:

Yanli Lin, Qixin Leng, Min Zhan, Feng Jiang (2018) A plasma long noncoding RNA signature for early detection of lung cancer. *Transl Oncol* 11 (5): 1225 – 1231

Shuai Yuan, Ying Xiang, Xiaoping Guo, Yao Zhang, Chengying Li, Weijia Xie, Na Wu, Long Wu, Tongjian Cai, Xiangyu Ma, Zubin Yu, Li Bai, Yafei Li (2020) Circulating long noncoding RBNA s act as Diagnostic biomarkers in non-small cell lung cancer. *Front Oncol* 10: 537120

Die präsentierten Artikel zeigen, dass der gewählte Ansatz Lungenkrebs mittels lncRNAs zu detektieren prinzipiell geeignet ist. Unterschiede in den Ergebnissen sind dabei auf unterschiedlichen eingesetzten Methoden zurückzuführen.

Biophysik:

Nicolas Coudray, Paolo Santiago Ocampo, Theodore Sakellaropoulos, Navneet Narula, Matija Snuderl, David Feny, Andre L Moreira, Narges Razavian, Aristotelis Tsirigos (2018) Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature medicine* 24 (10): 1559 – 1567

In Bezug auf PURE Lunge 2 zeigt diese Veröffentlichung, dass Imaging-Verfahren in Kombination mit maschinellem Lernen ein großes Potential für eine bessere und schnellere Bewertung von Tumorgewebe haben. Da dieser Ansatz auf Basis von klassischen histologischen Bildern in direkter Konkurrenz zu unserer Methodik steht, möchten wir

diese Arbeit nicht unerwähnt lassen. Im Gegensatz zu den in dieser Veröffentlichung beschriebenen Ansätzen mit tiefen neuronalen Netzen, erreicht die Marker-freie spektrale Gewebeanalyse aus PURE Lunge 2 mit klassischen maschinellen Lernen eine höhere Erkennungsgenauigkeit und bessere Auflösung der räumlichen Heterogenität der betrachteten Mutationen. Der von uns verfolgte Ansatz soll nun auf Basis der gewonnenen Daten analog zu dieser Veröffentlichung auf tiefe neuronale Netze übertragen werden. Die notwendigen Algorithmen wurden in den letzten Jahren durch unsere Bioinformatik entwickelt. Hierdurch versprechen wir uns eine weitere Verbesserung in der Erkennungsgenauigkeit, die die an histopathologischen Bildern durchgeführten Arbeiten noch deutlicher übertreffen wird.

6. Bewertung der Ergebnisse hinsichtlich des Forschungszwecks/-ziels, Schlussfolgerungen

Eine Vielzahl der Marker Kandidaten aus dem Vorgängerprojekt konnte in dem neuen Kollektiv von Patientinnen und Patienten bestätigt werden. Insbesondere in den gewebebasierten Analysen konnten die Ergebnisse zur Unterscheidung von Tumor- und Nichttumorproben repliziert sowie zusätzlich Tumorsubtypen und Mutationen differenziert werden. Für die Verifizierung von Markern in Körperflüssigkeiten (Blutplasma, Speichel) zeigten sich unterschiedliche Ergebnisse: Die DNA-Analyse konnte die Marker aus dem Gewebe nur unzureichend im Plasma bestätigen, trotz Anwendung deutlich sensitiverer Methoden im Verlauf der Studie. In den verschiedenen epigenetischen Analysen bewährten sich zwar einige Marker für die Unterscheidung von Patientinnen und Patienten mit und ohne eine Tumorerkrankung, jedoch konnte die überwiegende Zahl der Marker nicht ausreichend zwischen Patienten mit Krebs und solchen ohne Krebs aber mit benignen Lungenerkrankungen (insbesondere Personen mit COPD) unterscheiden. Bei der Proteomanalyse wurden neue Marker Kandidaten identifiziert, die auch eine Abgrenzung zur COPD möglich erscheinen lassen (z. B. SerpinA3, VWF, F2, IGHV3-15). Diese Marker Kandidaten müssen jedoch noch verifiziert werden.

