

Epigenetische Alterationen des Lungenkarzinoms:
Diagnostische Nutzung und Induktion durch
E-Zigarettenkonsum

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Swetlana Scheufele

Kiel, 2018

Erste Gutachterin:

Prof. Dr. rer. nat. Manuela Dittmar

Zweiter Gutachter:

Prof. Dr. rer. nat. Ole Ammerpohl

Tag der mündlichen Prüfung:

17.04.2018

Inhaltsverzeichnis

| | |
|---|----|
| Abkürzungen | 6 |
| 1. Einleitung | 8 |
| 1.1 Rolle der Epigenetik bei der Entstehung multifaktorieller Erkrankungen..... | 8 |
| 1.1.1 DNA-Methylierung | 9 |
| 1.1.2 RNA-basierte Mechanismen | 11 |
| 1.1.3 Modifikation von Histonen..... | 11 |
| 1.1.4 Organisation des Zellkerns | 13 |
| 1.2 Lungenkrebs | 15 |
| 1.2.1 Molekularbiologie und Epigenetik des Lungenkarzinoms..... | 16 |
| 1.2.2 Diagnostik des Lungenkarzinoms, molekulare Möglichkeiten der Lungenkrebsfrüherkennung..... | 17 |
| 1.2.3 Rolle des Zigarettenkonsums bei der Krankheitsentstehung | 18 |
| 1.3 Zielsetzung | 21 |
| 2. Material | 22 |
| 2.1 Probenmaterial | 22 |
| 2.1.1 OP-Präparate | 22 |
| 2.1.2 Plasma-Proben..... | 23 |
| 2.1.3 Zellen | 24 |
| 2.2 Reagenzien..... | 25 |
| 2.3 Kits..... | 26 |
| 2.4 Chemikalien | 27 |
| 2.5 <i>Beads</i> | 27 |
| 2.6 Verbrauchsmaterialien | 28 |
| 2.7 Instrumente und Geräte | 28 |
| 2.8 Software | 30 |
| 3. Methoden | 31 |
| 3. 1 Methoden im Projekt „Entwicklung des Plasma <i>Panels</i> “ | 31 |
| 3.1.1 Nachweis der aberranten DNA-Methylierung in primärem Tumorgewebe..... | 33 |
| 3.1.2 Nachweis der aberranten DNA-Methylierung in Blutplasma..... | 33 |
| 3.1.2.1 Extraktion der zirkulierenden, zellfreien DNA (zfdNA) aus Blutplasma | 34 |
| 3.1.2.2 Quantifizierung und Qualitätskontrolle (QC) der extrahierten zfdNA | 35 |
| 3.1.2.3 Bisulfitkonvertierung der zfdNA | 36 |
| 3.1.2.4 <i>Whole Genome Bisulfite Sequencing</i> (WGBS)..... | 37 |
| 3.1.2.4.1 Erstellen der WGBS <i>Libraries</i> | 39 |
| 3.1.2.4.2 Sequenzierung der WGBS <i>Libraries</i> | 42 |

| | |
|--|----|
| 3.1.2.5 Bioinformatische Auswertung der WGBS Ergebnisse | 42 |
| 3.1.2.5.1 Qualitätskontrolle der Rohdaten..... | 45 |
| 3.1.2.5.2 Datenprozessierung..... | 45 |
| 3.1.2.5.3 Auswertung der prozessierten Daten | 46 |
| 3.1.2.5.4 Korrelationsanalysen | 48 |
| 3.1.2.5.5 Auswählen der CpG <i>Loci</i> für das Plasma <i>Panel</i> | 48 |
| 3.1.3 Weitere Komponenten des Plasma <i>Panels</i> (<i>in silico</i> Datenanalysen) | 49 |
| 3.1.3.1 Die prognostische Studie..... | 49 |
| 3.1.3.2 Die bivalente Chromatin Studie | 50 |
| 3.2 Methoden im Projekt „Validierung des Plasma <i>Panels</i> “ | 50 |
| 3.2.1 Anreicherung von differentiell methylierten Regionen | 51 |
| 3.2.2 Sequenzierung des Plasma <i>Panels</i> | 54 |
| 3.2.3 Bioinformatische Auswertung der Sequenzierdaten..... | 55 |
| 3.2.3.1 Qualitätskontrolle und Prozessieren der Rohdaten | 55 |
| 3.2.3.2 Auswertung der prozessierten Daten..... | 55 |
| 3.2.3.3 Erstellen eines Klassifikators | 56 |
| 3.3 Methoden im Projekt „Multi-OMICS Studie zur Untersuchung von Risiken der E-Zigaretten“ | 58 |
| 3.3.1 Zellkultur..... | 58 |
| 3.3.1.1 A549..... | 58 |
| 3.3.1.2 AECII..... | 59 |
| 3.3.2 Herstellung der Zigaretten- (CSE) und E-Zigarettenextrakte (eCSE) | 60 |
| 3.3.3 Stimulation der Lungenzelllinien mit CSE und eCSE | 62 |
| 3.3.4 Messung der Apoptose..... | 62 |
| 3.3.4.1 „ApoONE® Homogeneous Caspase-3/7 Assay“ | 62 |
| 3.3.4.2 „IncuCyte ZOOM Live Cell Analysis System“ | 63 |
| 3.3.5 Bestimmung der Viabilität | 63 |
| 3.3.6 RNA-Sequenzierung (RNA-Seq.) | 64 |
| 3.3.6.1 RNA-Extraktion..... | 66 |
| 3.3.6.2 Herstellung der RNA-Seq. <i>Libraries</i> , Sequenzierung | 66 |
| 3.3.6.3 Bioinformatische Auswertung der RNA-Seq. Daten | 66 |
| 3.3.6.3.1 Qualitätskontrolle und Prozessieren der Daten | 67 |
| 3.3.6.3.2 Auswertung der prozessierten Daten | 69 |
| 3.3.7 <i>Assay for Transposase-Accessible Chromatin Using Sequencing</i> (ATAC-Seq.)...70 | |
| 3.3.7.1. Herstellung der ATAC-Seq. <i>Libraries</i> | 71 |
| 3.3.7.2 Sequenzierung der ATAC <i>Libraries</i> | 72 |

| | |
|--|-----------|
| 3.3.7.3 Bioinformatische Auswertung der Sequenzierdaten..... | 72 |
| 3.3.7.3.1 Qualitätskontrolle und Prozessieren der Daten | 73 |
| 3.3.7.3.2 Auswertung der prozessierten Daten | 73 |
| 3.3.8 „Infinium MethylationEPIC BeadChip“(EPIC BC) | 76 |
| 3.3.8.1 DNA-Extraktion..... | 77 |
| 3.3.8.2 Bisulfitkonvertierung | 79 |
| 3.3.8.3 Hybridisierung des EPIC BC..... | 79 |
| 3.3.8.4 Bioinformatische Auswertung der EPIC BC Daten | 81 |
| 3.3.9 Fluoreszenzmikroskopie | 81 |
| 3.3.10 Durchflusszytometrie | 82 |
| 4. Ergebnisse | 83 |
| 4.1 Ergebnisse des Projekts „Entwicklung des Plasma <i>Panels</i> “ | 83 |
| 4.1.1 Detektion der tumor- und entitätsspezifischen DNA-Methylierung in primärem Tumorgewebe | 84 |
| 4.1.2 Detektion der tumor- und entitätsspezifischen DNA-Methylierung in Blutplasma .. | 85 |
| 4.1.3 Korrelationsanalysen der verwendeten Methoden zur genomweiten Detektion von DNA-Methylierungsmustern..... | 88 |
| 4.2 Ergebnisse des Projekts „Validierung des Plasma <i>Panels</i> “ | 93 |
| 4.2.1 Erstellen der NGS <i>Libraries</i> | 93 |
| 4.2.2 Sequenzierung und Datenanalyse..... | 94 |
| 4.2.3 Erstellen eines Klassifikators | 101 |
| 4.3 Ergebnisse des Projekts „multi-OMICS Studie zur Untersuchung von Risiken der E-Zigaretten“ | 103 |
| 4.3.1 Funktionelle und morphologische Befunde | 103 |
| 4.3.2 Molekularbiologische und epigenetische Befunde | 108 |
| 4.3.2.1 Auswertung von RNA-Seq. Daten..... | 108 |
| 4.3.2.2 Auswertung von ATAC-Seq. Daten..... | 110 |
| 4.3.2.3 Auswertung von EPIC BC Daten | 110 |
| 4.3.2.4 Die integrative Analyse von RNA-Seq., ATAC-Seq. und EPIC BC Daten | 113 |
| 4.3.2.4.1 Induktion der Apoptose nach der Behandlung mit nikotinhaltigen CSE und eCSE..... | 113 |
| 4.3.2.4.2 Erhöhte Zellproliferation und Aneuploidie nach Behandlung der A549 Tumorzelllinie mit dem nikotinfreiem eCSE | 114 |
| 4.3.2.4.3 Weitere Ergebnisse der integrativen Datenanalyse..... | 115 |

| | |
|---|-----|
| 5. Diskussion | 118 |
| 5.1 Entwicklung und Validierung des Plasma <i>Panels</i> | 118 |
| 5.2 multi-OMICS Studie zur Untersuchung von Risiken der E-Zigaretten..... | 125 |
| 6. Zusammenfassung | 135 |
| 7. Summary | 137 |
| 8. Literaturverzeichnis | 139 |
| 9. Anhang | 167 |
| 9.1 Entwicklung und Validierung des Plasma <i>Panels</i> | 167 |
| 9.2 multi-OMICS Studie zur Untersuchung von Risiken der E-Zigaretten..... | 168 |
| 9.3 Tabellen..... | 171 |
| 9.4 Ethikvotum | 200 |
| 10. Danksagung | 207 |
| 11. Publikationsliste | 208 |
| 11.1 Publizierte Originalarbeiten | 208 |
| 11.2 Eingereichte Originalarbeiten | 208 |
| 11.3 Poster | 209 |
| 12. Eidesstattliche Erklärung | 210 |

Abkürzungen

| | |
|---------------|--|
| 5-caC | 5-Carboxylcytosin |
| 5-fC | 5-Formylcytosin |
| 5-hmC | 5-Hydroxymethylcytosin |
| 5-mC | 5-Methylcytosin |
| ACRK | ATP-abhängige Chromatin- <i>Remodeling</i> -Komplexe |
| AECII | Alveolarepithelzellen Typ II (engl. <i>alveolar epithelial type II cells</i>) |
| A.K. | Adenokarzinom |
| A.S.K. | adenosquamöses Karzinom |
| ATAC-Seq. | engl. <i>assay for transposase-accessible chromatin using sequencing</i> |
| BER | Basenexzisionsreparatur |
| bp | Basenpaar |
| cDNA | komplementäre DNA (engl. <i>complementary DNA</i>) |
| CpG | Cytosin-(Phosphat-)Guanin-Dinukleotid |
| CSE | Zigarettenextrakt (engl. <i>cigarette smoking extract</i>) |
| CT | Computertomographie |
| DNMT | DNA-Methyltransferase |
| dNTP | Desoxyribonukleosidtriphosphat |
| dsDNA | doppelsträngige DNA (engl. <i>double stranded DNA</i>) |
| DZL | Das Deutsche Zentrum für Lungenforschung |
| EPIC BC | „Infinium MethylationEPIC BeadChip“ |
| eCSE | E-Zigarettenextrakt (engl. <i>e-cigarette smoking extract</i>) |
| FACS | Durchflusszytometrie (engl. <i>fluorescence activated cell sorting</i>) |
| gDNA | genomische DNA |
| G.K. | großzelliges Karzinom |
| GO | Genontologie |
| HM 450K | „Illumina Infinium Human Methylation450K BeadChip“ |
| HME | Histon-modifizierende Enzyme |
| I | Intensität |
| K. | Kontrolle |
| kb | Kilo-Basenpaar (engl. <i>kilo base pair</i>) |
| K.K. | Karzinoid |
| kNN | engl. <i>k-nearest neighbors</i> |
| LAD | engl. <i>lamina associated domain</i> |
| M <i>Bead</i> | <i>Bead</i> für das methylierte Allel |
| miRNA | mikro-RNA (engl. <i>micro RNA</i>) |

| | |
|---------------|--|
| mRNA | Boten-RNA (engl. <i>messenger RNA</i>) |
| MRT | Magnetresonanztomographie |
| N | Nikotin |
| NAD | engl. <i>nucleolus-associated chromatin domain</i> |
| ncRNA | nicht kodierende RNA (engl. <i>non-coding RNA</i>) |
| NGS | engl. <i>next generation sequencing</i> |
| NPC | engl. <i>nuclear pore complex</i> |
| NSCLC | nicht-kleinzelliges Lungenkarzinom (engl. <i>non-small cell lung cancer</i>) |
| nt | Nukleotide |
| PCA | Hauptkomponentenanalyse (engl. <i>principal component analysis</i>) |
| PCR | Polymerase-Kettenreaktion (engl. <i>polymerase chain reaction</i>) |
| PE | engl. <i>pair end</i> |
| PET | Positronen-Emissions-Tomographie |
| P.K. | Plattenepithelkarzinom |
| QC | Qualitätskontrolle (engl. <i>quality check</i>) |
| RNA-Seq. | RNA-Sequenzierung |
| rRNA | ribosomale RNA |
| RT | engl. <i>random trees</i> |
| SAM | S-Adenosylmethionin |
| SBE | engl. <i>single base extension</i> |
| SCLC | kleinzelliges Lungenkarzinom (engl. <i>small cell lung cancer</i>) |
| SE | engl. <i>single end</i> |
| SNP | Einzelnukleotid-Polymorphismus (engl. <i>single nucleotide polymorphism</i>) |
| Sp. | Spender |
| snRNA | engl. <i>small nuclear RNA</i> |
| ssDNA | einzelsträngige DNA (engl. <i>single stranded DNA</i>) |
| SVM | engl. <i>support vector machines</i> |
| TAD | engl. <i>topological associated domain</i> |
| TDG | Thymin-DNA-Glykosylase |
| U <i>Bead</i> | <i>Bead</i> für das unmethylierte Allel |
| WGBS | genomweite Bisulfitsequenzierung (engl. <i>whole genome bisulfite sequencing</i>) |
| WHO | Weltgesundheitsorganisation (engl. <i>World Health Organization</i>) |
| zfDNA | zirkulierende, zellfreie DNA |

1. Einleitung

1.1 Rolle der Epigenetik bei der Entstehung multifaktorieller Erkrankungen

Als multifaktoriell werden Erkrankungen bezeichnet, für deren Entstehung die genetische Veranlagung allein nicht ausreicht. Dabei bestimmen die Umweltfaktoren und die Lebensweise des Individuums (z.B. Umweltverschmutzung, ungesunde Ernährung oder Rauchen), ob es zum Krankheitsausbruch kommt oder nicht. Auch der Verlauf der Erkrankung kann davon beeinflusst werden. Die bekanntesten Beispiele sind Diabetes, koronare Herzkrankheit und Krebs [1-6].

Die Umwelteinflüsse können das Epigenom des Betroffenen modellieren. Dabei bleibt die eigentliche Nukleotidsequenz des DNA-Strangs unverändert. Vielmehr aktivieren oder deaktivieren epigenetische Mechanismen Gene und führen somit zu Veränderungen der Genexpression [7-14]. Epigenetische Modifikationen sind sowohl stabil als auch reversibel und können sowohl mitotisch als auch zum Teil meiotisch vererbt werden [15, 16].

Doch die Epigenetik spielt nicht nur während der Krankheitsentstehung eine Rolle, sie steuert zahlreiche zelluläre Prozesse, wie z.B. Differenzierung, Metabolismus und Zelltod [17-19]. Ein durchschnittlicher Mensch besteht aus etwa 10^{13} Zellen, die sich in 220 Zelltypen klassifizieren lassen [20, 21]. Obwohl alle Zellen im Körper ein identisches Erbgut besitzen, unterscheiden sie sich morphologisch und funktionell voneinander. Hierfür ist das Epigenom mitverantwortlich, das für jeden Zelltyp spezifisch ist [22].

Alle Zellen entwickeln sich aus einer pluripotenten Stammzelle. Während der Embryonalentwicklung teilt sich die Stammzelle und differenziert in einzelne Zelltypen. Dabei werden in einigen Zelltypen bestimmte Gene eingeschaltet, während sie bei anderen ausgeschaltet werden. Das bedeutet, dass die Unterschiede in Genexpressionsmustern die Charakteristika einer Zelle bestimmen [17, 23, 24]. Dabei wird die Genaktivität auf unterschiedlichen Ebenen von zahlreichen epigenetischen Mechanismen gesteuert. Die wichtigsten sind: DNA-Methylierung, RNA-basierte Mechanismen, Histonmodifikationen, und Organisation des Zellkerns (Abb. 1) [25-27]. In folgenden Abschnitten werden sie jeweils näher erläutert.

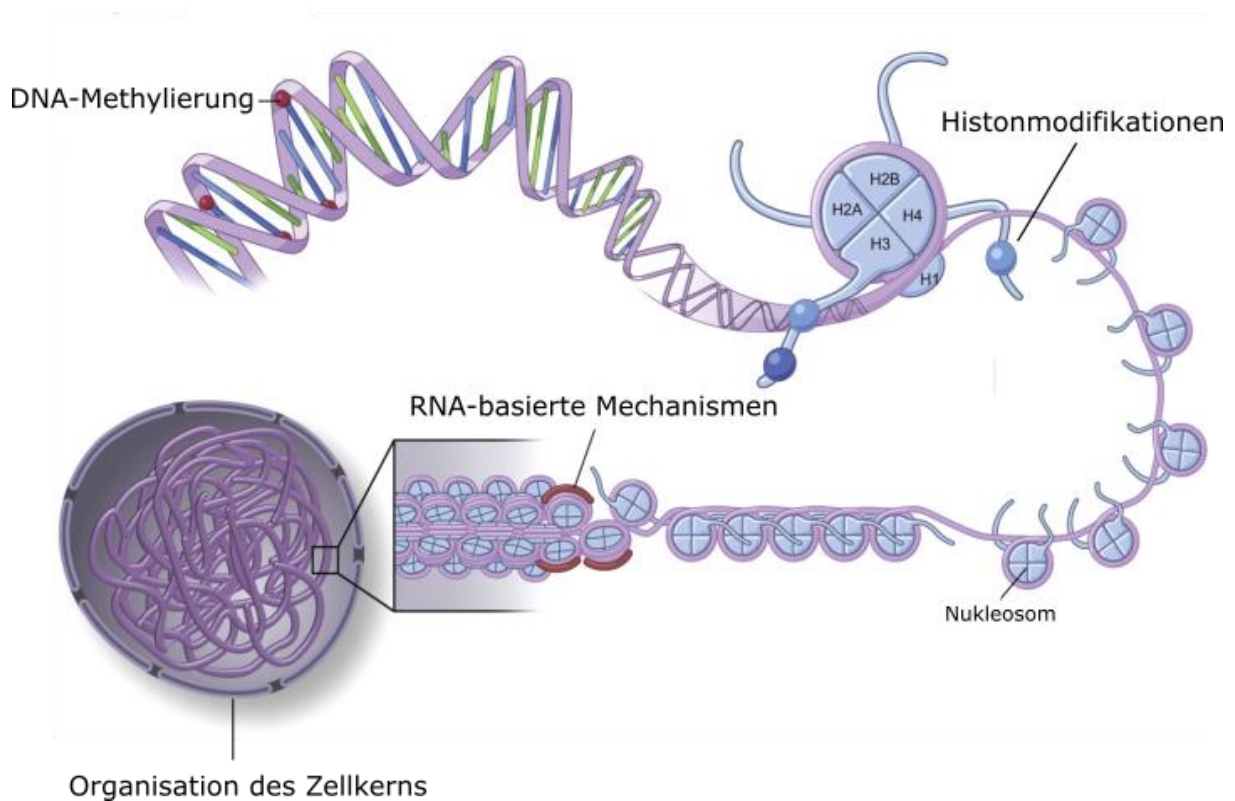


Abb. 1: Durch die epigenetischen Mechanismen wird die zelluläre Genexpression gesteuert. Die vier wichtigsten Mechanismen sind dabei: DNA-Methylierung, Modifikation von Histonen, RNA-basierte Regulation sowie Organisation des Zellkerns. Die Abbildung wurde verändert nach [27].

1.1.1 DNA-Methylierung

Bei der DNA-Methylierung handelt es sich um eine stabile, kovalente Modifikation des Cytosins. Sie ist reversibel und für viele zelluläre Prozesse zuständig. Die wichtigsten sind Expressionskontrolle, Genomintegrität, Abwehr molekularer Parasiten (z.B. Retroviren) sowie genomische Prägung [28-31].

Die Reaktion wird von DNA-Methyltransferasen katalysiert und ist in der Abb. 2A dargestellt. Dabei wird eine Methylgruppe an das fünfte Kohlenstoffatom des Pyrimidinrings addiert, als Methylgruppendonor dient S-Adenosylmethionin (SAM). In Säugetieren sind drei unterschiedliche Methyltransferasen für die *de novo* und Erhaltungsmethylierung zuständig [32].

DNMT3A und DNMT3B sind für die *de novo* DNA-Methylierung bestimmter genomischer Regionen während der frühen embryonalen Entwicklung verantwortlich. In adulten, somatischen Geweben werden diese Enzyme entweder gar nicht oder kaum exprimiert [33].

DNMT1 ist für die Aufrechterhaltung des bestehenden Methylierungsmusters zuständig. Während der Zellteilung wird das Erbgut verdoppelt. Das wiederum führt zunächst zur Bildung eines hemimethylierten DNA-Doppelstrangs. Das bedeutet, dass der eine Strang der Doppelhelix der ursprüngliche ist und das entsprechende Methylierungsmuster aufweist, während der andere DNA-Einzelstrang neu synthetisiert wurde und zunächst gar keine Methylierungsmuster hat. Die DNMT1 erkennt den hemimethylierten DNA-Doppelstrang und nutzt den ursprünglichen DNA-Einzelstrang als *Template*, um den neu-synthetisierten DNA-Einzelstrang an entsprechenden Stellen zu methylieren (Abb. 2B) [34-37].

Die DNA-Methylierung erfolgt überwiegend an Cytosinen in CpG-Dinukleotiden. Im humanen Genom sind ca. 70–80% aller CpG-Dinukleotide methyliert, die meisten davon (ca. 98%) liegen im Genom ungleichmäßig verteilt vor. Eine Ausnahme stellen die sogenannten CpG-Inseln dar, welche die übrigen 2% der CpG-Dinukleotide beinhalten. Hierbei handelt es sich um DNA-Regionen mit einer Mindestlänge von 500bp, deren CpG-Dichte mehr als 55% beträgt. Innerhalb eines Gens sind CpG-Inseln häufig in Promotorbereichen lokalisiert [38-41]. Werden mehrere Cytosine innerhalb eines Promotors methyliert, so kann die Transkriptionsmaschinerie zum Teil aufgrund der sterischen Hinderung nicht mehr binden [42]. Des Weiteren werden Methyl-CpG-bindende Proteine (MBD1, MBD2 und MeCP2) rekrutiert, die eine zusätzliche Umstrukturierung des Chromatins bewirken. Auf diese Weise kann die DNA-Methylierung die transkriptionelle Repression eines Gens bewirken [43, 44].

Wie bereits erwähnt, ist DNA-Methylierung auch reversibel. Die Demethylierung mehrerer Cytosine im Promotorbereich kann das reprimierte Gen wieder aktivieren. Dieser Vorgang wird von TET-Enzymen katalysiert. Dabei wird 5-Methylcytosin (5-mC) zunächst zu 5-Hydroxymethylcytosin (5-hmC) und dann über 5-Formylcytosin (5-fC) zu 5-Carboxylcytosin (5-caC) oxidiert. Anschließend schneidet Thymin-DNA-Glykosylase (TDG) 5-fC sowie 5-caC aus dem DNA-Strang heraus. Im Laufe der darauffolgenden Basenexzisionsreparatur (BER) wird der DNA-Strang durch unmethyliertes Cytosin aufgefüllt (Abb. 2C). Alternativ kann auch eine spontane Decarboxylierungsreaktion von 5-caC zu unmethyliertem Cytosin erfolgen [45].

1.1.2 RNA-basierte Mechanismen

Innerhalb einer Zelle existieren neben der proteinkodierenden mRNA zahlreiche regulatorische RNA-Moleküle. Die sogenannte *non-coding RNA* (ncRNA) wird ebenfalls durch die Transkription synthetisiert, jedoch nicht in Proteine translatiert. Die ncRNA Moleküle sind überwiegend für die Regulation der Genexpression zuständig [46]. Die Ausnahmen stellen die ribosomalen (rRNA) sowie die sogenannten *small nuclear RNA* Moleküle (snRNA) dar, die wichtige Bestandteile von Ribozymen und Spleißosomen sind [47-49].

Unterschiedliche ncRNA wirken auf die Genexpression entweder aktivierend oder reprimierend. In beiden Fällen lagern sie sich zunächst an die komplementären Abschnitte der DNA an (z.B. in Promotorbereichen). Soll ein Gen aktiviert werden, können bestimmte Transkriptionsfaktoren rekrutiert werden. Ist das Gegenteil der Fall, können ncRNA Moleküle z.B. DNA-Methyltransferasen dirigieren. Der Promotorbereich wird in diesem Fall methyliert und die Transkription des entsprechenden Gens inhibiert [50, 51]. Die RNA-Interferenz stellt einen weiteren Mechanismus der transkriptionellen Inaktivierung dar. Dabei bindet ncRNA die komplementären mRNA-Moleküle und führt in weiterem Verlauf zu deren Degradierung. Somit steht die entsprechende mRNA für die Proteinbiosynthese nicht mehr zur Verfügung [52, 53].

1.1.3 Modifikation von Histonen

Das humane Genom umfasst $3,3 \times 10^9$ bp. Nur ein Bruchteil des Erbguts kodiert für Proteine und ist einem Gen zuzuordnen. Regionen mit hoher Gen-Dichte liegen meist offen vor und werden als Euchromatin bezeichnet. Gen-arme Regionen sind dagegen verpackt, hier spricht man von Heterochromatin [54, 55].

Für eine effiziente DNA-Verpackung werden zunächst aus jeweils zwei Kopien der Histone H2A, H2B, H3 und H4 globuläre Histonoktamere gebildet. Um ein Histonoktamer können 146bp DNA gewickelt werden. Der Histon-DNA-Komplex wird auch als Nukleosom bezeichnet. Der Histon H1 stabilisiert das Nukleosom, indem er die DNA an den Eintritts- bzw.- Austrittspunkten fixiert. Die benachbarten Nukleosomen werden durch kurze DNA-Abschnitte, die sogenannten *Linker*, verbunden [56-59]. Diese Struktur wird als 10nm-Faser bezeichnet, im Elektronenmikroskop sieht sie wie eine Perlenschnur aus. Im nächsten Schritt wird die 10nm-Faser schraubenförmig zu einer 30nm-Faser aufgewunden. Sechs bis acht Nukleosomen sind für eine Windung notwendig. Die Anordnung der 30nm-Faser in 60-150kb

Schleifen wird als Solenoid bezeichnet. Während der Mitose kondensiert das Solenoid, die dabei entstehenden Strukturen werden als Chromosomen bezeichnet (Abb. 3) [60-63].

Damit ein Gen transkribiert werden kann, müssen die jeweiligen Promotor- und *Enhancer*-Bereiche für die Transkriptionsmaschinerie zugänglich sein. Um das zu ermöglichen verfügen Zellen über Histon-modifizierende Enzyme (HME) sowie ATP-abhängige Chromatin-*Remodeling*-Komplexe (ACRK) [64].

Für beide spielen die N-Terminalen Histonenden, die sogenannten Histonschwänze, eine wichtige Rolle. Hierbei handelt es sich um Bereiche des Histonoktamers, die keine Sekundärstrukturen aufweisen und somit unterschiedlichste Konformationen annehmen können [65]. Innerhalb der Histonschwänze gibt es Regionen, die über elektrostatischen Wechselwirkungen, Wasserstoffbrückenbindungen sowie Van-der-Waals-Kräfte mit den kleinen bzw. großen Furchen der negativ geladenen DNA-Doppelhelix interagieren und somit zur Stabilität des Nukleosoms beitragen können [65, 66]. Des Weiteren sind Regionen vorhanden, die flexibel bzw. mobil bleiben [66].

HME können diese flexiblen Regionen kovalent modifizieren [67]. In diesem Zusammenhang spricht man von dem sogenannten „Histon-Code“, das die Acetylierung, Methylierung, Phosphorylierung, Ubiquitinierung, Sumoylierung oder Biotinylierung bestimmter Aminosäuren beinhaltet [67, 68]. Je nach Modifikation kann die DNA-Histon-Wechselwirkung verstärkt oder gelockert werden [68, 69]. Falls die DNA-Histon-Wechselwirkung in einem Promotorbereich gelockert wird, kann dort die Transkriptionsmaschinerie binden und die entsprechenden Gene ablesen. Ist das Gegenteil der Fall, so wird die entsprechende Genregion verpackt und inaktiviert (Abb. 3) [70-74]. So führt zum Beispiel eine Methylierung des Histons 3 an den Lysinresten 9 und 27 (H3K9, H3K27) zu transkriptionellen Repression. Dagegen kommt es durch Methylierung der Lysinreste 4, 36 und 79 (H3K4, H3K36, H3K79) zu einer Genaktivierung [75].

Bei den ACRK handelt es sich um sehr große Proteine, die bis zu 17 Untereinheiten enthalten können. ACRK werden in vier funktionell und strukturell unterschiedliche Familien unterteilt: SWI/SNF, ISWI, NURD/MI-2/CHD und INO80. Eins haben sie jedoch alle gemeinsam: eine ATPase Untereinheit [76]. Die Enzyme hydrolysieren ATP, um die DNA-Histon-Wechselwirkung zu reduzieren. Dadurch sind sie in der Lage Nukleosome zu verschieben, zu ersetzen oder ganz aufzulösen [64, 76, 77]. Nachfolgend rekrutieren sie den transkriptionellen Apparat zu dem freigelegten Gen und ermöglichen somit dessen Transkription. Um ein aktives Gen wieder abzuschalten, können ACRK die freiliegende Promotorregion wieder um ein Histonoktamer wickeln [64].

1.1.4 Organisation des Zellkerns

Beim Betrachten der Zellkernorganisation fällt als erstes die Lamina auf. Dabei handelt es sich um eine dichte, proteinhaltige Schicht, die auf der Innenseite der Kernhülle lokalisiert ist. Sie ist dafür zuständig, dass aktive Gene in das Innere des Zellkerns gezogen werden, während inaktive Gene in der Peripherie bleiben. Bei den regulatorisch wichtigsten Bestandteilen handelt es sich um Lamine A, B1, B2 und C, die mit bestimmten, 0,1-10Mb langen DNA-Regionen, den sogenannten *lamina associated domains* (LADs), interagieren können [78-84].

Der Nucleolus bildet eine weitere auffällige Struktur innerhalb des Zellkerns. Um ihn herum sind die *nucleolus-associated chromatin domains* (NADs) lokalisiert. NADs sind genarm, und bestehen hauptsächlich aus repetitiven DNA-Sequenzen. Für die Transkriptionsmaschinerie sind sie schlecht zugänglich [85-87].

Zusätzlich zu NADs und LADs wird häufig von *topological associated domains* (TADs) gesprochen. Hierbei handelt es sich um Chromatinstrukturen, in deren Mitte sich reprimierte oder auch zellspezifische Gene befinden, die wiederum von aktiven Haushaltsgenen umgeben sind [88-90].

Die Chromosomen sind die umfangreichste Komponente des Zellkerns und befinden sich in bestimmten Bereichen. Jedem Chromosom ist eine für ihn spezielle *chromosome territory* zugeordnet. Dabei befinden sich genreiche, aktive Chromatinbereiche in der Peripherie, dagegen reprimierte Bereiche in der Mitte der jeweiligen *chromosome territory* [91, 92].

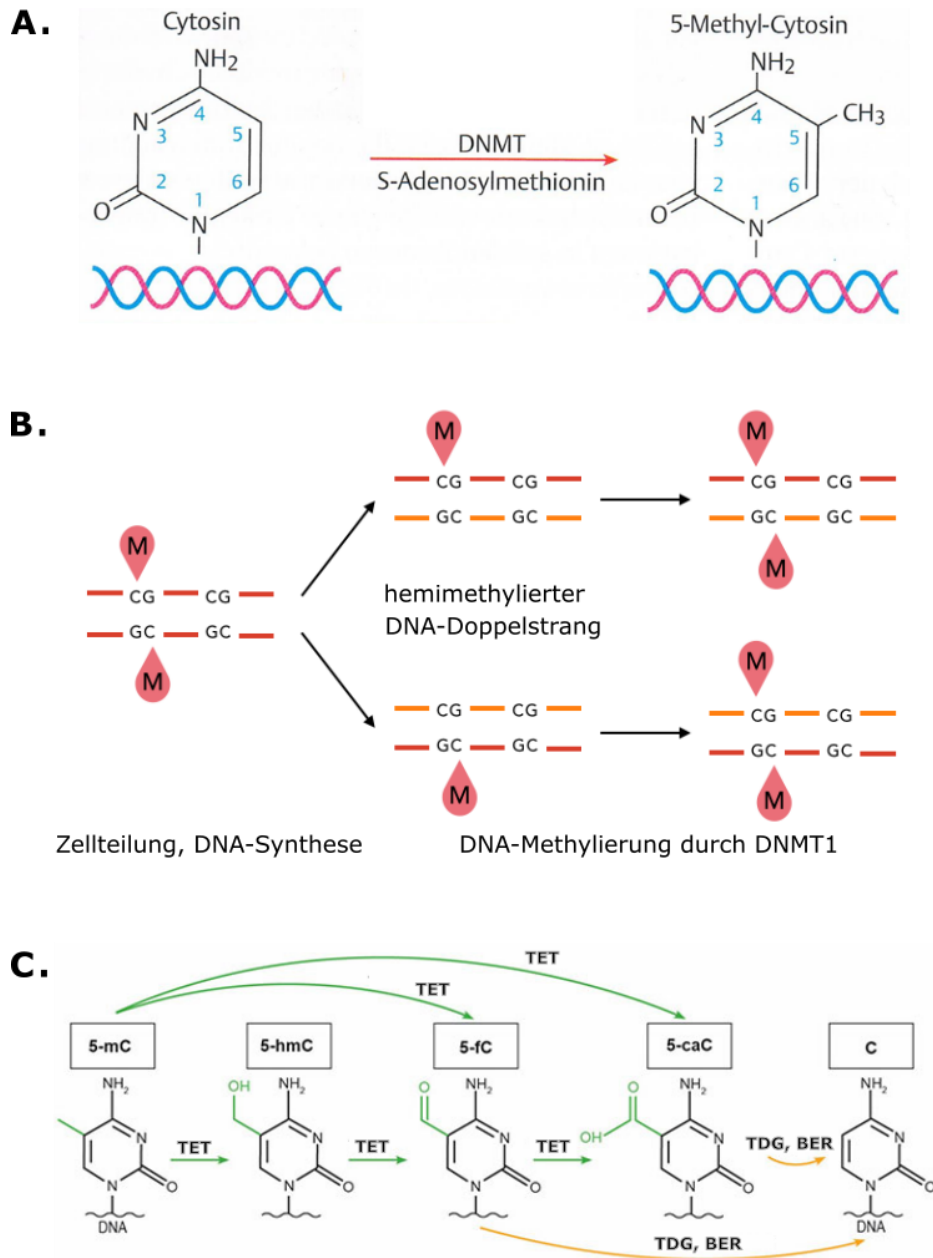


Abb. 2: **A.** Die DNA-Methylierung des Cytosins erfolgt an fünfter Position des Pyrimidinrings. Die Reaktion wird von DNA-Methyltransferasen (DNMT) katalysiert, als Methylgruppendedor dient dabei S-Adenosylmethionin (SAM). **B.** Um das DNA-Methylierungsmuster nach der Zellteilung aufrechtzuerhalten, ist die DNMT1 notwendig. Das Enzym erkennt den hemimethylierten DNA-Doppelstrang und verwendet den ursprünglichen, methylierten DNA-Einzelstrang (rot) als *Template* um den neu synthetisierten DNA-Einzelstrang (orange) an den richtigen Stellen zu methylieren (M). **C.** Während der DNA-Demethylierung oxidieren TET-Enzyme das 5-Methylcytosin (5mC) zu 5-Hydroxymethylcytosin (5-hmC). 5-hmC wird dann über 5-Formylcytosin (5-fC) zu 5-Carboxycytosin (5-caC) umgesetzt (grüne Pfeile). Thymin-DNA-Glykosylase (TDG) schneidet im nächsten Schritt 5-fC sowie 5-caC aus dem DNA-Strang heraus. Während der darauffolgenden Basenexzisionsreparatur (BER) wird der DNA-Strang durch ein unmethyliertes Cytosin aufgefüllt. Alternativ kann eine spontane Decarboxylierungsreaktion von 5-caC zu einem unmethylierten Cytosin erfolgen. Die Abbildungen wurden verändert nach [32, 93, 94].

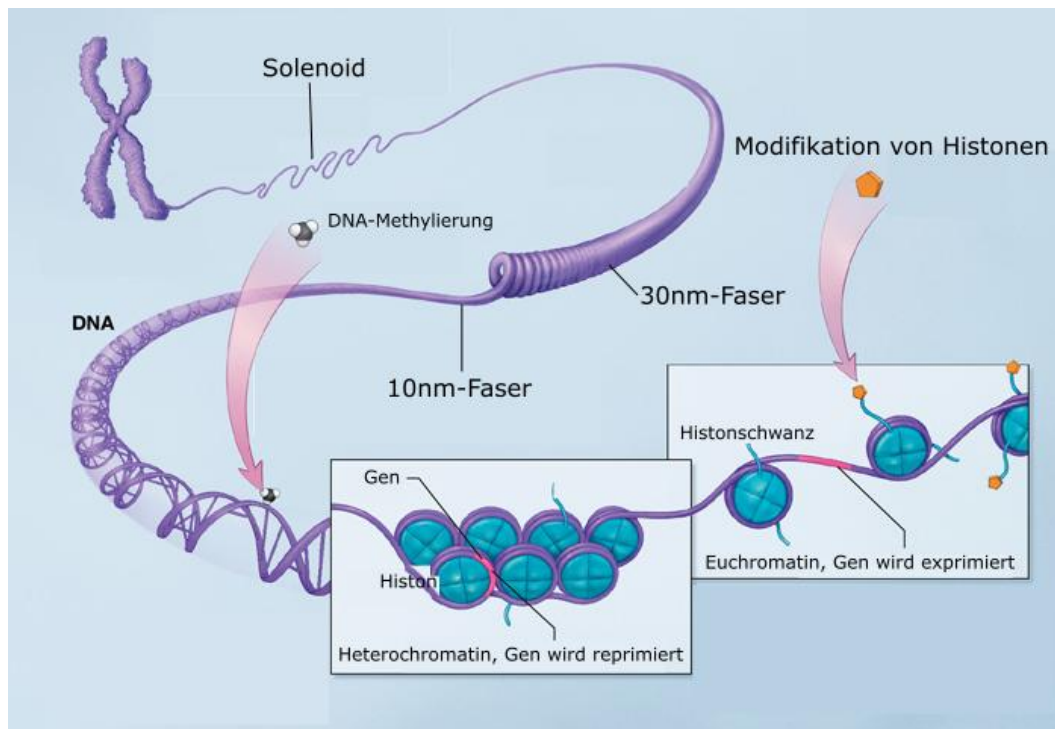


Abb. 3: Um eine effiziente DNA-Verpackung zu gewährleisten, können 146bp lange DNA-Regionen um ein Histonoktamer gewickelt werden. Das globuläre Histonoktamer besteht aus je zwei Kopien der Histone H2A, H2B, H3 und H4. Die N-terminalen Enden der Histone, die sogenannten Histonschwänze, ragen heraus und können chemisch modifiziert werden. Die Modifikationen können aktivierend oder reprimierend wirken. Nach einer aktivierenden Histonmodifikation wird die DNA-Histon-Wechselwirkung gelockert, die entsprechende Genregion wird freigelegt und ist nachfolgend für die Transkriptionsmaschinerie zugänglich. Dagegen führt eine reprimierende Histonmodifikation zu einer stärkeren DNA-Histon-Wechselwirkung. Ein aktives Gen kann somit verpackt und reprimiert werden. Mehrere benachbarte Nucleosomen bilden die 10nm-Faser, die schraubenförmig zu einer 30nm-Faser gewunden werden kann. Um die Solenoid-Struktur zu erhalten, wird die 30nm-Faser in 60-150kb lange Schleifen organisiert. Während der Mitose kondensiert das Solenoid zu Chromosomen. Die Abbildung wurde verändert nach [95].

1.2 Lungenkrebs

Lungenkrebs ist weltweit die zweithäufigste Krebsart bei Männern und Frauen. In Deutschland werden jährlich ca. 52.500 Neuerkrankungen registriert. Das mittlere Erkrankungsalter liegt für Männer bei 70 und für Frauen bei 69 Jahren. Dabei wird zwischen dem kleinzelligen (engl. *small cell lung cancer*; SCLC) und nicht-kleinzelligen (engl. *non-small cell lung cancer*; NSCLC) Lungenkarzinom unterschieden [96, 97]. NSCLC ist deutlich häufiger und tritt bei 85% der betroffenen Patienten auf. Des Weiteren wird NSCLC in mehrere Subentitäten unterschieden, davon sind die häufigsten Adeno- und Plattenepithelkarzinom [98].

Dass die Symptome der Erkrankung meistens sehr spät auftreten, spiegelt sich in einer schlechten Prognose wieder. Die 5-Jahres Überlebensrate liegt bei 15 %. Meistens klagen die betroffenen Patienten über einen plötzlich auftretenden und immer schlimmer werdenden Husten, Müdigkeit, Abgeschlagenheit oder plötzliche Gewichtsabnahme [99].

1.2.1 Molekularbiologie und Epigenetik des Lungenkarzinoms

Lungenkrebs ist eine heterogene Erkrankung. So können z.B. Mutationen innerhalb von *KRAS*, *EGFR*, *BRAF*, *MEK1*, *MET*, *HER2*, *ALK*, *ROS1*, *RET*, *FGFR1*, *DDR2*, *MET*, *PTEN*, *LKB1*, *RB1*, *CDKN2A* oder *TP53* Genen die Entstehung eines primären Lungenkarzinoms induzieren [100]. Zusätzlich akkumulieren im Laufe der Tumor-Evolution die sogenannten *Passenger*-Mutationen, die zur verschiedenen Subklonen führen können [101].

Zusätzlich zu Mutationen spielen während der Tumor-Evolution auch Epimutationen eine entscheidende Rolle [102]. So werden z.B. Promotoren innerhalb bestimmter Tumorsuppressorgenen hypermethyliert, was wiederum in deren transkriptioneller Repression resultiert [103, 104]. Dieses Phänomen ist durch die Überexpression von DNA-Methyltransferasen begleitet [105, 106]. Besonders häufig wurde in der Literatur Promotorhypermethylierung innerhalb der *P16INK4A*, *RASSF1A*, *APC*, *RARB*, *CDH1*, *CDH13*, *DAPK*, *FHIT* und *MGMT* Gene beschrieben [107].

Die genomweite Hypomethylierung des NSCLC ist mit der genomischen Instabilität assoziiert. Eine gezielte Hypomethylierung der Gene konnte bisher nur bei *MAGEA3/6*, *TKTL1*, *BORIS*, *DDR1*, *YWHAZ* sowie *TMSB10* identifiziert werden [107, 108].

Des Weiteren weisen maligne Lungentumore häufig eine veränderte Histon-Acetylierung an den Positionen H4K5, H4K8, H4K12 und H4K16 auf. Auch der globale Anteil an H4K20me3 ist in NSCLC geringer als in gesundem Lungengewebe [107].

Zusätzlich kann es zu aberranten ncRNA-Expression kommen, wie z.B. *MIR196A*, *MIR200B*, *MALAT1* sowie *HOTAIR* [109-111].

1.2.2 Diagnostik des Lungenkarzinoms, molekulare Möglichkeiten der Lungenkrebsfrüherkennung

Laut nationalen und internationalen Empfehlungen wird bei einer Verdachtsdiagnose der betroffene Patient zunächst einer umfassenden körperlichen Untersuchung unterzogen. Nachfolgend wird der Brustkorb durch bildgebende Verfahren wie z.B. Röntgen oder Computertomographie (CT) untersucht [99, 112, 113]. Falls dabei ein Tumor detektiert wird, wird eine Bronchoskopie empfohlen. Dabei wird die Lunge gründlich analysiert sowie eine Biopsie des Tumors entnommen. Die Biopsie wird weiteren histologischen und molekulargenetischen Analysen unterzogen [113-115].

Während der histologischen Untersuchung wird festgestellt, ob der Tumor bösartig ist. Falls dies der Fall ist, wird dessen Entität ermittelt. Um die optimale Therapie zu identifizieren werden zusätzlich molekulargenetische sowie bildgebende Verfahren herangezogen [116-118]. Vor allem die bildgebenden sowie endoskopischen Verfahren können aufgrund der Radioaktivität und Invasivität für die betroffenen Patienten belastend sein [96]. Hier wäre es wünschenswert, zukünftig nicht invasive auf *Liquid Biopsies* basierte Verfahren verwenden zu können.

Auch in Hinsicht auf Lungenkrebsfrüherkennung hat die Methode den bildgebenden Verfahren gegenüber einen deutlichen Vorteil. Das Detektionslimit der radiologischen Verfahren liegt bei sieben bis 10mm, das entspricht einem Tumor, der bereits eine Milliarde Zellen beinhaltet. Mit *Liquid Biopsies* können Tumore viel früher, ab 50 Millionen Zellen erkannt werden [119].

Bei *Liquid Biopsies* werden dem Patienten einige Milliliter Blut entnommen. Aus dem Blutplasma oder Blutserum kann anschließend zirkulierende, zellfreie DNA (zfdDNA) isoliert werden. Im menschlichen Körper entsteht die zfdDNA im Laufe apoptotischer sowie nekrotischer Prozesse. Dabei wird zelluläre, genomische DNA (gDNA) durch DNAsen in 167bp lange Fragmente gespalten und im Blutkreislauf freigesetzt [120, 121].

Bei Patienten, die an malignen Erkrankungen leiden, ist in der Gesamtmenge an zfdDNA zusätzlich Tumor-DNA enthalten. Je nach Entität bzw. Stadium der Erkrankung, kann die zfdDNA-Menge stark variieren. Sie enthält jedoch alle diagnostisch, therapeutisch und prognostisch relevanten Informationen [122-124].

Zusätzlich zu genetischen Mutationen eines Tumors können auch Epimutationen analysiert werden. Besonders interessant ist in diesem Zusammenhang die DNA-Methylierung. Das DNA-Methylierungsmuster ist gewebespezifisch und ändert sich bereits in frühen Phasen der Tumor-Evolution [8, 38, 41]. Des Weiteren machte die Studie des *GNAS1 Locus* deutlich,

dass die zfDNA-Methylierung im Blut stabil bleibt. Sie wird weder modifiziert noch verfälscht und eignet sich somit als Biomarker in der klinischen Diagnostik [125].

Mehrere Studien haben bereits das diagnostische Potential der DNA-Methylierung deutlich gemacht. So zeigte die *SOX17*-Studie beim Magenkarzinom, dass das Gesamtüberleben der Patientenkohorte mit der nachgewiesenen Menge an methylierter *SOX17*-zfDNA korrelierte [126]. Eine Studie mit Patientinnen, die an Mammakarzinom litten, zeigte eine signifikante Hypermethylierung des *CST6*-Gens [127]. Liggett et.al. ist es gelungen, anhand des DNA-Methylierungsmusters das Pankreaskarzinom von seiner Vorstufe, der chronischen Pankreatitis, unterscheiden zu können [128].

Auch beim nicht-kleinzelligen Lungenkarzinom wurden von mehreren Arbeitsgruppen Veränderungen des DNA-Methylierungsmusters beschrieben. So konnten z.B. Balgkouranidou et al. eine signifikante Hypermethylierung des *BRMS1*-Gens bei Patienten mit Bronchialkarzinom nachweisen [129]. 2016 detektierten Marwitz, Scheufele et.al DNA-Hypomethylierung innerhalb der *CTLA4*- sowie *PDCD1*-Gene. Auf Transkriptom-Ebene waren diese Gene überexprimiert. Da es sich hier um wichtige *Checkpoint*-Regulatoren handelt, hat diese Arbeit starke therapeutische Relevanz [130].

Das diagnostische Potential der DNA-Methylierung wird auch am Beispiel des „Epi proLung“ Assays („Epigenomics AG“, Deutschland) deutlich. Dabei wird das zfDNA-Methylierungsmuster der *SHOX2* und *PTGER4* Gene analysiert. Bei einer Spezifität von 90%, beträgt die Sensitivität 67% [131]. Für ein zuverlässiges Lungenkrebs-*Screening* reicht die Empfindlichkeit des „Epi proLung“ Tests nicht aus. Bisher gibt es keine weiteren auf *Liquid Biopsies* basierte Verfahren, die eine zuverlässige präventive Lungenkrebsfrüherkennung ermöglichen [96].

1.2.3 Rolle des Zigarettenkonsums bei der Krankheitsentstehung

Viele ungünstige *Lifestyle* Faktoren können zur Entstehung des NSCLC beitragen. Die bekanntesten sind Zigarettenrauch, Asbest, Dieselruß sowie Feinstaub am Arbeitsplatz [132-136].

Der Konsum von Zigaretten ist die häufigste Ursache von Lungenkrebs bei Männern und Frauen. Die Korrelation zwischen der Häufigkeit von Bronchialkarzinomen und Rauchen ist seit Jahren beobachtet worden [137]. Der Zigarettenrauch führt zum Auftreten verschiedenster Mutationen, wie z.B. in *EGFR*, *TP53* oder *KRAS* Genen [138-144]. Er wirkt jedoch nicht nur auf genetischer Ebene, sondern beeinflusst auch epigenetische Regulatoren

und Mechanismen [145, 146]. So berichten mehrere Studien von massiver Veränderung des DNA-Methylierungsmusters bei Rauchern [147, 148].

Aufgrund der gesundheitlichen Nachteile, die mit Zigarettenkonsum einhergehen, wechseln immer mehr Raucher zu E-Zigaretten. Einige Forscher und Ärzte halten die E-Zigarette für deutlich gesünder als Tabakprodukte und gehen davon aus, dass der E-Zigarettenkonsum die Lungenkrebsrate zukünftig erheblich senken wird [149-151].

Der Aufbau einer E-Zigarette ist in der Abbildung 4A dargestellt. Optisch ähneln einige Geräte sogar der klassischen Zigarette. Im Inneren befinden sich jedoch das eigentliche *Liquid*, der Verdampfer, ein Lithium-Ionen-Akku, eine LED-Leuchte und ein Mikroprozessor. Beim Zug am Mundstück registrieren die eingebauten Sensoren einen Unterdruck. Das wiederum führt zur Aktivierung des Akkus, der den Verdampfer mit notwendiger elektrischer Spannung versorgt. Die Heizspirale des Verdampfers wird schnell erhitzt und führt zum Verdampfen des *Liquids* [152]. Aufgrund der steigenden Beliebtheit ist heutzutage ein großes Angebot an *Liquid*-Aromen auf dem Markt erhältlich [153].

Das *Liquid* besteht hauptsächlich aus Glycerin, Propylenglykol und Aromen. Beim Erhitzen reagiert das Glycerin-Propylenglykol-Gemisch jedoch zu gesundheitsschädlichen chemischen Verbindungen, unter anderem entstehen hohe Mengen an giftigen Acrolein und Formaldehyd (Abb. 4B). Deshalb raten einige internationale Studien und die Weltgesundheitsorganisation (engl. *World Health Organisation*, WHO) vom E-Zigarettenkonsum ab [154-158]. Das führte dazu, dass in Bahrain, Singapur, Thailand und Arabischen Emiraten der Verkauf von E-Zigaretten verboten wurde [159].

Die Markteinführung der E-Zigarette ist noch nicht lange her. Um die tatsächlichen Risiken der E-Zigarette abzuschätzen, fehlen Ergebnisse aus Langzeitstudien. Des Weiteren sind kaum biologische multi-OMICS Studien vorhanden, die die Wirkung des E-Zigaretten dampfs auf den Phänotyp, das Genom und das Epigenom einer Lungenzelle analysieren. Diese Untersuchungen wären jedoch entscheidend, um das tatsächliche Gefährdungspotential der E-Zigaretten einschätzen zu können [155, 157, 160].

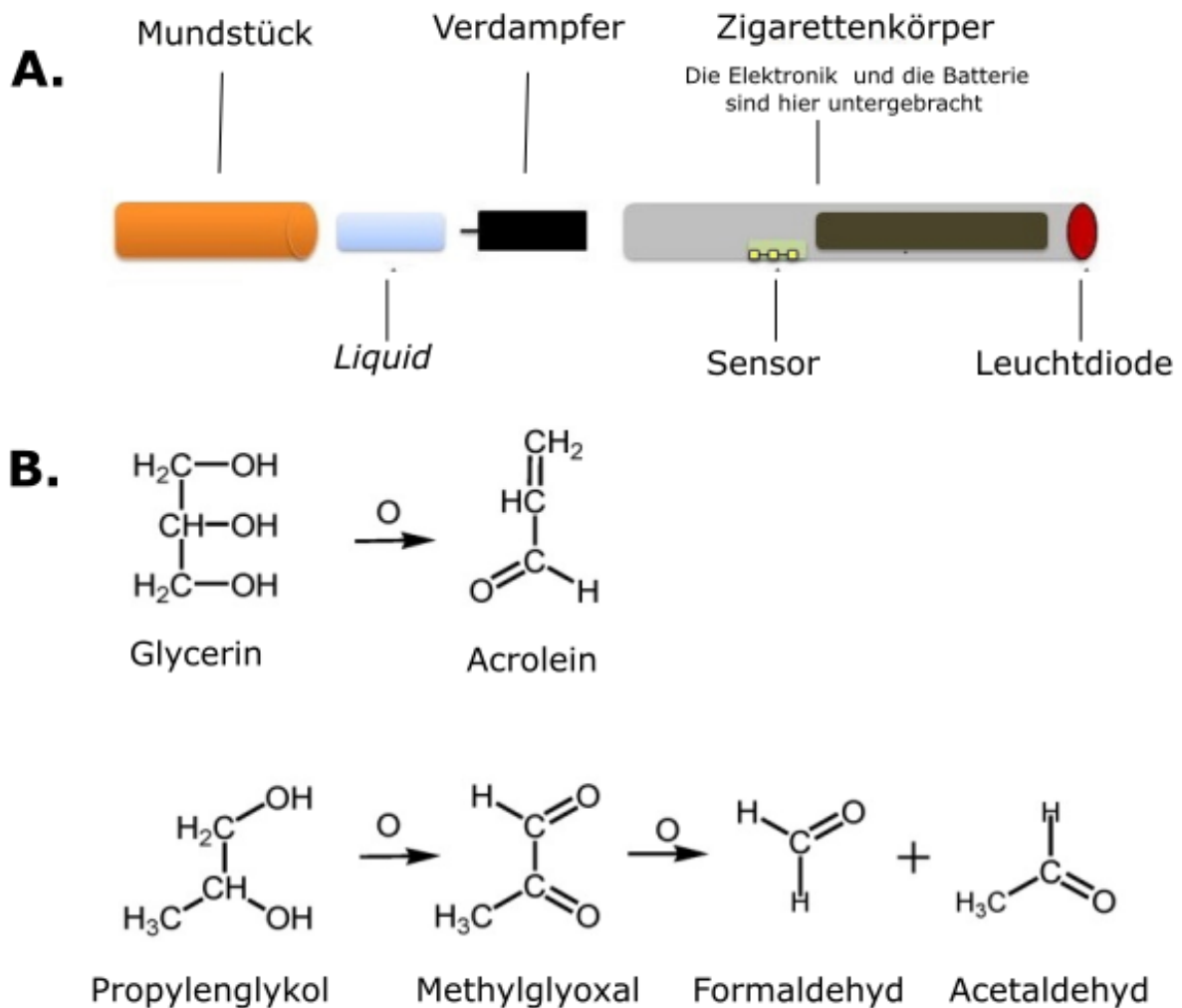


Abb. 4: **A.** Die E-Zigarette ist aus mehreren Bestandteilen aufgebaut. Beim Ziehen am Mundstück entsteht Unterdruck, der von Sensoren im Zigarettenkörper registriert wird. Die Sensoren aktivieren den Akku, der den Verdampfer mit elektrischer Spannung versorgt. Der Verdampfer erhitzt das *Liquid*, dabei entsteht der E-Zigaretteendampf. Gleichzeitig wird das Glimmen der Zigarette durch die eingebaute Leuchtdiode simuliert. **B.** Beim Verdampfen des *Liquids* kommt es zur Oxidation des Glycerin-Propylenglykol-Gemisches. Dabei entstehen giftige chemische Verbindungen, wie z.B. Acrolein und Formaldehyd. Die Abbildung wurde verändert nach [157, 161].

1.3 Zielsetzung

Lungenkrebs ist die zweithäufigste Krebsart bei Männern und Frauen. Da die Erkrankung häufig erst in fortgeschrittenen Stadien diagnostiziert wird, ist Lungenkrebs mit einer schlechten Prognose assoziiert. Bisher gibt es keine Möglichkeiten der routinemäßigen Lungenkrebsfrüherkennung. In den letzten Jahren machten mehrere wissenschaftliche Studien das Potential von epigenetischen Biomarkern in der klinischen Diagnostik deutlich. Zu den stabilsten tumorspezifischen Epimutationen zählen Veränderungen der DNA-Methylierungssignatur, die bereits früh in der Tumor-Evolution auftreten. Mehrere Studien konnten auch umfangreiche, genomweite Veränderungen der DNA-Methylierungssignatur bei Rauchern nachweisen. Der Tabakkonsum zählt zu den häufigsten Ursachen von Lungenkrebs, weshalb immer mehr Raucher zu E-Zigaretten wechseln. Um jedoch das Gefährdungspotential der E-Zigaretten einschätzen zu können, fehlen zum einen Ergebnisse aus Langzeitstudien, zum anderen sind kaum biologische Studien vorhanden, welche die Wirkung des E-Zigarettdampfs auf die Entstehung von Epimutationen analysieren. Die vorliegende Doktorarbeit wurde im Rahmen des Deutschen Zentrums für Lungenforschung (DZL) angefertigt, dabei wurden folgende Ziele verfolgt:

1. Das erste Ziel war die Entwicklung eines NGS-basierten Lungenkrebs(früh)erkennungstests, der unter Verwendung von *Liquid Biopsies* maligne Lungentumore auch in frühen Stadien zuverlässig identifiziert und Informationen bezüglich der Entität, des Stadiums und Prognose der betroffenen Patienten liefert. Die für dieses Vorhaben benötigten Technologien (insbesondere das *whole genome bisulfite sequencing* (WGBS) von zirkulierender zellfreier DNA sowie die anschließende Datenanalyse) mussten etabliert und ggf. validiert werden.
2. Das nächste Ziel war die Validierung des entwickelten Plasma *Panels*. Hierfür sollten im Rahmen einer Pilot-Studie *Liquid Biopsies* einer kleinen Patientenkohorte mit Hilfe des *Panels* analysiert werden.
3. Ein weiteres Ziel war die Untersuchung der durch E-Zigarettenkonsum induzierten, molekularen Veränderungen in Lungenzellen. Dazu sollten verschiedene funktionelle, morphologische sowie genomweite molekulargenetische und -epigenetische Analysen durchgeführt werden.

2. Material

2.1 Probenmaterial

2.1.1 OP-Präparate

Die OP Präparate wurden im Rahmen des DZL Konsortiums von der "LungenClinic Grosshansdorf GmbH", (Deutschland) bereitgestellt. Alle IDs wurden nach einem Zufallsverfahren anonymisiert, sodass es nicht mehr möglich ist, einzelne Patienten zu identifizieren. Die Studie wurde von der Ethikkommission der Universität Lübeck befürwortet (Ref. 12-220). Insgesamt wurden 40 Tumorgewebe sowie deren korrespondierende Kontrollen (40 Patienten) unter Verwendung des „Illumina Infinium Human Methylation450K BeadChips“ analysiert (Tab. 1).

Tab. 1: Studienkollektiv bestehend aus 40 Lungenkarzinompatienten.

| Patienten ID | Entität | Patienten ID | Entität |
|--------------|-------------------------|--------------|------------------------|
| A.K.1 | Adenokarzinom | N.K.1 | NSCLC |
| A.K.2 | Adenokarzinom | P.K.1 | Plattenepithelkarzinom |
| A.K.3 | Adenokarzinom | P.K.2 | Plattenepithelkarzinom |
| A.K.4 | Adenokarzinom | P.K.3 | Plattenepithelkarzinom |
| A.K.5 | Adenokarzinom | P.K.4 | Plattenepithelkarzinom |
| A.K.6 | Adenokarzinom | P.K.5 | Plattenepithelkarzinom |
| A.K.7 | Adenokarzinom | P.K.6 | Plattenepithelkarzinom |
| A.K.8 | Adenokarzinom | P.K.7 | Plattenepithelkarzinom |
| A.K.9 | Adenokarzinom | P.K.8 | Plattenepithelkarzinom |
| A.K.10 | Adenokarzinom | P.K.9 | Plattenepithelkarzinom |
| A.K.11 | Adenokarzinom | P.K.10 | Plattenepithelkarzinom |
| A.K.12 | Adenokarzinom | P.K.11 | Plattenepithelkarzinom |
| A.K.13 | Adenokarzinom | P.K.12 | Plattenepithelkarzinom |
| A.K.14 | Adenokarzinom | P.K.13 | Plattenepithelkarzinom |
| A.K.15 | Adenokarzinom | P.K.14 | Plattenepithelkarzinom |
| A.S.K.1 | adenosquamöses Karzinom | P.K.15 | Plattenepithelkarzinom |
| A.S.K.2 | adenosquamöses Karzinom | P.K.16 | Plattenepithelkarzinom |
| A.S.K.3 | adenosquamöses Karzinom | P.K.17 | Plattenepithelkarzinom |
| G.K.1 | großzelliges Karzinom | P.K.18 | Plattenepithelkarzinom |
| K.K.1 | Karzinoid | P.K.19 | Plattenepithelkarzinom |

Abkürzungen: Adenokarzinom (A.K.), adenosquamöses Karzinom (A.S.K.), großzelliges Karzinom (G.K.), Karzinoid (K.K.), nicht-kleinzelliges Lungenkarzinom (N.K.), Plattenepithelkarzinom (P.K.)

2.1.2 Plasma-Proben

Die Plasma-Proben wurden im Rahmen des DZL Konsortiums von der "LungenClinic Grosshansdorf GmbH", (Deutschland) bereit gestellt. Alle IDs wurden nach einem Zufallsverfahren anonymisiert, sodass es nicht mehr möglich ist, einzelne Patienten zu identifizieren. Die Studie wurde von der Universität Lübeck befürwortet (Ref. 12-238). Für die WGBS sowie Entwicklung des Plasma *Panels* wurden 28 Proben genutzt (Tab. 2). Für die Validierung des *Panels* wurde Plasma von 12 Patienten untersucht (Tab. 3).

Tab. 2: Proben für WGBS sowie anschließende Entwicklung des Plasma *Panels*.

| Patienten ID | Entität | Stadium |
|--------------|--|---------|
| K.1 bis K.19 | Kontrollen (Patienten ohne maligne Erkrankungen) | - |
| A.K.16 | Adenokarzinom | IV |
| A.K.17 | Adenokarzinom | IV |
| A.K.18 | Adenokarzinom | IV |
| A.K.19 | Adenokarzinom | IV |
| A.K.20 | Adenokarzinom | IV |
| P.K.20 | Plattenepithelkarzinom | IIIA |
| P.K.21 | Plattenepithelkarzinom | IIIB |
| P.K.22 | Plattenepithelkarzinom | IIIB |
| P.K.23 | Plattenepithelkarzinom | IIIB |

Abkürzungen: Kontrolle (K.), Adenokarzinom (A.K), Plattenepithelkarzinom (P.K.)

Tab. 3: Proben für die Validierung des Plasma *Panels*.

| Patienten ID | Entität | Stadium |
|--------------|--|---------|
| K.20-K.22 | Kontrollen (Patienten ohne maligne Erkrankungen) | - |
| A.K.21 | Adenokarzinom | IV |
| A.K.22 | Adenokarzinom | IV |
| A.K.23 | Adenokarzinom | IIa |
| A.K.24 | Adenokarzinom | IIa |
| P.K.24 | Plattenepithelkarzinom | IV |
| P.K.25 | Plattenepithelkarzinom | IV |
| P.K.26 | Plattenepithelkarzinom | IV |
| P.K.27 | Plattenepithelkarzinom | IIa |
| P.K.28 | Plattenepithelkarzinom | IB |

Abkürzungen: Kontrolle (K.), Adenokarzinom (A.K), Plattenepithelkarzinom (P.K.)

2.1.3 Zellen

- A549 Zelllinie

Die Zellen wurden bei „DSMZ“ (Deutschland) bestellt, die Katalognummer lautete ACC 107. Bei den Zellen handelte es sich um eine artifizielle Tumorzelllinie, die 1972 aus einem Bronchialkarzinom etabliert wurde. Der genetische Fingerabdruck der Zelllinie wurde unter Verwendung einer Multiplex-PCR laut „DSMZ“-Angaben verifiziert [162].

- AECII Zellen (engl. *alveolar epithelial type II cells*)

Die Extraktion primärer Lungenepithelzellen aus dem humanen OP-Material erfolgte im „Institut für klinische und experimentelle Pathologie“ („Forschungszentrum Borstel“, Deutschland). Alle IDs wurden nach einem Zufallsverfahren anonymisiert, sodass es nicht mehr möglich ist, einzelne Patienten zu identifizieren. Die Studie wurde von der Ethikkommission der Universität Lübeck befürwortet (Ref. 17-350), die Spender sind in der Tab. 4 zusammengefasst.

Tab. 4: Primäre AECII Zellen wurden aus dem Lungengewebe folgender Patienten extrahiert.

| Assay | CSE | | eCSE (+N) | | eCSE (-N) | |
|-----------------------|-----------|-----------|-----------|-----------|-----------|-----------|
| | Versuch 1 | Versuch 2 | Versuch 1 | Versuch 2 | Versuch 1 | Versuch 2 |
| Zellviabilität | Sp.1 | Sp.3 | Sp.1 | Sp.4 | Sp.5 | Sp.7 |
| Apoptose | Sp.2 | Sp.3 | Sp.8 | Sp.2 | Sp.2 | Sp.6 |
| ATAC-Seq. | Sp.1 | Sp.3 | Sp.1 | Sp.4 | Sp.5 | Sp.7 |
| EPIC BC | Sp.1 | Sp.3 | Sp.1 | Sp.9 | Sp.5 | Sp.6 |
| RNA-Seq. | Sp.1 | Sp.3 | Sp.1 | Sp.9 | Sp.5 | Sp.7 |

Abkürzungen: Zigarettenextrakt (CSE), E-Zigarettenextrakt (eCSE), Nikotin (N), Spender (Sp.)

2.2 Reagenzien

| Bezeichnung | Hersteller |
|---|--|
| „3R4F – Kentucky University Reference Cigarettes“ | „University of Kentucky“, USA |
| „Accutase Cell Detachment Solution“ | „Merck Millipore“, USA |
| „ActinGreen™ 488 ReadyProbes®“ | „Thermo Fisher Scientific“, USA |
| „Alexa Fluor® 647 anti-H2A.X (Ser139) Phospho (Ser139) Antibody“ | „Biolegend“, USA |
| „ApoONE® Homogeneous Caspase-3/7 Assay“ | „Promega“, USA |
| „BIOCOLL“ | „Biochrome GmbH“, Deutschland |
| „Bovine Serum Albumin“ (BSA) | „Thermo Fisher Scientific“, USA |
| „CD45 MicroBeads, human“ | „Miltenyi Biotec GmbH“, Deutschland |
| „CellTiter-Blue® Cell Viability Assay“ | „Promega“, USA |
| „DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride)“ | „Thermo Fisher Scientific“, USA und „BioLegend“, USA |
| „DIPSE-eGo-Cigarette“: mit Nikotin (24 mg/ml) ohne Nikotin (0mg/ml) | „DIPSE“, Deutschland |
| „Dispase II (Neutral Protease, Grade II)“ | „Roche“, Schweiz |
| „DNase I Recombinant, RNase-free“ (DNase) | „Sigma-Aldrich“, USA |
| „DMEM–Dulbecco's Modified Eagle Medium from Gibco®“ (DMEM) | „Thermo Fisher Scientific“, USA |
| „EDTA, Disodium Salt, Dihydrate, Molecular Biology Grade“ (EDTA) | „Merck Millipore“, USA |
| „Heat Inactivated Fetal Bovine Serum from Gibco®“ (FBS) | „Thermo Fisher Scientific“, USA |
| „IncuCyte® Annexin V Reagent“ | „Essen Bioscience“, USA |
| „LD Columns“ | „Miltenyi Biotec GmbH“, Deutschland |
| „Millipak Filter (Porengröße 0.22µm)“ | „Merck Millipore“, USA |
| „Mounting Medium“ | „Thermo Fisher Scientific“, USA |
| Nylon-Filter (Porengröße 100µm, 50µm und 20µm) | „Merck Millipore“, USA |
| „PBS - Phosphate-Buffered Saline from Gibco®“ (PBS) | „Thermo Fisher Scientific“, USA |

| | |
|-------------------------------------|---------------------------------|
| „Perm Buffer III“ | „BD Biosciences“, USA |
| “Trypsin/EDTA Solution from Gibco®“ | „Thermo Fisher Scientific“, USA |
| „Vybrant® Dil Cell-Labeling“ | „Thermo Fisher Scientific“, USA |

2.3 Kits

| Bezeichnung | Hersteller |
|---|---------------------------------|
| „Accel-NGS® Methyl-Seq DNA Library Kit“ | „Swift Biosciences“, USA |
| „Agilent 2100 High Sensitivity DNA Kit“ | „Agilent“, USA |
| „EZ DNA Methylation Kit“ | „Zymo Research“, USA |
| „EZ DNA Methylation-Gold™ Kit“ | „Zymo Research“, USA |
| „Infinium MethylationEPIC BeadChip Kit“ | „Illumina“, USA |
| „Methyl-Seq Set A Indexing Kit“ | „Swift Biosciences“, USA |
| „MiSeq Reagent Kit v3 (150-cycle)“ | „Illumina“, USA |
| „NEBNext High-Fidelity 2x PCR Master Mix“ | „Life Technologies“, USA |
| „Nextera DNA Library Preparation Kit“ | „Illumina“, USA |
| „PME free-circulating DNA Extraction Kit“ | „Analytik Jena AG“, Deutschland |
| „PhiX Control v3“ | „Illumina“, USA |
| „Ribo-Zero Gold rRNA Removal Kit (Human/Mouse/Rat)“ | „Illumina“, USA |
| „RNeasy Mini Kit“ | „Qiagen“, Deutschland |
| „SeqCap Epi Enrichment Kit“ | „Roche“, Schweiz |
| „SeqCap Epi Accessory Kit“ | „Roche“, Schweiz |
| „SeqCap EZ Hybridization and Wash Kit“ | „Roche“, Schweiz |
| „SeqCap EZ HE-Oligo Kit A“ | „Roche“, Schweiz |
| „SeqCap EZ Pure Capture Bead Kit“ | „Roche“, Schweiz |
| „TruSeq Stranded mRNA Library Prep Kit“ | „Illumina“, USA |
| „Qiagen MinElute PCR Purification Kit“ | „Qiagen“, Deutschland |
| „Qubit dsDNA High Sensitivity Assay Kit“ | „Thermo Fisher Scientific“, USA |
| „QIAamp DNA Mini Kit“ | „Qiagen“, Deutschland |
| „QIAGEN Proteinase K“ | „Qiagen“, Deutschland |

2.4 Chemikalien

| Bezeichnung | Hersteller |
|-------------------------------------|--|
| 2-Propanol | „Sigma-Aldrich“, USA |
| „Ampuwa“ (dH ₂ O) | „Fresenius Kabi AG“, Deutschland |
| Di-Natriumhydrogenphosphat-Dihydrat | „Merck Millipore“, USA |
| Ethanol für die Molekularbiologie | „Merck Millipore“, USA |
| Glucose | „Merck Millipore“, USA |
| HEPES | „Merck Millipore“, USA |
| IGEPAL CA-630 | „Sigma-Aldrich“, USA |
| Kaliumchlorid | „Merck Millipore“, USA |
| Kalziumchlorid-Dihydrat | „Merck Millipore“, USA |
| 2-Mercaptoethanol | „Sigma-Aldrich“, USA |
| Magnesiumchlorid | „Sigma-Aldrich“, USA |
| Magnesiumsulfat-Heptahydrat | „Merck Millipore“, USA |
| Natriumchlorid | „Sigma-Aldrich“, USA |
| Natronlauge (1N) | „Merck KGaA“, Deutschland |
| Natriumazid | „Carl Roth GmbH & Co. KG“, Deutschland |
| Paraformaldehyd | „Carl Roth GmbH & Co. KG“, Deutschland |
| Tris Hydrochlorid | „Sigma-Aldrich“, USA |
| Triton X-100 | „Carl Roth GmbH & Co. KG“, Deutschland |

2.5 Beads

| Bezeichnung | Hersteller |
|-----------------------|------------------------|
| „Agencourt AMPure XP“ | „Beckman Coulter“, USA |
| „SPRI Beads“ | „Beckman Coulter“, USA |

2.6 Verbrauchsmaterialien

| Bezeichnung | Hersteller |
|-------------------------------------|---|
| 6-Well-Platte | „Corning“, USA |
| 8er Deckelkette, flach | „Sarstedt AG&Co“, Deutschland |
| 96-Well-Platte | „Corning“, USA |
| 96-Well-Lagerungsplatte (0,8ml) | „Thermo Fisher Scientific“, USA |
| Deckgläser | „VWR“, USA |
| „CryoPure“ Gefäß (1.8ml) | „Sarstedt AG&Co“, Deutschland |
| Dichtungsmatten für 96-Well-Platten | „Thermo Fisher Scientific“, USA |
| Fluoreszenzfilter | „Fa. AHF Analysentechnik AG“, Deutschland |
| „Multiply® -µStrip“ (0,2µl) | „Sarstedt AG&Co“, Deutschland |
| „QIAshredder Spin Column“ | „Qiagen“, Deutschland |
| Petrischale (92 x16mm) | „Sarstedt AG&Co“, Deutschland |
| Reaktionsgefäße (1,5ml, 15ml, 50ml) | „Sarstedt AG&Co“, Deutschland |

2.7 Instrumente und Geräte

| Bezeichnung | Hersteller |
|---|---------------------------------|
| „Agilent 2100 Bionalyzer“ | „Agilent“, USA |
| „Agilent Priming Station“ | „Agilent“, USA |
| Biologische Sicherheits- und Reinluftwerkbank | „Thermo Fisher Scientific“, USA |
| Brutschrank „Galaxy B“ | „Eppendorf AG“, Deutschland |
| DNA/RNA UV-Cleaner box UVC/T-M-AR | „Kisker GbR“, Deutschland |
| „DynaMag™-2“ Magnet | „Thermo Fisher Scientific“, USA |
| „DynaMag™-PCR“ Magnet | „Thermo Fisher Scientific“, USA |
| „Heraeus Megafuge 40R“ | „Thermo Fisher Scientific“, USA |
| „Heraeus Megafuge 1 L-R“ | „Thermo Fisher Scientific“, USA |
| „Heraeus Multifuge X1R“ | „Thermo Fisher Scientific“, USA |
| „High-Speed Microplate Shaker“ | „Illumina“, USA |
| „HiSeq4000“ | „Illumina“, USA |

| | |
|--|--|
| Hybridisierungssofen | „Illumina“, USA |
| „IKA® MS3 Vortexer“ | „IKA®-Werke GmbH & Co. KG“, Deutschland |
| „Illumia iScan“ | „Illumina“, USA |
| „IncuCyte ZOOM Live Cell Analysis System“ | „Essen Bioscience“, USA |
| „MiSeq“ | „Illumina“, USA |
| „NextSeq500“ | „Illumina“, USA |
| Plattenschüttler „MS3 Basic“ | „IKA®-Werke GmbH & Co. KG“, Deutschland |
| „Photozoom Inverted Microscope“ | „Cambridge Instruments“, England |
| „Qubit 2.0“ Fluorometer | „Thermo Fisher Scientific“, USA |
| „Rack Chamber Te-Flow“ | „Tecan Schweiz AG“, Schweiz |
| „Reax 2000“ Reagenzschüttler | „Heidolph Instruments GmbH & Co. KG“, Deutschland |
| „Sprout“ Minizentrifuge | „BIOZYM Gesellschaft für Enzymtechnologie mbH“, Deutschland |
| „ Synergy HT Multi-Mode Microplate Reader“ | „BioTek Instruments, Inc.“, USA |
| „Thermoblock Thermomixer Comfort“ | „Eppendorf AG“, Deutschland |
| <i>Thermocycler</i> „T3000“ | „Biometra GmbH“, Deutschland |
| Vakuumkonzentrator „Jouan RC10-10“ | „Thermo Fisher Scientific“, USA |
| Vakuumpumpe | „Thermo Fisher Scientific“, USA |
| Wasserbad | „Mettler GmbH“, Deutschland |
| „Zeiss–Axioskop“ | „Zeiss AG“, Deutschland |

2.8 Software

| Software | Hersteller |
|--|--|
| „BamUtil“ (Version 1.0.13) | „University of Michigan School of Public Health“, USA |
| „Bedtools“ (Version 2.2.6) | „The University of Utah“, USA |
| „Bisulfite Analysis Toolkit“ (Version 0.1) | „Interdisziplinäres Zentrum für Bioinformatik, Universität Leipzig“, Deutschland |
| „Cutadapt“ (Version 1.9.1) | „TU Dortmund“, Deutschland |
| „DoDas“ (Version 1.0) | „Interdisziplinäres Zentrum für Bioinformatik, Universität Leipzig“, Deutschland |
| „edgeR“ (Version 3.18.1) | „Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research“, Australien |
| „FastQC“ (Version 0.11.15) | „Babraham Institute“, England |
| „Illumina® GenomeStudio“ (Version 2.0.3) | „Illumina“, USA |
| „IncuCyte® S3 Base Software“ | „Essen Bioscience“, USA |
| „Isis“ | „MetaSystems GmbH“, Deutschland |
| „Java“ (Version 1.8.0) | „Oracle“, USA |
| „MACS2“ (Version 2.1.1) | „Python Software Foundation“, USA |
| „MobaXterm Personal Edition“ | „Mobatek“, Frankreich |
| „Python“ (Version 3.3.6) | „Python Software Foundation“, USA |
| „R“ (Version 3.2.0) | „R Foundation for Statistical Computing“, Österreich |
| „Perl“ (Version 5.26.0) | „The Perl Foundation“, USA |
| „Picard“ (Version 2.13.2) | „Broad Institute of MIT and Harvard“, USA |
| „Samtools“ (Version 1.3.1) | „Wellcome Trust Sanger Institute“, England und „Broad Institute of MIT and Harvard“, USA |
| „StrandNGS“ (Version 3.1) | „Avadis®“, USA |
| „Qlucore Omics Explorer“ (Version 3.2) | „Qlucore“, Schweden |
| „Segemehl“ (Version 0.2.0) | „Interdisziplinäres Zentrum für Bioinformatik, Universität Leipzig“, Deutschland |
| „WebGestalt“ | „Graduate School in Genome Science and Technology“, USA und „University of Tennessee-Oak Ridge National Laboratory“, USA |

3. Methoden

3. 1 Methoden im Projekt „Entwicklung des Plasma *Panel*s“

Um eine nichtinvasive Lungenkrebsdiagnostik zu ermöglichen, sollte im Rahmen dieser Arbeit ein geeignetes *Panel* zur DNA-Methylierungsanalyse im Blutplasma entwickelt werden. Die Entwicklung des Plasma *Panel*s erfolgte in drei voneinander unabhängigen Ansätzen (Abb. 5).

Im ersten Ansatz sollte zunächst geprüft werden, ob die DNA-Methylierung sich generell als Biomarker eignet (siehe Abschnitt 3.1.1). Hierfür wurden im Rahmen einer DZL-Studie 40 Lungenkarzinome sowie deren korrespondierenden Kontrollen unter Verwendung des „Illumina Infinium Human Methylation450K BeadChips“ (HM 450K) analysiert. Die Methode identifizierte deutliche, tumorspezifische DNA-Methylierungssignaturen. Die Datensätze wurden für die vorliegende Doktorarbeit zur Verfügung gestellt. Als nächstes wurden, wie im Abschnitt 3.1.1 beschrieben, die Regionen mit den stärksten DNA-Methylierungsunterschieden ermittelt und in das Plasma *Panel* aufgenommen.

Im zweiten Ansatz wurde untersucht, ob tumorspezifische DNA-Methylierungssignaturen auch im Blutplasma der betroffenen Patienten detektiert werden können (siehe Abschnitt 3.1.2). Dafür wurde zirkulierende, zellfreie DNA aus dem Plasma von Adeno- (n=5) und Plattenepithelkarzinompatienten (n=4) extrahiert und anschließend *gepoolt*. Als Kontrolle diente Plasma einer tumorfreien Patientenkohorte (n=19). Die detaillierten Informationen zu den Patienten sind im Abschnitt 2.1.2 zusammengestellt. Durch das *Poolen* wurden individuelle DNA-Methylierungsmuster weitgehend eliminiert, dagegen die allgemeinen tumor- bzw. lungenspezifischen Signaturen hervorgehoben. Dann wurden die zfDNA-*Pools* einer genomweiten Bisulfitsequenzierung unterzogen (engl. *whole genome bisulfite sequencing* (WGBS), siehe Abschnitt 3.1.2.4). Die Methode detektierte mehrere tausend aberrant methylierte CpG *Loci*, die nicht nur tumor- sondern auch entitätsspezifisch waren. Davon wurden die geeignetsten Regionen für die Differenzierung für das Plasma *Panel* ausgewählt (siehe Abschnitt 3.1.2.5.5).

Im dritten Ansatz wurde das Plasma *Panel* mit 59 tumorspezifischen sowie prognostisch relevanten CpG *Loci* aus weiteren Studien ergänzt (siehe Abschnitt 3.1.3).

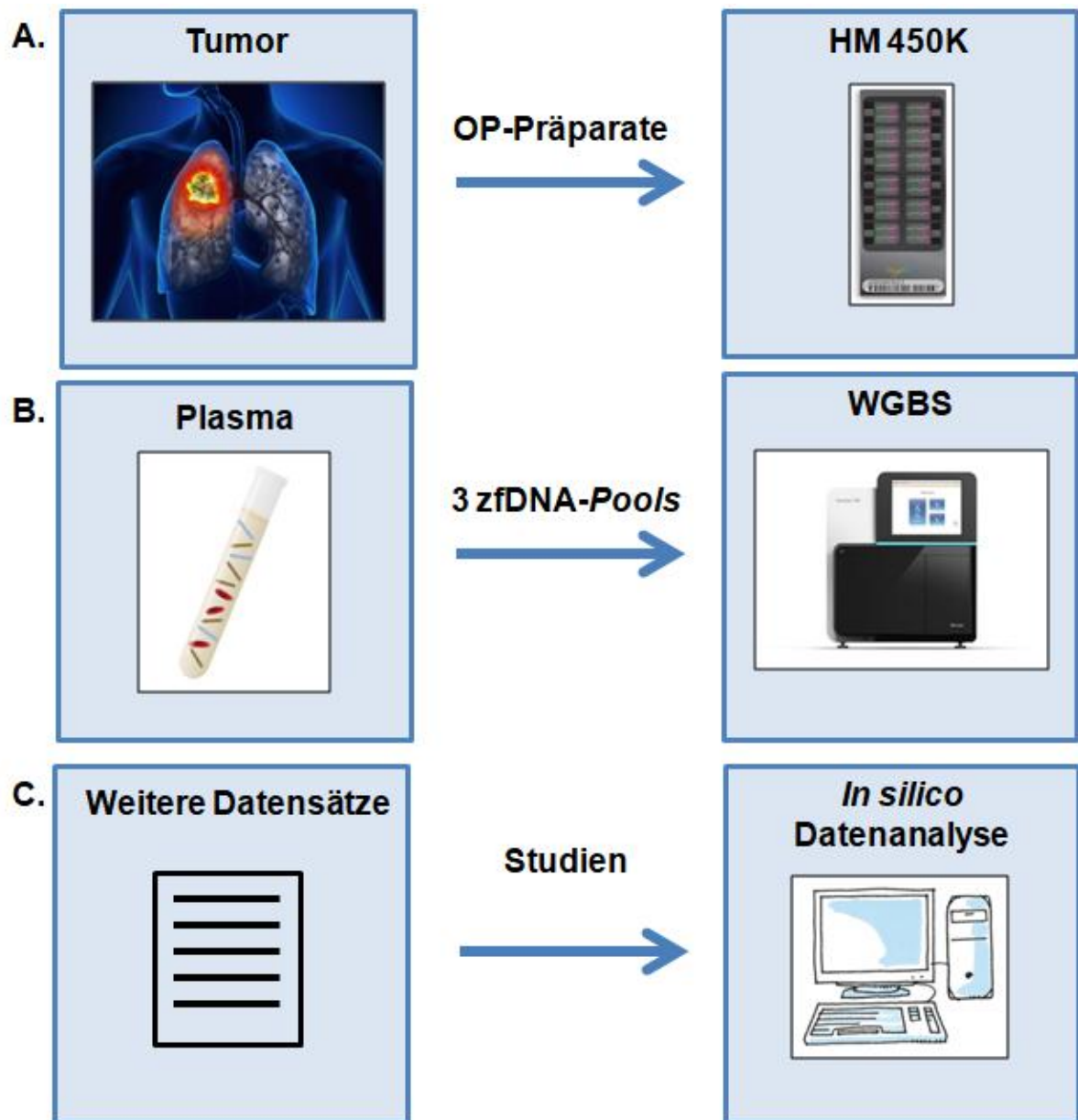


Abb. 5: Entwicklung des Plasma *Panel*s; **A.** Zunächst wurden unter Verwendung des „Illumina Infinium Human Methylation450K BeadChips“ (HM 450K) 40 OP-Präparate analysiert (Tumorgewebeproben und deren korrespondierenden Kontrollen). **B.** Dann wurde WGBS von 3 zfDNA Plasma-*Pools* (tumorfremie Kohorte (n=19), Patienten mit Adenokarzinom (n=5) und Plattenepithelkarzinom (n=4)) durchgeführt. **C.** Schließlich wurde das *Panel* mit differentiell methylierten Regionen aus weiteren Studien ergänzt. Die Abbildung wurde verändert nach [163-167].

3.1.1 Nachweis der aberranten DNA-Methylierung in primärem Tumorgewebe

Die DNA-Methylierungsdaten der HM 450K Analysen wurden im Rahmen einer DZL Studie erhoben und für diese Arbeit zur Verfügung gestellt.

Der HM 450K Datensatz enthielt Informationen über den Methylierungsstatus von 40 Lungenkarzinomen (Adeno- und Plattenepithelkarzinome) und deren korrespondierenden Kontrollen. Der Datensatz wurde mit der „Qlucore Omics Explorer“ Software [168] (Version 3.2, „Qlucore“, Schweden) ausgewertet und ergab:

- 1.) 897 CpG *Loci* (T-Test: $FDR < 1 \times 10^{-23}$, $\sigma/\sigma_{\max} > 0,4$), die zwischen dem Tumor- und gesundem Lungengewebe differenziert methyliert waren.
- 2.) 1.167 CpG *Loci* (T-Test: $FDR < 1 \times 10^{-4}$), die zwischen dem Adeno- und Plattenepithelkarzinomgewebe differenzierten.

Um die CpG *Loci* mit den stärksten DNA-Methylierungsunterschieden zu ermitteln, wurden die beiden Listen zunächst nach differentieller Methylierung größer 35% ($\text{avg.}\beta > 0,35$) gefiltert und unter Verwendung von „Bedtools“ [169] (Version 2.2.6, „The University of Utah“, USA) gegen das „HG19“ Referenzgenom annotiert. Alle CpG *Loci*, die innerhalb von häufigen SNPs ($\geq 1\%$ der Bevölkerung) lokalisiert waren und nicht proteinkodierend waren, wurden verworfen. Die übrig gebliebenen *Loci* wurden in das finale Plasma *Panel* aufgenommen (Abschnitt 9.3, Tab. 7).

3.1.2 Nachweis der aberranten DNA-Methylierung in Blutplasma

Die zirkulierende, zellfreie DNA ist zurzeit eine der vielversprechendsten Kandidaten in der nichtinvasiven Diagnostik von soliden Tumoren. Falls ein Patient an einer malignen Tumorerkrankung leidet, ist in der Gesamtmenge der zirkulierenden DNA auch die TumordNA enthalten, welche alle therapeutisch und prognostisch relevanten Informationen über die genetischen und epigenetischen Charakteristika des Tumors enthält [122-124, 170]. Da sie nur in einer sehr geringen Menge aus dem Blutplasma extrahiert werden kann, ist es umso wichtiger, eine Methode zu wählen, die sehr spezifisch und effizient die zfDNA anreichert, ohne dabei weitere Bestandteile des Plasmas zu isolieren [171].

Das in dieser Arbeit verwendete „PME free-circulating DNA Extraction Kit“ („Analytik Jena“, Deutschland, siehe Abschnitt 3.1.2.1) enthält ein Polymer, welches nur sehr spezifisch kurzsträngige dsDNA-Fragmente komplexiert. Der Polymer-zfDNA-Komplex wird anschließend ausgefällt und aufgereinigt. Nach der Aufreinigung kann die Komplexverbindung gelöst werden, die dabei freigegebene DNA wird in weiteren Schritten

vom Polymer gereinigt und aufkonzentriert. Das dabei entstehende Produkt ist sehr sauber und kann auch für empfindliche NGS-basierte Analysemethoden wie z.B. WGBS verwendet werden [172].

3.1.2.1 Extraktion der zirkulierenden, zellfreien DNA (zfDNA) aus Blutplasma

Das Blutplasma wurde in der „LungenClinic Grosshansdorf GmbH“ (Deutschland) präpariert und auf Trockeneis verschickt. Hierfür wurde das Vollblut innerhalb von 30min nach der Entnahme bei 1.500g 10min lang zentrifugiert. Nach der Zentrifugation wurde der Plasmaüberstand vorsichtig abpipettiert, auf „CryoPure“ Gefäße („Sarstedt AG&Co“, Deutschland) verteilt und sofort bei -80°C eingefroren.

Die eingefrorenen Plasmaproben wurden langsam unter lauwarmem Wasser aufgetaut und anschließend bei 4.500g 10min lang zentrifugiert. Das Pellet wurde verworfen, der klare Überstand in ein 10ml Röhrchen überführt und mit dem „PME free-circulating DNA Extraction Kit“ prozessiert. Die im Kit enthaltenen Arbeitsschritte sind in der Abb. 6 dargestellt.