Gewebebasierter Marker können in der Regel nur im Rahmen von therapeutischen oder diagnostischen Verfahren gewonnen werden, wenn das Risiko (des Fortschreitens) einer Erkrankung die invasiven Maßnahmen zur Gewebeentfernung und damit verbundene Risiken rechtfertigt. Im Gegensatz dazu sind die Marker in nicht oder minimal invasiv gewonnenen Körperflüssigkeiten wie Blut und Speichel prinzipiell für eine Früherkennung von Lungentumoren, bei der ein (Groß-)Teil der gescreenten Personen nicht erkranken wird, geeignet. Aufgrund der heterogenen Ergebnisse der plasma- und speichelbasierten Marker sowie der zusätzlichen Herausforderung der Abgrenzung von Krebs und gutartigen Lungenerkrankungen (insbesondere COPD) kann noch nicht abschließend bewertet werden, ob sich im Rahmen dieses Projektes Marker für eine anschließende Validierung mit seriellen, prädiagnostischen Proben von Lungenkrebskranken aus einer prospektiven Kohorte von Risikopersonen herausstellen lassen. Erst in einer solchen Langzeitstudie könnte die Eignung der Marker für die Früherkennung von Lungenkrebs letztlich beurteilt werden.

Zuvor könnten die folgenden, anknüpfenden Analysen zielführend sein:

- Die Auswertung von unterschiedlichen Marker-Kombinationen ist noch nicht abgeschlossen. Sowohl Marker-Kombinationen innerhalb der einzelnen molekularen Ebenen als auch Ebenen-übergreifend sollen weiterhin auch arbeitsgruppenübergreifend untersucht werden. Die Verknüpfung von CNV mit DNA-Methylierung und von miRNAs mit Proteinen haben gezeigt, dass Markerkombinationen aus unterschiedlichen molekularen Ebenen zur Verbesserung eines Panels beitragen können. Diese Analysen werden im Rahmen des Konsortiums weiterverfolgt.
- Die Proteinmarker, die bisher eine gute Unterscheidung von Tumorerkrankten (mit Adenokarzinom) und anderen Lungenerkrankungen und im Speziellen auch COPD-Kontrollen ermöglichten, sollten mit den verbleibenden Proben aus PURE Lunge 1 und 2 verifiziert werden.

Die verwendeten Methoden der Bioinformatik, vor allem die Methoden des *machine learning* und der Künstlichen Intelligenz, sind noch nicht vollständig ausgereizt. So sind zum einen weitere *feature importance*-Maße berechenbar, um die informationstragendsten und damit funktionell interessantesten Proteine zu bestimmen. Zum anderen sind die ausgelernten Klassifikatoren nahezu direkt (bis auf Normalisierungsschritte quantitativer Daten) auf sich neu vorstellende Patientinnen und Patienten anwendbar.