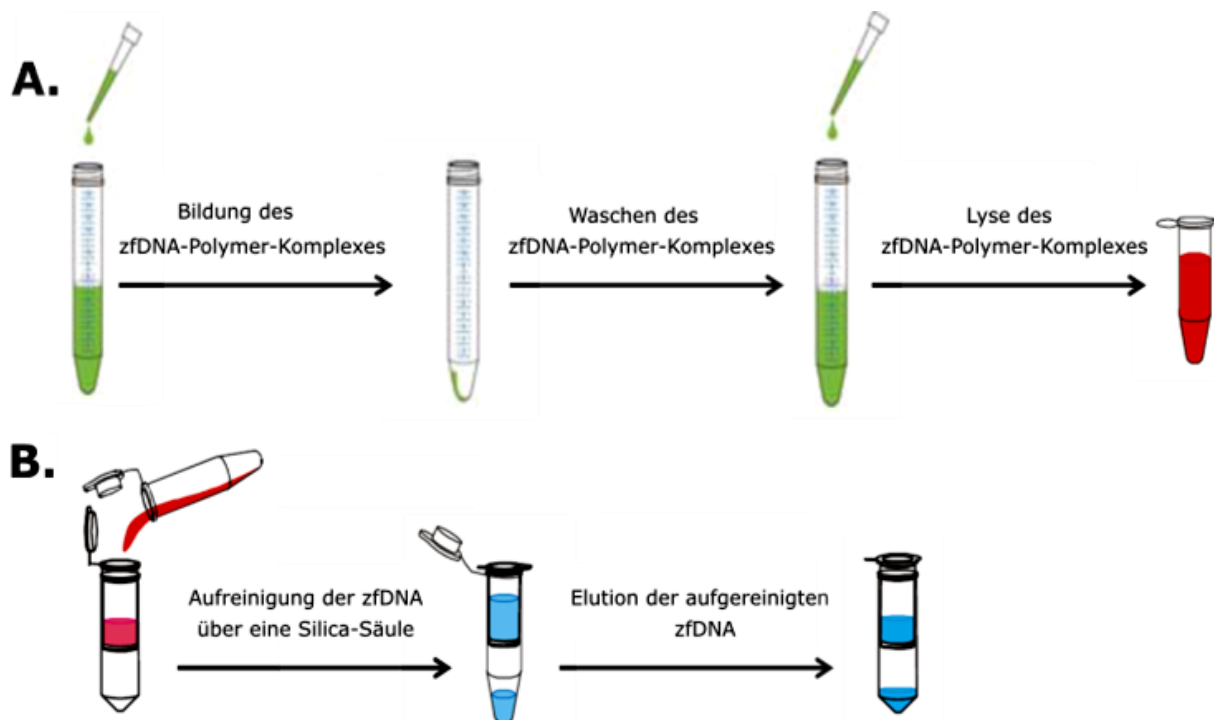


Abb. 6: Extraktion der zfDNA wurde unter Verwendung des „PME free-circulating DNA Extraction Kit“ durchgeführt. **A.** Dabei wurde die zfDNA an einen Polymer-Komplex gebunden und **B.** anschließend über eine Silica-Säule aufgereinigt. Die Abbildung wurde verändert nach [172].

Zunächst wurde der Polymer-zfDNA-Komplex erzeugt. Hierfür wurde die Plasmaprobe mit 100µl proprietärer „VCR-1“ und 600µl „VCR-2“ versetzt, kurz gemischt und 10min bei Raumtemperatur inkubiert. Der gebildete zfDNA-Polymer-Komplex wurde dann bei 4.500g pelletiert, im weiteren Verlauf in 5ml Wasser resuspendiert und bei 4.500g erneut pelletiert. Um die gefällte zfDNA freizusetzen, musste der Komplex im nächsten Schritt wieder aufgelöst werden. Hierfür wurde die Probe zunächst mit 400µl „SE Lysepuffer“ und 20µl „Proteinkinase A“ 15min bei 70°C inkubiert. Als nächstes erfolgte eine Aufreinigung der freigesetzten zfDNA. Hierfür wurde die Probe mit 200µl „SBS Puffer“ versetzt, gemischt, auf eine Silica-Säule gegeben und 1min bei 11.000g zentrifugiert. Die gebundene zfDNA wurde nun bei 11.000g mit 500µl „Puffer C“ und 650µl „Puffer BS“ gewaschen. Die Pufferreste wurden bei einem zusätzlichen Zentrifugationsschritt entfernt und die zfDNA anschließend in 30µl dH₂O eluiert.

3.1.2.2 Quantifizierung und Qualitätskontrolle (QC) der extrahierten zfDNA

Die zfDNA wurde fluorometrisch unter Verwendung des „Qubit dsDNA High Sensitivity Assay Kit“ („Thermo Fisher Scientific“, USA) quantifiziert. Hierfür wurde jeweils 1µl der Probe mit den 198µl „Qubit dsDNA HS Buffer“ sowie 1µl „Qubit dsDNA HS Reagent“ vermischt, 2min lang inkubiert und anschließend im „Qubit 2.0“ Fluorometer („Thermo Fisher Scientific“, USA) vermessen. Bei dem „Qubit dsDNA HS Reagent“ handelte es sich um einen Farbstoff, der unter normalen Bedingungen ein sehr schwaches Fluoreszenzsignal erzeugt. Im Vorhandensein von doppelsträngiger DNA (dsDNA) interkaliert es, verändert seine Struktur und erzeugt ein starkes Fluoreszenzsignal. Dabei wird weder einzelsträngige DNA (ssDNA) noch RNA gebunden. Somit korreliert die Signalintensität ausschließlich mit der in der Probe vorhandenen Menge an dsDNA [173, 174].

Die Qualität der extrahierten zfDNA wurde mit Hilfe des „Agilent 2100 High Sensitivity DNA Kit“ („Agilent“, USA) analysiert. Bei der Methode handelte es sich um eine Kapillar-Gelelektrophorese [175]. Zunächst musste das „Gel-Dye Mix“ vorbereitet werden. Dabei wurden 300µl der Gelmatrix mit 15µl des Farbstoffkonzentrats versetzt, gemischt und auf einen „Spin Filter“ gegeben. Die Zentrifugation erfolgte 10min lang bei 2.240g. Als nächstes wurde der DNA-Chip in der „Priming Station“ platziert und äquilibriert. Dafür wurden 9µl des „Gel-Dye Mix“ in das für den Äquilibriervorgang vorgesehene *Well* pipettiert. Der Stempel der „Priming Station“ wurde auf einen Milliliter justiert. Nachdem die „Priming Station“ fest verschlossen war, wurde der Stempel eine Minute lang heruntergedrückt. Schließlich wurden die restlichen *Wells* des Chips nach Angabe des Herstellers beladen. Der Chip wurde 1min inkubiert und direkt danach gemessen. Während der Inkubationszeit interkalierte ein im „Gel-

Dye Mix“ enthaltener fluoreszierender Farbstoff zwischen den Basen der dsDNA. Die dsDNA Fragmente wurden anschließend durch die mikroskopisch kleinen Kapillaren des „Agilent 2100 Bionalyzer“ („Agilent“, USA) gezogen, dabei je nach Fragmentgröße aufgetrennt und detektiert.

3.1.2.3 Bisulfitkonvertierung der zfdNA

Für die genomweite Analysen des DNA-Methylierungsmusters z.B. durch den HM 450K oder die WGBS, wird DNA einer genomweiten PCR-basierten Amplifikation unterzogen. Die DNA-Polymerasen können nicht zwischen den Cytosinen und 5-Methylcytosinen unterscheiden. Während der Reaktion werden alle 5-Methylcytosine durch Cytosine ersetzt. Die neu synthetisierten Stränge werden nicht erneut methyliert [176].

Um Cytosine von 5-Methylcytosinen unterscheiden zu können, wird die Probe vor der PCR einer Behandlung mit Natriumbisulfit unterzogen. Dieser Prozess wird als Bisulfitkonvertierung bezeichnet, dabei werden alle unmethylierten Cytosine in Uracile umgewandelt. Die methylierten Cytosine bleiben dagegen unter den gewählten Reaktionsbedingungen unverändert [177, 178]. Die Reaktion der Bisulfitkonvertierung ist in der Abb. 7 dargestellt.

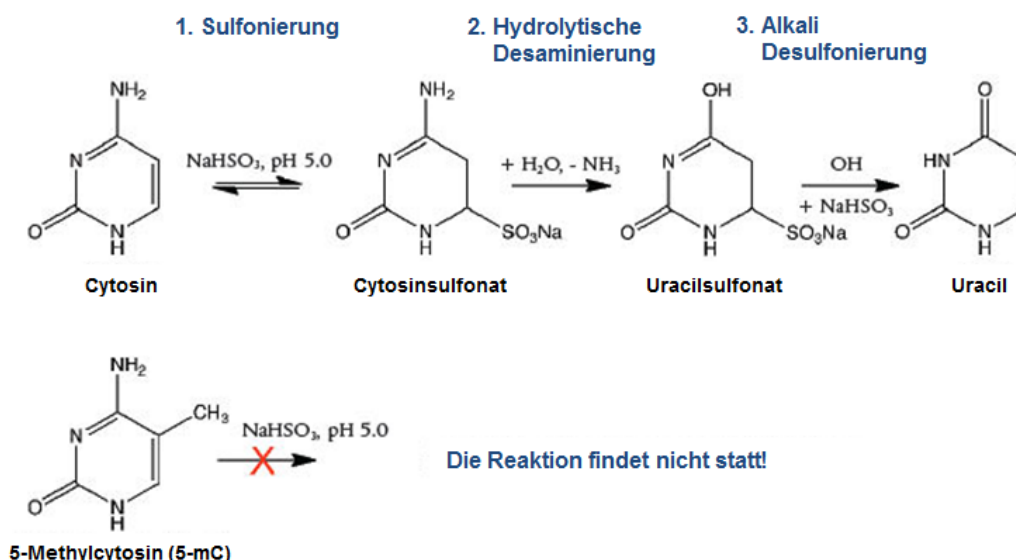


Abb. 7: Bisulfitkonvertierung läuft in drei Schritten ab. Zuerst erfolgt die Sulfonierungsreaktion der unmethylierten Cytosine, bei der das Cytosinsulfonat entsteht. Im zweiten Schritt wird das Cytosinsulfonat zu Uracilsulfonat desaminiert. Im dritten Schritt wird das Uracilsulfonat desulfoniert, was zur Entstehung von Uracil führt. Im Falle von 5-Methylcytosin, kann der erste Schritt, die Sulfonierung, nicht erfolgen. Deswegen läuft die Bisulfitkonvertierung nicht ab, die methylierte Cytosine bleiben somit unverändert. Die Abbildung wurde verändert nach [179].

Zuerst erfolgt die Sulfonierungsreaktion von Cytosin zu Cytosinsulfonat. Dabei werden die Bisulfitionen an die 5-6 Doppelbindung des Cytosins addiert. Dann wird das entstandene Cytosinsulfonat hydrolytisch zu einem Uracilsulfonat desaminiert. Schließlich erfolgt eine Alkalibehandlung, bei der die Sulfonylgruppe vom Uracilsulfonat abgespalten wird. Das wiederum führt zur Bildung des Uracils [180].

Die Bisulfitkonvertierung der zfDNA erfolgte mit dem „EZ DNA Methylation-Gold™ Kit“ („Zymo Research“, USA). Dafür wurden 10ng der zuvor extrahierten zfDNA in 20µl Wasser gelöst, mit 130µl „CT“-Konvertierungsreagenz versetzt und im *Thermocycler* bei folgendem Programm prozessiert: 10min 98°C, 2,5h 64°C, bis zu 20h bei 4°C. Im nächsten Schritt wurden die bisulfitkonvertierten Proben desulfoniert und aufgereinigt. Hierfür wurden sie mit 600µl „M-Binding Buffer“ versetzt, auf die „Zymo-Spin™ IC“ Säulen pipettiert und bei 10.000g 30s lang zentrifugiert. Dann wurden 100µl „M-Wash Buffer“ auf die Säulen gegeben. Die Säulen wurden 30s lang bei 10.000g zentrifugiert und für 20min mit 200µl „M-Desulphonation Buffer“ behandelt. Nach anschließender 30s langer Zentrifugation bei 10.000g wurden die „Zymo-Spin™ IC“ Säulen mit 200µl „M-Wash Buffer“ gewaschen, zur Entfernung verbliebener Flüssigkeiten für 30s bei 10.000g zentrifugiert und die DNA mit 15µl „Elution Buffer“ bei 10.000g 30s lang eluiert.

3.1.2.4 Whole Genome Bisulfite Sequencing (WGBS)

Um das zfDNA-Methylierungsprofil genomweit analysieren zu können, wurden die zuvor bisulfitkonvertierten Proben einer WGBS unterzogen. WGBS ist eine NGS-basierte Methode (engl. *next generation sequencing*). Heutzutage existieren zahlreiche Technologien, die NGS ermöglichen [181, 182]. Jedoch die am meisten verbreitete und auch hier verwendete NGS Technologie bietet die Firma „Illumina“ (USA) an (Abb. 8). Die zugrunde liegende Sequenzierreaktion ist fluoreszenzbasiert und erfolgt auf einem Glasträger, auch *Flowcell* genannt. Um die DNA-Fragmente an der *Flowcell* zu immobilisieren, werden zunächst spezielle „Illumina“ Adapter (kurze Oligonukleotide) ligiert. Nachfolgend wird die Probe einer Denaturierungsreaktion unterzogen. Da sich auf der *Flowcell* nicht nur die Adapterbindestellen, sondern auch *Primer* befinden, kommt es zum "Umknicken" des zu sequenzierenden ssDNA-Fragments. Während der nachfolgenden PCR-Reaktion werden die DNA-Stränge vervielfältigt. Dieser Vorgang wird als *Bridge Amplification* bezeichnet. Dabei entstehen durch die fortschreitende Amplifikation an begrenzten Positionen die sogenannten Sequenziercluster, die nachfolgend dissoziieren. Nach der Clusterbildung erfolgt die eigentliche Sequenzierreaktion, bei der DNA-Basen eingebaut werden, die je nach eingebauter Base Fluoreszenzsignale unterschiedlicher Wellenlängen erzeugen. Nach

jedem abgeschlossenem Einbauzyklus werden diese Fluoreszenzsignale detektiert und liefern somit die Informationen über die Basenabfolge innerhalb eines *Reads* [183].

Je nach gewünschtem Durchsatz, können unterschiedliche „Illumina“ Plattformen genutzt werden. Für die Sequenzierung sehr kleiner bakterieller bzw. viraler Genome sowie gezielter Regionen, sogenannten *Panels*, ist die schnellere und preisgünstigere „MiSeq“ Plattform ausreichend. Ist die Fragestellung jedoch genomweit, dann sollte die Sequenzierung auf den „NextSeq 500“ oder „HiSeq“ Plattformen erfolgen [184].

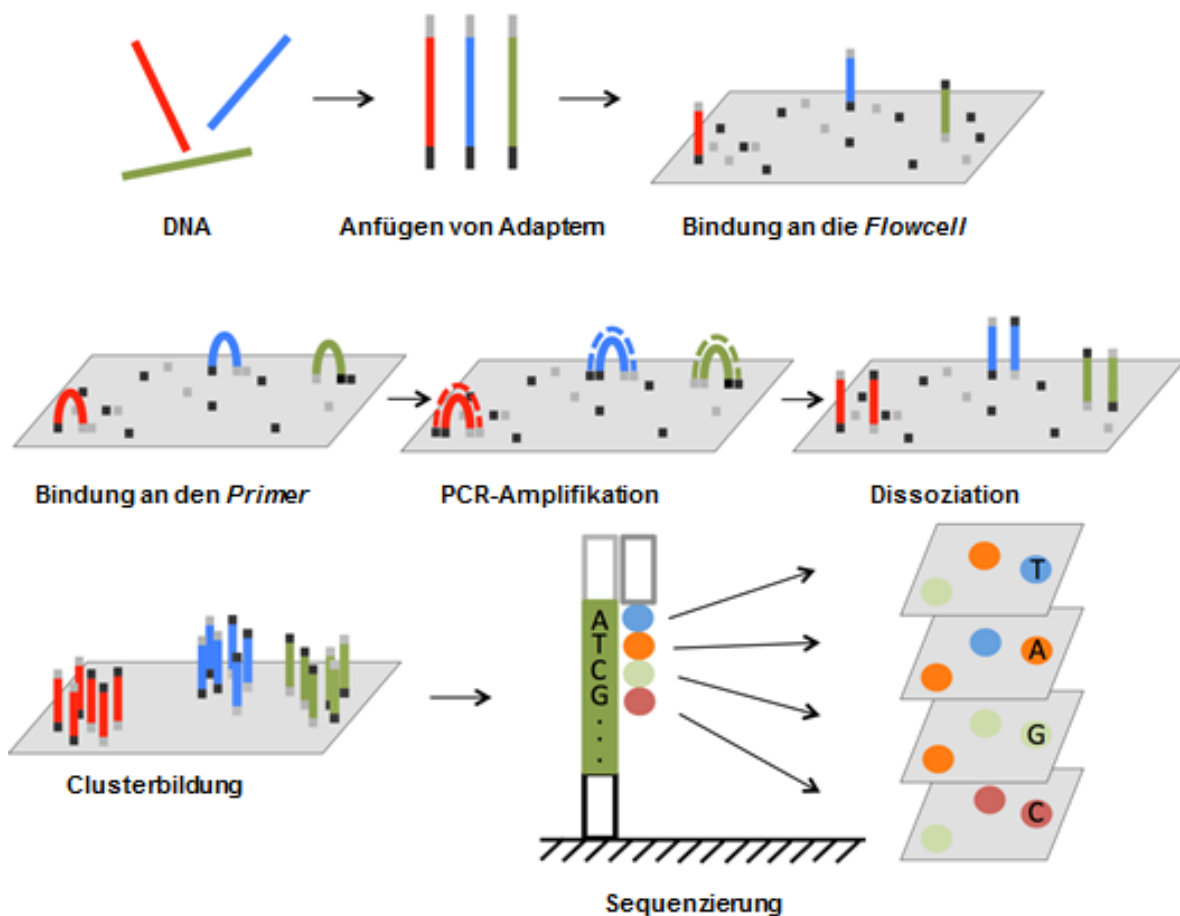


Abb. 8: Das Prinzip des NGS von Firma „Illumina“ (USA); An die DNA-Fragmente werden zunächst „Illumina“ Adapter ligiert. Dann wird die Probe denaturiert und an der *Flowcell* immobilisiert. Die nachfolgende PCR-Amplifikation führt zur Clusterbildung. Schließlich erfolgt die fluoreszenzbasierte Sequenzierreaktion, verändert nach [185].

3.1.2.4.1 Erstellen der WGBS *Libraries*

Während der Bisulfitkonvertierung wird DNA durch die verwendeten Reagenzien (siehe Abschnitt 3.1.2.3) sehr stark beansprucht und somit zu einem hohen Anteil degradiert [186]. Deswegen verwenden die herkömmlichen WGBS Protokolle sehr hohe Mengen an DNA, mindestens 500ng [187]. Da die zellfreie, zirkulierende DNA zum einen von Anfang an schon sehr stark fragmentiert ist und zum anderen nur in einer sehr geringen Menge gewonnen werden kann, ist die Herstellung von WGBS *Libraries* mit herkömmlichen Kits zurzeit unmöglich [121, 188].

Deshalb wurde für folgende Experimente das „Accel-NGS[®] Methyl-Seq DNA Library Kit“ („Swift Biosciences“, USA) etabliert. Das Kit wurde speziell für WGBS der zfdDNA entwickelt. Bereits bei zfdDNA-Mengen von weniger als 10ng können damit komplexe WGBS *Libraries* generiert werden. Die zentrale Rolle spielt dabei das Enzym „Adaptase“, welches einen 10nt langen Überhang am 3'-Ende der bisulfitkonvertierten ssDNA anfügt. Dieser Überhang in Verbindung mit patentierter Chemie ermöglicht ein besseres Ligieren der Sequenzieradapter und somit eine effizientere *Library* Herstellung (Abb. 9) [189, 190].

Wie in der Abb. 9 dargestellt wurde die *Library* Herstellung in vier Schritten durchgeführt: Behandlung mit dem Enzym „Adaptase“, Extension, Ligation, PCR.

Für die Behandlung mit dem Enzym „Adaptase“ wurden 10ng bisulfitkonvertierter zfdDNA in 15µl Wasser aufgenommen und bei 95°C für 2min denaturiert. Dann wurden zu der Probe 25µl des „Adaptase Reaction Mix“ gegeben, vorsichtig gemischt und im *Thermocycler* prozessiert (Programm 1, Abb. 10).

Als nächstes erfolgte die Extension, hierfür wurde die Probe mit 44µl „Extension Reaction Mix“ versetzt, vorsichtig gemischt und im *Thermocycler* inkubiert (Programm 2, Abb. 10). Das Produkt wurde mit „SPRI Beads“ („Beckman Coulter“, USA) aufgereinigt (Tab. 5).

Danach erfolgte die Ligation, für die 15µl des Produkts mit 15µl „Ligation I Reaction Mix“ versetzt und im *Thermocycler* prozessiert wurden (Programm 3, Abb. 10). Auch bei diesem Schritt wurde das fertige Produkt unter Verwendung von „SPRI Beads“ („Beckman Coulter“, USA) aufgereinigt (Tab. 5).

Schließlich wurde die PCR durchgeführt. Dabei wurden pro Probe 5µl des jeweiligen Index sowie 25µl des „Indexing PCR Reaction Mix“ hinzugegeben. Die fertige PCR-Reaktion wurde im *Thermocycler* inkubiert (Programm 4, Abb. 10) und über die „SPRI Beads“ („Beckman Coulter“, USA) aufgereinigt (Tab. 5).

Die fertigen WGBS *Libraries* wurden wie im Abschnitt 3.1.2.2 beschrieben quantifiziert und auf ihre Qualität überprüft.

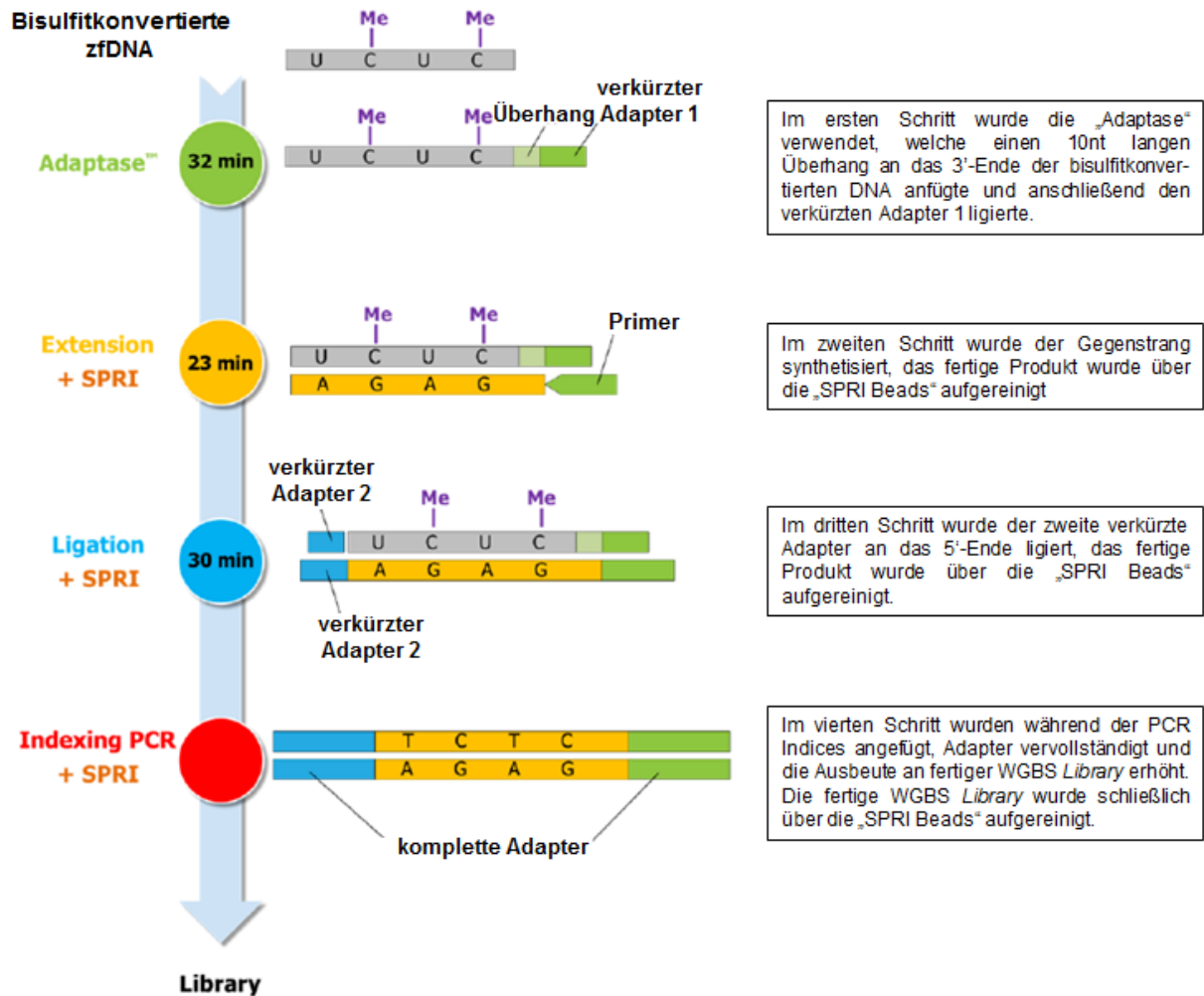


Abb. 9: Die WGBS *Libraries* der konvertierten zfDNA-Proben wurden mit dem „Accel-NGS® Methyl-Seq DNA Library Kit“ („Swift Biosciences“, USA) erstellt. Dabei wurden während der ersten drei Schritte verkürzte Adapter an die bisulfitkonvertierte zfDNA angefügt, welche in einem anschließenden PCR-Schritt vervollständigt wurden. Die Abbildung wurde verändert nach [191].

Programm 1:

| Temperatur | Zeit |
|------------|-------|
| 37°C | 15min |
| 95°C | 2min |
| 4°C | ∞ |

Programm 3:

| Temperatur | Zeit |
|------------|-------|
| 25°C | 15min |
| 4°C | ∞ |

Programm 2:

| Temperatur | Zeit |
|------------|------|
| 98°C | 1min |
| 62°C | 2min |
| 65°C | 5min |
| 4°C | ∞ |

Programm 4:

| Temperatur | Zeit |
|--------------|------|
| 98°C | 30s |
| PCR-Zyklen: | |
| 98°C | 10s |
| 60°C | 30s |
| 68°C | 1min |
| → 7-9 Zyklen | |
| 4°C | ∞ |

Abb. 10: Die WGBS *Libraries* wurden unter Verwendung der Programme 1 bis 4 hergestellt, der Deckel des *Thermocyclers* war dabei vorgeheizt.

„SPRI Beads“ Aufreinigung

Die Proben wurden in 1,5ml Eppendorf Reaktionsgefäße überführt und im vorgeschriebenen Verhältnis mit „SPRI Beads“ („Beckman Coulter“, USA) versetzt (Tab. 5). Dann wurden die Proben gemischt und 5min bei Raumtemperatur inkubiert. Da die *Beads* magnetisch waren, konnte zum Pelletieren das Prinzip der magnetischen Separation genutzt werden. Hierfür wurden die Reaktionsgefäße auf einem magnetischen Ständer platziert und dann 2min lang bei Raumtemperatur inkubiert. Nach der Inkubation wurde der Überstand entfernt, die *Beads* mit je 500µl 80%igen Ethanol („Merck Millipore“, USA) zwei Mal gewaschen und anschließend an der Luft getrocknet. Sobald das Ethanol verdampft war, wurden die Proben vom magnetischen Ständer entnommen. Die „SPRI Beads“ wurden in der vorgeschriebenen Menge proprietärer „Low EDTA TE“ Puffer resuspendiert (Tab. 5) und 2min lang bei Raumtemperatur inkubiert. Schließlich wurden die Proben wieder auf den magnetischen Ständer platziert. Nach ca. 2min erfolgte eine vollständige Trennung des Überstands und der „SPRI Beads“. Der Überstand enthielt das aufgereinigte Produkt, wurde abpipettiert und für den nächsten Schritt verwendet.

Tab. 5: Proben- und Reagenzvolumina für die Aufreinigungsschritte mit den „SPRI Beads“.

| Schritt | Probe | „SPRI Beads“ | „Low EDTA TE“ Puffer |
|-----------|-------|--------------|----------------------|
| Extension | 84µl | 101µl | 15µl |
| Ligation | 30µl | 36µl | 20µl |
| PCR | 50µl | 40µl | 20µl |

3.1.2.4.2 Sequenzierung der WGBS *Libraries*

Die Sequenzierung der WGBS *Libraries* erfolgte auf der „NextSeq 500“ Plattform („Illumina“, USA) im „TATAA-Biocenter“ (Göteborg, Schweden). Dabei wurden vier 76 *pair end* (PE) Läufe im Hochdurchsatzmodus durchgeführt.

3.1.2.5 Bioinformatische Auswertung der WGBS Ergebnisse

Wie schon vorher beschrieben, konnten die WGBS *Libraries* aufgrund der starken Fragmentierung und geringen Mengen an zfDNA nicht mit herkömmlichen Protokollen angefertigt werden. Die mit dem „Accel-NGS[®] Methyl-Seq DNA Library Kit“ („Swift Biosciences“, USA) hergestellten zfDNA *Libraries* wiesen somit eine andere Komplexität und Fragmentverteilung auf, als die herkömmlichen WGBS *Libraries*. Deswegen musste auch eine geeignete bioinformatische Auswertepipeline etabliert werden, um die Daten optimal analysieren zu können.

Generell müssen mehrere Schritte etabliert werden, um WGBS Daten auswerten zu können (Abb. 11). Zunächst wird die Qualität der Rohdaten überprüft. Hierfür wird am häufigsten die „FastQC“ Software (Version 0.11.15, „Babraham Bioinformatics“, England) verwendet (siehe Abschnitt 3.1.2.5.1). Die Software visualisiert die Qualität der Sequenzierung, Längenverteilung und Zusammensetzung der *Reads*. Des Weiteren werden Informationen über mögliche Adapterkontaminationen sowie Anzahl an Kmeren und PCR-Duplikaten geliefert [192, 193]. Als Kmere werden Sequenzen mit einer Mindestlänge von zwei Nukleotiden bezeichnet, die sich in den Rohdaten immer wieder wiederholen [194, 195].

Falls die „FastQC“ Software zufriedenstellende Ergebnisse liefert, erfolgt das Trimmen der Adaptersequenzen [196]. Bei den zfDNA *Libraries* musste außerdem der von der „Adaptase“ erzeugte 10nt lange Überhang aus den Rohdaten eliminiert werden (siehe Abschnitt 3.1.2.5.2).

Nach dem Trimmen können die *Reads* gegen ein Referenzgenom der Wahl angeordnet werden, dieser Vorgang wird auch als *Alignment* bezeichnet (siehe Abschnitt 3.1.2.5.3). Für das *Alignment* sind viele Algorithmen verfügbar [197]. Je nach Beschaffenheit der WGBS *Library* muss der passende ausgewählt und optimiert werden. Hierfür kann die *Mapping* Effizienz analysiert werden [198]. Dabei wird berechnet wie viel Prozent an analysierten *Reads* dem Referenzgenom zugeordnet werden können. Für die herkömmlichen WGBS *Libraries* wird am häufigsten der „Bismarck“ Algorithmus verwendet [199]. Bei den hier beschriebenen zfDNA *Libraries* lieferte „Bismarck“ (Version 0.15.0, „Babraham Institute“,

England) jedoch keine zufriedenstellende Ergebnisse (*Mapping* Effizienz von ca. 70%). Deshalb wurden weitere Algorithmen ausgetestet. Die besten Ergebnisse mit einer *Mapping* Effizienz von mindestens 98% lieferte der „Segemehl“ Algorithmus (Version 0.2.0, „Interdisziplinäres Zentrum für Bioinformatik, Universität Leipzig“, Deutschland) [200].

Nach dem *Alignment* werden die Daten nach CpG Kontext sowie der gewünschten *Coverage* (mindestens vierfach) mit dem „Bisulfite Analysis Toolkit“ (Version 0.1, „Interdisziplinäres Zentrum für Bioinformatik, Universität Leipzig“, Deutschland) gefiltert und erst dann für das *Peak Calling* verwendet [201] (siehe Abschnitt 3.1.2.5.3). Die *Coverage*, auch Sequenziertiefe genannt, gibt an, wie häufig eine Position beim Sequenzieren abgelesen wurde [202]. Z.B. sagt eine mittlere *Coverage* von 100fach aus, dass jede sequenzierte Base im Durchschnitt 100 Mal abgelesen wurde. Das *Peak Calling* ist der eigentliche Schritt, in dem der Methylierungsstatus des jeweiligen CpG berechnet wird. Dabei werden alle *Reads* angeschaut, die ein bestimmtes CpG enthalten, das Cytosin zu Uracil Verhältnis wird berechnet und das Ergebnis als eine Zahl zwischen 0 und 1 ausgegeben, wobei 0 einer Methylierung von 0% und 1 einer Methylierung von 100% entspricht [201]. Die herkömmlichen *Libraries* haben eine durchschnittliche *Coverage* von 30 bis 40fach, darauf sind auch die herkömmlichen Methoden für das *Peak Calling* ausgelegt. Die zfDNA *Libraries* hatten aufgrund der geringeren Komplexität eine durchschnittliche *Coverage* von 8 bis 10fach. Dementsprechend musste auch das Filtern und *Peak Calling* mit dem „Bisulfite Analysis Toolkit“ optimiert werden.

Sobald die DNA-Methylierungsraten feststehen, können weitere, spezifische Analysen je nach Fragenstellung in einer Programmiersprache der Wahl erfolgen. Für die hier beschriebenen Analysen wurden „R“ (Version 3.2.0, „R Foundation for Statistical Computing“, Österreich), „Perl“ (Version 5.26.0, „The Perl Foundation“, USA) und „Python“ (Version 3.3.6, „Python Software Foundation“, USA) [203-205] verwendet (siehe Abschnitt 3.1.2.5.3). Die Analysen erfolgten in Kooperation mit Dr. Helene Kretzmer („Interdisziplinäres Zentrum für Bioinformatik, Universität Leipzig“, Deutschland), die im Rahmen ihrer Dissertation das oben beschriebene „Bisulfite Analysis Toolkit“ entwickelte und für diese Arbeit zur Verfügung stellte.

Da die hier beschriebenen Analysen eine sehr hohe Rechenkapazität benötigten, erfolgten sie auf dem „NEC-HPC-Linux-Cluster“ der „Christian-Albrechts-Universität zu Kiel“ (Deutschland). Zugang zum Vorrechner erfolgte über eine SSH-Verbindung unter Verwendung der „MobaXterm Personal Edition“ Software („Mobatek“, Frankreich).

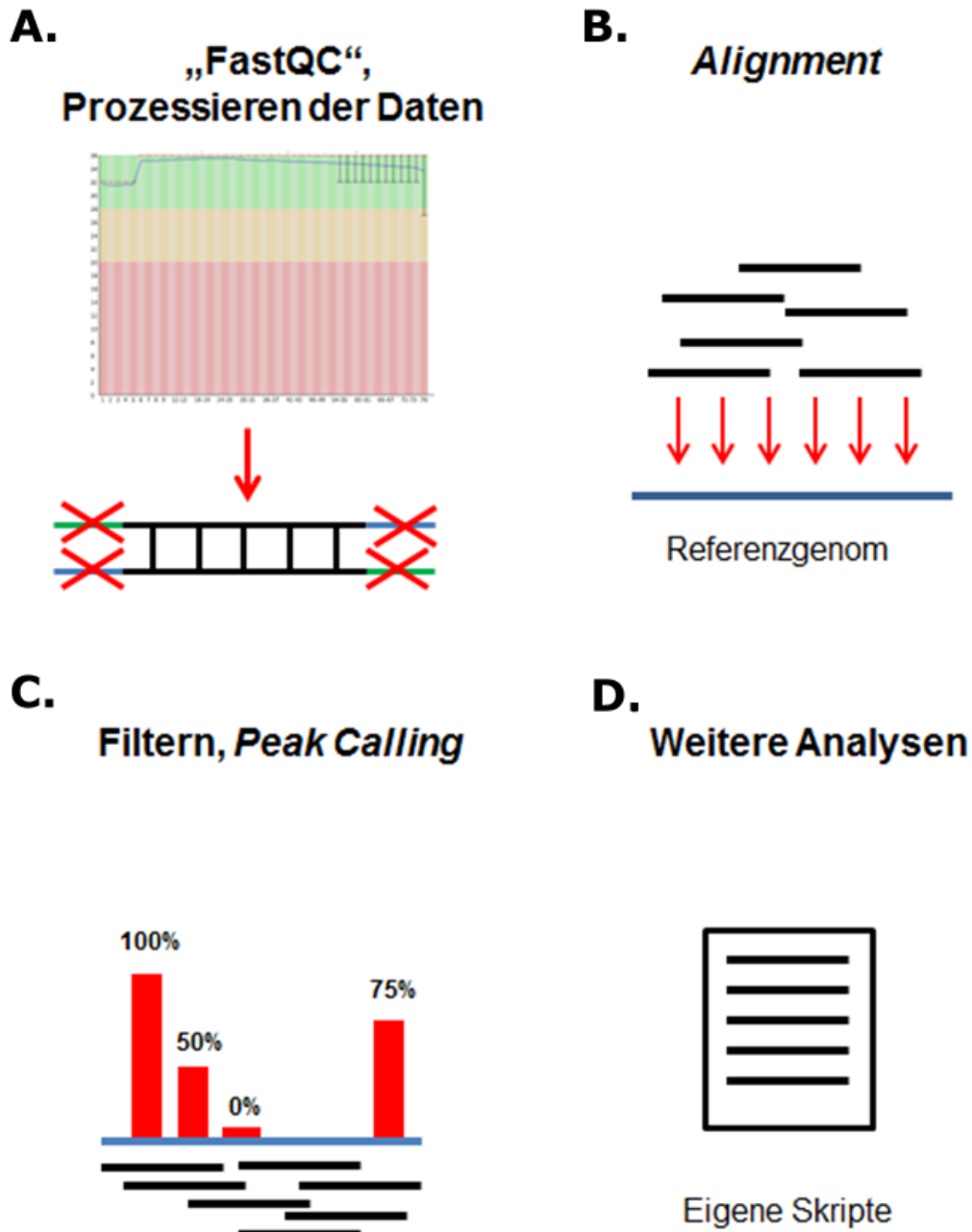


Abb. 11: Die Auswertung der WGBS Sequenzierdaten erfolgte in mehreren Schritten. **A.** Zunächst wurden die Daten einer QC unterzogen und anschließend prozessiert. **B.** Dann wurden die prozessierten Daten gegen ein Referenzgenom (z.B. „HG19“) *aligned* und anschließend **C.** zum Berechnen der DNA-Methylierungsraten verwendet. Die Positionen, an denen eine Methylierungsrate ermittelt wurde, wurden dann nach bestimmten Kriterien gefiltert (z.B. *Coverage* und CpG Kontext) und schließlich **D.** weiteren Analysen unter Verwendung eigener Skripte unterzogen.

3.1.2.5.1 Qualitätskontrolle der Rohdaten

Die Rohdaten wurden im „FastQ“-Format geliefert. Hierbei handelt es sich um ein text-basiertes Format, das zum Speichern der *Reads* sowie dazugehöriger Qualitätsparameter verwendet wird [206]. Um die Qualität der Sequenzierung zu überprüfen, wurde die „FastQC“ Software verwendet [207].

3.1.2.5.2 Datenprozessierung

Die Rohdaten wurden unter Verwendung der „Cutadapt“ Software (Version 1.9.1, „TU Dortmund“, Deutschland) [208] prozessiert. Dabei wurden zwei Schritte durchgeführt.

1.) Eliminierung der überpräsentierten Sequenzen

Während der Sequenzierung wurden die ersten 76 Basen jedes DNA-Fragments von beiden Enden abgelesen (76 PE Sequenzierung). Die mit dem „Accel-NGS® Methyl-Seq DNA Library Kit“ generierten *Libraries*, enthielten DNA-Fragmente unterschiedlicher Länge. Das bedeutet, wenn ein DNA-Fragment kürzer als 152bp war, wurden auch die „Illumina Adapter“ bzw. die *Flowcell* sequenziert. Das resultierte im Vorhandensein von „NNNNNNNNNN“-Sequenzen. Da im weiteren Verlauf der Datenanalyse das *Alignment* der dazugehörigen und sonst qualitativ guten *Reads* aus diesem Grund verhindert werden würde, mussten die überpräsentierten Sequenzen entfernt werden. Der Befehl lautete:

```
cutadapt -q 20 -o 5 --minimum-length 30 -a GATCGGAAGAG -A AGATCGGAAGAG -o  
<Name_Read_1>.clipped.fastq.gz -p <Name_Read_2>.clipped.fastq.gz  
<Name_Read_1>.fastq.gz <Name_Read_2>.fastq.gz &><Name>.clipping.stats
```

2.) Beseitigung des von der „Adaptase“ erzeugten Überhangs

Während der Herstellung der WGBS *Library* wurde das Enzym „Adaptase“ verwendet, welches am 3'-Ende des zweiten *Reads* einen Überhang mit niedriger Komplexität erzeugte. Dieser Bereich würde genauso wie die überpräsentierte Sequenzen beim späteren *Alignment* stören und musste somit entfernt werden. Der Befehl lautete:

```
cutadapt --minimum-length 25 -u 11 -o <Name_Read_2>.clipped.trimmed.fastq.gz -p  
<Name_Read_1>.clipped.trimmed.fastq.gz <Name_Read_2>.clipped.fastq.gz  
<Name_Read_1>.clipped.fastq.gz
```

3.1.2.5.3 Auswertung der prozessierten Daten

Die anschließende Datenanalyse erfolgte mit dem „Bisulfite Analysis Toolkit“ [201]. Die Funktion der modular aufgebauten Software ist in der Abb. 12 dargestellt.

Das *Alignment* wurde gegen das „HG19“ Referenzgenom durchgeführt. Dabei wurde „Segemehl“ verwendet, dessen Prinzip auf der Suche nach einem optimalen Treffer im Referenzgenom beruht [209]. Die maximal erlaubte Anzahl an Ungenauigkeiten pro *Read* (z.B. Insertionen, Deletionen, Punktmutationen) betrug 10%. Alle Treffer, die diesen Schwellenwert unterschritten, wurden zu dem semi-globalen *Alignment* zugelassen. Letztendlich wurden nur die *Reads* mit einer Genauigkeit von mindestens 90% in einer finalen „BAM“-Datei aufgelistet und für die weiteren Analysen verwendet. Das „BAM“-Format ist eine komprimierte Version der „SAM“-Datei, ein text-basiertes Format, das zum Speichern von Ergebnissen des *Alignments* vom Algorithmus generiert wird [210]. Die statistische Auswertung der *Mapping* Effizienz erfolgte mit dem „BAT_mapping_stat“ Modul [201].

Schließlich wurden alle *Reads*, die zu einer Probe gehörten, in ein „BAM“-File mit dem „BAT_merging“ Modul zusammengeführt [201]. Die überlappenden Sequenzen wurden mit dem „ClipOverlap“ (BamUtil Version 1.0.13) Modul eliminiert [211]. Die Befehle lauteten:

```
perl BAT_mapping.pl -g hg19.fa -i hg19 -p <Name_Read_1>.clipped.trimmed.fastq.gz -q  
<Name_Read_2>.clipped.trimmed.fastq.gz -t 16 -tmp <Folder> --segemehl segemehl.x -o  
<Folder>/<Name>
```

```
perl BAT_mapping_stat.pl --bam <Name>.bam --fastq <Name>.clipped.trimmed.fastq.gz -b >  
<Name>.stat
```

```
perl BAT_merging.pl -o <Name>.bam --bam <file_1>.bam,<file_2>.bam, ..., <file_n>.bam
```

```
bamUtil_1.0.13/bamUtil/bin/bam ClipOverlap --in <Name>.bam --out <Name>.nooverlap.bam
```

Im nächsten Schritt wurde die DNA-Methylierung mit Hilfe von „BAT_calling“ detektiert. Das Modul erzeugt eine „VCF“-Datei. Hierbei handelt es sich um eine Textdatei, die nur Informationen über die detektierten DNA-Methylierungsraten, *Coverage*, Anzahl an abgedeckten Nukleotiden und den Sequenzkontext enthält [212]. Im weiteren Verlauf der Analysen wurde diese Datei nach CpG Kontext und einer *Coverage* von mindestens achtfach gefiltert. Dabei wurden Abbildungen generiert und weitere „VCF“- sowie „BedGraph“-Dateien erzeugt [201].

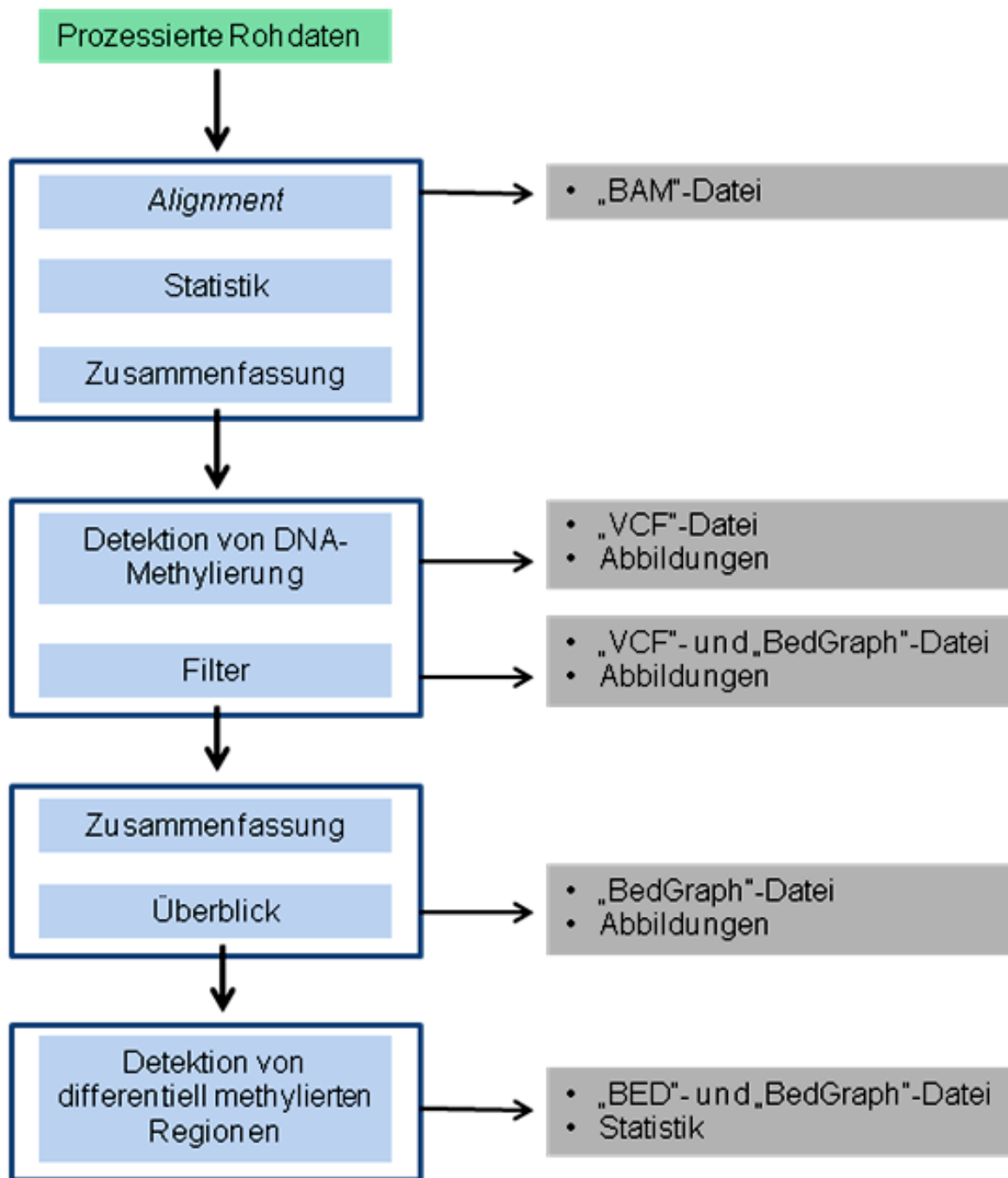


Abb. 12: Das „Bisulfite Analysis Toolkit“ wurde verwendet, um das *Alignment* gegen das „HG19“ Referenzgenom durchzuführen. Des Weiteren erfolgte Detektion von DNA-Methylierungsraten, differentiell methylierter Regionen sowie das Erstellen von Übersichtsgraphiken.

Als nächstes wurde das „BAT_summarize“ Modul verwendet, welches die Mittelwerte detektierter DNA-Methylierungsraten zweier Gruppen ermittelte. Die berechneten DNA-Methylierungsraten sowie die genomischen Koordinaten der Cytosine wurden in eine textbasierte „BedGraph“-Datei geschrieben, die im späteren Verlauf für die Identifikation differentiell methylierter Regionen verwendet wurde [201].

Die Visualisierung der DNA-Methylierung pro Gruppe erfolgte unter Verwendung des „BAT_overview“ Moduls [201]. Die Befehle lauteten:

```
BAT_calling.pl -d hg19.fa -q <Name>.nooverlap.bam --haarz segemehl_0_2_0/segemehl/ haarz.x -o <Folder>
```

```
BAT_filter_vcf.pl --vcf <Name>.nooverlap.vcf.gz --out <Name>_CG_cov_final --context CG --MDP_min 8 --MDP_max 50
```

```
BAT_summarize.pl --in1 Adeno_CG_cov.bedgraph,PEKA_CG_cov.bedgraph --in2 Control_CG_cov.bedgraph -l cancer,control --h1 Adeno,PEKA --h2 Control --out pilot --cs hg19.chrom.sizes --bgbw bedGraphToBigWig
```

```
Rscript BAT_overview.R -i pilot_cancer_control.txt -o pilot_overview.pdf -p cancer -q control
```

3.1.2.5.4 Korrelationsanalysen

Im Rahmen dieser Arbeit wurden Daten von zwei Methoden zur genomweiten Untersuchung von DNA-Methylierungsmustern verwendet: WGBS und *Methylation Array* (HM 450K).

Für die Korrelationsanalysen wurde die „Bedtools“ Software verwendet [169, 213, 214]. Das „Bedtools Intersect“ Modul liest sowohl die WGBS als auch HM 450K Ergebnisse ein, prüft sie auf Überlappung und schreibt die überlappenden CpG *Loc*i in eine neue „BED“-Datei. Bei dem „BED“-Format handelt es sich um eine Textdatei. Jede Zeile der Datei enthält genomische Koordinaten eines CpG. Die Spalten sind durch ein Tabulatorzeichen getrennt [215]. Die „BED“-Datei wurde nachfolgend direkt in „R“ geladen und der „Pearson“ Korrelationsanalyse unterzogen (p -Wert < 0,01). Die Visualisierung der Ergebnisse erfolgte ebenfalls in R.

3.1.2.5.5 Auswählen der CpG *Loc*i für das Plasma *Panel*

Die WGBS Daten wurden wie beschrieben ausgewertet. Die mit dem „BAT_summarize“ Modul generierte „BedGraph“-Datei enthielt drei Gruppen (Kontrolle, Adenokarzinom, Plattenepithelkarzinom) mit jeweils 11.289.424 Positionen pro Gruppe. Die „BedGraph“-Datei wurde in zwei Listen unterteilt. Die erste Liste enthielt 29.877 *Loc*i, die DNA-Methylierungsunterschiede zwischen der Tumor- und Kontrollgruppen zeigten. Die zweite Liste enthielt 76.374 CpG *Loc*i, die unterschiedlich jeweils in Adeno- und Plattenepithelkarzinomgruppen methyliert waren. Als differentiell methyliert wurden die Regionen bezeichnet, welche einen DNA-Methylierungsunterschied von mindestens 15% aufwiesen.

Als nächstes wurden die beiden Listen nach Chromosomen sortiert und mit dem „HG19“ Referenzgenom annotiert. Die CpG *Loci*, die auf Chromosomen X, Y und M (mitochondriales Chromosom) sowie innerhalb von häufigen SNPs ($\geq 1\%$ der Bevölkerung) lokalisiert waren und nicht proteinkodierend waren, wurden verworfen.

Die verbliebenen CpG *Loci* mussten eines der drei Kriterien erfüllen, um in das Plasma *Panel* aufgenommen zu werden:

- 1.) differentiell methyliertes CpG wurde von beiden Methoden detektiert (WGBS und HM 450K),
- 2.) differentiell methyliertes CpG liegt innerhalb eines Clusters bestehend aus mindestens zwei weiteren differentiell methylierten CpG *Loci*, alle CpG *Loci* des Clusters sind entweder hypo- oder hypermethyliert, der Abstand zwischen den CpG *Loci* beträgt 2 bis 20 Nukleotide,
- 3.) es handelt sich um ein CpG mit der höchsten differentiellen DNA-Methylierungsrate (>0.8).

Die DNA-Regionen, die eines der drei Kriterien erfüllten, wurden in das Plasma *Panel* aufgenommen (siehe Abschnitt 9.3, Tab. 7). Alle verwendeten Aufrufe sind im Abschnitt 9.1 detailliert beschrieben.

3.1.3 Weitere Komponenten des Plasma *Panels* (*in silico* Datenanalysen)

3.1.3.1 Die prognostische Studie

Zusätzlich zu den diagnostisch bzw. therapeutisch relevanten Informationen (z.B. Stadium und Tumorentität) sollte das *Panel* auch prognostische Informationen beinhalten. Deswegen wurde es um 33 CpG *Loci* erweitert, die im Rahmen einer klinischen DZL-Studie erhoben wurden. Der Titel der Studie lautete: „Comprehensive characterization of non-small cell lung cancer (NSCLC) by integrated clinical and molecular analysis“.

Der zur Verfügung gestellte HM 450K-Datensatz enthielt Informationen über den DNA-Methylierungsstatus von insgesamt 41 Lungenkarzinomen. Die Patienten wurden je nach Überlebensdauer klassifiziert. Dabei wurden 28 Patienten zu der prognostisch günstigen (Überlebensdauer länger als 15 Monate) und 13 zu der ungünstigen Gruppe (Überlebensdauer kürzer als 13 Monate) gezählt. Die 33 in das *Panel* aufgenommene CpG *Loci* konnten beide Gruppen anhand des DNA-Methylierungsmusters voneinander trennen und enthielten somit für die Prognose relevanten Informationen.

3.1.3.2 Die bivalente Chromatin Studie

Zusätzlich zu den WGBS und HM 450K Ergebnissen wurden 26 differentiell methylierte Regionen aus der Studie zu bivalentem Chromatin in Tumoren [216] in das Plasma *Panel* aufgenommen.

Die bivalenten Promotoren tragen sowohl aktivierende als auch reprimierende Histonmodifikationen, die vor allem während der Zelldifferenzierungsprozesse eine wichtige Rolle spielen. In Tumorzellen sind sie häufig fehlerhaft reguliert. Während der Studie wurden WGBS- und HM 450K-Datensätze verschiedener Tumorproben und -zelllinien (n=7000) analysiert [216].

3.2 Methoden im Projekt „Validierung des Plasma *Panels*“

Das fertige Plasma *Panel* enthielt 638 differentiell methylierte Regionen (Abschnitt 9.3, Tab. 7), wurde von der Firma „Roche“ (Schweiz) synthetisiert und auf Trockeneis verschickt. Hierbei handelte es sich um ein nach Kundenwunsch synthetisiertes „SeqCap Epi Enrichment Kit“ („Roche“, Schweiz). Laut Hersteller war das *Panel* sowohl für die Analyse von Gewebeproben als auch zirkulierender, zellfreier DNA geeignet und musste zunächst im Rahmen einer Pilotstudie validiert werden. Hierfür wurde vom DZL Blutplasma von 12 Patienten zur Verfügung gestellt. Davon waren drei Patienten zum Zeitpunkt der Untersuchung gesund bzw. tumorfrei (Kontrollgruppe) und neun litten an nicht-kleinzelligem Lungenkarzinomen unterschiedlicher Stadien (Tumorgruppe). Die detaillierten Informationen zu den Patienten sind im Abschnitt 2.1.2 zusammengestellt.

Die Validierung erfolgte in mehreren Schritten. Zunächst wurde das Validierungsmaterial, die zirkulierende, zellfreie DNA, vorbereitet. Die Extraktion aus dem Plasma, Quantifizierung, Qualitätskontrolle (QC) sowie Bisulfitkonvertierung erfolgten wie bereits in Abschnitten 3.1.2.1-3.1.2.3 beschrieben.

10ng der konvertierten zfDNA wurden dann für die Herstellung der Libraries verwendet. Die Library Präparation erfolgte in zwei Schritten. Im ersten Schritt wurde wie im Abschnitt 3.1.2.4 beschrieben von jeder Probe eine WGBS *Library* hergestellt, die Informationen über das gesamte zfDNA-Methylom des entsprechenden Patienten enthielt. Da jedoch im späteren Verlauf nur die 638 differentiell methylierte Regionen sequenziert und analysiert werden sollten, wurden sie im zweiten Schritt aus dem gesamten Methylom extrahiert und angereichert. Das geschah unter Verwendung des „SeqCap Epi Enrichment Kit“, dessen Bestandteil das von „Roche“ synthetisierte Plasma *Panel* war (siehe Abschnitt 3.2.1).

Die fertige *Library* wurde einer QC unterzogen sowie quantifiziert (siehe Abschnitt 3.1.2.2) und anschließend auf dem „MiSeq“ („Illumina“, USA) sequenziert (siehe Abschnitt 3.2.2). Die Sequenzierdaten wurden im „FastQ“-Format gespeichert und mussten anschließend analysiert werden (siehe Abschnitt 3.2.3). Hierfür wurde die bioinformatische Pipeline aus dem Abschnitt 3.1.2.5 angepasst. Denn diesmal sollte nicht das gesamte Methylom, sondern nur die 638 bestimmten Regionen analysiert werden.

Die Ergebnisse wurden schließlich für die Entwicklung eines Klassifikators verwendet, der nachfolgend die DNA-Methylierungsmuster interpretierte und diagnostisch sowie klinisch relevante Informationen über den gesundheitlichen Zustand eines Patienten lieferte (siehe Abschnitt 3.2.3.3).

3.2.1 Anreicherung von differentiell methylierten Regionen

Für die Extraktion und Anreicherung von 638 differentiell methylierten Regionen aus dem gesamten zfDNA-Methylom wurde das „SeqCap Epi Enrichment Kit“ verwendet. Einer der Bestandteile des Kits war das designte Plasma *Panel* (siehe Abschnitt 9.3, Tab. 7). Die dort enthaltenen Oligonukleotide, auch „Capture Probes“ genannt, hybridisierten an die differentiell methylierte Regionen und konnten im weiteren Verlauf angereichert und amplifiziert werden (Abb. 13).

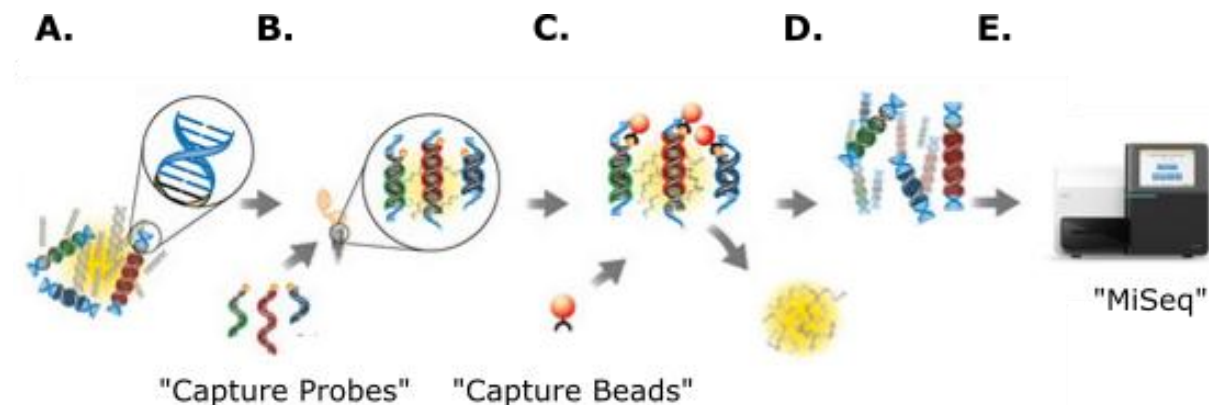


Abb. 13: Das „SeqCap Epi Enrichment Kit“ war in mehrere Schritte unterteilt. **A.** zunächst wurden wie im Abschnitt 3.1.2.4 beschriebenen WGBS *Libraries* hergestellt und äquimolar *gepoolt*. **B.** Die 638 differentiell methylierten Regionen wurden dann an „Capture Probes“ hybridisiert, **C.** unter Verwendung von „Capture Beads“ angereichert und schließlich **D.** in einer PCR-Reaktion amplifiziert. **E.** Die fertigen NGS *Libraries* wurden dann quantifiziert, einer QC unterzogen und auf dem „MiSeq“ sequenziert. Die Abbildung wurde verändert nach [217].

Hybridisierungsreaktion

Die 12 hergestellten WGBS *Libraries* wurden äquimolar *gepoolt* und zunächst für eine Hybridisierungsreaktion vorbereitet. Hierfür wurden 1µg des WGBS *Library*-Pools mit 10µl „Bisulfite Capture Enhancer“, 1µl „SeqCap HE Universal Oligo“ und 1µl „SeqCap HE Index Oligo“ in ein 1,5ml Reaktionsgefäß mit einem kleinen Loch im Deckel pipettiert. Die Probe wurde in einem Vakuumkonzentrator solange eingedampft, bis ein klares weißliches Pellet zu sehen war. Die „SeqCap HE Universal“ und „SeqCap HE Index“ Oligonukleotide wurden in einem Überschuss hinzugegeben (1µl entsprach 1.000pmol) und dienten dazu, die freiliegenden WGBS Universal- und Indexadapter zu binden. Somit sollte verhindert werden, dass die WGBS Adapter die nachfolgende Hybridisierungsreaktion stören.

Für die eigentliche Hybridisierungsreaktion wurden 7,5µl zweifacher „Hybridisation Buffer“ sowie 3µl „Hybridisation Component A“ direkt auf das Pellet gegeben, 10s gemischt, kurz zentrifugiert und 10min bei 95°C inkubiert. Dann wurde die Probe in ein 0,2µl Reaktionsgefäß überführt, mit 4,5µl „Capture Probes“ versetzt, gut gemischt und bei 47°C 72 Stunden lang in einem *Thermocycler* inkubiert. Der Deckel des *Thermocyclers* war auf 57°C vorgeheizt. Die „Capture Probe“ wurde von „Roche“ speziell für dieses Projekt synthetisiert (siehe Abschnitt 9.3, Tab. 7). Sie enthielt 638 unterschiedliche Oligonukleotide, die komplementär zu den untersuchten differentiell methylierten Regionen waren und im Laufe der Hybridisierungsreaktion diese gezielt banden.

Anreicherung und Waschen der hybridisierten „Capture Probes“

Im nächsten Schritt wurden die gebundenen „Capture Probes“ angereichert und mehrfach gewaschen. Hierfür wurden mehrere Waschpuffer sowie die „Capture Beads“ nach Angaben des Herstellers vorbereitet.

Die hybridisierte Probe wurde mit 100µl „Capture Beads“ versetzt, kurz gemischt und 45min bei 47°C im *Thermocycler* inkubiert. Der Deckel des *Thermocyclers* war auf 57°C vorgeheizt. Um das Absetzen der *Beads* zu verhindern, wurden die Proben alle 15min kurz aus dem *Thermocycler* entnommen und gemischt. Bei den hier verwendeten „Capture Beads“ handelte es sich um Streptavidin-*Beads*, die mit den biotinylierten „Capture Probes“ interagierten.

Nach der Inkubation wurden die Proben aus dem *Thermocycler* entnommen und die „Capture Beads“ mehreren Waschschritten unterzogen. Das Trennen der *Beads* vom Puffer erfolgte jedes Mal bei Raumtemperatur unter Verwendung des „DynaMag™-PCR“ Magnets („Thermo Fisher Scientific“, USA).

Im ersten Teil des Waschprotokolls wurden nur Puffer verwendet, die zuvor auf 47°C vorgewärmt wurden. Dabei wurde die Probe mit 100µl einfachem „Wash Buffer I“ versetzt, kurz gemischt und mit Hilfe eines Magneten pelletiert. Der Überstand wurde verworfen, die *Beads* in 200µl einfachen „Stringent Wash Buffer“ gelöst, 5min lang im Thermocycler bei 47°C inkubiert und wieder mit Hilfe eines Magneten pelletiert. Der Überstand wurde wieder verworfen und die *Beads* zwei weitere Male mit 200µl einfachen „Stringent Wash Buffer“ gewaschen.

Der zweite Teil des Waschprotokolls erfolgte komplett bei Raumtemperatur, dementsprechend mussten die hierfür verwendeten Puffer auf Raumtemperatur äquilibriert werden. Zunächst wurden die zuvor bei 47°C gewaschene „Capture Beads“ in 200µl einfachen „Wash Buffer I“ gelöst, 2min lang gemischt und mit Hilfe eines Magneten pelletiert. Der Überstand wurde verworfen, die *Beads* mit 200ml einfachem „Wash Buffer II“ versetzt, 1min lang gemischt und wieder mit Hilfe eines Magneten pelletiert. Auch hier wurde der Überstand verworfen, die *Beads* in 200ml „Wasch Buffer III“ gelöst, kurz gemischt und schließlich auf dem Magneten vom Überstand getrennt.

Für die darauffolgende Elution wurden 50µl dH₂O direkt auf die *Beads* gegeben, 2min lang bei Raumtemperatur inkubiert und mit Hilfe eines Magneten pelletiert. Der Überstand wurde aus dem Reaktionsgefäß vorsichtig abpipettiert und für alle weiteren Schritte verwendet.

Amplifikation der angereicherten differentiell methylierten Regionen

Nach dem Waschen erfolgte die Amplifikation der angereicherten, differentiell methylierten Regionen. Hierfür wurden zu den 20µl des Eluats 25µl zweifachen „KAPA HiFi HotStart Ready Mix“ („Roche“, Schweiz) sowie 5µl „Post LM PCR Oligonukleotide“ („Roche“, Schweiz) gegeben, gut gemischt und unter Verwendung von folgendem PCR-Programm im *Thermocycler* mit vorgeheiztem Deckel amplifiziert:

Schritt 1: 45s 98°C

Schritt 2: 15s 98°C

Schritt 3: 30s 60°C

Schritt 4: 30s 72°C

Schritt 5: Wiederholung der Schritte zwei bis vier 15 weitere Male

Schritt 6: 60s 72°C

Schritt 7: Pause bei 4°C

Aufreinigung der angereicherten und amplifizierten differentiell methylierten Regionen

Die amplifizierten Regionen wurden nachfolgend unter Verwendung der „AmpureXP“ *Beads* („Beckman Coulter“, USA) aufgereinigt. Hierfür wurden die *Beads* zunächst auf Raumtemperatur äquilibriert. Die Probe wurde in ein 1,5ml Reaktionsgefäß überführt. Zu 50µl Probe wurden 50µl dH₂O und 180µl „AmpureXP“ *Beads* gegeben. Die Probe wurde kurz gemischt, 15min lang bei Raumtemperatur inkubiert, kurz zentrifugiert und auf den „DynaMag™-2“ Magneten („Thermo Fisher Scientific“, USA) gestellt. Der Überstand wurde verworfen und die *Beads* zwei Mal mit jeweils 200µl frisch zubereitetem 80%igem Ethanol gewaschen. Dann wurden die *Beads* 15min lang bei Raumtemperatur getrocknet. Um die *Libraries* zu eluieren, wurden 52µl dH₂O auf die trockenen *Beads* pipettiert. Die *Beads* wurden gut gemischt, 2min lang bei Raumtemperatur inkubiert und wieder auf „DynaMag™-2“ gestellt. Der Überstand wurde vorsichtig abpipettiert und für die Quantifizierung, QC (siehe Abschnitt 3.1.2.2) sowie Sequenzierung auf dem „MiSeq“ verwendet.

3.2.2 Sequenzierung des Plasma Panels

Die Sequenzierung der NGS *Library* aus angereicherten, differentiell methylierten Regionen erfolgte auf dem „MiSeq“.

Hierfür musste die hergestellte *Library* zunächst auf 4nM verdünnt und denaturiert werden. Dann wurden 5µl der 4nM *Library* in ein 1,5ml Reaktionsgefäß überführt, mit 5µl 0,2N NaOH versetzt, kurz gemischt, 1min lang bei 280g zentrifugiert und 5min bei Raumtemperatur inkubiert. Die denaturierte *Library* wurde dann mit 990µl „Buffer HT1“ („Illumina“, USA) versetzt und wieder gut gemischt. Das ergab eine 20pM *Library*, die anschließend mit „Buffer HT1“ auf 4pM verdünnt und mit 10% „PhiX“ („Illumina“, USA) versetzt wurde.

Schließlich wurde eine „MiSeq 150 V3“ Kassette („Illumina“, USA) mit der fertigen Probe beladen und in einem 76 PE Lauf sequenziert.

3.2.3 Bioinformatische Auswertung der Sequenzierdaten

3.2.3.1 Qualitätskontrolle und Prozessieren der Rohdaten

Die Daten wurden wie in Abschnitten 3.1.2.5.1 und 3.1.2.5.2 beschrieben einer „FastQC“ Analyse unterzogen und anschließend prozessiert [207].

3.2.3.2 Auswertung der prozessierten Daten

Die prozessierten Daten wurden wie in Abschnitt 3.1.2.5.3 beschrieben gegen das „HG19“ Referenzgenom mit dem „Segemehl“ Algorithmus *aligned*. Die PCR-Duplikate wurden unter Verwendung von „Samtools“ (Version 1.3.1, „Wellcome Trust Sanger Institute“, England, „Broad Institute of MIT and Harvard“, USA) entfernt [210]. Der Befehl lautete:

```
samtools rmdup -S <Name>.bam <Name>_wo_dup.bam
```

Die DNA-Methylierungsraten innerhalb der sequenzierten Regionen wurden mit dem „BAT_calling“ Modul berechnet und mit dem „BAT_filter_vcf“ Modul nach dem CpG Kontext und einer *Coverage* von mindestens achtfach gefiltert [201] (siehe Abschnitt 3.1.2.5.3). Schließlich wurden die Daten gegen die Regionen des Plasma *Panels* annotiert. Die Aufrufe lauteten:

```
for i in *.vcf.gz; do o=`echo $i | sed 's/.vcf.gz/_CG.vcf.gz/'; echo $i $o; perl BAT_filter_vcf --vcf $i --out $o --context CG; done
```

```
for i in *_CG.vcf.gz; do  
o=`echo $i | sed 's/_CG.vcf.gz/_CG.cov.region.vcf.gz/'`  
echo $i $o  
zcat $i | grep "#" >tmp.vcf; bedtools intersect -u -b OID44445_hg19_07mar2017_primary_targets.bed -a $i >>tmp.vcf  
gzip tmp.vcf  
perl BAT_filter_vcf --vcf tmp.vcf.gz --out $o --context CG --MDP_min 8 --MDP_max 200  
rm tmp.vcf.gz  
done
```

```
bedtools unionbedg -filler NA -header -names <sample_1> ... <sample_n> -i  
<name_sample_1>_wo_dup_CG.cov.region.bedgraph ... <name_sample_n>_wo_dup_CG.cov.region.  
bedgraph > <name>.bed
```


3.2.3.3 Erstellen eines Klassifikators

Mit Hilfe des Plasma *Panels* sollte das DNA-Methylierungsmuster eines Patienten analysiert werden. Daraus sollte geschlossen werden, ob ein Patient an einem malignen Lungentumor erkrankt ist. Falls ja, sollten aus dem DNA-Methylierungsprofil Informationen über die Entität des Tumors und die Prognose des betroffenen Patienten abgeleitet werden.

Hierfür musste ein Klassifikator erstellt werden, der in der Lage ist, die Ergebnisse der in den Abschnitten 3.2.3.1 und 3.2.3.2 beschriebenen Pipeline, schnell und zuverlässig zu interpretieren. Ein Klassifikator, auch *Predictive Modelling* genannt, ist ein Beispiel für das supervidierte Lernen [218, 219]. Das Ziel eines Klassifikators ist es, nach dem Erhalt von Variablen (z.B. DNA-Methylierungsmustern) und einer Annotation, zunächst ein Modell zu erstellen, das später in der Lage ist, die Variablen von unabhängigen Proben richtig zu klassifizieren (Abb. 14).

Die Software „Qlucore Omics Explorer“ bietet mehrere Möglichkeiten an, unter Verwendung von DNA-Methylierungsdaten einen für die jeweilige Fragestellung optimalen Klassifikator zu erstellen. Dabei kann aus drei Algorithmen gewählt werden: „k-Nearest Neighbors Algorithm“ (kNN), „Support Vector Machines“ (SVM) und „Random Trees“ (RT). Bei kNN wird eine Klassenzuordnung anhand der Berücksichtigung von k nächsten Nachbarn vorgenommen [220, 221]. SVM beschreibt jedes Objekt durch einen Vektor in einem Vektorraum. Innerhalb des Vektorraums wird eine Hyperebene so gesetzt, dass sie als Trennfläche zwischen den Gruppen agiert und sie in zwei Klassen unterteilt [222, 223]. RT besteht aus mehreren unkorrelierten Entscheidungsbäumen, die während des Lernprozesses generiert wurden. Jeder Baum fällt eine Entscheidung, die Klasse mit den meisten Stimmen entscheidet letztendlich über die endgültige Klassifikation [224, 225].

Generell ist es schwierig für eine neue Fragestellung im Voraus die Aussage treffen zu können, welcher Algorithmus die optimalen Ergebnisse liefern wird. Daher wurden alle drei verfügbaren Algorithmen getestet, um den jeweils besten für die jeweilige Kategorie zu finden.

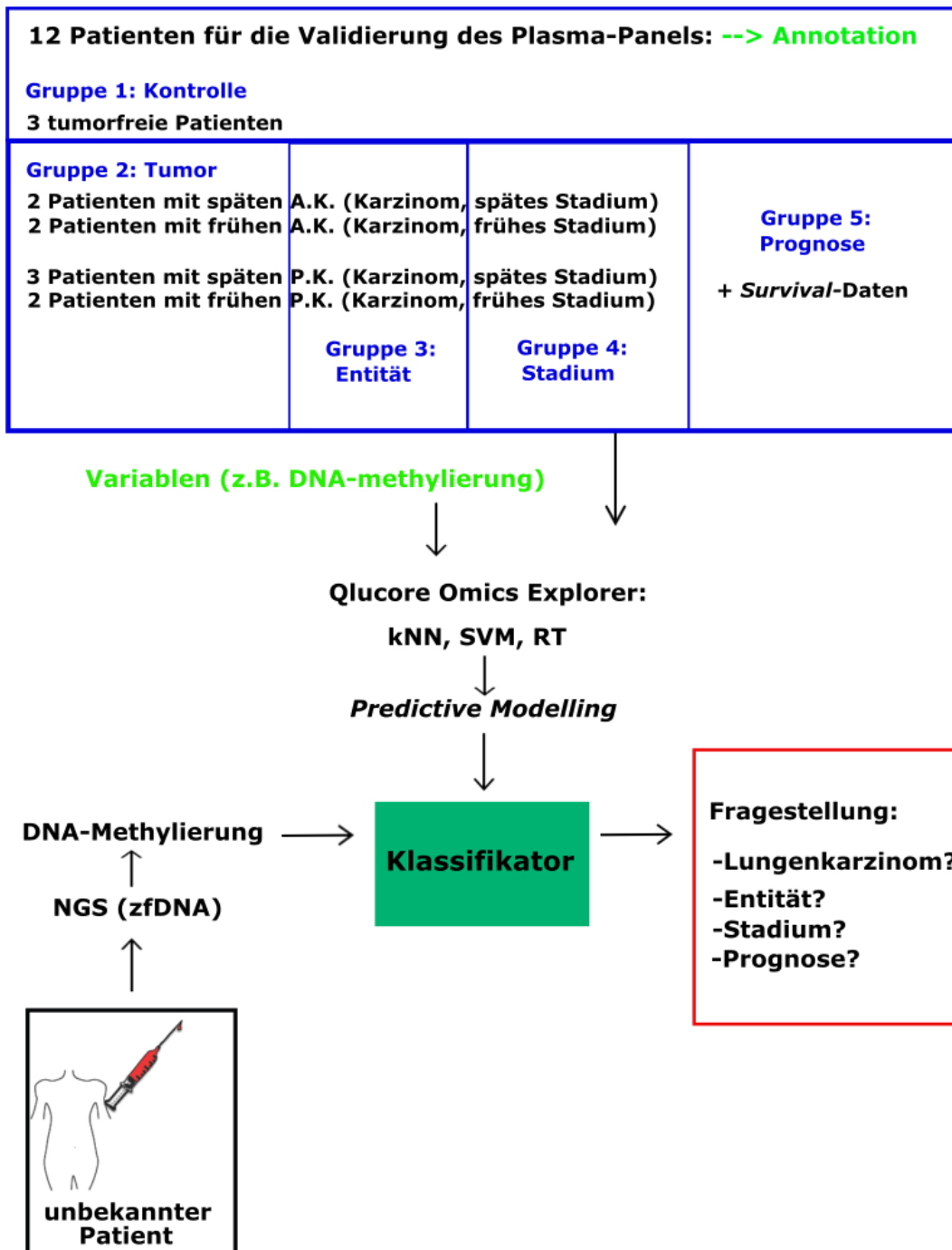


Abb. 14: Das Funktionsprinzip eines Klassifikators; Aus den Daten der Validierungskohorte (12 Patienten) wird zunächst ein Annotationsfile generiert, welcher zusätzlich mit den ermittelten DNA-Methylierungsraten der im Plasma *Panel* enthaltenen Regionen (siehe Abschnitt 9.3, Tab. 7) in die „Qlucore Omics Explorer“ Software geladen wird. Die DNA-Methylierungsdaten (Variablen) und das Annotationsfile werden von implementierten Algorithmen („k-Nearest Neighbors Algorithm“ (kNN), „Support Vector Machines“ (SVM) und „Random Trees“ (RT)) dazu verwendet, ein optimales Model zu erstellen. Dieser Vorgang wird als *Predictive Modelling* bezeichnet. Nachdem der optimale Klassifikator generiert ist, ist dieser in der Lage, das zfDNA-Methylierungsmuster eines unbekanntes Patienten zu analysieren und somit eine Diagnose zu stellen (Adenokarzinom (A.K.), Plattenepithelkarzinom (P.K.)).

3.3 Methoden im Projekt „Multi-OMICS Studie zur Untersuchung von Risiken der E-Zigaretten“

Heutzutage gewinnen die E-Zigaretten immer mehr an Beliebtheit. Viele Raucher wechseln zu E-Zigaretten, weil sie von einigen Wissenschaftlern und Ärzten als gesündere Alternative suggeriert werden [153, 226, 227]. Im Rahmen dieser Arbeit sollte eine multi-OMICS Studie entwickelt und durchgeführt werden, um kurzfristige Auswirkungen der E-Zigaretten auf menschliche Lungenzellen zu beschreiben.

Der Aufbau der multi-OMICS Studie ist in der Abb. 15 dargestellt. Im Laufe der Studie wurden sowohl primäre Zellen aus menschlicher Lunge als auch eine gut charakterisierte Lungenzelllinie, die A549, analysiert (siehe Abschnitt 3.3.1). Die Zellen wurden mit unterschiedlich konzentrierten Zigaretten- sowie E-Zigarettenextrakten behandelt (engl. *cigarette smoking extract* (CSE), *e-cigarette smoking extract* (eCSE)). Des Weiteren wurde nicht nur ein nikotinhaltiger E-Zigarettenextrakt, sondern auch ein nikotinfreier untersucht (siehe Abschnitt 3.3.3). Die Zellen wurden 48 Stunden lang inkubiert. Während der Inkubation wurden die Apoptoseraten gemessen (siehe Abschnitt 3.3.4). Nach der Inkubation wurde zunächst die Zellviabilität ermittelt (siehe Abschnitt 3.3.5). Zeitgleich wurden RNA-, Chromatin- und DNA-Extraktionen durchgeführt, die im weiteren Verlauf zu genomweiten Analysen der Genexpression (siehe Abschnitt 3.3.6), des Chromatinstatus (siehe Abschnitt 3.3.7) und der DNA-Methylierung (siehe Abschnitt 3.3.8) verwendet wurden. Zusätzlich wurde die Morphologie der Zellen mit Hilfe der Fluoreszenzmikroskopie (siehe Abschnitt 3.3.9) sowie Durchflusszytometrie-Messungen (siehe Abschnitt 3.3.10) analysiert.

3.3.1 Zellkultur

3.3.1.1 A549

A549 ist eine der am häufigsten verwendeten humanen Lungenzelllinien. Sie wurde 1972 aus dem Lungenkarzinom eines 58-Jährigen, männlichen Kaukasiers isoliert [162].

Die A549 Zellen sind adhärent und wurden bei 37°C und 5% CO₂ im Medium, bestehend aus 90% DMEM („Thermo Fisher Scientific“, USA) und 10% FBS („Thermo Fisher Scientific“, USA) kultiviert. Das Passagieren der Zellkultur erfolgte alle drei Tage, dabei wurden die Zellen 1:10 gesplittet.

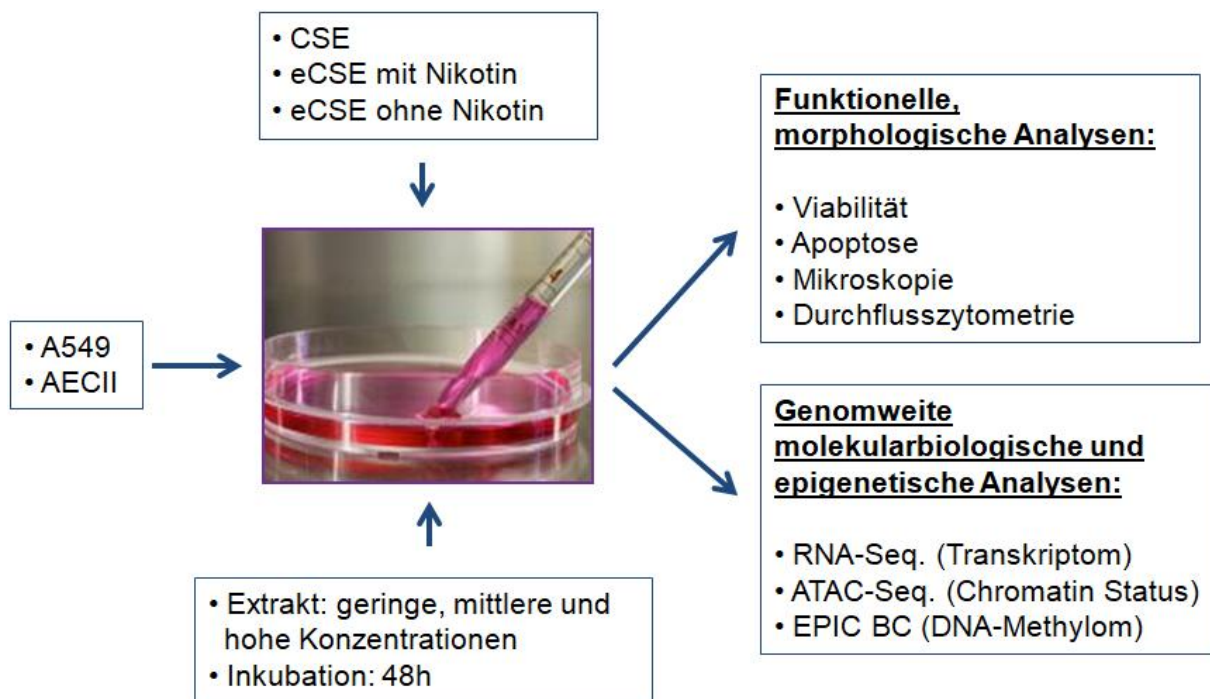


Abb. 15: Während der multi-OMICS Studie wurden A549 Zelllinie und primäre Zellen mit Zigaretten- und E-Zigarettenextrakten unterschiedlicher Konzentrationen 48 Stunden lang behandelt. Währenddessen wurden die Apoptoseraten gemessen. Nach der Inkubation erfolgten zahlreiche funktionelle, morphologische sowie genomweite molekulare Analysen der stimulierten Zellen. Dabei wurde zum einen die Zellviabilität untersucht und zum anderen gleichzeitig RNA-, Chromatin sowie DNA extrahiert. Die extrahierten Proben wurden verwendet, um das Transkriptom, Chromatinstatus und DNA-Methylom genomweit zu analysieren. Hierfür wurden RNA-Sequenzierung, ATAC-Sequenzierung sowie „Infinium MethylationEPIC BeadChip“ (EPIC BC) durchgeführt. Des Weiteren wurde die Morphologie der behandelten A549 Zellen unter Verwendung von Fluoreszenzmikroskopie und Durchflusszytometrie analysiert.

3.3.1.2 AECII

Bei den AECII Zellen (*alveolar epithelial type II cells*) handelt es sich um primäre Lungenepithelzellen. Die Extraktion aus dem humanen OP-Material erfolgte im „Institut für klinische und experimentelle Pathologie“ („Forschungszentrum Borstel“, Deutschland) wie in [228] beschrieben.

Das frische Lungengewebe wurde zunächst mit sterilem einfachen PBS („Thermo Fisher Scientific“, USA) gespült. Um das restliche Blut zu entfernen, wurde das Gewebe seziiert und mit dem „AECII“ Puffer behandelt (1,9 mmol/l Kalziumchlorid-Dihydrat, 1,3 mmol/l Magnesiumsulfat-Heptahydrat, 136 mmol/l Natriumchlorid, 6,1 mmol/l Kaliumchlorid, 3,2

mmol/l Di-Natriumhydrogenphosphat-Dihydrat, 6,1 mmol/l Glucose, 9,9 mmol/l HEPES; „Merck-Millipore“, USA).

Um das Gewebe zu zersetzen, wurde zunächst „AECII“ Puffer mit 2mg/ml „Dispase II“ („Roche“, Schweiz) versetzt. Das Gewebestück wurde dann in den angesetzten Puffer gelegt und eine Stunde lang bei 37°C und 5% CO₂ inkubiert. Nach der Inkubation wurde die Zellsuspension unter Verwendung von Nylon-Filtern absteigender Porengröße (100µm, 50µm und 20µm; „Merck Millipore“, USA) gefiltert und 15min lang bei 478g zentrifugiert. Das Zellpellet wurde dann in 50ml „AECII“-Puffer resuspendiert, mit 0,001%iger „Accutase Cell Detachment Solution“ („Merck Millipore“, USA) sowie 20µg/ml „DNAse I“ („Roche“, Schweiz) versetzt und zu 10ml „BIOCOLL“ („Biochrome GmbH“, Deutschland) gegeben. Die Probe wurde 25min bei 478g zentrifugiert. Die Interphase wurde vorsichtig abpipettiert und dann bei 478g 15min mit „AECII“-Puffer gewaschen.

Um die Zellen des Immunsystems zu entfernen, wurden 5×10^7 Zellen ausgesät und 20min bei 37°C inkubiert. Während der Inkubation adhärten die Zellen des Immunsystems, die AECII blieben dagegen in Lösung und wurden nachfolgend vorsichtig abpipettiert. Um die restlichen CD45 positiven Zellen zu entfernen, wurden 1×10^7 Zellen in 80µl „MACS“-Puffer resuspendiert (PBS pH 7,2 mit 0.5% BSA und 2mmol/l EDTA; „Thermo Fisher Scientific“, USA), mit 20µL „CD45 MicroBeads, Human“ („Miltenyi Biotec GmbH“, Deutschland) versetzt und 15min unter Rotation inkubiert. Anschließend erfolgte die Depletion unter Verwendung von „LD“-Säulen („Miltenyi Biotec GmbH“, Deutschland). Die verbliebenen AECII Zellen wurden schließlich im Medium, bestehend aus 90% DMEM und 10% FBS, resuspendiert und bei Raumtemperatur nach Kiel transportiert. Die Kultivierung der AECII Zellen in Kiel erfolgte bei 37°C und 5% CO₂.

3.3.2 Herstellung der Zigaretten- (CSE) und E-Zigarettenextrakte (eCSE)

Die Zigaretten und E-Zigarettenextrakte wurden im „Institut für klinische und experimentelle Pathologie“ („Forschungszentrum Borstel“, Deutschland) hergestellt und auf Trockeneis verschickt.

Da die handelsüblichen Zigaretten je nach Charge stark in ihrer Zusammensetzung variieren können, wurden im Rahmen dieser Arbeit „3R4F – Kentucky University Reference Cigarettes“ („University of Kentucky“, USA) verwendet. Die Referenzzigaretten enthielten pro Zigarette 9,4mg Teer sowie 0,7mg Nikotin. Die Herstellung des CSE erfolgte, wie in Abb. 16 dargestellt. Eine Zigarette wurde in der Apparatur befestigt und innerhalb von 5min verbrannt, der dabei entstandene Zigarettenrauch wurde von einem Schlauch in einen

Kolben abgeführt. Das offene Ende des Schlauchs tauchte in DMEM. Um eine 100%ige CSE-Lösung zu erhalten, wurden drei Zigaretten in 30ml DMEM aufgenommen und dann unter Verwendung eines „Millipak Filters“ (Porengröße 0.22µm, „Merck Millipore“, USA) sterilisiert.

Die Herstellung des eCSE erfolgte ähnlich wie des CSE, nur dass anstelle der klassischen Zigarette die „DIPSE-eGo-Cigarette“ („DIPSE“, Deutschland) verwendet wurde. Das *Liquid* innerhalb der E-Zigarette wurde für fünf Sekunden erhitzt. Nach dem Erhitzen entstand der E-Zigarettenrauch, der wie bereits oben beschrieben in einen Schlauch entwich. Nach 25s wurde der Vorgang wiederholt. Um eine 100%ige eCSE-Lösung zu bekommen, wurde die E-Zigarette insgesamt 10 Mal erhitzt. Insgesamt wurden im Rahmen dieser Arbeit zwei verschiedene *Liquids* untersucht: eCSE mit Nikotin (24 mg/ml) und eCSE ohne Nikotin (0mg/ml) („DIPSE“, Deutschland).