- Bereits im Verlauf der Studie wurde eine Vielzahl von Methoden optimiert, um eine bessere und schnellere Detektion der Marker zu ermöglichen. Im Rahmen der DNA-Analyse konnten Mutationen im Gewebe nicht wie erwartet in der ctDNA aus Plasma entdeckt werden. Dabei ist noch nicht klar, ob einfach nicht genügend ctDNA oder ctDNA-Fragmente im Plasma vorhanden waren, diese nicht ausreichend detektiert wurden oder die Verdünnung durch zirkulierende Normal-DNA zu groß war. Die Anwendung weiter verbesserter Detektionsmethoden könnte zur Klärung dieser Frage hilfreich sein. Zuvor sollten jedoch noch alle restlichen Proben mit der bereits eingesetzten, sensitiveren Methode der ddPCR analysiert werden.
- Für die DNA-Methylierung und die CNV-Analytik ist ebenfalls anzunehmen, dass eine weitere Optimierung der Methoden (Isolation, Konvertierung, Detektion) notwendig ist, um kleine Mengen an ctDNA vor dem Hintergrund verdünnender Normal-DNA besser im Plasma detektieren zu können. Zudem werden ausreichend spezifische Marker für die Unterscheidung von COPD-Fällen benötigt. Die Ergebnisse mit den kleinzelligen Lungentumoren haben gezeigt, dass dies grundsätzlich möglich ist.
- Einen weiteren möglichen Forschungsansatz stellt ein Follow-up aller Teilnehmenden hinsichtlich ihrer Prognose und des Krankheitsverlaufs dar, um prognostische Marker zu identifizieren, die den weiteren Krankheitsverlauf der Patientinnen und Patienten vorhersagen können. Zusätzlich könnte im Rahmen eines solchen Follow-ups das Risiko der COPD-Kontrollen, einen Lungenkrebs zu entwickeln, bestimmt werden. Von den 90 COPD-Kontrollen kann bei ca. 15% ein Tumor innerhalb von 2-3 Jahren erwartet werden (ca. n=14). Von diesen Teilnehmenden würde jedoch nur ein Teil das entsprechende Markerspektrum aufweisen, so dass die statistische Aussagekraft für diese Fragestellung eingeschränkt ist. Es ist zudem fraglich, ob dieses die fehlende Differenzierung zwischen Lungenkrebs und COPD erklären kann.
- Abgesehen von der Früherkennung könnte eine Entwicklung der biospektroskopischen Gewebeanalyse zur praktischen Anwendung im Rahmen von Bronchoskopien die Diagnostik unterstützen. Allerdings muss erst noch gezeigt werden, dass die biospektroskopische Methodik alleine zur Krebsdiagnose anhand von EBUS einen Vorteil gegenüber der relativ voraussetzungslosen und schnellen Zytologie bringt. Der Vorteil der Biospektroskopie besteht aus pathologischer Sicht insbesondere in der Gleichzeitigkeit von Analysen, z.B. von Tumorsubtypen und Mutationen, auch an kleinsten Gewebeproben, ohne diese zu verbrauchen.
- Weiterhin sollen, basierend auf dem vollständigen Kollektiv unter Berücksichtigung der vorliegenden Daten der Teilnehmenden und des Follow-up, Proben zur Label-freien Gewebesammlung für die Proteomik selektiert werden, um für die Früherkennung und Diagnostik interessante klinische Fragestellungen zu adressieren. Dies ist erst jetzt nach erfolgreicher Etablierung und Verifizierung des Workflows möglich und sinnvoll.
- Die Follow-up Daten sollen in der biospektroskopischen Gewebeanalyse auch in die bereits gestartete Etablierung von Deep Learning Modellen einfließen, um eine Prognostik und eventuelle Prädiktion des Therapieansprechens zu ermöglichen.

7. Aktueller Umsetzungs- und Verwertungsplan

Schritte zur Umsetzungs- und Verwertung sind zurzeit noch nicht vorgesehen.

Folgende Publikationen wurden veröffentlicht:

Brik A., Weber DG., Casjens S., Rozynek P., Meier S., Behrens T., Stamatis G., Darwiche K., Theegarten D., Brüning T., Johnen G. (2020). Digital PCR for the Analysis of *MYC* Copy Number Variation in Lung Cancer. *Dis Markers*. 2020:4176376. doi: 10.1155/2020/4176376.

Goertzen N., Pappesch R., Fassunke J., Brüning Th., Ko Y.-D., Schmidt J., Großerueschkamp F., Buettner R., Gerwert K. (2021). Quantum cascade laser-based infrared imaging as a label-free and automated approach to determine mutations in lung adenocarcinoma. *Am J Pathol*. 2021; 191(7):1269-1280, DOI: 10.1016/j.ajpath.2021.04.013.

Weitere Publikationen sind nach den abschließenden Auswertungen geplant.

8. Anhang/Anhänge

Erklärung für das Berichtswesen in Kooperationsprojekten

Zeitpläne (geplant und durchgeführt)

Unterschriftenseite verpflichtend für Kooperationsprojekte

Projektnummer: FB-0259

Titel: Verifizierung von neuen molekularen Markern für die Frühdiagnostik von Lungentumoren zum Nutzen der nachgehenden Vorsorge

Erklärung für das Berichtswesen in Kooperationsprojekten

Hiermit erklären die Unterzeichnenden, dass der zum 27.07.2021 vorgelegte

Abschlussbericht mit allen Kooperationspartnern abgestimmt ist.