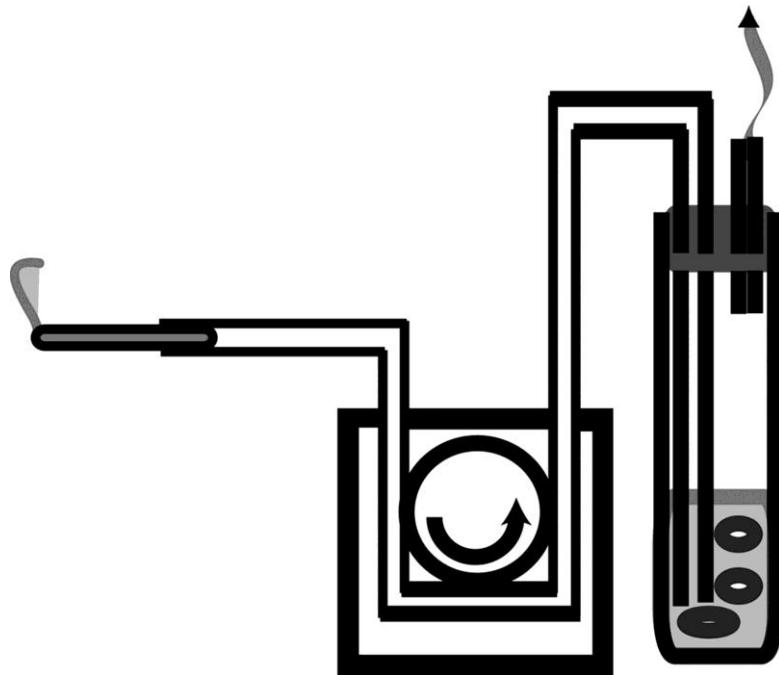


Abb. 16: Um einen Zigaretten oder E-Zigarettenextrakt herzustellen wurden je nach Fragestellung „3R4F – Kentucky University Reference Cigarettes“ („University of Kentucky“, USA) oder „DIPSE-eGo-Cigarette“ („DIPSE“, Deutschland) in der Apparatur montiert und verdampft. Der Rauch wurde in einen Schlauch abgeführt, der in einen Kolben führte. Das offene Ende des Schlauchs tauchte in 30ml DMEM. Um eine 100%ige CSE-Lösung zu erzeugen, wurden drei Zigaretten verbraucht. Für die Herstellung einer 100%igen eCSE-Lösung wurde das *Liquid* innerhalb der E-Zigarette 10 Mal für jeweils fünf Sekunden erhitzt. Anschließend wurden die Extrakte unter Verwendung eines Filters (Porengröße 0.22µm) sterilisiert. Die Abbildung stammt aus [229].

3.3.3 Stimulation der Lungenzelllinien mit CSE und eCSE

Für alle Experimente wurde ein Zellmedium angesetzt, bestehend aus 90% DMEM und 10% FBS. Um die Apoptose- und Viabilitätsmessungen durchzuführen, wurden pro *Well* einer 96-*Well*-Platte („Corning“, USA) 3.000 Zellen sowie 100µl Zellmedium pipettiert. Für Fluoreszenzmikroskopie, RNA-Seq., ATAC-Seq. und EPIC BC wurden 6-*Well*-Platten („Corning“, USA) verwendet. Pro *Well* wurden je 89.000 Zellen und 2ml Zellmedium gegeben. Als nächstes wurden die 96- und 6-*Well*-Platten 24 Stunden lang bei 37°C und 5% CO₂ inkubiert, währenddessen adhärten die A549 und AECII Zellen. Dann wurde das alte Zellmedium vorsichtig abpipettiert und mit dem frisch zubereiteten CSE- bzw. eCSE haltigem Medium verschiedener Konzentrationen ersetzt. Die Inkubation behandelter Zellen erfolgte 48 Stunden lang bei 37°C und 5% CO₂.

Für die A549 Zelllinie wurden für die Behandlungen mit CSE und eCSE (mit Nikotin) folgende Konzentrationen gewählt: 0% als Kontrolle, 0,1% als niedrige Konzentration, 1% als mittlere Konzentration und 5% als hohe Konzentration. Da die primäre AECII Zelllinie sehr empfindlich auf die nikotinhaltigen Extrakte reagierte, mussten die Konzentrationen an CSE und eCSE (mit Nikotin) wie folgt angepasst werden: 0% als Kontrolle, 0,01% als niedrige Konzentration, 0,1% als mittlere Konzentration und 5% als hohe Konzentration. Die Behandlung mit nikotinfreiem eCSE erfolgte bei beiden Zelllinien gleich: 0% als Kontrolle, 0,1% als niedrige Konzentration, 1% als mittlere Konzentration und 5% als hohe Konzentration. Im Ergebnisteil sind die verwendeten Konzentrationsangaben spezifiziert.

3.3.4 Messung der Apoptose

3.3.4.1 „ApoONE® Homogeneous Caspase-3/7 Assay“

Die Apoptose wurde unter Verwendung von „ApoONE® Homogeneous Caspase-3/7 Assay“ („Promega“, USA) drei Stunden nach der Behandlung bei A549 sowie AECII Zellen gemessen. Pro *Well* einer 96-*Well*-Platte wurden hierfür 100µl proprietäres „ApoONE®-Reagenz“ pipettiert. Dabei handelte es sich um einen Lysepuffer, der unter anderem ein nicht fluoreszierendes Caspase 3/7 Substrat enthielt (Rhodamine 110, bis-(N-CBZ-l-aspartyl-l-glutamyl-l-valyl-aspartic acid amide) (Z-DEVD-R110)). Während der 60-minütigen Zellyse, wurden die Caspasen 3 und 7 freigesetzt, die nachfolgend das DEVD-Peptid des Substrats abspalteten. Das dabei entstehende fluoreszierende Produkt konnte bei einer Wellenlänge von 499Ex/521Em quantifiziert werden (Abb. 17). Die Reaktion verlief bei Raumtemperatur

unter Lichtausschluss [230]. Alle Messungen erfolgten als technische Triplikate. Alle Versuche wurden mindestens ein weiteres Mal repliziert.

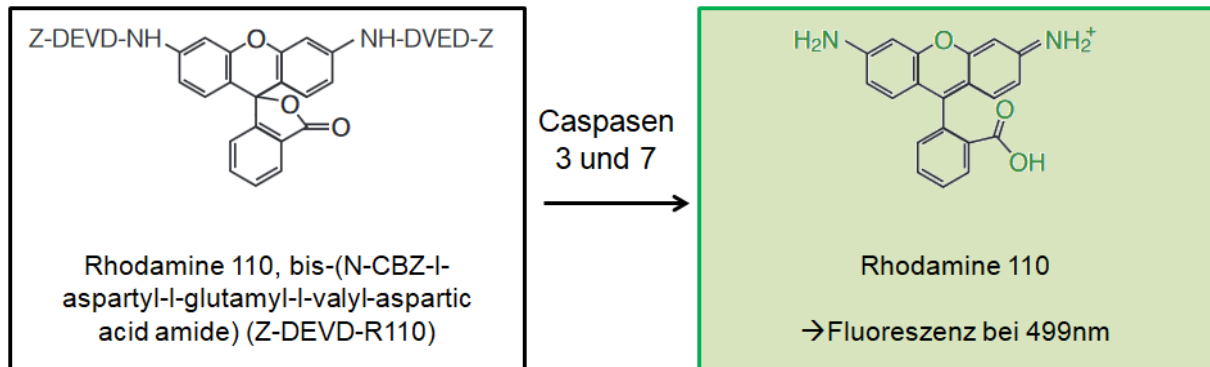


Abb. 17: Das im „ApoONE® Homogeneous Caspase-3/7 Assay“ enthaltene „ApoONE®-Reagenz“ bewirkte die Zellyse mit anschließender Freisetzung der Caspasen 3 und 7. Die Caspasen spalteten das nicht fluoreszierende Substrat so, dass Rhodamin 110 gebildet wurde. Bei einer Anregung mit Licht der Wellenlänge 499nm, fluoreszierte Rhodamin 110 grün und konnte bei 521nm detektiert werden. Die Abbildung wurde verändert nach [230].

3.3.4.2 „IncuCyte ZOOM Live Cell Analysis System“

Die Messung der Apoptoseraten in Echtzeit erfolgte unter Verwendung von „IncuCyte ZOOM Live Cell Analysis System“ („Essen Bioscience“, USA). Nach der Stimulation der A549 Zelllinie mit den Zigaretten- bzw. E-Zigarettenextrakten wurden pro Well einer 96-Well-Platte 50µl „IncuCyte® Annexin V Reagent“ pipettiert („Essen Bioscience“, USA). Die Zellen wurden 48 Stunden lang bei 37°C und 5% CO₂ inkubiert. Falls eine Zelle apoptotisch wurde, fluoreszierte das „IncuCyte® Annexin V Reagent“. Die Detektion der Apoptosesignale mit dem „IncuCyte ZOOM Live Cell Analysis System“ erfolgte stündlich. Die ermittelten Daten wurden unter Verwendung der „IncuCyte® S3 Base Software“ („Essen Bioscience“, USA) ausgewertet und visualisiert. Alle Messungen erfolgten als technische Triplikate. Alle Versuche wurden mindestens ein weiteres Mal repliziert.

3.3.5 Bestimmung der Viabilität

Die Messungen der Zellviabilität erfolgten mit dem „CellTiter-Blue® Cell Viability Assay“ („Promega“, USA) 48 Stunden nach der Zellstimulation. Hierfür wurde pro Well einer 96-Well-Platte 5µl „CellTiter-Blue®“ Reagenz pipettiert und 60min bei 37°C und 5% CO₂ inkubiert. Während der Inkubation reduzierten lebende Zellen das blaue Resazurin zum violetten Resorufin. Dieser Prozess fand in toten Zellen nicht statt, sodass der Anteil an

lebenden Zellen innerhalb eines *Wells* mit der Menge des gebildeten Resorufins korrelierte. Das Resorufin konnte fluorometrisch bei einer Wellenlänge von 560Ex/590Em quantifiziert werden (Abb. 18) [231].

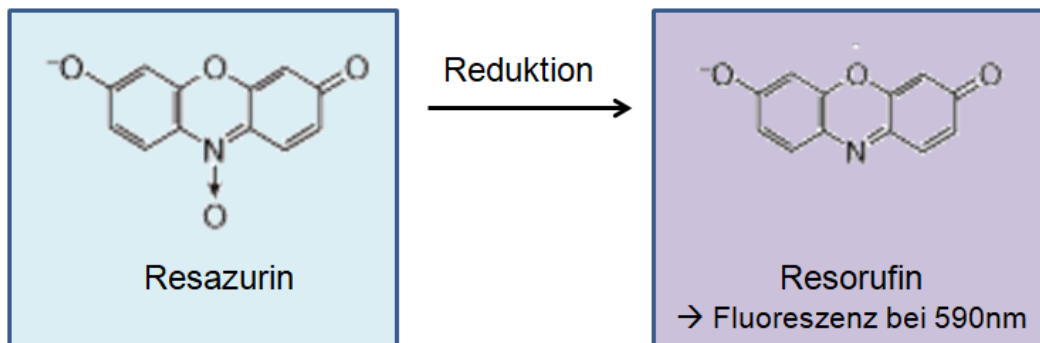


Abb. 18: Im „CellTiter-Blue® Cell Viability Assay“ war unter anderem das blau fluoreszierende Reazurin enthalten. Das Resazurin wurde von lebenden Zellen zum violett fluoreszierenden Resorufin umgewandelt. Die Menge des im *Well* einer 96-*Well*-Platte gebildeten Resorufins konnte bei einer Wellenlänge von 590nm quantifiziert werden und korrelierte mit der Anzahl an lebenden Zellen. Die Abbildung wurde verändert nach [231].

3.3.6 RNA-Sequenzierung (RNA-Seq.)

Die RNA-Sequenzierung ist eine NGS-basierte Methode zur genomweiten Untersuchung der Genexpression [232]. Im Gegensatz zur DNA-Sequenzierung (siehe Abschnitt 3.1.2.4) wird die RNA Probe vor *Library* Herstellung einer Depletion der ribosomalen RNA (rRNA) sowie cDNA Synthese unterzogen.

Der erste Schritt dient dazu, die ribosomale RNA zu eliminieren, die bis zu 90% der Gesamtprobe ausmachen kann [233]. Die rRNA ist zwar für die Translation essentiell, spielt jedoch bei der Transkription keine Rolle. Die rRNA-Depletion im Rahmen dieser Arbeit erfolgte unter Verwendung von „Ribo-Zero Gold rRNA Removal Kit (Human/Mouse/Rat)“ („Illumina“, USA), der nicht nur cytoplasmatische sondern auch die mitochondriale rRNA entfernt [234, 235]. Dabei wird rRNA an magnetische *Beads* gebunden und so effizient aus der Probe eliminiert (Abb. 19) [236].

Damit die im Abschnitt 3.1.2.4 beschriebene NGS *Library* Herstellung erfolgen kann, muss die verwendete Probe doppelsträngig sein. Daher wird im zweiten Schritt die in der Probe verbliebene, aufgereinigte und fragmentierte mRNA zunächst in die cDNA umgeschrieben. Im weiteren Verlauf wird der zweite DNA-Strang synthetisiert. Um die Sequenzieradapter zu ligieren werden die 3'-Enden der DNA-Stränge zunächst adenyliert. Schließlich werden die dUTPs enzymatisch entfernt und durch dTTPs ersetzt, die Probe wird in einer PCR-Reaktion

amplifiziert, quantifiziert und schließlich sequenziert. Für den hier beschriebenen Prozess wurde das „TruSeq Stranded mRNA Library Prep Kit“ („Illumina“, USA) verwendet (Abb. 19) [236].

Die genomweite Sequenzierung der hergestellten *Libraries* wurde von der Firma „Fasteris SA“ (Schweiz) auf dem „HiSeq4000“ („Illumina“, USA) durchgeführt (siehe Abschnitt 3.3.6.2).

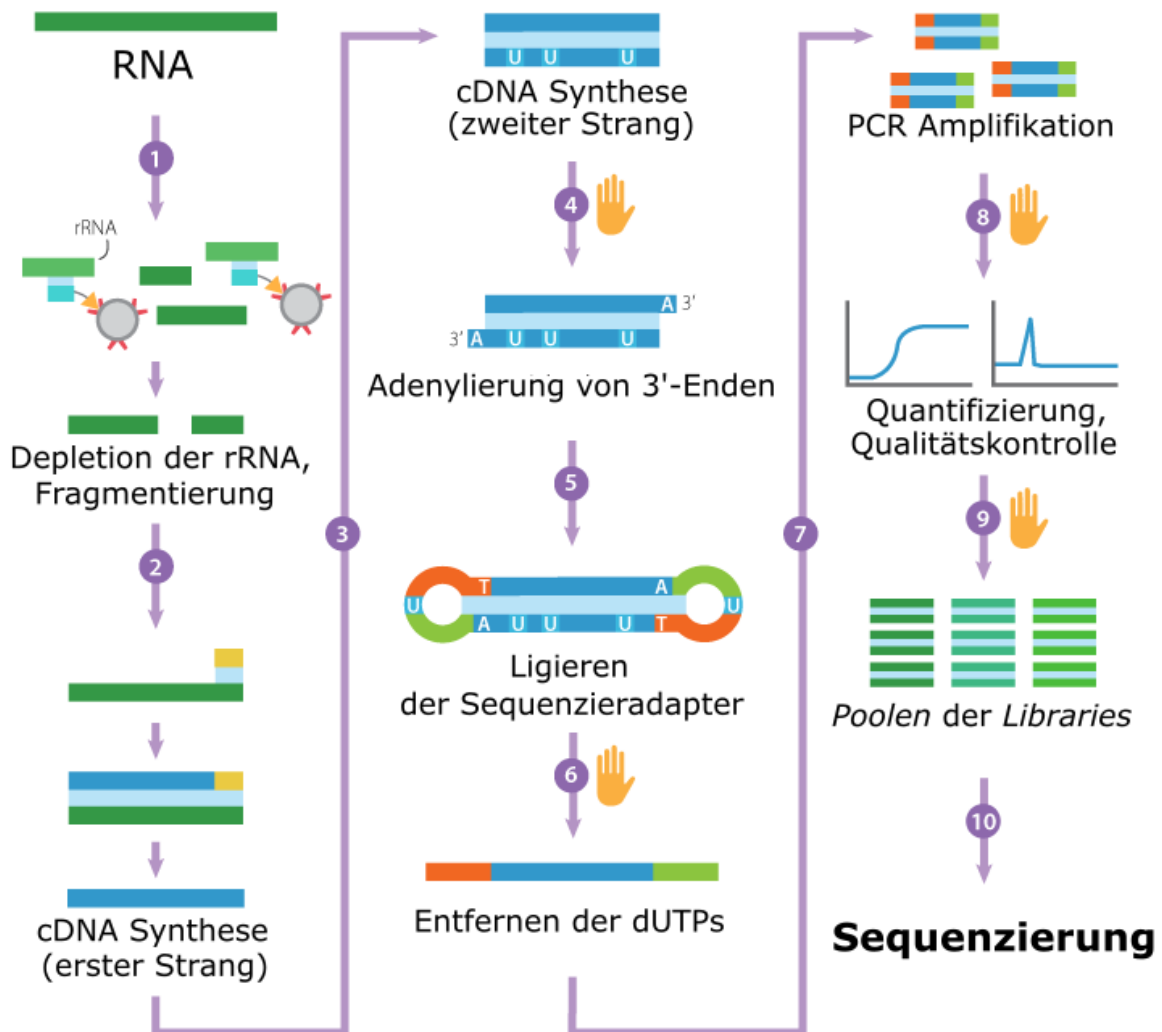


Abb. 19: Die Herstellung der RNA-Seq. *Libraries* erfolgte in mehreren Schritten. Zunächst wurde unter Verwendung des „Ribo-Zero Gold rRNA Removal Kit (Human/Mouse/Rat)“ die rRNA aus der Probe entfernt. Die verbliebene mRNA wurde mit dem „TruSeq Stranded mRNA Library Prep Kit“ aufgereinigt, fragmentiert und zur cDNA-Synthese verwendet. Im weiteren Verlauf wurde der zweite DNA-Strang erzeugt. Die 3'-Enden beider DNA-Stränge wurden adenyliert und zum Ligieren der Sequenzieradapter verwendet. Schließlich erfolgte eine enzymatische Entfernung von dUTPs, sowie PCR-Amplifikation der Probe. Die fertigen *Libraries* wurden dann quantifiziert, einer Qualitätskontrolle unterzogen und auf einer „HiSeq4000“ Plattform genomweit sequenziert. Die Abbildung wurde verändert nach [236].

3.3.6.1 RNA-Extraktion

Die RNA-Extraktion erfolgte unter Verwendung von „RNeasy Mini Kit“ („Qiagen“, Deutschland). Um die Zellen zu lysieren, wurden 350µl „Buffer RLT“ sowie 3,5µl 2-Mercapthoethanol („Sigma-Aldrich“, USA) auf ein Zellpellet bestehend aus 100.000 Zellen pipettiert. Der „Buffer RLT“ enthielt hohe Mengen an Guanidiniumthiocyanat, was eine sofortige Inaktivierung der freigesetzten RNasen bewirkte. Das Zelllysat wurde dann auf die „QIAshredder Spin Column“ („Qiagen“, Deutschland) überführt und 2min lang bei 10.000g homogenisiert. Dann wurde die Probe mit 350µl 70%igen Ethanol versetzt („Merck Millipore“, USA), vorsichtig resuspendiert, auf die „RNeasy Spin Column“ („Qiagen“, Deutschland) transferiert und 15s lang bei 8.000g zentrifugiert. Bei der „RNeasy Spin Column“ handelte es sich um eine Silica-Säule, die bevorzugt RNA-Fragmente länger als 200bp immobilisierte. Das Waschen der Säule erfolgte in drei Schritten. Zunächst wurde die Säule mit 700µl „Buffer RW1“ versetzt und 15s lang bei 8.000g gewaschen. Dann wurden auf die Säule 500µl „Buffer RPE“ pipettiert und wieder 15s bei 8.000g zentrifugiert. Schließlich wurden auf die Säule weitere 500µl „Buffer RPE“ gegeben und 2min bei 8.000g zentrifugiert. Im nächsten Schritt wurde die „RNeasy Spin Column“ in ein neues 2ml Auffanggefäß überführt und eine weitere Minute bei 8.000g zentrifugiert. Die Elution erfolgte mit 50µl RNase freiem Wasser mit anschließender einminütiger Zentrifugation bei 8.000g.

3.3.6.2 Herstellung der RNA-Seq. *Libraries*, Sequenzierung

Die Herstellung der RNA-Seq. *Libraries* sowie die 150bp *single end* (SE) Sequenzierung auf der „HiSeq4000“ Plattform erfolgten durch „Fasteris SA“ (Genf, Schweiz) nach firmeneigenen Protokollen. Pro Probe wurden ca. 10-20 Millionen *Reads* generiert.

3.3.6.3 Bioinformatische Auswertung der RNA-Seq. Daten

Um die RNA-Seq. Daten auszuwerten, mussten mehrere Schritte etabliert werden (Abb. 20).

Wie bereits im Abschnitt 3.1.2.5 beschrieben, wurde zunächst mit Hilfe der „FastQC“ Software die Qualität der Rohdaten überprüft [207]. Dann wurden unter Verwendung von „Cutadapt“ die Adaptersequenzen eliminiert [208] sowie das *Alignment* gegen das „HG19“ Referenzgenom durchgeführt. Auch in diesem Fall wurde hierfür der „Segemehl“ Algorithmus gewählt [200, 209] (siehe Abschnitte 3.3.6.3.1, 3.3.6.3.2).

Nach dem *Alignment* erfolgte die Identifikation differentiell exprimierter Gene. Hierfür wurde zunächst die „DoDas“ Software (Version 1.0, „Interdisziplinäres Zentrum für Bioinformatik“, Deutschland) verwendet (siehe Abschnitt 3.3.6.3.2). Dabei wurde pro Transkript die Anzahl an *Reads* detektiert. Dieser Parameter wird auch als *Count* bezeichnet. Die *Counts* wurden im nächsten Schritt in die „edgeR“ Software (Version 3.18.1, „The Walter and Eliza Hall Institute of Medical Research“, „The University of Melbourne“, Australia) geladen und normalisiert (siehe Abschnitt 3.3.6.3.2). Die Normalisierung erfolgte unter Verwendung der negativen Binomialverteilung. Dabei wurden viele Faktoren wie z.B. Transkriptlängen, CG-Gehalt oder *Coverage* berücksichtigt. Die „edgeR“ Software ist vor allem für ihre Sensitivität bekannt, sodass auch sehr schwach exprimierte Transkripte zuverlässig detektiert werden konnten [237].

Im letzten Schritt wurden unter Verwendung von „R“ [203] nur die Transkripte ausgewählt, die sowohl bei der A549 Zelllinie als auch bei primären AECII Zellen nach der Behandlung mit Zigaretten- bzw. E-Zigarettenextrakten differentiell exprimiert waren (siehe Abschnitt 3.3.6.3.2). Die verwendeten Aufrufe sind im Abschnitt 9.2 detailliert beschrieben. Die Analysen erfolgten auf dem „NEC-HPC-Linux-Cluster“ in Kooperation mit Dr. Helene Kretzmer.

Um die zelluläre Funktion identifizierter Gene zu untersuchen, wurde schließlich eine *Gene Ontology* (GO) Analyse durchgeführt. Hierfür wurde die „WebGestalt“ Software („Graduate School in Genome Science and Technology“, „University of Tennessee-Oak Ridge National Laboratory“, USA) verwendet [238-240] (siehe Abschnitt 3.3.6.3.2).

3.3.6.3.1 Qualitätskontrolle und Prozessieren der Daten

Die RNA-Seq. Rohdaten wurden im „FastQ“-Format geliefert. Qualitätskontrolle und Prozessieren der Daten erfolgten wie bereits in Abschnitten 3.1.2.5.1 und 3.1.2.5.2 beschrieben. Da es sich diesmal um eine *single end* Sequenzierung handelte, wurde der „Cutadapt“- Befehl wie folgt geändert:

```
cutadapt -q 20 -o 5 --minimum-length 35 -a AGATCGGAAGAGC -o <Name>.clipped.fastq.gz  
<Name>.fastq.gz &><Name>.clipping.stats
```


3.3.6.3.2 Auswertung der prozessierten Daten

Das *Alignment* gegen das „HG19“ Referenzgenom erfolgte wie im Abschnitt 3.1.2.5.3 beschrieben unter Verwendung von „Segemehl“. Der Aufruf wurde wie folgt modifiziert:

```
segemehl.x -s -S -t 10 -d hg19.fa -i hg19.idx -q <Name>.clipped.fastq.gz -o <Folder>/<Name>.sam  
-u out.unmapped.fq
```

Um die *Counts* zu ermitteln, wurden die *Alignment* Ergebnisse im „BAM“-Format gespeichert und in das „DoDas“ Tool geladen. Der Befehl lautete:

```
python pre_segemehl.py -t 10 -d hg19.fa -a gencode.v19.genes.bed -l DoDAS_liste.bed
```

Die ermittelten *Counts* wurden dann in das „edgeR“ Programm geladen und für die Identifikation differentiell exprimierter Transkripte verwendet. Die verwendeten Skripte sind im Abschnitt 9.2 detailliert beschrieben. Schließlich wurden nur Transkripte mit einem p-Wert kleiner als 0,05 ausgegeben. Die Befehle lauteten:

```
sh run.edgeR.sh <control> <treatment>  
less file | perl -ane 'if($F[3]<=0.05){print $_}'
```

Um besonders starke Effekte des Rauchens auf die Genregulation analysieren zu können, wurden unter Verwendung von „R“ nur die Transkripte ausgewählt, die bei beiden Zelllinien nach der Behandlung mit Zigaretten- bzw. E-Zigarettenextrakten differentiell exprimiert waren. Der „log2 fold-change“-Wert betrug dabei mindestens 0,25. Die hierfür verwendeten Befehle sind im Abschnitt 9.2 dargestellt.

Um die Rolle der identifizierten Gene innerhalb einer Zelle zu verstehen, erfolgte eine GO Analyse unter Verwendung des „WebGestalt“ Tools. Die optimierten Parameter lauteten:

- **Organismus:** „*hsapiens*“
- **Methode:** „*overpresentation enrichment analysis (ORA)*“
- **Datenbank:** „*gene ontology: biological process, molecular function*“
- **ID:** „*ensembl_gene_id*“
- **Referenz:** „*genome_protein-coding*“

Die CSE sowie eCSE (ohne Nikotin) Ergebnisse wurden nach FDR kleiner 0,05 gefiltert. Da dieses Vorgehen bei eCSE (mit Nikotin) keine signifikanten Ergebnisse lieferte, erfolgte in diesem Fall das Filtern nach p-Wert kleiner 0,001.

3.3.7 Assay for Transposase-Accessible Chromatin Using Sequencing (ATAC-Seq.)

Die ATAC-Sequenzierung (engl. *assay for transposase-accessible chromatin using sequencing* (ATAC-Seq.)) ist eine NGS-basierte Methode zur genomweiten Untersuchung des Chromatinstatus. Im Gegensatz zur DNA-Sequenzierung (Abschnitt 3.1.2.4) werden hier nur die offenen Chromatinbereiche während der *Library* Herstellung angereichert und darauffolgend sequenziert. Hierfür wird das Enzym „Tn5 Transposase“ („Illumina“, USA) verwendet [241].

Das Prinzip der ATAC-Sequenzierung ist in der Abbildung 21 dargestellt. Dabei erfolgt zunächst die Isolierung und Lyse der Zellkerne. Das dabei freigesetzte Chromatin wird dann mit der „Tn5 Transposase“ behandelt. Das Enzym bindet und schneidet offenes Chromatin, auch Euchromatin genannt. Gleichzeitig werden an beide Enden der so entstandenen DNA-Fragmente Sequenzieradapter ligiert. Das geschlossene Chromatin, oder Heterochromatin, bleibt unprozessiert und kann im weiteren Verlauf die *Flow Cell* nicht binden. Somit wird nach dem Amplifizieren nur das Euchromatin sequenziert [242] (siehe Abschnitt 3.3.7.1).

Da es sich hier um eine genomweite Untersuchung des Chromatinstatus handelte, erfolgte die Sequenzierung unter Verwendung des „HiSeq4000“ (siehe Abschnitt 3.3.7.2).

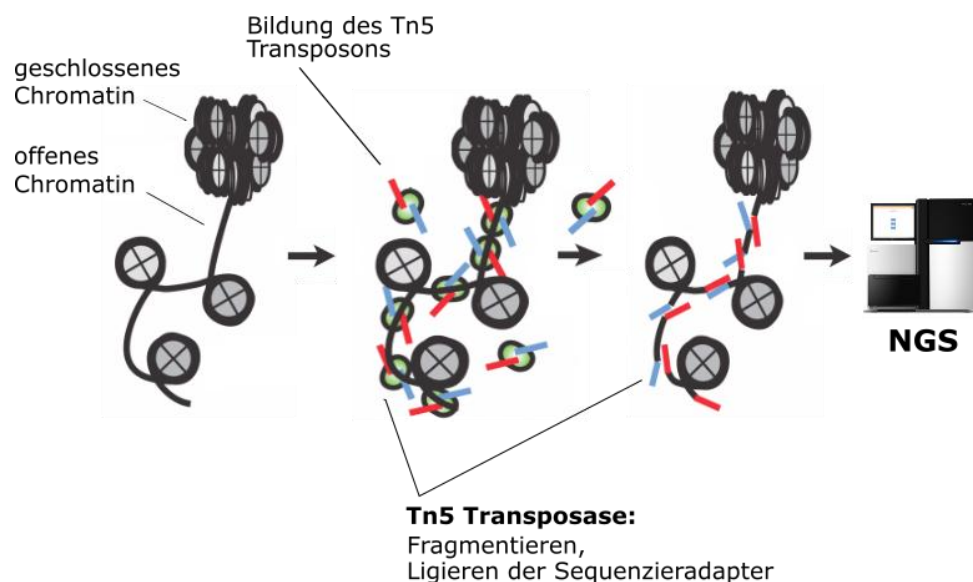


Abb. 21: Die Herstellung der ATAC-Seq. *Library* erfolgt in mehreren Schritten. Zunächst werden die Zellkerne isoliert und lysiert. Das freigesetzte Chromatin wird mit „Tn5 Transposase“ behandelt. Das gebildete Transposom bindet und schneidet offene Chromatinbereiche. Gleichzeitig werden vom Enzym Sequenzieradapter ligiert. Im weiteren Verlauf können die so gebildeten DNA-Fragmente amplifiziert und sequenziert werden. Die Abbildung wurde verändert nach [241].

3.3.7.1. Herstellung der ATAC-Seq. *Libraries*

Die Herstellung der ATAC-Seq. *Libraries* erfolgte nach Protokoll von Buenrostro *et.al.* [242] unter Verwendung des „Nextera DNA Library Preparation Kits“ („Illumina“, USA).

Zunächst wurden 25.000 bis 50.000 Zellen in 500µl einfacher PBS aufgenommen und fünf Minuten lang bei 500g und 4°C zentrifugiert.

Im nächsten Schritt wurden die Zellen lysiert. Hierfür wurde das Zellpellet in 500µl Lysepuffer (10mM Tris-HCl (pH 7,4), 10mM NaCl, 3mM MgCl₂, 0.1% IGEPAL CA-630; „Sigma-Aldrich“, USA) resuspendiert und 10min lang bei 500g und 4°C zentrifugiert. Während der Zentrifugation bildeten die freigesetzten Zellkerne ein weißliches Pellet.

Der Überstand wurde vorsichtig abpipettiert und das Pellet mit 50µl Transposasepuffer versetzt. Der Transposasepuffer bestand aus 25µl H₂O, 22,5µl „TD 2x Reaction Buffer“ sowie 2,5µl „TDE1 Nextera Tn5 Transposase“ („Nextera DNA Library Preparation Kit“). Die darauffolgende 30-minütige Transposasereaktion erfolgte unter leichtem Schütteln bei 37°C.

Die freigesetzten DNA-Fragmente wurden dann unter Verwendung des „Qiagen MinElute PCR Purification Kits“ („Qiagen“, Deutschland) aufgereinigt. Alle Zentrifugationsschritte erfolgten bei 10.000g. Zunächst wurden zu den 50µl der Probe 250µl „Buffer PB“ gegeben, gut gemischt und auf die „MinElute“ Silica-Säule pipettiert. Die Säule wurde dann eine Minute lang zentrifugiert. Der Durchfluss wurde verworfen. Um die Silica-Säule zu waschen, wurden darauf 750µl „Buffer PE“ pipettiert und für eine Minute zentrifugiert. Der Durchfluss wurde verworfen und der Zentrifugationsschritt wiederholt. Als nächstes wurde die Säule in ein sauberes 1,5ml Reaktionsgefäß überführt, mit 10µl „Buffer EB“ (10mM Tris·Cl, pH 8,5) versetzt und für eine Minute bei Raumtemperatur inkubiert. Die Elution erfolgte schließlich eine Minute lang bei 10.000g.

Die aufgereinigten DNA-Fragmente wurden unter Verwendung von PCR amplifiziert. Hierfür wurden 10µl DNA mit 10µl H₂O, 25µl „NEBNext High-Fidelity 2x PCR Master Mix“ („Life Technologies“, USA), 2,5µl 25µM „Custom Nextera PCR Primer 1“ und 2,5µl 25µM „Custom Nextera PCR Primer 2“ (Tab. 6) versetzt. Die PCR-Reaktion erfolgte im *Thermocycler*, das Programm war wie folgt aufgebaut: 30s 98°C, 10 Zyklen bestehend aus 10s 98°C, 30s 63°C sowie 1min 72°C.

Das PCR Produkt wurde dann, wie bereits beschrieben, unter Verwendung des „Qiagen MinElute PCR Purification Kits“ aufgereinigt. Die anschließende Quantifizierung und Qualitätskontrolle der hergestellten *Libraries* erfolgte unter Verwendung von „Qubit dsDNA

High Sensitivity Assay“ sowie „Agilent 2100 High Sensitivity DNA“ Kits entsprechend der Vorschriften der Hersteller, wie im Abschnitt 3.1.2.2 beschrieben.

Tab. 6: Die verwendeten „Custom Nextera PCR Primer“ [241].

| Primer | Sequenz |
|---------------------------------|---|
| „Custom Nextera PCR Primer 1“ | AATGATACGGCGACCACCGAGATCTACACTCGTC GGCAGCGTCAGATGTG |
| „Custom Nextera PCR Primer 2.1“ | CAAGCAGAAGACGGCATACGAGATTCGCCTTAGT CTCGTGGGCTCGGAGATGT |
| „Custom Nextera PCR Primer 2.2“ | CAAGCAGAAGACGGCATACGAGATCTAGTACGGT CTCGTGGGCTCGGAGATGT |
| „Custom Nextera PCR Primer 2.3“ | CAAGCAGAAGACGGCATACGAGATTTCTGCCTGT CTCGTGGGCTCGGAGATGT |

3.3.7.2 Sequenzierung der ATAC *Libraries*

Die Sequenzierung der ATAC-Seq. *Libraries* wurde auf der „HiSeq4000“ Plattform bei „BGI“ (Hongkong, China) durchgeführt. Dabei wurden pro *Lane* drei *Libraries gepoolt* und anschließend im 50 PE Modus sequenziert.

3.3.7.3 Bioinformatische Auswertung der Sequenzierdaten

Für die Auswertung der ATAC-Seq. Rohdaten wurden mehrere Schritte etabliert (Abb. 22).

Die ersten Schritte umfassten das Überprüfen der Sequenzierqualität, Prozessieren der Rohdaten (siehe Abschnitt 3.3.7.3.1) sowie das anschließende *Alignment* gegen das „HG19“ Referenzgenom (siehe Abschnitt 3.3.7.3.2). Hierfür wurden „FastQC“, „Cutadapt“ und „Segemehl“ verwendet.

Genauso wie andere DNA *Footprinting* Methoden, entstehen beim ATAC-Seq. systematische Fehler, die sogenannten *Bias*, die im nächsten Schritt eliminiert wurden. Beim ersten *Bias* handelte es sich um PCR-Duplikate. Um dieses Problem zu lösen, wurde die „Picard“ Software (Version 2.13.2, „Broad Institute“, USA) optimiert [243, 244] (siehe Abschnitt 3.3.7.3.2). Der zweite *Bias* wurde durch die „Tn5 Transposase“ verursacht. Die Transposon-Dimere fügen die Sequenzieradapter nicht direkt nebeneinander an, sondern in

einem Abstand von neun Basenpaaren. Um also weiterhin den Startpunkt eines *Reads* als Mittelpunkt jeweiliger Transposonbindestelle behandeln zu können, mussten alle *Reads* des positiven DNA-Strangs um vier Nukleotide *upstream*, und alle *Reads* des negativen DNA-Strang um fünf Nukleotide *downstream* verschoben werden (siehe Abschnitt 3.3.7.3.2).

Nachdem die *Bias* korrigiert wurden, konnte das *Peak Calling* unter Verwendung der „MACS2“ Software (Version 2.1.1.20160309, „Python Software Foundation“, USA) erfolgen [245] (siehe Abschnitt 3.3.7.3.2). Wie bereits beschrieben wird bei ATAC-Seq. nur das offene Chromatin, das Euchromatin, sequenziert. Das geschlossene Chromatin, das Heterochromatin, wird von der „Tn5 Transposase“ nicht prozessiert. Diese Tatsache wird beim Betrachten der Abbildung 23 deutlich. Innerhalb des Genoms gibt es Stellen, die kaum *Reads* haben (Heterochromatin), und Bereiche, die sehr viele *Reads* haben (Euchromatin). Dabei sind die *Reads* so angeordnet, dass *Peaks* entstehen. „MACS2“ Software ist in der Lage solche *Peaks* zu detektieren, zu normalisieren und schließlich deren Koordinaten in einer „BED“-Datei zusammenzufassen. Die „BED“-Datei kann dann im späteren Verlauf für integrative Analysen sowie Visualisierung mit „R“ verwendet werden. Die bioinformatische Auswertung der Rohdaten erfolgte auf dem „NEC-HPC-Linux-Cluster“ in Kooperation mit Dr. Helene Kretzmer.

3.3.7.3.1 Qualitätskontrolle und Prozessieren der Daten

Die ATAC-Seq. Rohdaten wurden im „FastQ“-Format geliefert. Qualitätskontrolle und Prozessieren der Daten erfolgten wie bereits in Abschnitten 3.1.2.5.1, 3.1.2.5.2 beschrieben. Der „Cutadapt“- Befehl lautete wie folgt:

```
cutadapt -a CTGTCTCTTATACACATCT -A CTGTCTCTTATACACATCT -q 20 --minimum-length 25  
-o <Name_Read_1>.cutadapt.fq.gz -p < Name_Read_2>.cutadapt.fq.gz < Name_Read_1>.fq.gz <  
Name_Read_2>.fq.gz
```

3.3.7.3.2 Auswertung der prozessierten Daten

Das *Alignment* gegen das „HG19“ Referenzgenom erfolgte wie im Abschnitt 3.1.2.5.3 beschrieben. Der Befehl für die ATAC-Seq. Daten wurde wie folgt modifiziert:

```
segemehl.x -s -t 5 -d hg19.fa -i hg19.idx -q <Name_Read_1>.cutadapt.fq.gz -p <Name_Read_2>  
.cutadapt.fq.gz -o <Folder>/<Name>.sam -u out.unmapped.fq
```

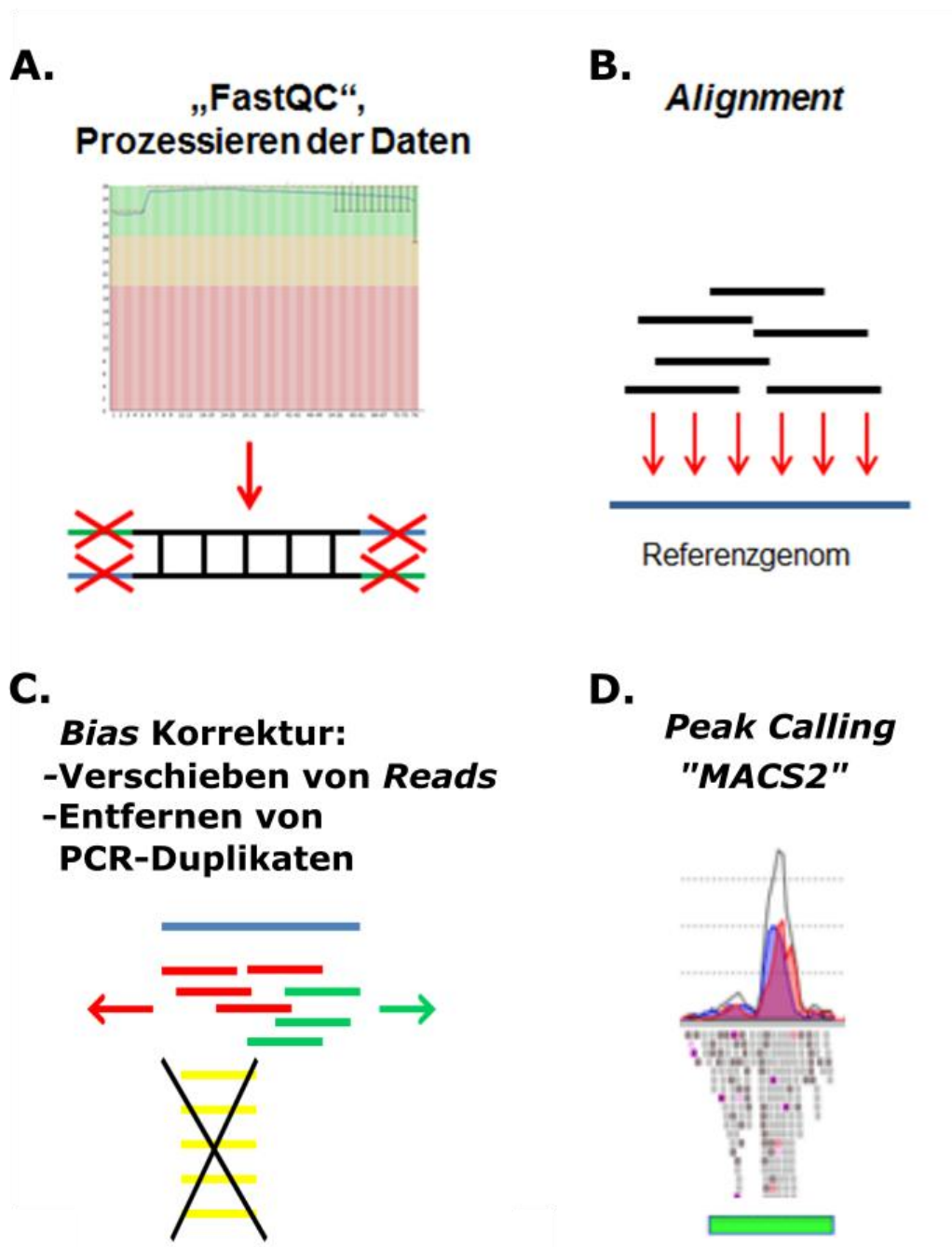


Abb. 22: Die Auswertung der ATAC-Seq. Rohdaten erfolgte in mehreren Schritten. **A.** Zunächst wurden die Rohdaten unter Verwendung von „FastQC“ einer Qualitätskontrolle unterzogen und anschließend mit „Cutadapt“ prozessiert. **B.** Die prozessierten Daten wurden dann gegen das „HG19“ Referenzgenom *aligned*. **C.** Um die *Bias* Korrektur durchzuführen, wurden zunächst PCR-Duplikate (gelb markiert) unter Verwendung von „Picard“ eliminiert. Alle *Reads* des positiven DNA-Stranges (rot markiert) wurden um vier Nukleotide *upstream* verschoben, dagegen *Reads* des negativen DNA-Stranges (grün markiert) um 5nt *downstream*. **D.** Schließlich wurde der „MACS“ *Peak Caller* zum Identifizieren des Euchromatins verwendet.

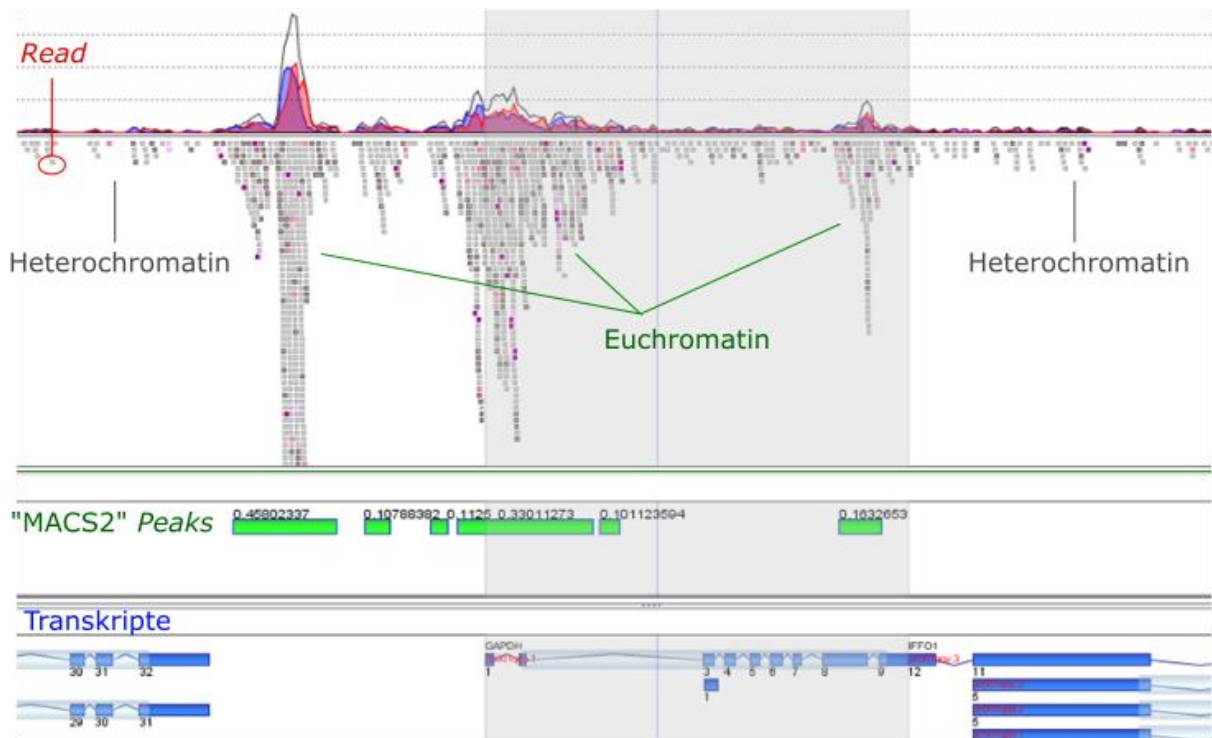


Abb. 23: Für die Visualisierung der ATAC-Seq. Daten wurde die „StrandNGS“ Software (Version 3.1, „Avadis[®]“, USA) verwendet. Die *Reads* wurden gegen das „HG19“ Genom *aligned*. Die geschlossenen Bereiche des Chromatins (Heterochromatin) enthielten so gut wie keine *Reads*. Die offenen Bereiche des Chromatins (Euchromatin) enthielten viele *Reads*, bildeten die sogenannten *Peaks* und konnten mit „MACS2“ Software identifiziert werden (grüne Balken entsprachen der Länge eines *Peaks*).

Die *Alignments* wurden in „BAM“-Files ausgegeben. Um die PCR-Duplikate zu entfernen wurde die „Picard“-Software verwendet [243, 244]. Die Software analysierte die 5'-Enden sowie die Orientierungen jedes *Read*-Paares. Die *Read*-Paare mit gleichen 5'-Koordinaten und Orientierungen wurden als Duplikate markiert. Das *Read*-Paar mit der besten Sequenzierqualität wurde beibehalten, der Rest dagegen aus dem Datensatz entfernt [243]. Eine Zusammenfassung über die Anzahl an entfernten PCR-Duplikaten sowie Informationen über die Größe der übriggebliebenen *Library*, wurde in einer Textdatei ausgegeben. Der verwendete „Picard“-Befehl war:

```
java -jar picard.jar MarkDuplicates I=<Name>.bam O=<marked_duplicates_Name>.bam M=<marked_dup_metrics_Name>.txt REMOVE_DUPLICATES=true ASSUME_SORTED= true VALIDATION_STRINGENCY=LENIENT
```

Im nächsten Schritt wurden alle *Reads* des positiven DNA-Strangs um vier Nukleotide *upstream*, und alle *Reads* des negativen DNA-Strangs um fünf Nukleotide *downstream*

Bei dem „Infinium I“ Assay werden zwei *Bead* Typen unterschieden, einer für den methylierten und einer für den unmethylierten CpG *Locus*. Die *Beads* für den unmethylierten *Locus* tragen am 3'-Ende des *Linkers* ein Adenin, die *Beads* für den methylierten *Locus* dagegen ein Guanin. So kann die zuvor bisulfitkonvertierte, fragmentierte und denaturierte DNA-Probe entsprechend ihrem Methylierungsstatus mit dem komplementären *Bead* hybridisieren. Anschließend kommt es zur sogenannten *Single Base Extension* (SBE), bei der ein Fluorophor eingebaut wird. Nach dem Scannen des Chips kann das Signal dem *Bead* mit dem entsprechenden Methylierungsstatus zugeordnet werden (Abb. 24A) [246, 247].

Bei dem „Infinium II“ Assay wird pro CpG *Locus* nur ein *Bead* Typ verwendet. Unabhängig vom Methylierungsstatus der Probe wird das DNA-Fragment an das *Bead* hybridisiert. Wenn das CpG methyliert vorliegt, wird am 3'-Ende des *Linkers* ein Guanin eingebaut, falls nicht ein Adenin. Beide Nukleotide sind mit unterschiedlichen Fluorophoren markiert und können somit nachfolgend während des Scans voneinander unterschieden werden (Abb. 24B) [246, 247].

Die DNA-Methylierungsrate eines CpG *Locus* wird als „avg.beta“-Wert angegeben. Hierfür wird das Verhältnis aus der Intensität (I) des methylierten Allels zu der Summe der Intensitäten des methylierten und unmethylierten Allels gebildet.

$$"avg. beta" = \frac{I_{methyliert}}{I_{methyliert} + I_{unmethyliert} + 100}$$

Der „avg.beta“-Wert ist demnach eine Zahl zwischen 0 und 1, wobei 0 einer Methylierung von 0% und 1 einer Methylierung von 100% entspricht [248].

3.3.8.1 DNA-Extraktion

Die DNA-Extraktion erfolgte unter Verwendung des „QIAamp DNA Mini Kits“ („Qiagen“, Deutschland).

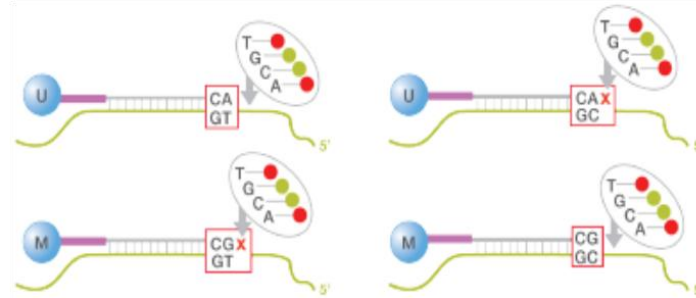
Hierfür wurden 150.000 Zellen fünf Minuten lang bei 300g pelletiert. Der Überstand wurde verworfen, das Zellpellet in 200µl einfacher PBS resuspendiert und mit 20µl „QIAGEN Proteinase K“ („Qiagen“, Deutschland) versetzt. Um die Zellen zu lysieren, wurde zu der Probe 200µl „Buffer AL“ gegeben. Dann wurden die Zellen 15s lang gemischt und für 10min bei 56°C inkubiert. Im nächsten Schritt erfolgte die DNA-Fällung. Hierfür wurde die Probe kurz zentrifugiert, mit 200µl absolutem Ethanol versetzt und anschließend gut gemischt. Um

die gefällte DNA aufzureinigen, wurde die Probe auf eine „QIAamp Mini Spin“-Säule („Qiagen“, Deutschland) pipettiert. Nach anschließender einminütiger Zentrifugation bei 6.000g, wurde die Säule in ein frisches Auffanggefäß transferiert, mit 500µl „Buffer AW1“ versetzt und wieder eine Minute lang bei 6.000g zentrifugiert. Dann wurden auf die Säule 500µl „Buffer AW2“ pipettiert. Nach einer drei Minuten langen Zentrifugation bei 10.000g, wurde die Säule in ein weiteres sauberes Auffanggefäß überführt und eine weitere Minute bei 10.000 g zentrifugiert. Um die aufgereinigte DNA zu eluieren, wurde die „QIAamp Mini Spin“-Säule in einen 1,5ml Reaktionsgefäß platziert und mit 50µl dH₂O versetzt. Schließlich wurde die Säule für jeweils eine Minute bei Raumtemperatur inkubiert und bei 6.000g zentrifugiert. Die Quantifizierung der extrahierten Proben erfolgte unter Verwendung des „Qubit dsDNA High Sensitivity Assay Kit“ wie im Abschnitt 3.1.2.2 beschrieben.

A. Infinium I

unmethylierter CpG Locus

methylierter CpG Locus



B. Infinium II

unmethylierter CpG Locus

methylierter CpG Locus



Abb. 24: Das EPIC BC umfasst 850.000 CpG Loci (rote Markierung) und basiert sowohl auf dem „Infinium I“ als auch „Infinium II“ Assay. **A.** Das „Infinium I“ Assay beinhaltet Beads für unmethylierte (U) und methylierte Loci (M). Die Beads für das unmethylierte CpG tragen am 3'-Ende ein Adenin, dagegen die Beads für das methylierte ein Guanin. Die bisulfitkonvertierte DNA (grüne Markierung) hybridisiert entsprechend des Methylierungsstatus der Region an das passende Bead. Während der darauffolgenden SBE kommt es zum Einbau eines Fluorophors, welches nach dem Scannen des Chips den Methylierungsstatus des jeweiligen CpG definiert. **B.** Für das „Infinium II“ Assay wird nur ein Bead Typ verwendet. Nach der DNA-Hybridisierung erfolgt auch in diesem Fall die SBE. Dabei wird bei einem unmethylierten CpG Locus am 3'-Ende des Linkers ein Adenin und bei einem methylierten CpG Locus ein Guanin eingebaut. Beide Nukleotide sind mit unterschiedlichen Fluorophoren markiert. Beim anschließenden Scannen lassen die Fluoreszenzsignale auf den Methylierungsstatus der Loci zurückschließen. Die Abbildung wurde verändert nach [246].

3.3.8.2 Bisulfitkonvertierung

Die Bisulfitkonvertierung der extrahierten gDNA Proben wurde mit dem „EZ DNA Methylation Kit“ („Zymo Research“, USA) durchgeführt. Das Prinzip der Bisulfitkonvertierung wurde bereits im Abschnitt 3.1.2.3 beschrieben.

Dafür wurden 500ng DNA in 45µl dH₂O gelöst, mit 5µl „M-Dilution Buffer“ versetzt und 15min lang bei 37°C inkubiert. 100µl „CT Conversion Reagent“ wurden zu der Probe pipettiert und bei 50°C 12 Stunden lang inkubiert. Nach der Inkubation wurden die Proben für 10min auf Eis inkubiert, mit 400µl „M-Binding Buffer“ versetzt, auf die „Zymo-SpinTM IC“-Säule pipettiert und bei 10.000g 30s lang zentrifugiert. Dann wurde die Säule mit 100µl „M-Wash Buffer“ für 30s bei 10.000g gewaschen und 20min lang mit 200µl „M-Desulphonation Buffer“ behandelt. Als nächstes wurde der „M-Desulphonation Buffer“ während der 30s langen Zentrifugation bei 10.000g entfernt. 200µl „M-Wash Buffer“ wurden auf die „Zymo-SpinTM IC“-Säule pipettiert. Anschließend wurde die Säule 30s lang bei 10.000g zentrifugiert. Schließlich erfolgte die Elution der konvertierten DNA-Probe. Hierfür wurden auf die Säule 10µl „M-Elution Buffer“ gegeben, die darauffolgende Zentrifugation erfolgte für 30s bei 10.000g.

3.3.8.3 Hybridisierung des EPIC BC

Zunächst erfolgte eine genomweite Amplifikation der bisulfitkonvertierten DNA. Hierfür wurden 500ng der Probe auf 4µl Volumen mit dH₂O verdünnt und anschließend in eine 96-Well-Platte („Thermo Fisher Scientific“, USA) pipettiert. Dann wurde die Probe mit 20µl „MA1“ sowie 4µl 0,1N NaOH versetzt, eine Minute lang bei 1.600rpm geschüttelt, für eine weitere Minute bei 280g zentrifugiert und anschließend 10min lang bei Raumtemperatur inkubiert. Nach der Inkubation wurden zu der Probe 68µl „RPM“ sowie 75µl „MSM“ pipettiert. Die Platte wurde wieder je eine Minute lang bei 1.600rpm geschüttelt, bei 280g zentrifugiert und anschließend für 24 Stunden im Hybridisierungsofen („Illumina“, USA) inkubiert.

Nach der Inkubation wurde die Probe enzymatisch fragmentiert. Hierfür wurde die 96-Well-Platte zunächst für eine Minute bei 280g zentrifugiert. Dann wurden zu der Probe 50µl „FMS“ pipettiert. Die Platte wurde für jeweils eine Minute bei 1.600rpm geschüttelt und bei 280g erneut zentrifugiert. Anschließend erfolgte eine einstündige Inkubation bei 37°C.

Nach der Inkubation erfolgte die Präzipitation fragmentierter DNA. Hierfür wurde die Probe zunächst mit 100µl „PM1“ versetzt, eine Minute lang bei 1.600rpm geschüttelt, für 5min bei 37°C inkubiert und eine weitere Minute bei 280g zentrifugiert. Dann wurden zu der Probe 300µl absolutes 2-Propanol („Sigma-Aldrich“, USA) pipettiert. Anschließend wurde die Platte

10 Mal invertiert, 30min bei 4°C inkubiert und schließlich für 20min bei 4°C sowie 3.000g zentrifugiert. Danach wurde der Überstand durch schnelles Umdrehen und Ausklopfen der 96-Well-Platte entfernt. Um das vollständige Verdampfen der Flüssigkeit zu gewährleisten, wurde die Platte mit der Öffnung nach unten eine Stunde lang auf einem sauberen Reagenzglasständer getrocknet.

Um anschließend das DNA-Pellet aufzulösen, wurden pro Well 46µl „RA1“ pipettiert. Die 96-Well-Platte wurde in den Hybridisierungssofen gestellt und eine Stunde lang bei 48°C inkubiert. Dann wurde die Platte jeweils eine Minute lang bei 1.800rpm geschüttelt und bei 280g zentrifugiert.

Im nächsten Schritt erfolgte die eigentliche Hybridisierungsreaktion. Hierfür wurde die DNA-Probe 20min bei 95°C denaturiert. Währenddessen wurden die „Infinium MethylationEPIC BeadChips“ („Illumina“, USA) auf Raumtemperatur äquilibriert. Die Hybridisierungskammern wurden nach Anweisungen des Herstellers zusammengebaut und mit jeweils 400µl „PB2“ befüllt. Nach der Denaturierungsreaktion wurde die 96-Well-Platte für 30min bei Raumtemperatur abgekühlt und anschließend eine Minute lang bei 280g zentrifugiert. Der EPIC BC wurde mit 15µl der jeweiligen Probe beladen, in einer Hybridisierungskammer platziert und für 24 Stunden bei 48°C im Hybridisierungssofen inkubiert. Dann wurde der EPIC BC für 30min auf Raumtemperatur abgekühlt und anschließend durch mehrmaliges Schwenken in jeweils 200 ml „PB1“ gewaschen.

Die anschließende SBE erfolgte bei 44°C und bestand aus drei Schritten. Im ersten Schritt wurden unspezifisch gebundene sowie nicht hybridisierte DNA-Fragmente eliminiert. Hierfür wurde der EPIC BC mit 150µl „RA1“ beladen und für 30s inkubiert. Dieser Vorgang wurde sechs Mal wiederholt. Im zweiten Schritt erfolgte die eigentliche SBE. Hierfür wurden auf den EPIC BC 450µl „XC1“ pipettiert und 10min lang inkubiert. Dieser Schritt wurde mit 450µl „XC2“ wiederholt. Dann wurden auf den EPIC BC 200µl „TEM“ gegeben und für 15min inkubiert. Nach einer erfolgreichen SBE, wurde das passende Nukleotid am 3'-Ende des *Linkers* eingebaut. Somit war die hybridisierte DNA für die Ermittlung des Methylierungsstatus des jeweiligen CpG *Locus* nicht mehr notwendig und konnte im dritten und letzten Schritt eliminiert werden. Hierfür wurde das EPIC BC jeweils für eine Minute ein Mal mit 450µl Formamid/1mM EDTA-Lösung („Illumina“, USA) und zwei Mal mit 450µl „XC3“ inkubiert.

Als nächstes erfolgte bei 32°C eine antikörperbasierte Färbung des EPIC BC. Dieser Schritt war erforderlich, um die Methylierungsraten der CpG *Loci* detektieren zu können. Dabei wurde der *BeadChip* zunächst zwei Mal wie folgt behandelt: 10min mit 250µl „STM“, zwei Mal jeweils eine Minute mit 450µl „XC3“, 10min 250µl „ATM“, zwei Mal jeweils eine Minute

mit 450µl „XC3“. Dann erfolgte eine weitere 10min lange Inkubation des *BeadChips* mit 250µl „STM“ sowie anschließende zweimalige Behandlung mit 450µl „XC3“. Um den EPIC BC nach dem Färbevorgang zu reinigen, wurde er unter Schwenken in 310ml „PB1“ sowie „XC4“ gewaschen. Als nächstes wurde der *BeadChip* in einem Vakuumexsikkator bei 680mbar für eine Stunde getrocknet. Um die Rückstände des „XC4“ von der Rückseite des EPIC BC zu entfernen, wurde er vorsichtig mit 2-Propanol behandelt.

Schließlich wurde der EPIC BC unter Verwendung des „Illumina iScan“ („Illumina“, USA) ausgelesen.

3.3.8.4 Bioinformatische Auswertung der EPIC BC Daten

Die Rohdaten wurden im „IDAT“-Format gespeichert und zunächst unter Verwendung der „Illumina® GenomeStudio“ Software (Version 2.0.3, „Illumina“, USA) gegen die internen Kontrollen des EPIC BC normalisiert. Nach dem Normalisieren ermittelte die Software die „avg.beta“- sowie p-Werte und annotierte sie anschließend gegen das „HG19“ Referenzgenom. Nachfolgend wurden die „avg.beta“-Werte nach einem p-Wert kleiner 0,01 gefiltert und im „TXT“-Format für alle weitere Analysen gespeichert [249-251].

Die „TXT“-Datei wurde dann in die „Qlucore Omics Explorer“ Software geladen [168]. Um die DNA-Methylierungsunterschiede zwischen den behandelten und unbehandelten Zellen zu detektieren, wurde ein in der Software implementierter „T-Test“ durchgeführt (p-Wert < 0,01, $\sigma/\sigma_{\max} = 0$). Anschließend wurden alle identifizierten CpG *Loci* nach einer Methylierungsdifferenz von mindestens 0,15 gefiltert.

Um die *Batch* Effekte ausschließen zu können wurde sowohl eine Hauptkomponentenanalyse (engl. *principal component analysis* (PCA)) mit „Qlucore Omics Explorer“ Software als auch eine „Pearson“ Korrelationsanalyse in „R“ (Abschnitt 3.1.2.5.4) durchgeführt.

3.3.9 Fluoreszenzmikroskopie

Die Fluoreszenzmikroskopie erfolgte unter Verwendung von „DAPI“, „ActinGreen™ 488 ReadyProbes®“ sowie „Vybrant® Dil Cell-Labeling“ Reagenzien („Thermo Fisher Scientific“, USA). Hierfür wurde pro *Well* einer 6-*Well*-Platte ein Deckgläschen („VWR“, USA) platziert. Darauf wurden 89.000 Zellen, wie im Abschnitt 3.3.3. beschrieben, ausgesät, mit Zigaretten- bzw. E-Zigarettenextrakten behandelt und anschließend 48h lang kultiviert. Für das Protokoll wurden mehrere Schritte etabliert.

Zunächst erfolgte die Färbung der Plasmamembran. Hierfür wurde das Zellmedium abpipettiert und durch 2,5µl „Vybrant® Dil Cell-Labeling Solution“ sowie 500µl DMEM ersetzt. Nach einer 20-minütigen Inkubation bei 37°C wurden die Proben drei Mal in je 500µl DMEM und anschließend drei weitere Male in je 500µl einfacher PBS gewaschen. Dann wurden die Zellen unter Verwendung von 500µl einer 4%igen Paraformaldehydlösung („Carl Roth GmbH & Co. KG“, Deutschland) 15min lang fixiert. Nach der Fixierung erfolgte die Zellpermeabilisierung. Hierfür wurden die Proben zunächst mit 500µl PBS gewaschen und dann mit einer 0,25%igen Triton X-100-Lösung („Carl Roth GmbH & Co. KG“, Deutschland) 5min lang behandelt. Als nächstes wurden die Proben ein weiteres Mal mit 500µl PBS gewaschen und anschließend mit 500µl einer 10%igen BSA-Lösung 30min lang bei 37°C inkubiert. In diesem Schritt wurden unspezifische Antikörperbindestellen blockiert, sodass als nächstes die Aktinfärbung erfolgen konnte. Hierfür wurden die Zellen drei Mal mit je 500µl PBS gewaschen, dann mit 500µl DMEM sowie einem Tropfen „ActinGreen™ 488 ReadyProbes® Reagent“ versetzt und 30min bei Raumtemperatur unter Lichtausschluss inkubiert. Nach der Inkubation wurde das Medium abgenommen und die Zellen wieder drei Mal mit je 500µl PBS gewaschen. Um die Zellkerne anzufärben, wurden die Proben mit 1µl „DAPI“ sowie 999µl PBS versetzt und 20min lang unter Lichtausschluss inkubiert. Schließlich wurden die Zellen wieder drei Mal mit je 500µl PBS gewaschen und unter Verwendung des „Mounting Medium“ („Thermo Fisher Scientific“, USA) konserviert. Die Fluoreszenzmikroskopie erfolgte mit dem „Zeiss–Axioskop“ („Zeiss AG“, Deutschland; Fluoreszenzfilter der „Fa. AHF Analysentechnik AG“, Deutschland) bei folgenden Wellenlängen (λ):

- „Vybrant® Dil Cell-Labeling Solution“: $\lambda_{\text{Ex.}}$ 549/ $\lambda_{\text{Em.}}$ 565 nm
- „ActinGreen™ 488 ReadyProbes® Reagent“: $\lambda_{\text{Ex.}}$ 495/ $\lambda_{\text{Em.}}$ 518 nm
- DAPI: $\lambda_{\text{Ex.}}$ 341/ $\lambda_{\text{Em.}}$ 452 nm

Die Auswertung wurde mit der Software „Isis“ („MetaSystems GmbH“, Deutschland) durchgeführt.

3.3.10 Durchflusszytometrie

Die Durchflusszytometrie wurde im „Institut für klinische und experimentelle Pathologie“ („Forschungszentrum Borstel“, Deutschland) durchgeführt. Um die Anzahl an Zellkernen pro Zelle zu bestimmen, erfolgte eine DAPI-Färbung. Hierfür wurden 89.000 Zellen pro Well einer 6-Well-Platte, wie im Abschnitt 3.3.3. beschrieben, ausgesät, mit Zigaretten- bzw. E-Zigarettenextrakten behandelt und anschließend 48h lang kultiviert. Dann wurden 500.000

Zellen in 500µl einfachen PBS resuspendiert und mit 500µl zweiprozentiger Paraformaldehydlösung 10min lang fixiert. Die fixierten Zellen wurden mit 1ml Waschpuffer (einfache PBS mit 1% FBS und 0,09% Natriumazid („Carl Roth GmbH & Co. KG“, Deutschland)) versetzt und für fünf Minuten bei 300g pelletiert. Im nächsten Schritt erfolgte die Permeabilisierung der Zellen. Hierfür wurde 1ml des „Perm III“ Puffers („BD Biosciences“, USA) langsam auf die Zellen gegeben. Die Probe wurde 30min bei 4°C inkubiert. Nach der Inkubation wurden die Zellen mit 3ml Waschpuffer versetzt und fünf Minuten lang bei 300g zentrifugiert. Nachdem dieser Schritt noch ein Mal wiederholt wurde, konnte die intrazelluläre Antikörperfärbung erfolgen. Hierfür wurden die Zellen in 95µl Waschpuffer resuspendiert, mit 5µl „Alexa Fluor® 647 anti-H2A.X Phospho (Ser139)“ („Biolegend“, USA) versetzt und für 30min bei Raumtemperatur unter Lichtausschluss inkubiert. Als nächstes wurde zu der Probe ein weiterer Milliliter des Waschpuffers gegeben, die anschließende Zentrifugation erfolgte fünf Minuten lang bei 300g. Dieser Schritt wurde ein weiteres Mal mit 2ml Waschpuffer wiederholt. Dann wurden die Zellkerne mit „DAPI“ gefärbt. Hierfür wurden 500µl 3µM „DAPI“ Lösung („BioLegend“, USA) auf das Zellpellet pipettiert. Die Probe wurde 15min bei Raumtemperatur unter Lichtausschluss inkubiert und danach sofort im Durchflusszytometer vermessen.

4. Ergebnisse

4.1 Ergebnisse des Projekts „Entwicklung des Plasma *Panels*“

Um die *Targets* für das Plasma *Panel* zu identifizieren, wurden mehrere Schritte durchgeführt. Zuerst wurden DNA-Methylierungssignaturen in 40 malignen Lungentumoren sowie deren korrespondierenden Kontrollen untersucht. Dann erfolgte eine Analyse von DNA-Methylierungssignaturen im Blutplasma von 9 Patienten. Davon litten 5 Patienten an Adeno- und 4 an Plattenepithelkarzinom der Lunge. Die 19 übrigen Patienten waren dagegen frei von malignen Erkrankungen und bildeten die Kontrollkohorte. Schließlich wurden zusätzliche Datensätze mehrerer zur Verfügung gestellter Studien ausgewertet, was das Identifizieren weiterer tumorspezifischer und prognostischer CpG *Loci* ermöglichte. Das von der Firma „Roche“ synthetisierte „SeqCap Epi“-Plasma *Panel* wurde anschließend im Rahmen einer Pilotstudie validiert.

4.1.1 Detektion der tumor- und entitätsspezifischen DNA-Methylierung in primärem Tumorgewebe

40 OP-Präparate und korrespondierende Kontrollen wurden unter Verwendung des „Illumina Infinium HumanMethylation450K BeadChips“ auf ihre genomweite DNA-Methylierung untersucht. Die Daten wurden im Rahmen des DZL erhoben und für diese Arbeit zur Verfügung gestellt.

Im Vergleich zum gesunden Lungengewebe wurden im malignen Tumorgewebe 898 aberrant methylierte CpG *Loci* identifiziert ($q < 1 \times 10^{-23}$, $\sigma/\sigma_{\max} > 0,4$; Abb. 25A). Adeno- und Plattenepithelkarzinome stellen die zwei häufigsten Entitäten des nicht-kleinzelligen Lungenkarzinoms dar. Eine Analyse ergab 1.167 zwischen den Tumorentitäten differentiell methylierte CpG *Loci* (FDR $< 1 \times 10^{-4}$; Abb. 25B).

Nachfolgend wurden im Rahmen der vorliegenden Doktorarbeit die CpG *Loci* ausgewählt, die eine zuverlässige Klassifikation von Lungentumoren aufgrund von Malignität und Entität ermöglichten. Hierfür wurden die im Abschnitt 3.1.1 beschriebenen bioinformatische Analysen durchgeführt, welche 287 CpG *Loci* ergaben. Diese *Loci* wurden in das Plasma *Panel* aufgenommen (Abschnitt 9.3, Tab. 7).

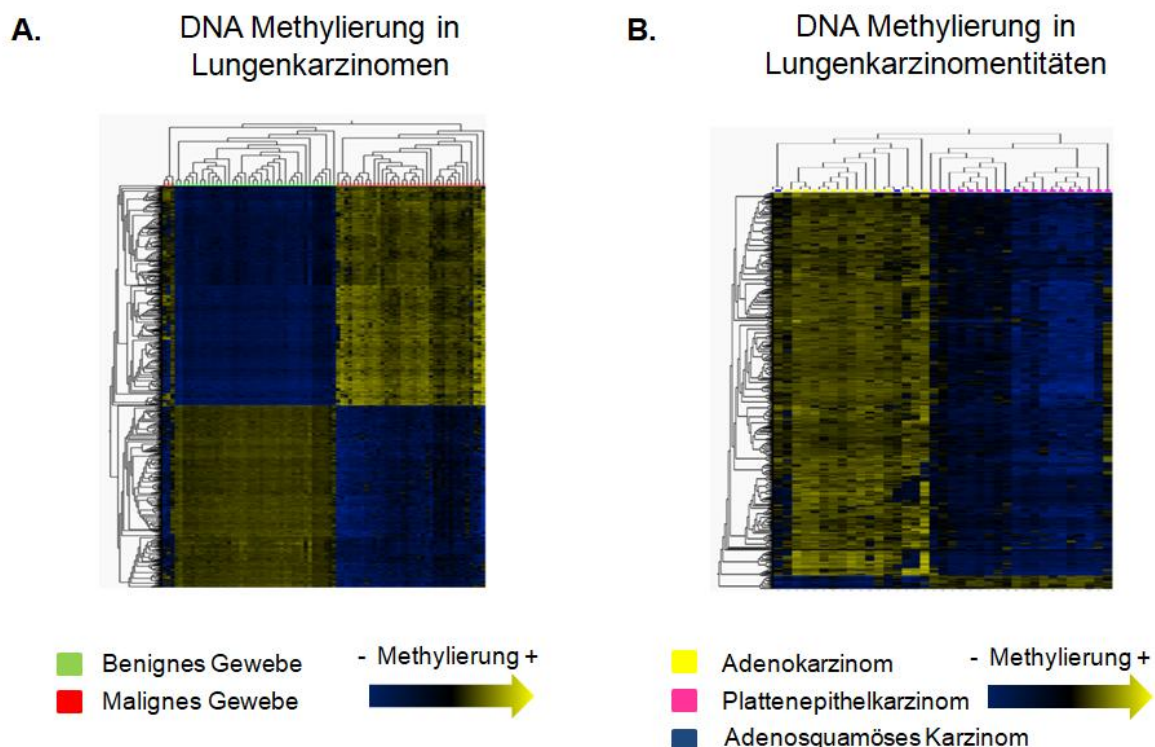


Abb. 25: Ergebnisse des HM 450K; hierarchische Clusteranalyse von 40 OP-Präparaten und deren korrespondierenden Kontrollen identifizierte **A.** 898 differentiell methylierte CpG *Loci* in Tumorproben ($q < 1 \times 10^{-23}$, $\sigma/\sigma_{\max} > 0,4$) und **B.** 1.167 differentiell methylierte CpG *Loci* in unterschiedlichen Lungenkarzinomentitäten (FDR $< 1 \times 10^{-4}$).

4.1.2 Detektion der tumor- und entitätsspezifischen DNA-Methylierung in Blutplasma

Wie in Abschnitt 3.1.2.2 beschrieben, wurde jede einzelne zellfreie, zirkulierende DNA Probe nach der Extraktion quantifiziert und einer strikten Qualitätskontrolle unterzogen. Die Gesamtmenge der extrahierten DNA betrug pro Probe 10 bis 30ng, davon wurden 1ng mit dem „Agilent 2100 Bioanalyzer“ analysiert. Dabei zeigten die Proben einen klaren *Peak* bei ca. 167bp. Die *Peaks* bei 35 und 10.380bp entsprachen den unteren bzw. oberen Markern (Abb. 26).

Nach der Bisulfitkonvertierung wurden die zfDNA Proben zur Herstellung von WGBS *Libraries* verwendet. Die fertigen *Libraries* wurden wiederum quantifiziert und nachfolgend einer Qualitätskontrolle unter Verwendung des „Agilent 2100 Bioanalyzers“ unterzogen. Alle Proben zeigten einen klaren *Peak* bei ca. 300bp und entsprachen somit den Sequenzieranforderungen (Abb. 27).

Die hergestellten WGBS *Libraries* wurden auf Trockeneis zu „TATAA Biocenter“ geschickt, dort *gepoolt* und je nach Probe mit einer mittleren *Coverage* von acht- bis zehnfach auf einer „NextSeq 500“ Plattform sequenziert (Abb. 28). Die Rohdaten wurden im „FastQ“-Format geliefert.

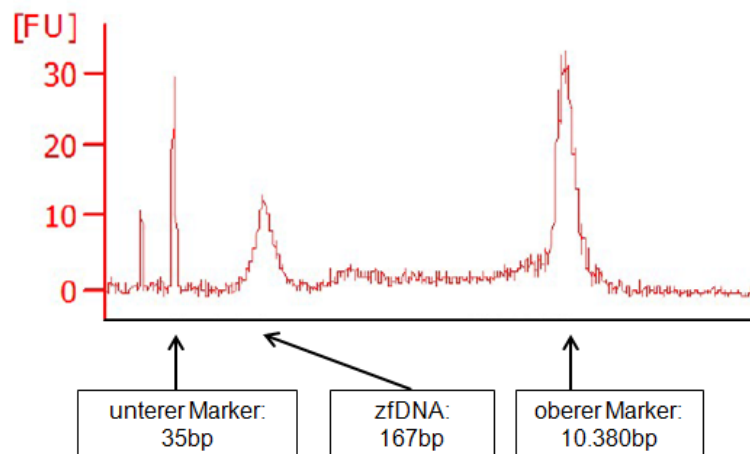


Abb. 26: 1ng der verwendeten zfDNA Proben wurden mit dem „Agilent 2100 High Sensitivity Assay Kit“ analysiert und wiesen folgendes Profil auf: ein 167bp großer, klarer *Peak* sowie keine sichtbaren Kontaminationen durch die genomische DNA. Die *Peaks* bei 35 und 10.380bp entsprachen den unteren bzw. oberen Markern des „Agilent 2100 High Sensitivity Assay Kit“ (engl. *fluorescent units* (FU)).

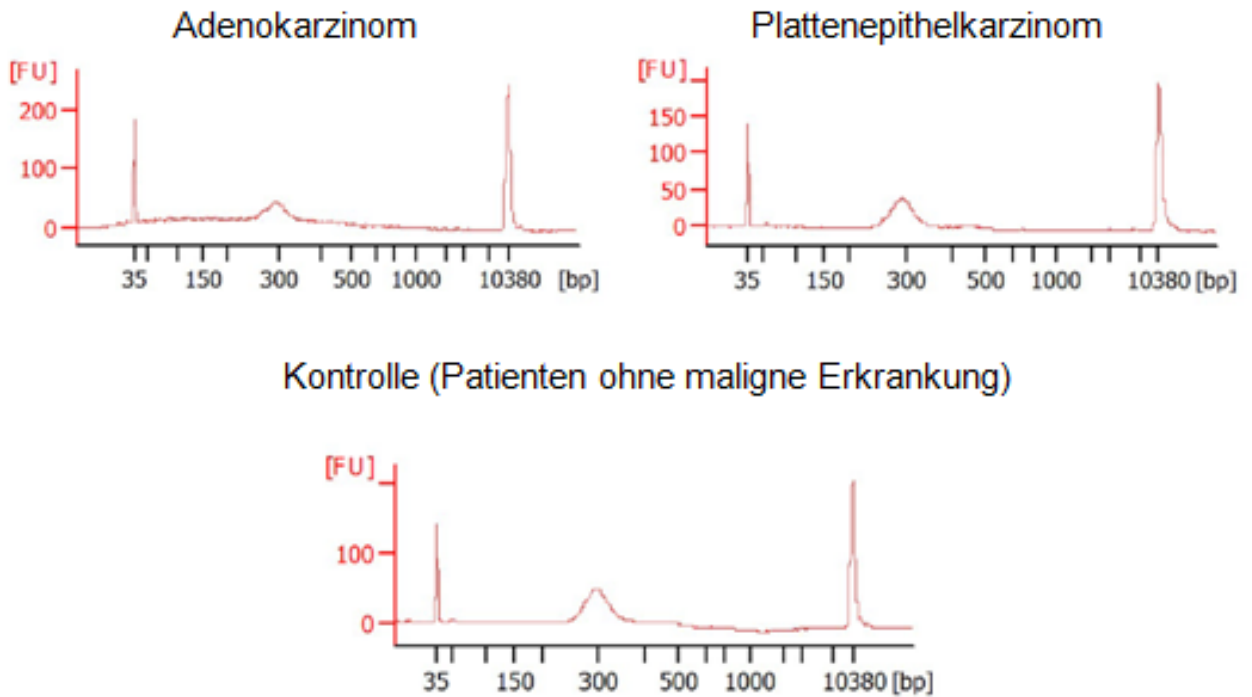


Abb. 27: Die „Agilent 2100 High Sensitivity Assay Kit“ Analyse der hergestellten WGBS *Libraries*. Alle *Libraries* zeigten Fragmente von 300bp Länge. Die Peaks bei 35 und 10.380bp entsprachen den oberen und unteren Markern des „Agilent 2100 High Sensitivity Assay Kit“ (engl. *fluorescent units* (FU)).

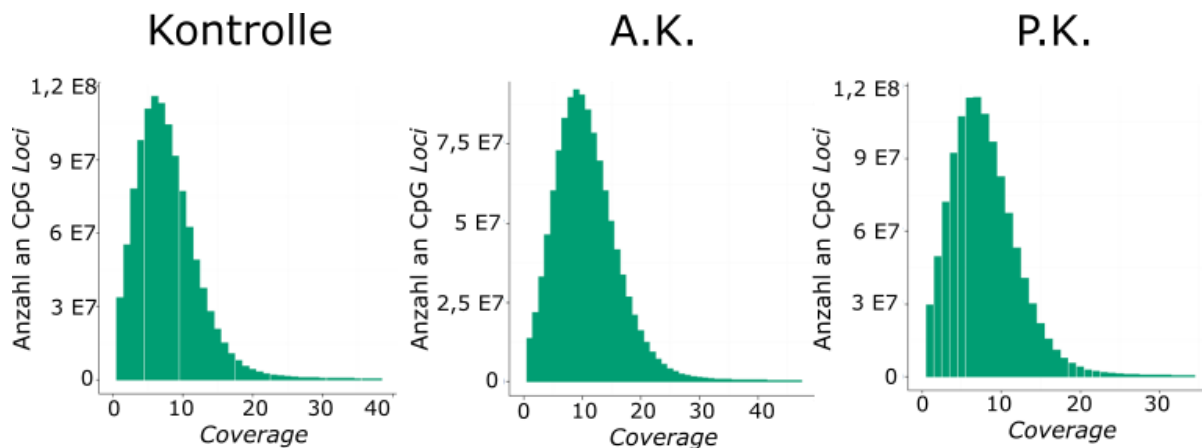


Abb. 28: Histogram-Plot der *Coverage* der im „TATAA Biocenter“ sequenzierten WGBS *Libraries*. Die Sequenzierung erfolgte auf der „NextSeq 500“ Plattform (Adenokarzinom (A.K.), Plattenepithelkarzinom (P.K.)).

Die Qualität der Rohdaten wurde unter Verwendung der „FastQC“ Software überprüft, die Ergebnisse sind in den Abbildungen 29 und 30 dargestellt.

Da die Proben 76 PE sequenziert wurden, lag die *Read* Länge wie erwartet bei 76bp. Innerhalb eines *Reads* betrug der Gehalt an Adaptern sowie nicht identifizierbaren Signalen 0% (Abb. 29A).

Die Genauigkeit der Sequenzierung wurde in „Phred“-Werten angegeben. Jeder „Phred“-Wert beschreibt, wie genau das Ablesen von Nukleotiden im Verlauf der Sequenzierung erfolgte. Die Rohdaten wiesen einen „Phred“-Score von über 30, was einer Genauigkeit von mehr als 99,9% entsprach (Abb. 29B). Des Weiteren konnte nur eine sehr geringe Menge an Kmeren detektiert werden. Als Kmere werden Sequenzen mit einer Mindestlänge von zwei Nukleotiden bezeichnet, die sich in den Rohdaten immer wieder wiederholen (Abb. 30B). Die Anzahl an PCR-Duplikaten lag bei nahezu 0% (Abb. 30A). Die Menge an PCR-Duplikaten wird ermittelt, indem die Prozentanzahl deduplizierter Sequenzen (rote Kurve) berechnet und mit der Anzahl aller Sequenzen (blaue Kurve) verglichen wird. Eine geringe Menge an Kmeren sowie PCR-Duplikaten weist auf eine gute *Library*- und Sequenzierqualität hin.

Des Weiteren wird in der Abb. 29C eine für WGBS typische Basenzusammensetzung sichtbar. Im Laufe der Bisulfidkonvertierung wurden die meisten unmethylierten Cytosine durch Thymin ersetzt. Deshalb lag der Thymingehalt der Rohdaten bei ca. 50% und der Cytosinegehalt bei nahezu 0%. Die Adenin- und Guaninzusammensetzung wurde während der Bisulfidkonvertierung nicht beeinflusst und lag bei jeweils 25%. Dieses Phänomen wird auch in der Abb. 29D deutlich: die Abweichung des von der Software ermittelten, theoretischen GC-Gehalts (blaue Kurve) vom experimentellen (rote Kurve).

Nachfolgend wurden die WGBS Rohdaten unter Verwendung der „Cutadapt“ Software prozessiert (siehe Abschnitt 3.1.2.5.2). Durch das Prozessieren wurden sowohl überpräsentierte Sequenzen (Abb. 29A), als auch der 10nt lange Überhang am Anfang von *Read 2* (Abb. 29C, gelbe Markierung) entfernt.

Die prozessierten Sequenzierdaten wurden dann in das „Bisulfite Analysis Toolkit“ geladen und mit dem dort implementierten „Segemehl“ Algorithmus gegen das „HG19“ Referenzgenom *aligned*. Die Effizienz des *Alignments* wird als *Mapping* Effizienz angegeben. Dabei wird ermittelt wie viel Prozent an *Reads* dem Referenzgenom zugeordnet werden können. In diesem Fall betrug die *Mapping* Effizienz des „Segemehl“ Algorithmus 98% bis 99% und war somit für alle weiteren Analysen geeignet.

Die *Alignments* der Kontroll-, Adenokarzinom- und Plattenepithelkarzinomgruppen wurden als nächstes in das „BAT_calling“ Modul geladen. Das Modul ermittelte DNA-

Methylierungsraten jeweiliger Cytosine. Die Cytosine, die innerhalb einer CpG-Region lagen und eine *Coverage* von mindestens achtfach aufwiesen, wurden dann unter Verwendung des „BAT_filtering“ Moduls identifiziert und für alle weiteren Analysen verwendet. In der Abbildung 31 sind drei analysierte Gruppen dargestellt: Patienten ohne maligne Erkrankungen (Kontrolle; Abb. 31A), Patienten mit Adenokarzinom (Abb. 31B) und Patienten mit Plattenepithelkarzinom (Abb. 31C). Pro Gruppe sind links die *Coverage* und rechts die DNA-Methylierungsraten der CpG *Loci* dargestellt. Mehr als 4 Millionen CpG *Loci* pro Gruppe erfüllten die Kriterien des *Filterings* und wurden im weiteren Verlauf mit dem „BAT_overview“ Modul analysiert. Die Ergebnisse zeigten deutlich, dass sowohl Lungenkarzinom- als auch Kontrollgruppe anhand der DNA-Methylierungsmuster voneinander unterschieden werden können (Abb. 32A). Des Weiteren wird eine genomweite Hypermethylierung der Lungenkarzinomgruppen im Vergleich zur Kontrollgruppe sichtbar (Abb. 32B,C).

Um die für die jeweilige Gruppe spezifische, differentiell methylierte Regionen zu detektieren, erfolgte das Filtern nach einem DNA-Methylierungsunterschied von mindestens 15%. Die Anzahl an differentiell methylierten CpG *Loci* im Plasma von Lungenkarzinompatienten betrug dabei 18.000 (Abb. 33A). Des Weiteren wurden 44.000 CpG *Loci* identifiziert, die je nach Entität in Adeno- und Plattenepithelkarzinompatienten differentiell methyliert waren (Abb. 33B). Diese *Loci* wurden wie im Abschnitt 3.1.2.5.5 beschrieben weiteren Analysen unterzogen und für das Erstellen des Plasma *Panels* verwendet. Das fertige Plasma *Panel* enthielt 638 differenziell methylierte Regionen (Abschnitt 9.3, Tab. 7) und wurde von der Firma „Roche“ synthetisiert.

4.1.3 Korrelationsanalysen der verwendeten Methoden zur genomweiten Detektion von DNA-Methylierungsmustern

Um die detektierten DNA-Methylierungsmuster in den OP-Präparaten mit denen im Blutplasma der Lungenkarzinompatienten zu vergleichen, wurde unter Verwendung von „R“ und „Bedtools“ eine „Pearson“ Korrelationsanalyse durchgeführt (siehe Abschnitt 3.1.2.5.4), welche je nach Probe eine Übereinstimmung von 71 bis 77% ergab ($p\text{-Wert} < 2,2 \times 10^{-16}$, Abb. 34).

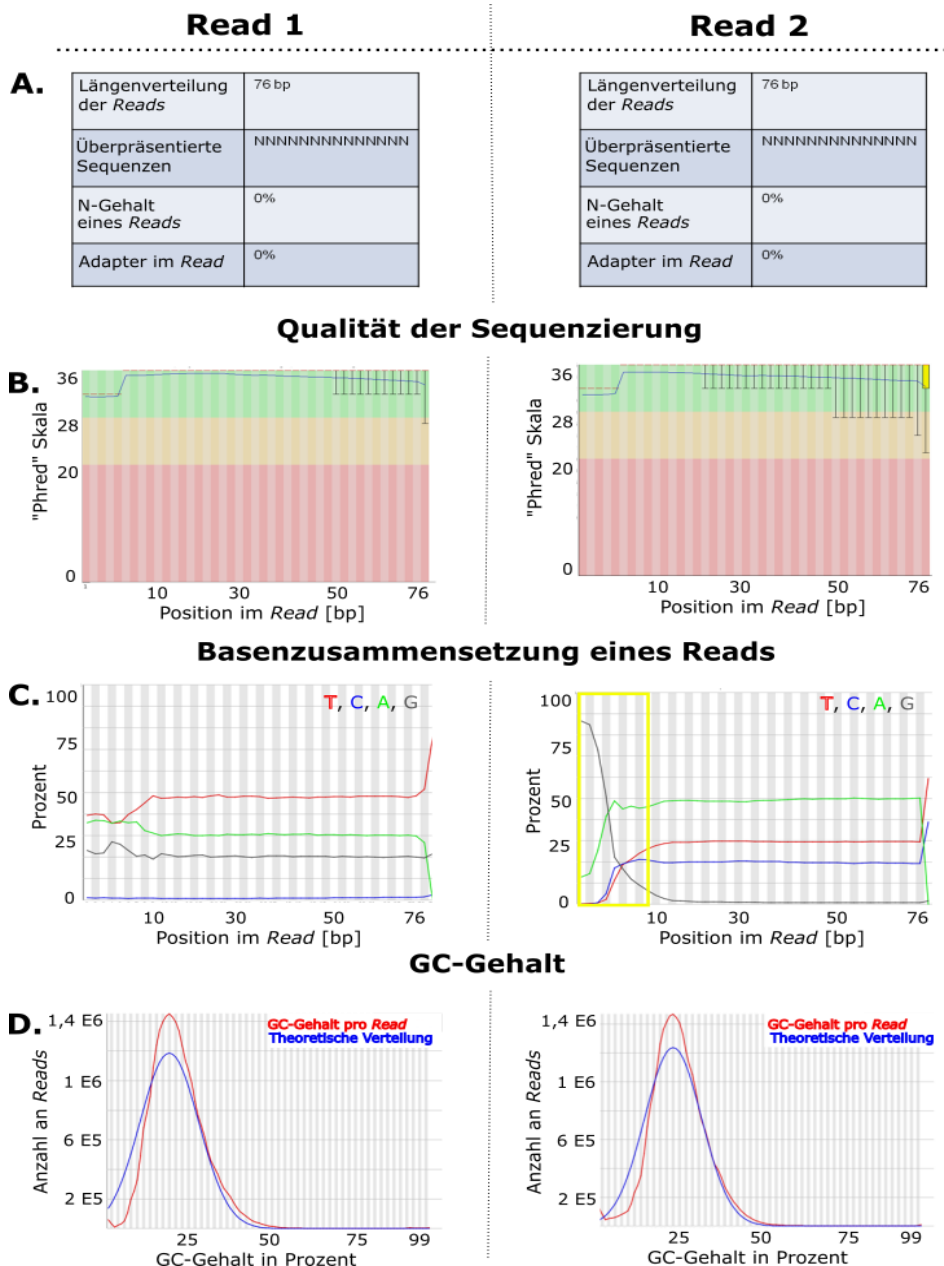


Abb. 29: Ergebnisse der mit der „FastQC“ Software durchgeführten Qualitätskontrolle der WGBS Rohdaten. **A.** Die Längen der Reads lagen bei 76bp, was bei einem 76 PE Lauf zu erwarten war. Der Adapter und N-Gehalt der Reads betrug 0%. Die überpräsenzierte Sequenzen wurden beim Prozessieren eliminiert. **B.** Die Qualität der Sequenzierung wurde in „Phred“-Werten angegeben und lag bei über 30, was einer Sequenziergenauigkeit von über 99,9% entsprach (Farbgebung: Grün=Qualität >99,9%, Gelb=Qualität 99-99,9%, Rot=Qualität<90%). **C.** Basenzusammensetzung eines Reads entsprach einer bisulfitkonvertierten Probe. *Read 1* verhielt sich komplementär zu *Read 2*. Während der Gehalt an Adenin (A; Grün) und Guanin (G; Grau) wie erwartet bei ca. 25% lag, wurden im Laufe der Bisulfitkonvertierung Cytosine durch Thyminen ausgetauscht, weshalb der Thymingehalt (T; Rot) viel höher als erwartet bei ca. 50% und Cytosinegehalt (C; Blau) viel niedriger bei nahezu 0% lag. Der durch „Adaptase“ erzeugte, 10nt lange Überhang des *Read 2* wurde gelb markiert. **D.** Auch der experimentell ermittelte (rote Kurve) GC-Gehalt einer WGBS *Library* wich wegen der Bisulfitkonvertierung vom theoretisch vorhergesagten (blaue Kurve) stark ab.