----- Datum: 02.08.2021
Name, Institution PROF. DR. VON-OSCHUN KO
JOHANNITER KLINIKEN BONN

Johanniter GmbH
Johanniter-Krankenhaus
Chefarzt Innere Abteilung
Prof. Dr. med. Y. Ko
Johanniterstraße 3-5 • 53113 Bonn

----- Datum:
Name, Institution

----- Datum:
Name, Institution

----- Datum:
Name, Institution

----- Datum:
Name, Institution

----- Datum:
Name, Institution

Unterschriftenseite verpflichtend für Kooperationsprojekte


Projektnummer: FB-0259

Titel: Verifizierung von neuen molekularen Markern für die Frühdiagnostik von Lungentumoren zum Nutzen der nachgehenden Vorsorge

Erklärung für das Berichtswesen in Kooperationsprojekten

Hiermit erklären die Unterzeichnenden, dass der zum 27.07.2021 vorgelegte

Abschlussbericht mit allen Kooperationspartnern abgestimmt ist.


----- Datum: **02. Aug. 2021**
Name, Institution

**INSTITUT FÜR PATHOLOGIE
UNIKLINIK KÖLN**
Direktor: Universitätsprofessor Dr. R. Büttner
Hausanschrift: Kerpener Str. 62
Postanschrift: 50924 Köln (Lindenthal)


----- Datum: **02. Aug. 2021**
Name, Institution

Molekularpathologische Diagnostik
Institut für Pathologie
Uniklinik Köln
Direktor: Universitätsprofessor Dr. R. Büttner
Kerpener Sr. 62, 50924 Köln

----- Datum:
Name, Institution

----- Datum:
Name, Institution

----- Datum:
Name, Institution

----- Datum:
Name, Institution

Unterschriftenseite verpflichtend für Kooperationsprojekte


Projektnummer: FB-0259

Titel: Verifizierung von neuen molekularen Markern für die Frühdiagnostik von Lungentumoren zum Nutzen der nachgehenden Vorsorge

Erklärung für das Berichtswesen in Kooperationsprojekten

Hiermit erklären die Unterzeichnenden, dass der zum 27.07.2021 vorgelegte

Abschlussbericht mit allen Kooperationspartnern abgestimmt ist.

Prof. Sitek, MPC Datum: 27.7.21
Name, Institution 

----- Datum:
Name, Institution

----- Datum:
Name, Institution

----- Datum:
Name, Institution

----- Datum:
Name, Institution

----- Datum:
Name, Institution

Unterschriftenseite verpflichtend für Kooperationsprojekte

Projektnummer: FB-0259

Titel: Verifizierung von neuen molekularen Markern für die Frühdiagnostik von Lungentumoren zum Nutzen der nachgehenden Vorsorge

Erklärung für das Berichtswesen in Kooperationsprojekten

Hiermit erklären die Unterzeichnenden, dass der zum 27.07.2021 vorgelegte

Abschlussbericht mit allen Kooperationspartnern abgestimmt ist.



Datum: 28.07.2021

Name, Institution

Prof. Dr. Klaus Gerwert
Zentrum für Proteindiagnostik
Geschäftsführender Gründungsdirektor

----- Datum:

Name, Institution

----- Datum:

Name, Institution

----- Datum:

Name, Institution

----- Datum:

Name, Institution

----- Datum:

Name, Institution

Unterschriftenseite verpflichtend für Kooperationsprojekte

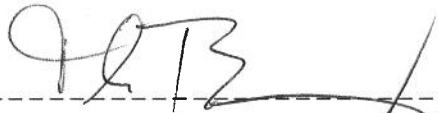
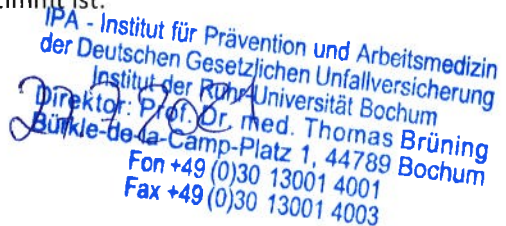
Projektnummer: FB-0259

Titel: Verifizierung von neuen molekularen Markern für die Frühdiagnostik von Lungentumoren zum Nutzen der nachgehenden Vorsorge



Erklärung für das Berichtswesen in Kooperationsprojekten

Hiermit erklären die Unterzeichnenden, dass der zum 27.07.2021 vorgelegte


Abschlussbericht mit allen Kooperationspartnern abgestimmt ist.