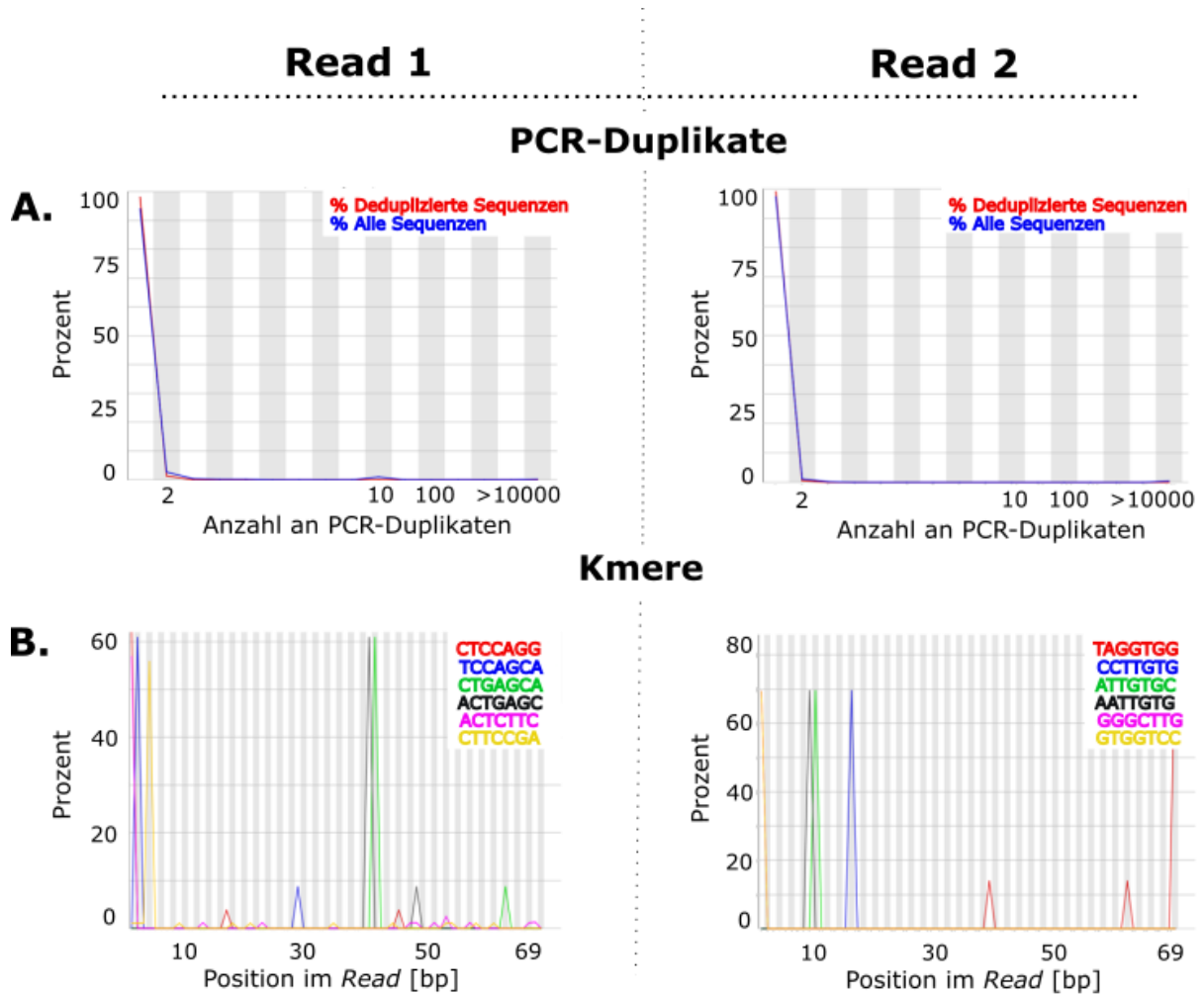


Abb. 30: Ergebnisse der mit „FastQC“ Software (Version 0.11.15, „Babraham Bioinformatics“, England) durchgeführten Qualitätskontrolle der WGBS Rohdaten. **A.** Da die Prozentanzahl an deduplizierten Sequenzen (rote Kurve) der Gesamtmenge aller Sequenzen (blaue Kurve) entsprach, lag die Menge an PCR-Duplikaten bei nahezu 0%. **B.** Innerhalb der *Reads* waren nur wenige Kmere enthalten (farbig dargestellte Sequenzen sowie entsprechend gefärbte *Peaks*).

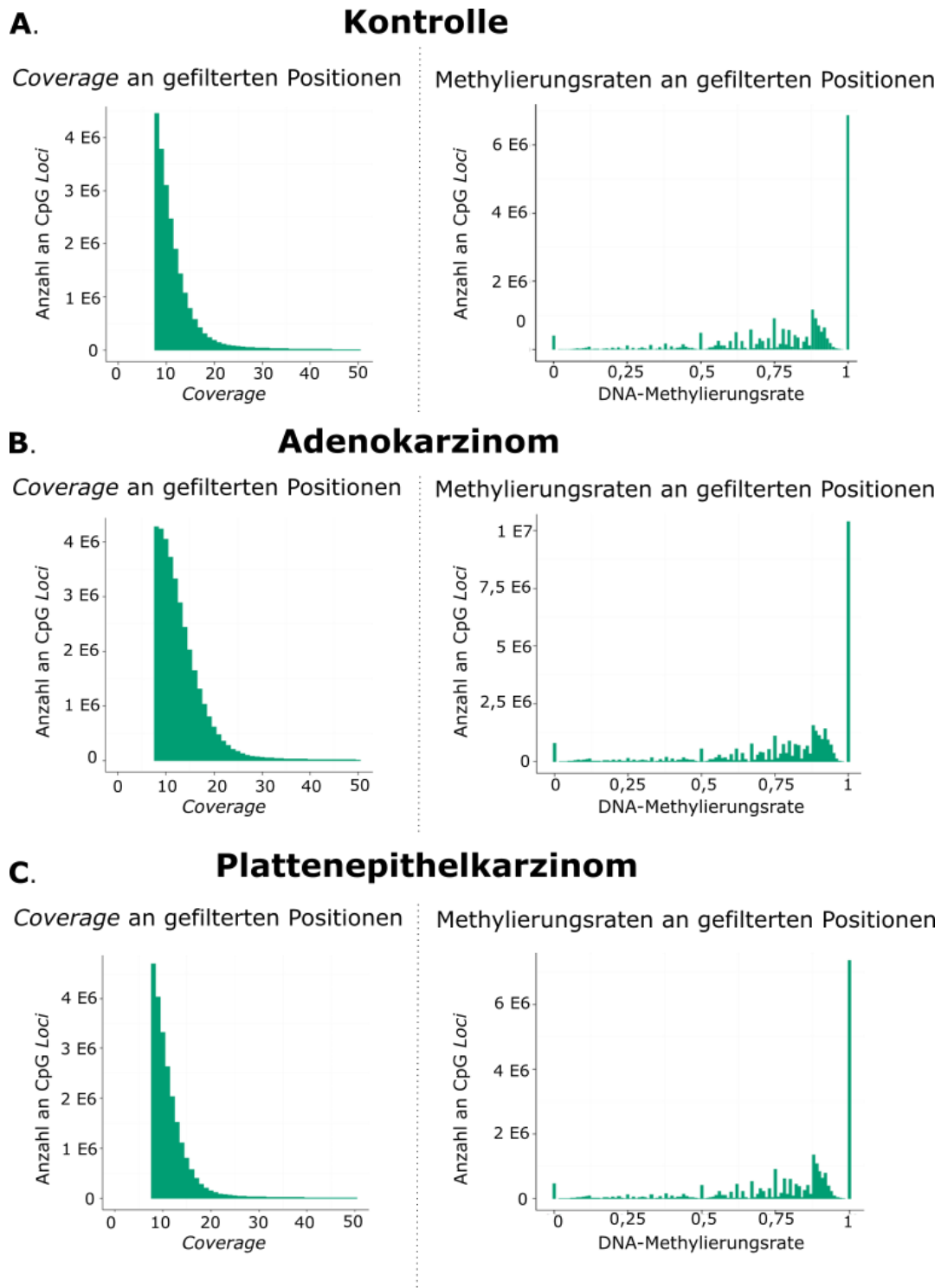


Abb. 31: Für die im Rahmen dieser Arbeit durchgeführten Analysen wurde das „Bisulfite Analysis Toolkit“ verwendet. Pro Gruppe wurden die DNA-Methylierungsraten aller sequenzierten Cytosine mit dem „BAT_calling“ Modul ermittelt: **A.** Kontrolle (tumorfremie Patientenkohorte) **B.** Adenokarzinomgruppe **C.** Plattenepithelkarzinomgruppe. Für alle anschließenden Analysen wurden unter Verwendung des „BAT_filtering“ Moduls nur diejenigen Positionen ausgewählt, die sowohl eine Coverage von mindestens achtfach hatten als auch innerhalb einer CpG-Region lokalisiert waren. Die Anzahl an CpG Loci, die diese Kriterien erfüllten sowie deren Methylierungsraten sind als Histogramme dargestellt.

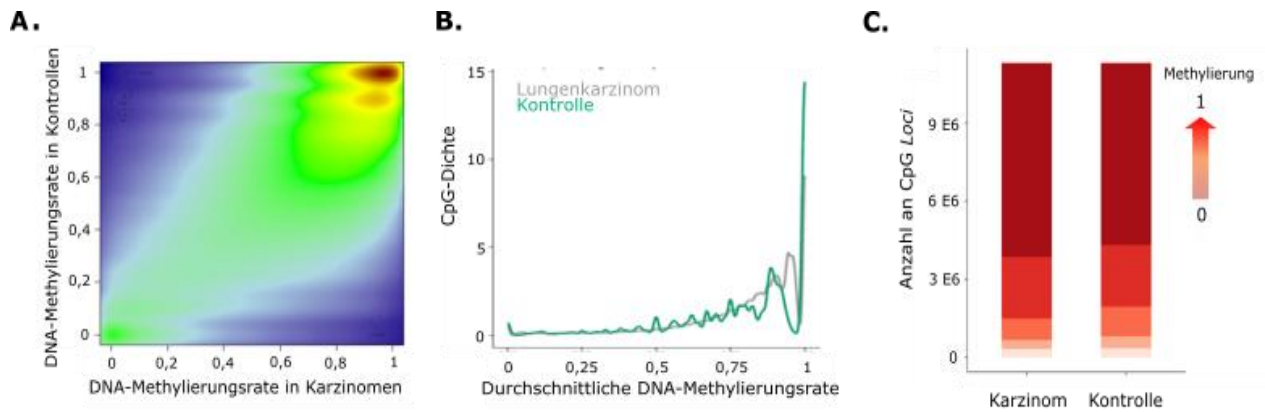


Abb. 32: Die mit den „BAT_calling“ und „BAT_filter_vcf“ Modulen ermittelten DNA-Methylierungsraten wurden in das „BAT_summarize“ Modul des „Bisulfite Analysis Toolkit“ geladen. **A.** Das Streudiagramm zeigt deutlich, dass die Lungenkarzinomgruppe anhand des DNA-Methylierungsmusters von der Kontrollgruppe (tumorfreie Patientenkohorte) unterschieden werden kann. **B.** Die mittlere sowie **C.** die gestaffelte Darstellungen der DNA-Methylierungsraten pro Gruppe verdeutlichen die genomweite Hypermethylierung der Lungenkarzinomgruppe im Vergleich zur Kontrollgruppe.

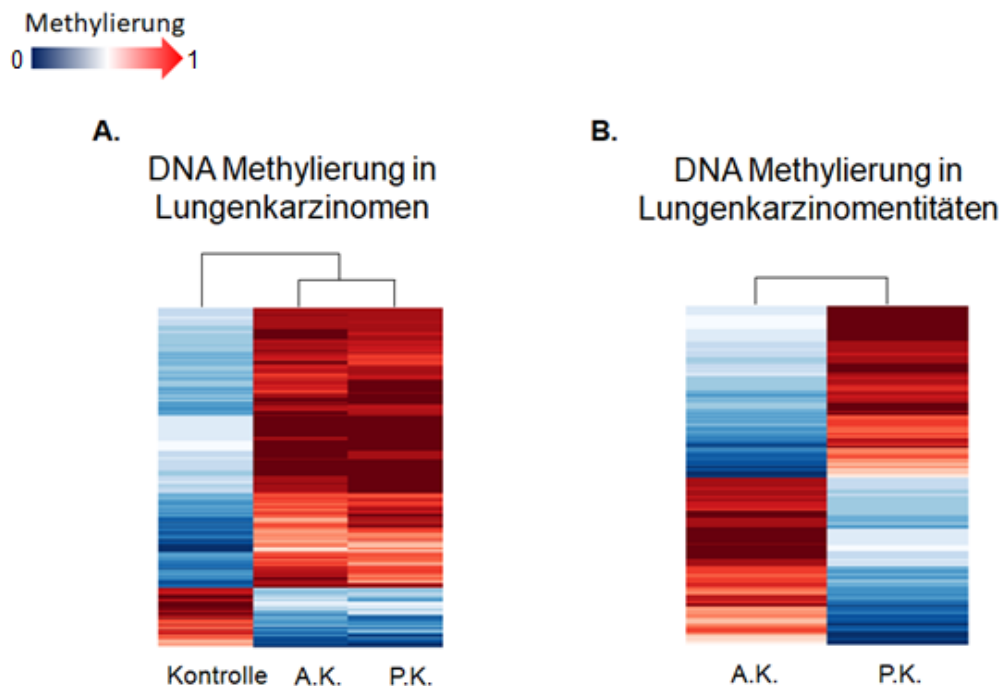


Abb. 33: Die ermittelten zfDNA-Methylierungsmuster wurden normalisiert und einer hierarchischen Clusteranalyse unterzogen. Dabei wurden **A.** 18.000 für das Lungenkarzinom und **B.** 44.000 für die jeweilige Entität spezifische differentiell methylierte CpG *Loci* identifiziert (Adenokarzinom (A.K.), Plattenepithelkarzinom (P.K.)).

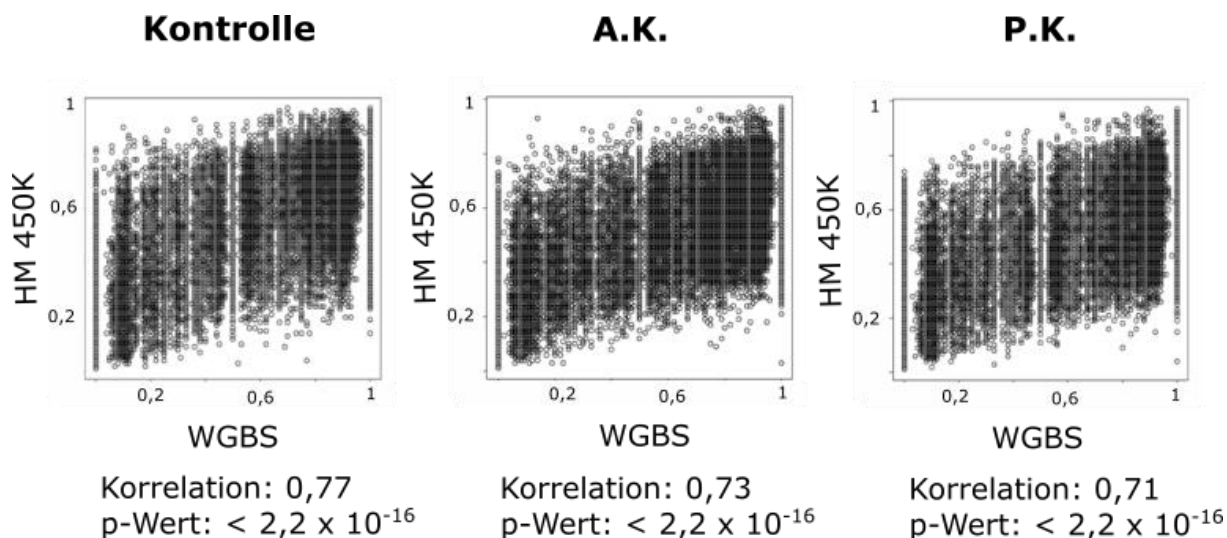


Abb. 34: „Pearson“ Korrelationsanalyse der mit beiden Methoden (HM 450K und WGBS) detektierten DNA-Methylierungswerte (Adenokarzinom (A.K.), Plattenepithelkarzinom (P.K.)).

4.2 Ergebnisse des Projekts „Validierung des Plasma Panels“

4.2.1 Erstellen der NGS Libraries

Zunächst wurden die extrahierten zfDNA-Proben wie im Abschnitt 3.1.2.2 beschrieben quantifiziert und einer Qualitätskontrolle unterzogen. Hierfür wurden jeweils 1ng der Proben mit dem „Agilent 2100 Bioanalyzer“ untersucht. Alle verwendeten zfDNA-Proben zeigten einen klaren *Peak* bei ca. 167bp. Nachfolgend wurden die Proben bisulfidkonvertiert und zur Herstellung von NGS *Libraries* verwendet. Wie im Abschnitt 3.2.1 beschrieben, erfolgte die Herstellung der *Libraries* in zwei Schritten.

Im ersten Schritt wurden WGBS *Libraries* hergestellt, welche die Informationen über das gesamte zfDNA-Methylom umfassten. Wie in der Abb. 35 zu sehen, zeigten alle 12 hergestellten WGBS *Libraries* einen klaren, großen *Peak* bei ca. 300bp. Bei den größeren 300 bis 1.000bp Peaks handelte es sich um die sogenannte *Daisy Chains*, aneinander hybridisierte ssDNA-Fragmente. Nach Angaben des Herstellers beeinflussen sie weder die darauffolgende Hybridisierungsreaktion noch die eigentliche Sequenzierung und müssen somit nicht eliminiert werden.

Im zweiten Schritt wurden die hergestellten WGBS *Libraries* quantifiziert, äquimolar *gepoolt* und mit dem „SeqCap Epi Enrichment Kit“ prozessiert. Das hier verwendete Kit enthielt die sogenannten „Capture Probes“, die speziell von „Roche“ für dieses Projekt synthetisiert wurden. Die „Capture Probes“ hybridisierten gezielt an die 638 Regionen des Plasma *Panels* (siehe Abschnitt 9.3, Tab. 7). Nach der Hybridisierung wurden die „Capture Probes“ samt

den gebundenen differentiell methylierten Regionen angereichert, gewaschen und amplifiziert. Die amplifizierte *Library* wurde dann quantifiziert und einer Qualitätskontrolle unterzogen. Das Ergebnis des „Agilent 2100 High Sensitivity DNA Kit“-Laufs ist in der Abb. 36 dargestellt. Die fertige *Library* wies einen hohen Peak bei ca. 300bp und entsprach somit den Sequenzieranforderungen des „MiSeq“.

4.2.2 Sequenzierung und Datenanalyse

Zunächst musste die Sequenzierung auf dem „MiSeq“ optimiert werden. Die Sequenzierung erfolgte in einem 76 PE Modus. D.h., dass von beiden Enden die ersten 76bp der sequenzierten DNA-Fragmente abgelesen wurden. Um die optimale Clusterdichte zu erzielen, wurde die *Library* auf 4pM verdünnt. Die hier beschriebene *Libraries* waren unbalanciert. Als unbalanciert werden *Libraries* bezeichnet, deren AT- bzw. GC-Konzentration weniger als 40% oder mehr als 60% beträgt. Solche *Libraries* weisen aufgrund ihrer Zusammensetzung meist nicht zufriedenstellende Sequenzierqualität auf. Um sie zu verbessern, kann die *Library* mit „PhiX Control V3“ versetzt werden. Die Konzentration an „PhiX“ muss je nach *Library* individuell angepasst werden. Die optimale Konzentration an „PhiX Control V3“ betrug in diesem Fall 10%.

Nach der Sequenzierung wurden die Daten im „FastQ“-Format gespeichert.

Die Qualität der Rohdaten wurde unter Verwendung der „FastQC“ Software überprüft, die Ergebnisse sind in den Abbildungen 37 und 38 dargestellt.

Aufgrund der 76 PE Sequenzierung, lag die *Read* Länge bei 76bp. Der Gehalt an Adaptoren sowie nicht identifizierbaren Signalen innerhalb eines *Reads* betrug 0% (Abb. 37A). Die Rohdaten wiesen einen „Phred“-Score von über 30, was einer Sequenziergenauigkeit von mehr als 99,9% entsprach (Abb. 37B). Die Basenzusammensetzung (Thymin-Gehalt bei ca. 50%, Cytosin-Gehalt bei nahezu 0%, Adenin- und Guanin-Gehalt bei 25%) wies auf eine erfolgreiche Bisulfitkonvertierung hin (Abb. 37C). Bei den ersten 10nt des zweiten *Reads* (Abb. 37C, gelbe Markierung) handelte es sich um einen durch das Enzym „Adaptase“ erzeugten Überhang. Das Abweichen des experimentell ermittelten vom theoretisch berechneten GC-Gehalt lag auch an der Bisulfitkonvertierung (Abb. 37D).

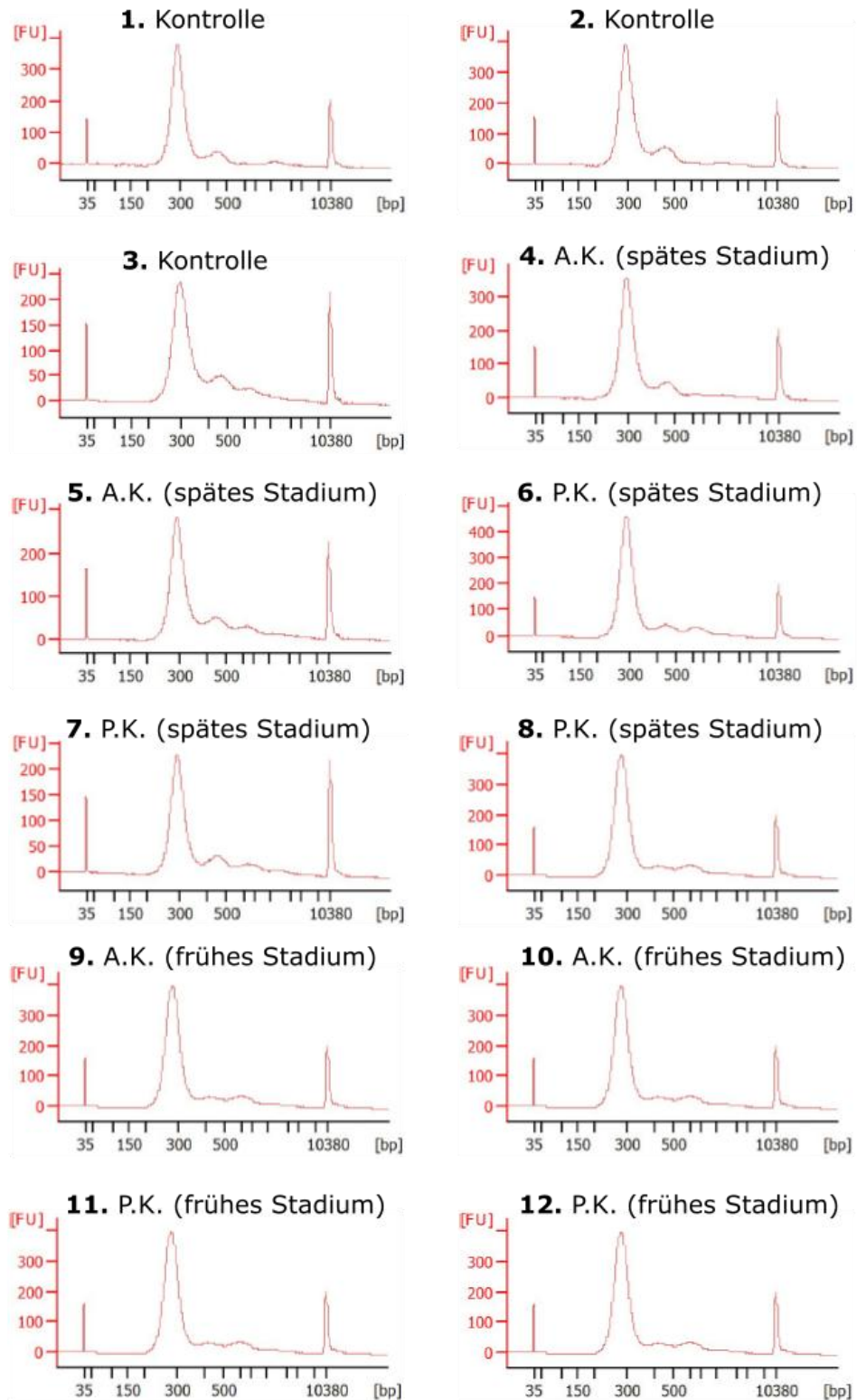


Abb. 35: Die „Agilent 2100 High Sensitivity Kit“-Analyse der hergestellten WGBS *Libraries*. Alle *Libraries* zeigten einen hohen *Peak* bei 300bp. Die *Peaks* bei 35 und 10.380bp entsprachen den oberen und unteren Markern (Adenokarzinom (A.K.), Plattenepithelkarzinom (P.K.), engl. *fluorescent units* (FU)).

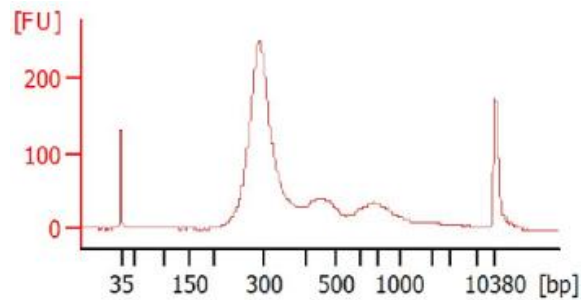


Abb. 36: Die 638 im Plasma *Panel* enthaltenen differentiell methylierte Regionen wurden mit dem „SeqCap Epi Enrichment Kit“ angereichert und amplifiziert. Die fertige NGS *Library* wurde mit dem „Agilent 2100 High Sensitivity Kit“ analysiert und zeigte einen hohen *Peak* bei ca. 300bp. Die *Peaks* bei 35 und 10.380bp entsprachen den oberen und unteren Markern (engl. *fluorescent units* (FU)).

Die Anzahl an PCR-Duplikaten lag bei ca. 15%. Die Anzahl an deduplizierten Sequenzen (rote Kurve) wich stark von der Gesamtmenge (blaue Kurve) ab (Abb. 38A). Das ist jedoch für ein *Panel* nicht ungewöhnlich. Im Gegensatz zu einer genomweiten Sequenzierung, wird bei einem *Panel* nur ein kleiner Bereich des Genoms sequenziert. Das führt zu einer sehr geringen Komplexität der *Library* und dementsprechend zur Entstehung von PCR-Duplikaten. Die Anzahl an Kmeren ist sehr gering (Abb. 38B) und für die weitere Auswertung nicht störend.

Zusammenfassend lässt sich sagen, dass die *Panel* Sequenzierdaten eine sehr gute Qualität aufwiesen. Um die Daten zu prozessieren wurden zwei Schritte durchgeführt. Zunächst wurde unter Verwendung der „Cutadapt“ Software der 10nt lange Überhang am Anfang von *Read 2* (Abb. 37C, gelbe Markierung) sowie Adapter entfernt. Dann wurden die PCR-Duplikate mit der „Samtools“-Software vollständig eliminiert.

Die prozessierten Sequenzierdaten wurden dann in das „Bisulfite Analysis Toolkit“ geladen. Das *Alignment* erfolgte mit „Segemehl“ gegen das „HG19“ Referenzgenom. Die *Mapping* Effizienz betrug mindestens 90%. D.h., mindestens 90% der Rohdaten konnten dem Referenzgenom zugeordnet werden. Die mittlere *Coverage*, also die Sequenziertiefe, betrug je nach Probe 10 bis 30fach (Abb. 39).

Im nächsten Schritt sollte die DNA-Methylierung detektiert werden. Hierfür wurden die 12 *Alignments* in das „BAT_calling“ Modul geladen. Die ermittelten Positionen wurden dann zunächst unter Verwendung der „Bedtools“ gegen das „HG19“ Referenzgenom annotiert. Dann wurden die methylierten Positionen mit dem „BAT_filtering“ Modul nach einer *Coverage* von mindestens achtfach gefiltert. Des Weiteren wurden mit dem Modul für das Erstellen eines Klassifikators nur solche Positionen ausgewählt, die sich zum einen in einer CpG Region befanden und zum anderen im Plasma *Panel* (Abschnitt 9.3, Tab. 7) aufgelistet waren (Abb. 40).

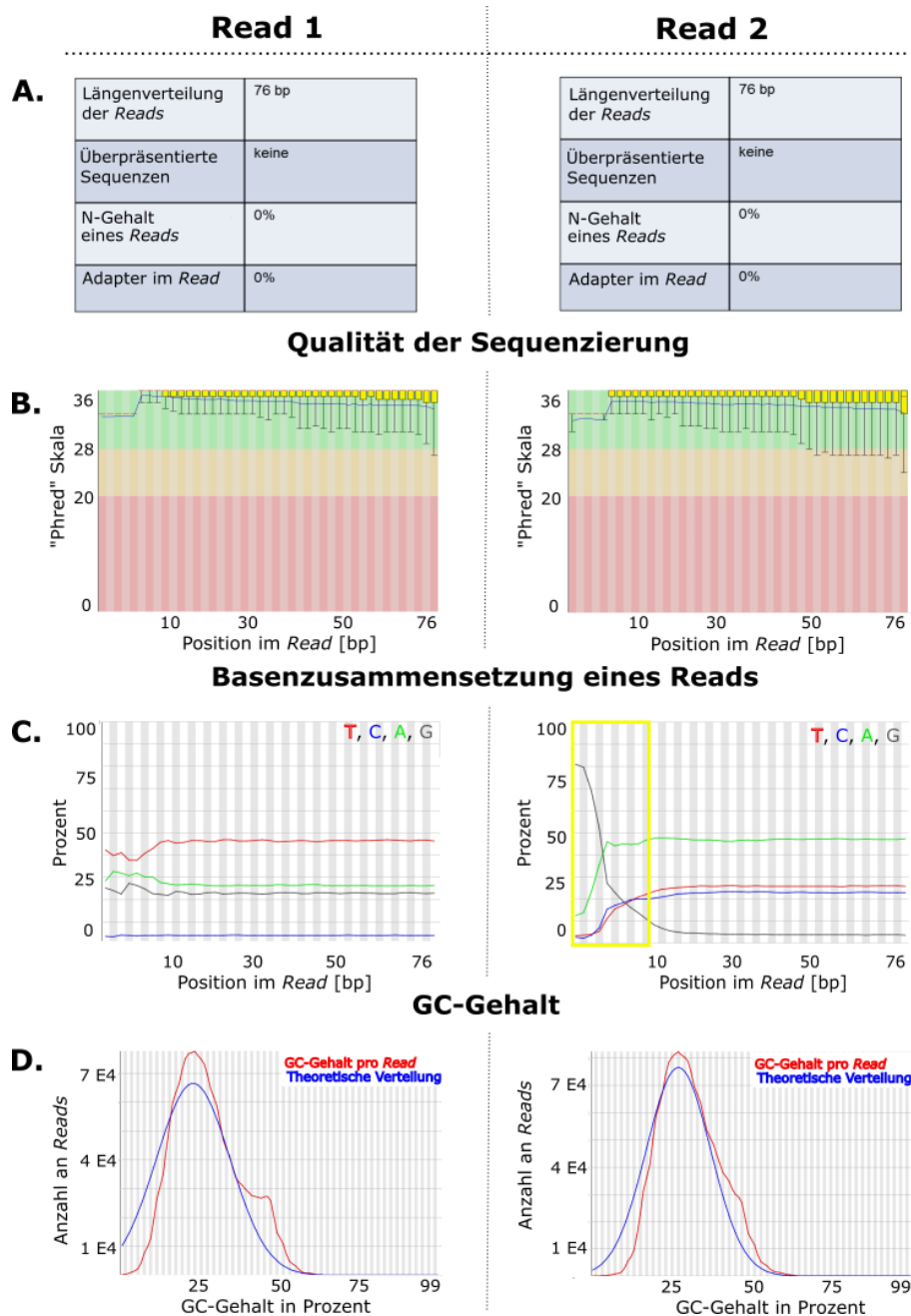


Abb. 37: Ergebnisse der mit der „FastQC“ Software durchgeführten Qualitätskontrolle der Plasma *Panel* Rohdaten **A.** Aufgrund der 76 PE Sequenzierung, lagen die Längen der *Reads* bei 76bp. Der Adapter und N-Gehalt der *Reads* betrug 0%. Überpräsenzierte Sequenzen konnten nicht detektiert werden. **B.** Die Sequenziergenauigkeit betrug 99,9%, was einem „Phred“-Wert von über 30 entsprach (Farbgebung: Grün=Qualität >99,9%, Gelb = Qualität 99-99,9%, Rot=Qualität <90%). **C.** Die Basenzusammensetzung wies auf eine erfolgreiche Bisulfitkonvertierung hin. Während der Gehalt an Adenin (A; Grün) und Guanin (G; Grau) bei ca. 25% lag, wurden im Laufe der Bisulfitkonvertierung Cytosine durch Thyminen ersetzt, weshalb der Thymingehalt (T; Rot) viel höher als erwartet bei ca. 50% und Cytosinegehalt (C; Blau) viel niedriger bei nahezu 0% war. *Read 1* verhielt sich komplementär zu *Read 2*. Der durch „Adaptase“ erzeugte, 10nt lange Überhang des *Read 2* wurde gelb markiert. **D.** Wegen der Bisulfitkonvertierung wich der experimentell ermittelte (rote Kurve) GC-Gehalt der *Library* vom theoretisch vorhergesagten (blaue Kurve) stark ab.

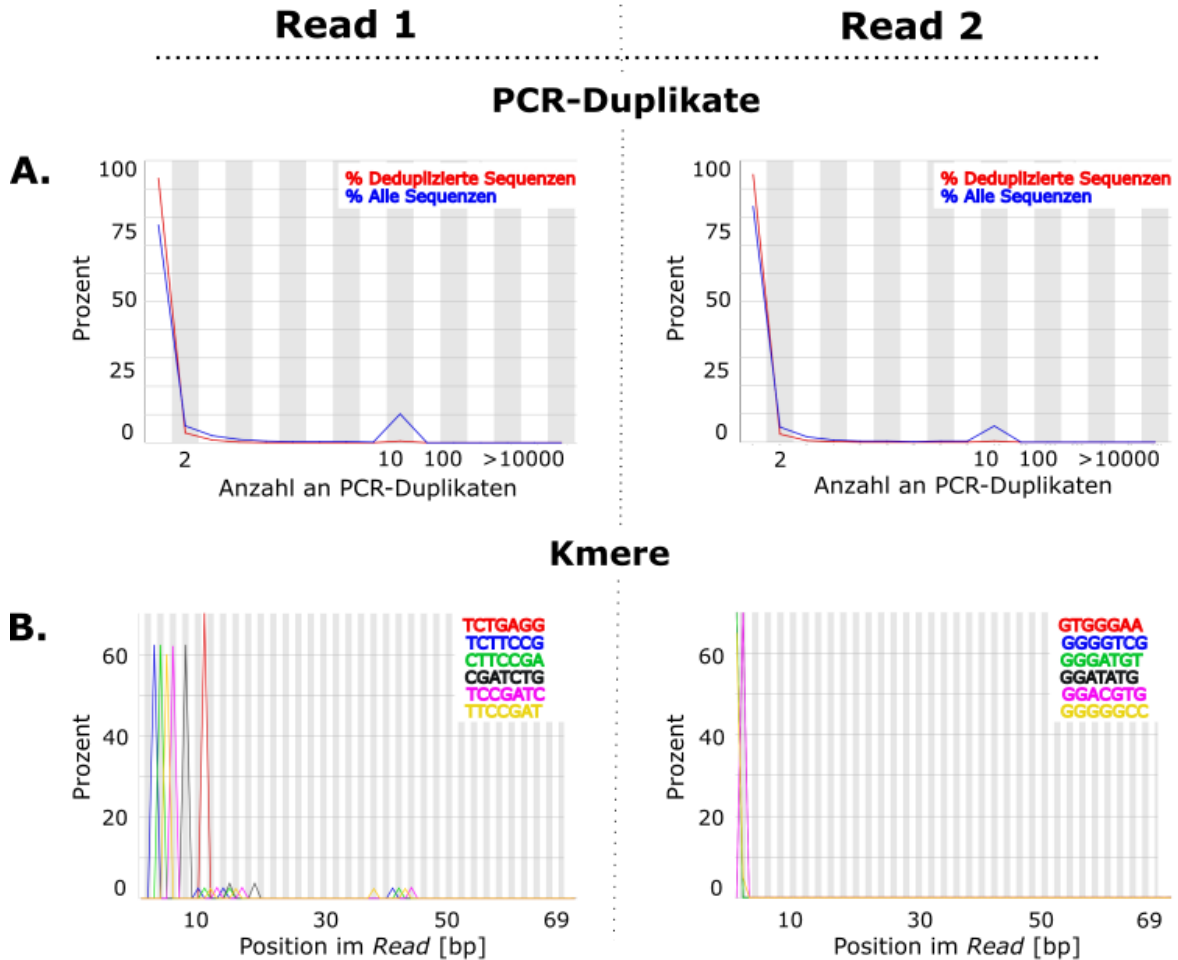


Abb. 38: Ergebnisse der mit „FastQC“ Software durchgeführten Qualitätskontrolle der Plasma *Panel* Rohdaten **A.** Die Menge an PCR-Duplikaten lag pro Probe bei ca. 15%. Die hohe Duplikationsrate trat aufgrund der niedrigen Komplexität der *Library* auf. Die Gesamtmenge aller Sequenzen (blaue Kurve) wich stark von der Prozentanzahl an deduplizierten Sequenzen (rote Kurve) ab. Die PCR-Duplikate wurden unter Verwendung von „Samtools“ entfernt. **B.** Innerhalb der *Reads* waren nur wenige Kmere enthalten (farbig dargestellte Sequenzen sowie entsprechend gefärbte *Peaks*).

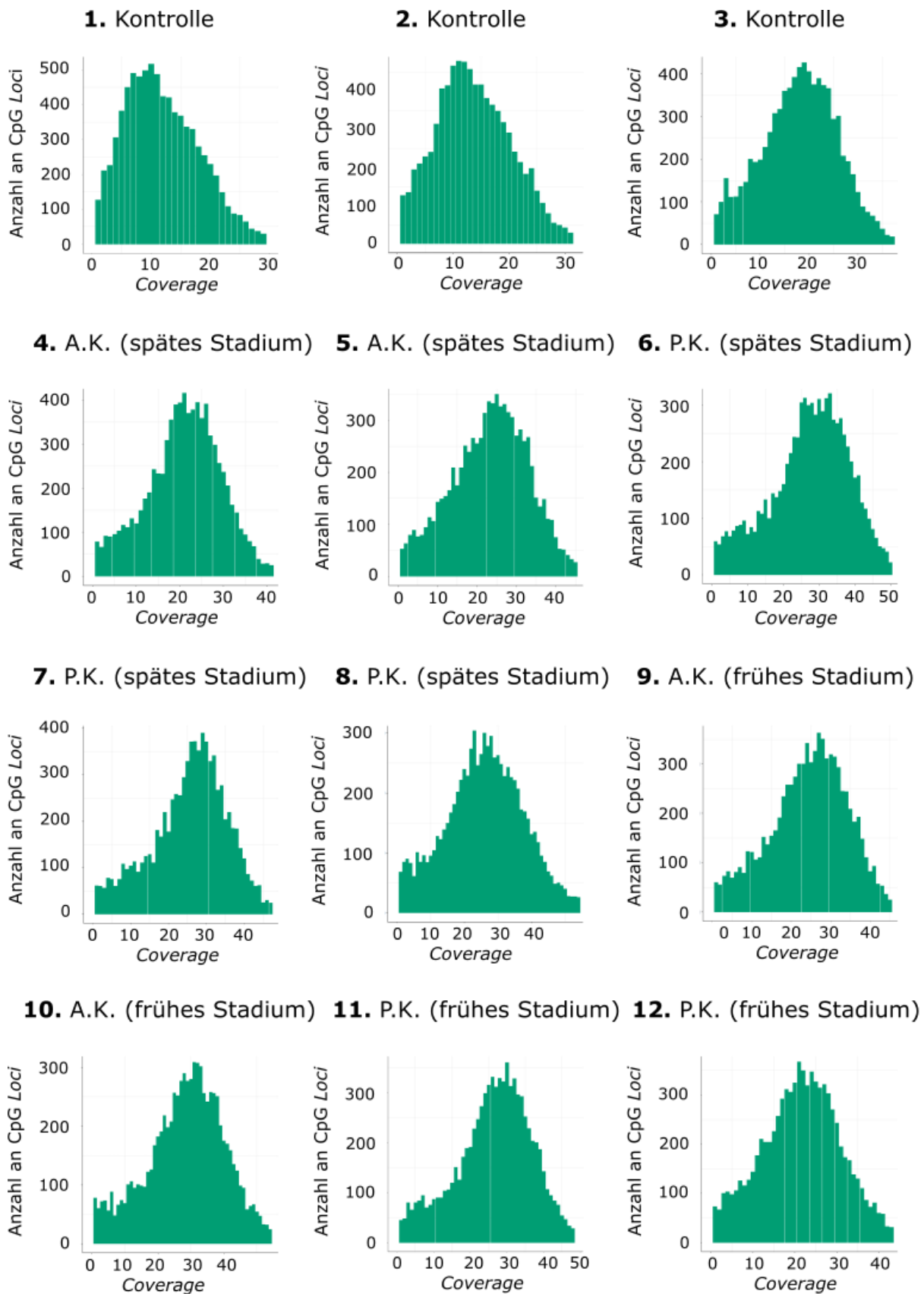


Abb. 39: Histogram-Plot der *Coverage* aller angereicherten und sequenzierten Regionen. Alle Proben wurden mit einer mittleren *Coverage* von 10 bis 30fach sequenziert. Die hier dargestellten Positionen wurden zum Ermitteln von DNA-Methylierungsraten verwendet. Die Berechnung erfolgte mit dem „BAT_Calling“ Modul.