----- Datum: 

Name, Institution Prof. Th. Brüning, IPA


----- Datum: 28.7.21 

Name, Institution Dr. Georg Johnen, IPA


----- Datum: 20.8.21

Name, Institution Prof. Th. Behrens, PURE, RUB

----- Datum:
Name, Institution

----- Datum:
Name, Institution

----- Datum:
Name, Institution

Zeitplan PURE Lunge 2 - Durchgeführte Arbeitspakete

Arbeitspakete (AP)	1. HJ 2017		2. HJ 2017				1. HJ 2018				2. HJ 2018				Erste kostenneutrale Verlängerung				Zweite kostenneutrale Verlängerung				Corona-bed. Verlängerung																													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
AP0: Vorbereitende Organisationsaufgaben																																																				
Ethikanträge (Bochum, Köln, Ärztekammer Nordrhein)																																																				
Datenschutzkonzept																																																				
AP1: Etablierung Abläufe, Optimierung Methoden, erste Proben u. Bestimmungen																																																				
Probanden rekrutieren (n=97 (geplant n=75)) / Organisation Probentransport																																																				
Blut-, Plasma-, Speichelproben gewinnen																																																				
Gewebeverarbeitung: Kryoasservierung, Formalinfixierung, Anfertigung der Schnittsets																																																				
Pathologische Begutachtung des Gewebes																																																				
Aufreinigung DNA aus Gewebeproben und der zirkulierenden DNA aus Plasmaproben																																																				
Charakterisierung genom. DNA / ctDNA mittels Parallelsequenzierung																																																				
SOPs optimieren und Eingang erster Proben																																																				
Anpassung der SOPs und Workflow für QCL-IR-Imaging																																																				
Bearbeitung erster Proben für SHP (Vorbereitung für die Analyse)																																																				
Optimierung statist. Studiendesign (Power-Analyse, Fallzahlabschätzung etc.)																																																				
Etablierung von Plasmaanalytik für Proteomics																																																				
Methylierung: DNA-Isolation, Vorauswahl 5 aus 50 Einzelassays PyroSeq Gewebe																																																				
RNA-Analyse: Nachweis pot. Targets in Plasma u. Speichel																																																				
CNV-Analyse: Verifizierung in Tumor- u. Nichttumorgewebe																																																				
AP2: Teil 1 der Hauptphase, Vorauswahl Marker, Zwischenauswertung																																																				
Probanden rekrutieren (n=141 (geplant n=175)) / Probentransport																																																				
Blut-, Plasma-, Speichelproben gewinnen																																																				
Gewebeverarbeitung: Kryoasservierung, Formalinfixierung, Anfertigung der Schnittsets																																																				
Pathologische Begutachtung des Gewebes																																																				
Aufreinigung DNA aus Gewebeproben und der zirkulierenden DNA aus Plasmaproben																																																				
Charakterisierung genom. DNA / ctDNA mittels Parallelsequenzierung																																																				
spektrale Vermessung der ersten Proben (FTIR, QCL-IR-Imaging)																																																				
Erweiterung der spektralen Datenbank																																																				
Lasermikrodissektion für Proteomik und Genomik (QCL-IR-Imaging)																																																				
Massenspektrometrische Analyse von Tumorproben nach Lasermikrodissektion																																																				
Abschluß Power-Analyse und Fallzahl-Abschätzung																																																				
Ausbau der Probenverwaltung																																																				
Methylierung: Verifizierung d. 5 besten Assays aus AP1 (Plasma, Speichel)																																																				
Methylierung: Entwicklung v. Deep-Sequencing Assays für Plasma-Proben																																																				
RNA-Analyse: Verifizierung der geeignetsten Assays aus AP1 (Plasma, Speichel)																																																				
CNV-Analyse: Verifizierung d. geeignetsten Assays aus AP1 (Tumor- u. Nichttumorgewebe)																																																				
Zwischenauswertung u. Vergleich der Daten aller Technologieplattformen																																																				