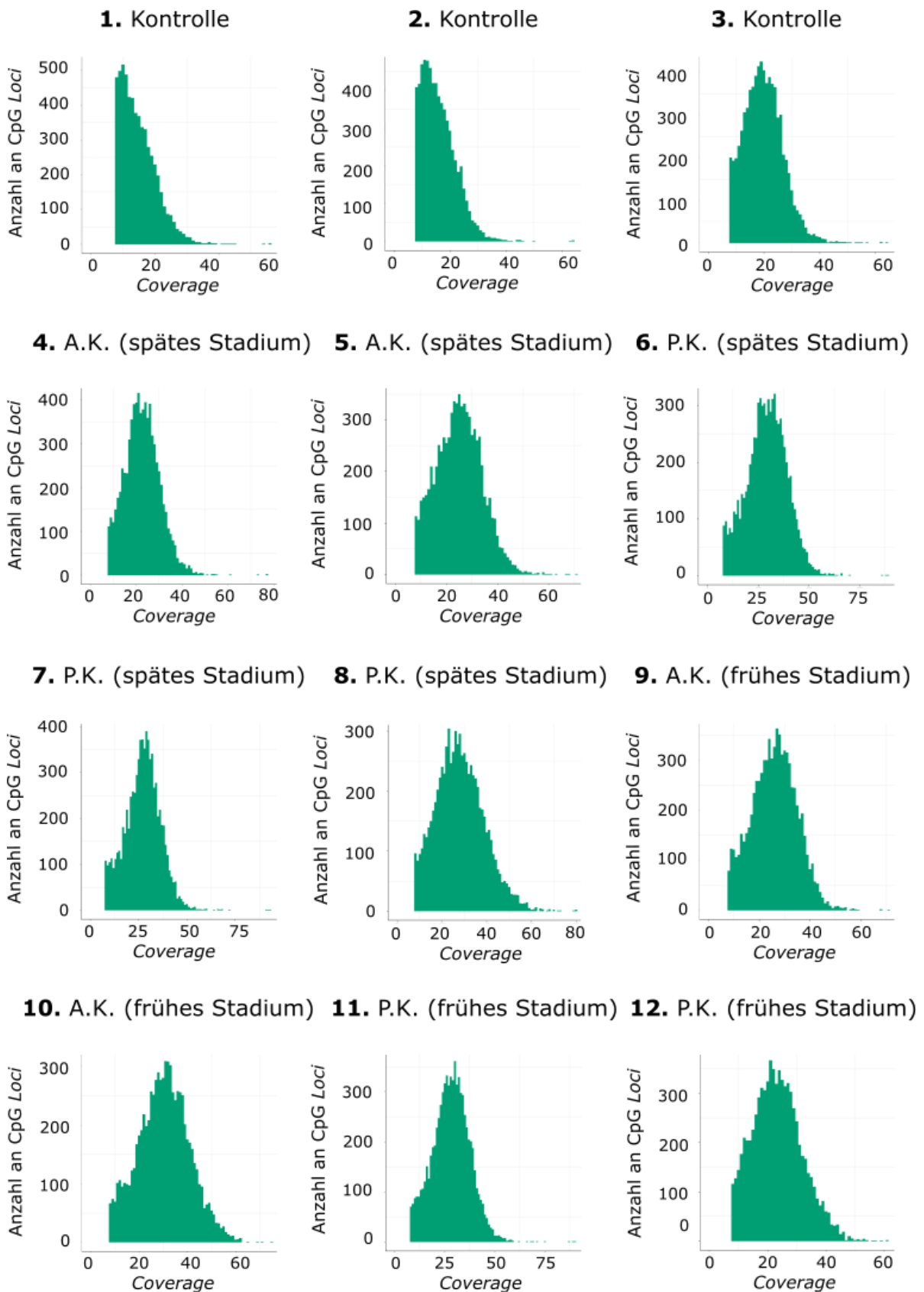


Abb. 40: Histogram-Plot der *Coverage* aller Positionen, die für das Erstellen des Klassifikators verwendet wurden. Um die Positionen zu ermitteln wurden die zuvor mit dem „BAT_Calling“ Modul detektierten, methylierten Positionen im nächsten Schritt mit dem „BAT_Filtering“ Modul nach einer *Coverage* von mindestens achtfach sowie CpG- und Plasma *Panel* Kontext gefiltert.

4.2.3 Erstellen eines Klassifikators

Die ermittelten zfDNA-Methylierungsraten wurden zum Erstellen eines Klassifikators verwendet. Wie in Abschnitt 3.2.3.3 beschrieben, wurde hierfür die „Glucose Omics Explorer“ Software verwendet, die folgende Klassifikationsalgorithmen enthielt: „k-Nearest Neighbors Algorithm“ (kNN), „Support Vector Machines“ (SVM) und „Random Trees“ (RT).

Das Plasma *Panel* wurde so designt, dass es in der Lage sein sollte, die Informationen bezüglich der Malignität, der Entität, des Stadiums und der Prognose zu liefern. Aufgrund zum Zeitpunkt dieser Arbeit noch laufender DZL interner Datenerfassungen und somit bisher nur unvollständig verfügbaren Metadaten, ließ sich keine Aussage bezüglich der Prognose der untersuchten Patientenkohorte machen. Alle weiteren Fragestellungen konnten durch die Wahl eines geeigneten Klassifikators zuverlässig beantwortet werden.

Um einen Klassifikator zu bewerten, wurden zwei Parameter betrachtet: die Richtigkeit (engl. *accuracy*) und die Komplexität. Die Richtigkeit eines Klassifikators wurde in Werten zwischen 0 und 1 angegeben, wobei 0 einer Richtigkeit von 0% und 1 einer Richtigkeit von 100% entsprach. Die Komplexität gab an, wie viele differentiell methylierte Positionen analysiert werden mussten, damit der Klassifikator diese Richtigkeit erzielte. Je weniger Positionen ausgewertet werden mussten, desto geeigneter war der Klassifikator für die Klinik. Denn mit der Anzahl an zu analysierenden Positionen steigen die Fehlerrate, Zeit und Kosten der Methode.

Die erste Fragestellung lautete, ob ein Patient generell an einem malignen Lungentumor litt. Hierfür lieferten sowohl der kNN als auch der RT Algorithmus eine Richtigkeit von 100%. Der RT Algorithmus benötigte für die Klassifikation 237 differentiell methylierte, im *Panel* enthaltene Positionen. Der kNN dagegen nur 10 Positionen, was ihn damit für diese Fragestellung als optimal qualifizierte (Abb. 41A).

Die Fragestellung bezüglich der Entität konnten alle drei Algorithmen mit einer Richtigkeit von 100% beantworten. Für die Berechnungen benötigte kNN 22 Positionen, SVM 22 Positionen und RT 10 Positionen. Somit war der RT Algorithmus für diese Fragestellung am besten geeignet (Abb. 41B).

Für die letzte Fragestellung bezüglich des Stadiums, war es am schwierigsten einen geeigneten Klassifikator zu wählen. Der SVM Algorithmus schaffte es, unter Verwendung von 523 Positionen die späten Tumorstadien mit einer 80%igen Richtigkeit zu differenzieren (Abb. 41C).

Alle Positionen und Klassifikationsparameter sind im Anhang detailliert beschrieben (siehe Abschnitt 9.3, Tab. 8-10).

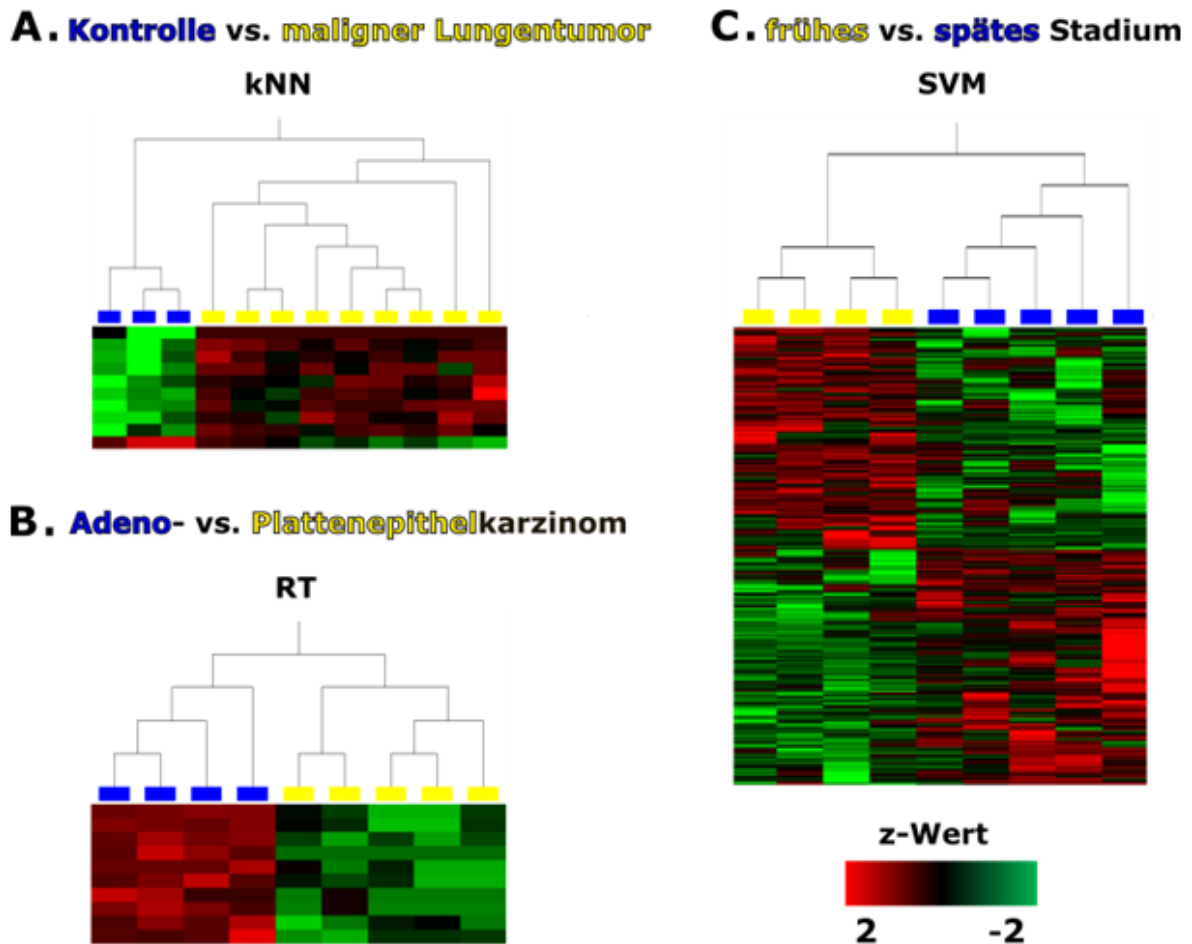


Abb. 41: Die ermittelten zfdNA-Methylierungsraten wurden in die „Qlucore Omics Explorer“ Software geladen und unter Verwendung folgender Klassifikationsalgorithmen analysiert: „k-Nearest Neighbors Algorithm“ (kNN), „Support Vector Machines“ (SVM) und „Random Trees“ (RT). **A.** Der kNN Algorithmus konnte unter Verwendung von 10 Positionen die gesunden (Kontrolle) von den an einem malignen Lungenkarzinom erkrankten Patienten unterscheiden. Sowohl die frühen (I, II) als auch die späten (III, IV) Lungenkarzinomstadien wurden mit 100%iger Richtigkeit klassifiziert. **B.** Der RT Algorithmus analysierte 10 Positionen, um mit einer 100%igen Richtigkeit die Entität des Tumors zu ermitteln. **C.** Die späten Tumorstadien (III, IV) konnten mit einer 80%igen Richtigkeit mit dem SVM Algorithmus identifiziert werden, dabei wurden 523 Positionen analysiert.

4.3 Ergebnisse des Projekts „multi-OMICS Studie zur Untersuchung von Risiken der E-Zigaretten“

Um die kurzfristigen Auswirkungen der E-Zigaretten auf menschliche Lungenzellen zu untersuchen, wurde eine multi-OMICS Studie designt und durchgeführt. Der Aufbau der Studie ist bereits im Abschnitt 3.3 detailliert beschrieben worden. In Kürze, A549 und primäre AECII Zellen wurden mit unterschiedlich konzentrierten Zigaretten- und E-Zigarettenextrakten behandelt. Drei Stunden nach der Behandlung erfolgte die Messung der Apoptoseraten. Nach 48 Stunden wurden unter Verwendung funktioneller, morphologischer sowie molekulargenetischer Methoden gleichzeitig zelluläre Veränderungen von Viabilität, Morphologie, Genexpression, Chromatinstatus und DNA-Methylierung erfasst.

4.3.1 Funktionelle und morphologische Befunde

Wie bereits im Abschnitt 3.3.3 beschrieben, wurden sowohl die A549 als auch die primäre AECII Zelllinien mit 0,1 bis 5%igen Zigaretten bzw. E-Zigarettenextrakten behandelt.

Zunächst wurden die nach der Behandlung stattfindenden Veränderungen der Zellviabilität und Apoptose unter Verwendung von „ApoONE® Homogeneous Caspase-3/7 Assay“ sowie „CellTiter-Blue® Cell Viability Assay“ untersucht. Die Ergebnisse sind in den Abbildungen 42 und 43 dargestellt. Drei Stunden nach der Stimulation der A549 und AECII Zellen mit hoch konzentrierten, nikotinhaltigen Zigaretten- sowie E-Zigarettenextrakten wurde im Vergleich zur Kontrolle eine deutlich höhere Apoptoserate detektiert. 48 Stunden nach der Behandlung kam es zu einem drastischen Abfall der Zellviabilität. Der nikotinfreie E-Zigarettenextrakt hatte in hohen Konzentrationen hingegen keinen Einfluss auf die Apoptoseraten sowie Zellviabilität der primären AECII Zellen. Des Weiteren reagierte die A549 Tumorzelllinie auf die Behandlung mit nikotinfreiem E-Zigarettenextrakt mit einer erhöhten Zellproliferation. Dabei nahm die Zellviabilität nach 48 Stunden um 20-40% zu. Eine Veränderung der Apoptoseraten konnte nicht detektiert werden.

Um diese Beobachtung zu bestätigen, wurde ein unabhängiges Experiment unter Verwendung des „IncuCyte ZOOM Live Cell Analysis System“ durchgeführt (Abb. 44). Bei dem Gerät handelt es sich um ein Mikroskop, das direkt in einen Brutschrank montiert wird und mit einer hochauflösenden CCD Kamera verbunden ist. Dadurch ist es möglich, die Konfluenz und Apoptose der Zellen 48 Stunden lang nach der Behandlung in Echtzeit zu verfolgen. Auch in diesem Experiment führte die Behandlung der A549 Zellen mit hoch konzentrierten, nikotinhaltigen Zigaretten- sowie E-Zigarettenextrakten nach wenigen Stunden zur Apoptose (Abb. 44, violette Signal). Nach 48 Stunden nahm die Konfluenz der

behandelten Zellen stark ab. Dagegen wurden keine Veränderungen der Apoptoseraten bei A549 Zellen detektiert, die mit dem hoch konzentriertem, nikotinfreien E-Zigarettenextrakt behandelt wurden. Im Vergleich zur Kontrolle nahm die Zellkonfluenz nach der Behandlung deutlich zu, was auf eine erhöhte Zellproliferation hinweisen könnte.

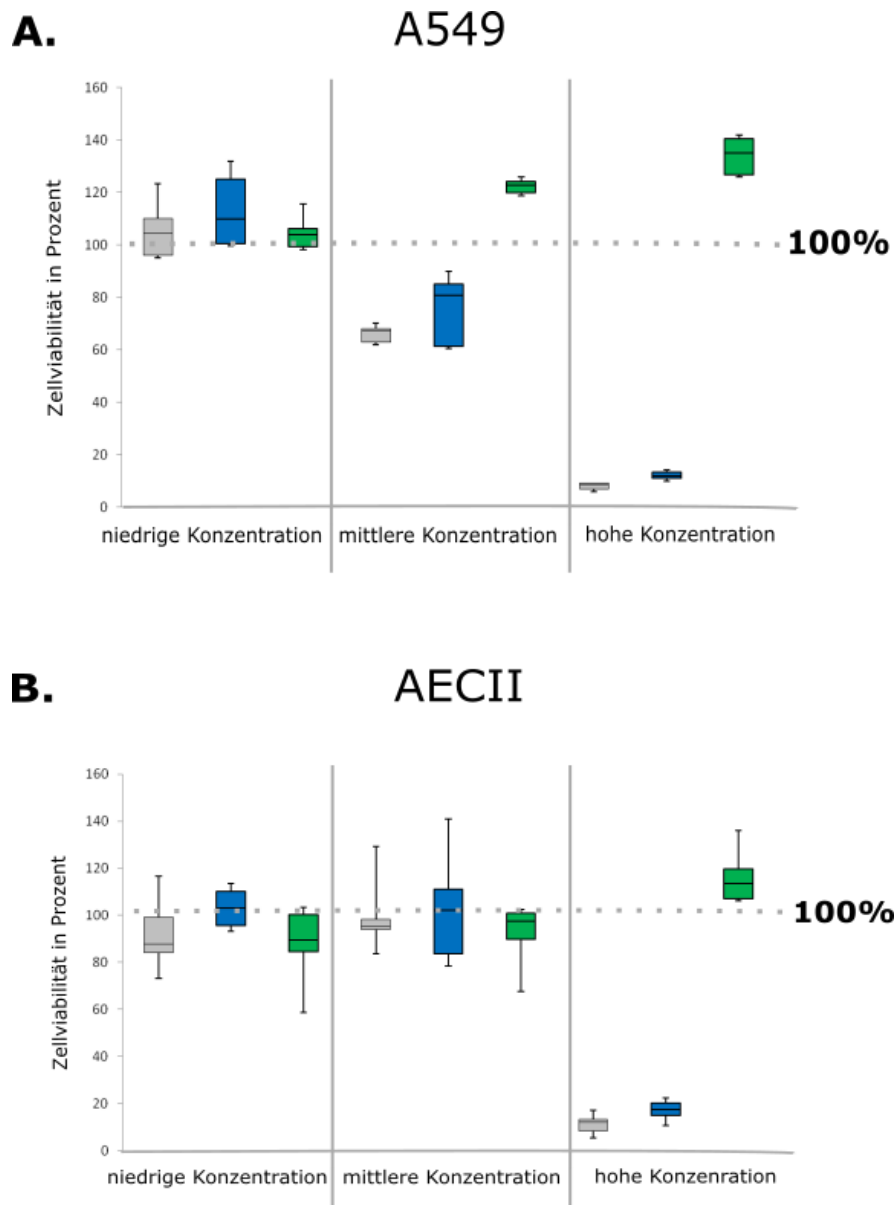


Abb. 42: Die Messung der Zellviabilität erfolgte unter Verwendung des „CellTiter-Blue® Cell Viability Assay“ 48 Stunden nach der Behandlung der Zellen mit niedrigen, mittleren und hohen Konzentrationen an Zigarettenextrakt (CSE, grau), E-Zigarettenextrakt mit Nikotin (eCSE mit Nikotin, blau) und E-Zigarettenextrakt ohne Nikotin (eCSE ohne Nikotin, grün). Die Viabilität beider Zelllinien nahm mit steigender CSE sowie eCSE (mit Nikotin) Konzentration ab. **A.** Nach der Behandlung der A549 Tumorzelllinie mit eCSE (ohne Nikotin) stieg deren Viabilität. **B.** Bei den primären AECII Zellen konnte keine Veränderung der Viabilität nach der Stimulation mit eCSE (ohne Nikotin) detektiert werden.

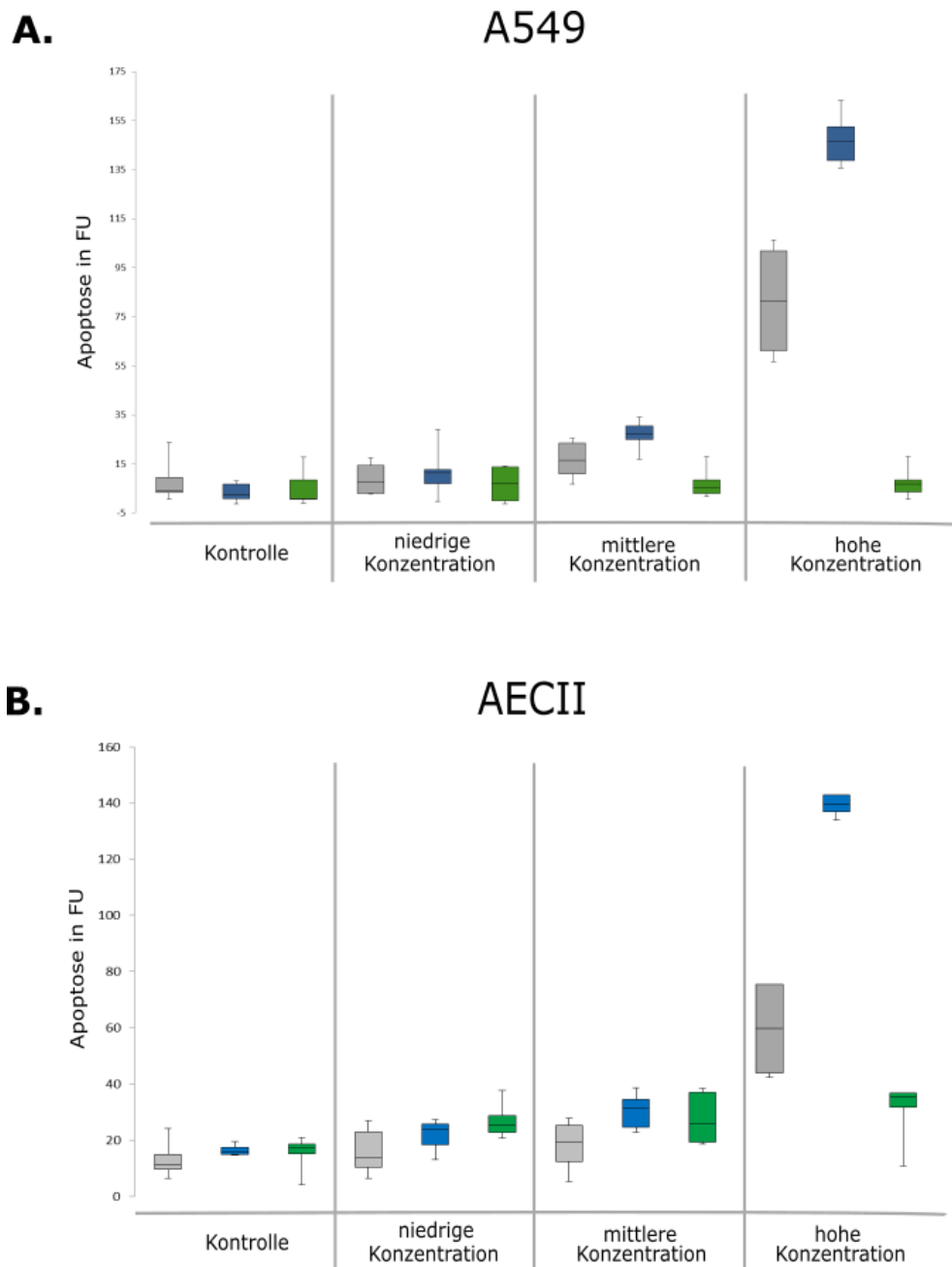
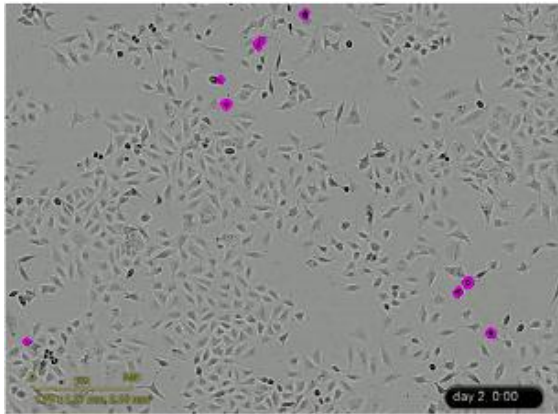
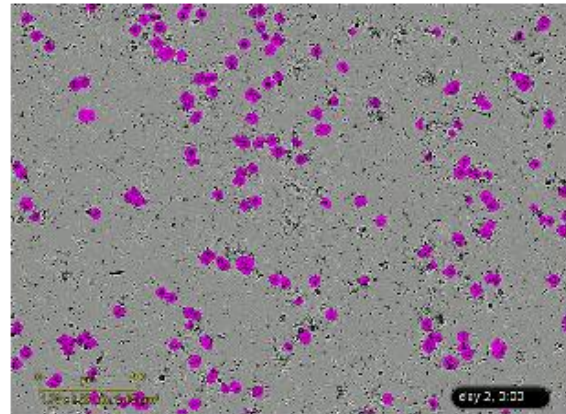


Abb. 43: Die Messung der Apoptose erfolgte unter Verwendung des „ApoONE® Homogeneous Caspase-3/7 Assay“ drei Stunden nach der Behandlung der Zellen mit niedrigen, mittleren und hohen Konzentrationen an Zigarettenextrakt (CSE, grau), E-Zigarettenextrakt mit Nikotin (eCSE mit Nikotin, blau) und E-Zigarettenextrakt ohne Nikotin (eCSE ohne Nikotin, grün). Die Ergebnisse wurden in engl. *fluorescent units* (FU) angegeben. Sowohl bei der **A.** A549 als auch **B.** AECII Zelllinie nahmen die Apoptoseraten mit steigender CSE sowie eCSE (mit Nikotin) Konzentration zu. Dagegen wurden nach der Behandlung der Zellen mit eCSE (ohne Nikotin) keine Veränderungen detektiert.

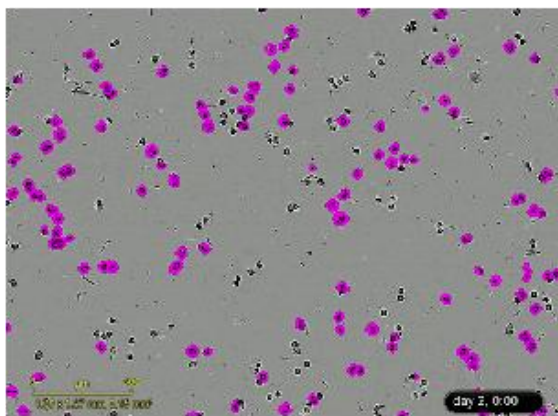
Kontrolle



CSE



eCSE
(mit Nikotin)



eCSE
(ohne Nikotin)

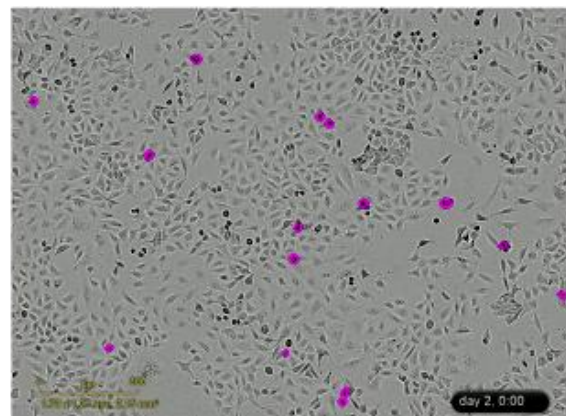


Abb. 44: Messung der Apoptose und Zellkonfluenz unter Verwendung des „IncuCyte ZOOM Live Cell Analysis System“; nach der Behandlung der A549 Zelllinie mit hoch konzentrierten, nikotinhaltenigen Zigaretten- (CSE) sowie E-Zigarettenextrakten (eCSE) stieg die Anzahl an Apoptosesignalen (violette Punkte) an, die Zellkonfluenz nahm ab. Dagegen wurde die Apoptose nach der Behandlung der Zellen mit hoch konzentriertem, nikotinfreiem E-Zigarettenextrakt (eCSE) nicht induziert, die Zellkonfluenz nahm im Vergleich zu Kontrolle zu.

Des Weiteren deuteten „IncuCyte ZOOM Live Cell Analysis System“ Aufnahmen auf morphologische Anomalien einiger mit dem nikotinfreiem E-Zigarettenextrakt behandelter Zellen hin. Aus diesem Grund wurde im nächsten Schritt die Fluoreszenzmikroskopie durchgeführt (Abb. 45). Dabei wurden die Zellkerne blau, die Plasmamembran orange und die Aktinfilamente grün angefärbt. Nach der Stimulation der A549 Zellen mit dem hoch konzentriertem, nikotinfreiem E-Zigarettenextrakt stieg im Vergleich zur Kontrolle die Anzahl multinuklearer, aneuploider Zellen. Diese Beobachtung konnte in einem unabhängigen FACS-Experiment bestätigt werden (Abb. 45).

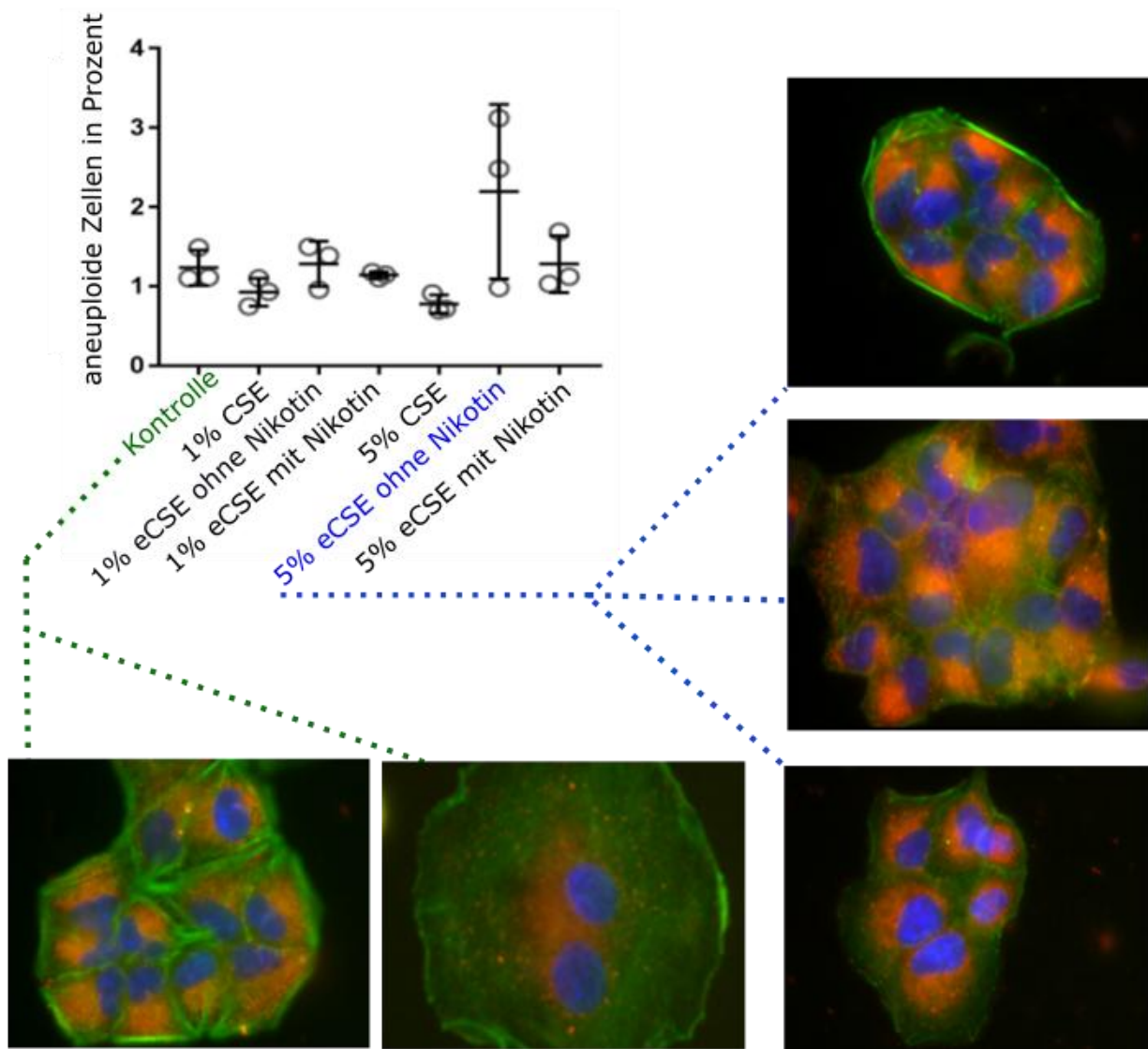


Abb. 45: Aneuploidie; die mit Zigaretten- (CSE) bzw. E-Zigarettenextrakten (eCSE) behandelten A549 Zellen wurden unter Verwendung des Fluoreszenzmikroskops untersucht. Dabei wurden die Zellkerne blau, Aktinfilamente grün und intrazelluläre Plasmamembran orange angefärbt. Die Ergebnisse wurden anhand einer FACS-Messung validiert. Im Vergleich zur Kontrolle war die Anzahl an aneuploiden Zellen nach der Behandlung mit hoch konzentriertem, nikotinfreiem eCSE deutlich erhöht. Des Weiteren waren pro aneuploide Zelle meistens mehr als drei Kerne enthalten, die zellulären Aktinstrukturen (grün) wirkten weniger ausgebildet.

4.3.2 Molekularbiologische und epigenetische Befunde

Die im vorherigen Abschnitt beschriebenen funktionellen und morphologischen Assays haben ergeben, dass E-Zigaretten Apoptose, Viabilität, Proliferation und Morphologie von Zellen beeinflussen können. Um zu erfahren, welche Mechanismen bzw. Gene dabei eine Rolle spielen, wurden folgende molekulargenetische Assays durchgeführt: RNA-Seq., ATAC-Seq. und EPIC BC. Die Rohdaten wurden zunächst unabhängig voneinander ausgewertet und dann einer integrativen Analyse unterzogen.

Die molekulargenetischen Analysen benötigen intakte RNA, DNA und Chromatin. Da aber die untersuchten nikotinhaltigen Zigaretten- und E-Zigarettenextrakte in hohen Konzentrationen zytotoxisch wirkten, wurden in diesem Fall mittel konzentrierte Extrakte eingesetzt.

Bei nikotinfreiem E-Zigarettenextrakt konnten hohe Konzentrationen beibehalten werden, da hier keine Zytotoxizität beobachtet wurde.

4.3.2.1 Auswertung von RNA-Seq. Daten

Mit Hilfe der RNA-Sequenzierung können Veränderungen der Genexpression detektiert werden. Die Methode wurde bereits im Abschnitt 3.3.6 ausführlich beschrieben. Die Sequenzierung erfolgte auf der „HiSeq4000“ Plattform. Die Daten wurden im „FastQ“-Format gespeichert und auf ihre Qualität unter Verwendung von „FastQC“ Software überprüft. Die Ergebnisse sind in der Abbildung 46 dargestellt.

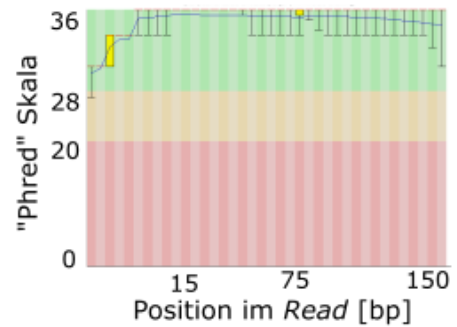
Da die Sequenzierung im 150 SE Modus durchgeführt wurde, lag die Länge der *Reads* bei 150bp (Abb. 46A). Die Rohdaten wiesen einen „Phred“-Score von über 30 auf, was einer Sequenziergenauigkeit von mehr als 99,9% entsprach (Abb. 46B). Die Basenzusammensetzung wirkte ausgeglichen, jede Base hatte einen Anteil von ca. 25%. (Abb. 46C). Der GC-Gehalt lag daher auch bei ca. 50% (Abb. 46D). Die Anzahl an Kmeren war sehr gering (Abb. 46F) und für alle weiteren Analysen nicht störend. Allerdings war die Anzahl an PCR-Duplikaten hoch (Abb. 46E), sodass sie im weiteren Verlauf unter Verwendung von „Samtools“ entfernt werden mussten.

Nachdem die Rohdaten gegen das „HG19“-Referenzgenom *aligned* wurden, konnte die differentielle Expression unter Verwendung von „edgeR“ detektiert werden. Die ermittelten Listen wurden schließlich in „WebGestalt“ geladen und einer GO Analyse unterzogen.

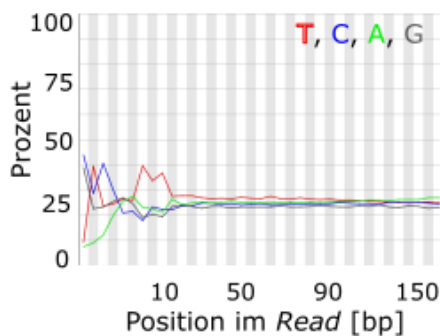
A. Parameter

| | |
|-------------------------------|--------|
| Längenverteilung der Reads | 150 bp |
| Reads mit schlechter Qualität | 0% |
| GC Gehalt | 48% |

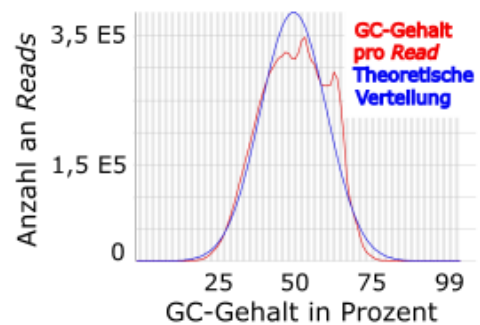
B. Qualität der Sequenzierung



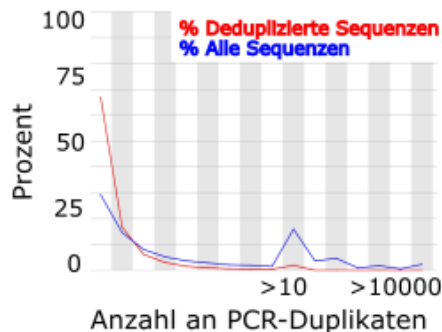
C. Basenzusammensetzung eines Reads



D. GC-Gehalt



E. PCR-Duplikate



F. Kmere

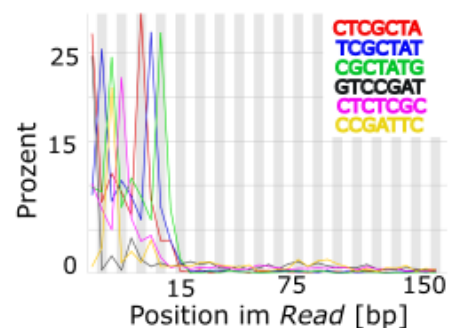


Abb. 46: Ergebnisse der mit der „FastQC“ Software durchgeführten Qualitätskontrolle der RNA-Seq. Rohdaten **A.** Die Länge der Reads betrug 150bp, was bei einem 150 SE Lauf zu erwarten war. **B.** Die Qualität der Sequenzierung wurde in „Phred“-Werten angegeben und lag bei über 30, was einer Sequenziergenauigkeit von über 99,9% entsprach (Farbgebung: Grün=Qualität >99,9%, Gelb = Qualität 99-99,9%, Rot=Qualität <90%) **C.** Die Basenzusammensetzung entsprach einer RNA-Seq. Library und lag pro Base bei ca. 25%. **D.** Der GC-Gehalt lag bei ca. 50%. **E.** Die Anzahl an PCR-Duplikaten war hoch und musste daher im weiteren Verlauf der Auswertung unter Verwendung von „Samtools“ eliminiert werden. **F.** Innerhalb der Reads waren nur wenige Kmere enthalten (farbig dargestellte Sequenzen sowie entsprechend gefärbte Peaks).

4.3.2.2 Auswertung von ATAC-Seq. Daten

Durch Verwendung von ATAC-Seq. können Veränderungen des Chromatinstatus erfasst werden. Die Methode wurde bereits im Abschnitt 3.3.7 ausführlich beschrieben. Auch in diesem Fall erfolgte die genomweite Sequenzierung auf der „HiSeq4000“ Plattform. Die Daten wurden im „FastQ“-Format gespeichert und auf ihre Qualität unter Verwendung von „FastQC“ Software überprüft. Die Ergebnisse sind in den Abbildungen 47 und 48 dargestellt.

Da es sich hierbei um eine 50 PE Sequenzierung handelte, lag die detektierte Länge der *Reads* bei 50bp (Abb. 47A). Ansonsten sahen die Ergebnisse den Qualitätsparametern einer RNA-Seq. *Library* ähnlich. Die Rohdaten wiesen einen „Phred“-Score von über 30 auf (Abb. 47B). Jede Base hatte einen Anteil von ca. 25% (Abb. 47C), sodass auch der GC-Gehalt ca. 50% betrug (Abb. 47D). Die Anzahl an Kmeren war zwar etwas höher als bei RNA-Seq. *Libraries*, aber dennoch sehr gering (Abb. 48B) und für weitere Analysen nicht störend. Die Anzahl an PCR-Duplikaten war hoch, was jedoch häufig in Verbindung mit ATAC-Seq. *Libraries* beschrieben wird (Abb. 48A) [252].

Nachdem die PCR-Duplikate unter Verwendung von „Picard“ eliminiert wurden, wurden die Rohdaten gegen das „HG19“-Referenzgenom *aligned*. Um das offene Chromatin zu detektieren wurde der „MACS2“ Algorithmus verwendet.

4.3.2.3 Auswertung von EPIC BC Daten

EPIC BC lieferte Informationen über den Methylierungsstatus von ca. 850.000 CpG *Loc*i. Die Methode wurde bereits im Abschnitt 3.3.8 ausführlich beschrieben.

Die Auswertung der Rohdaten erfolgte mit der „Illumina® GenomeStudio“ Software wie im Abschnitt 3.3.8.4 beschrieben. Die detektierten „avg.beta“-Werte zeigten eine bimodale Dichteverteilung mit der überwiegenden Zahl an Cytosinen entweder stark oder wenig methyliert. Des Weiteren wird deutlich, dass der EPIC BC in „avg.beta“-Bereichen unter 0,1 bzw. über 0,9 generell unpräzise Werte liefert. Das Hintergrundsignal war sehr gering, was auf eine gute Qualität der Rohdaten hindeutete (Abb. 49).

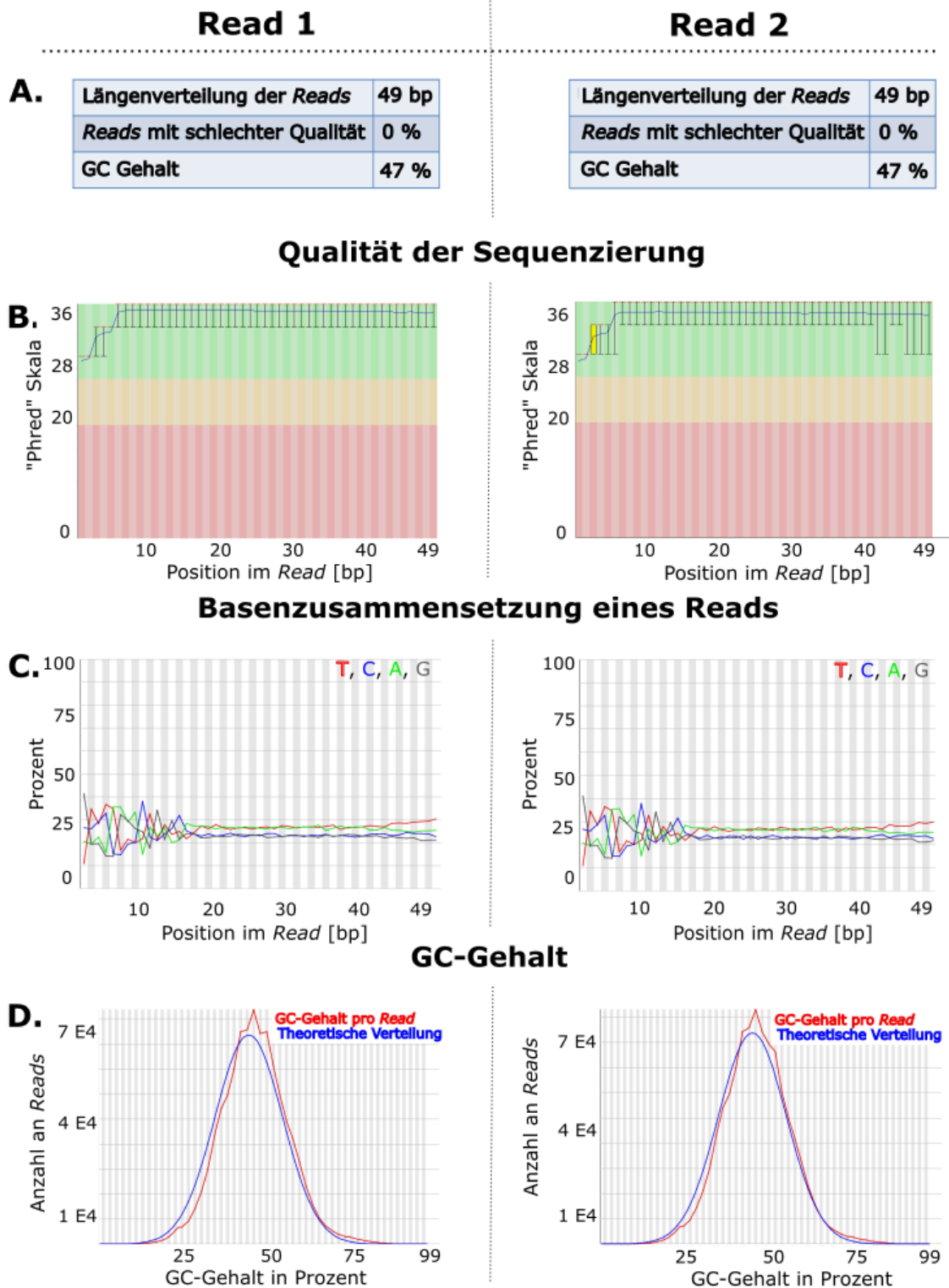


Abb. 47: Ergebnisse der mit der „FastQC“ Software durchgeführten Qualitätskontrolle der ATAC-Seq. Rohdaten **A.** Die Länge der Reads betrug 50bp, was bei einem 50 PE Lauf zu erwarten war. **B.** Die Qualität der Sequenzierung wurde in „Phred“-Werten angegeben und lag bei über 30, was einer Sequenziergenauigkeit von über 99,9% entsprach (Farbgebung: Grün=Qualität >99,9%, Gelb = Qualität 99-99,9%, Rot=Qualität <90%) **C.** Die Basenzusammensetzung entsprach einer ATAC-Seq. Library und lag pro Base bei ca. 25%. **D.** Der GC-Gehalt lag bei ca. 50%.

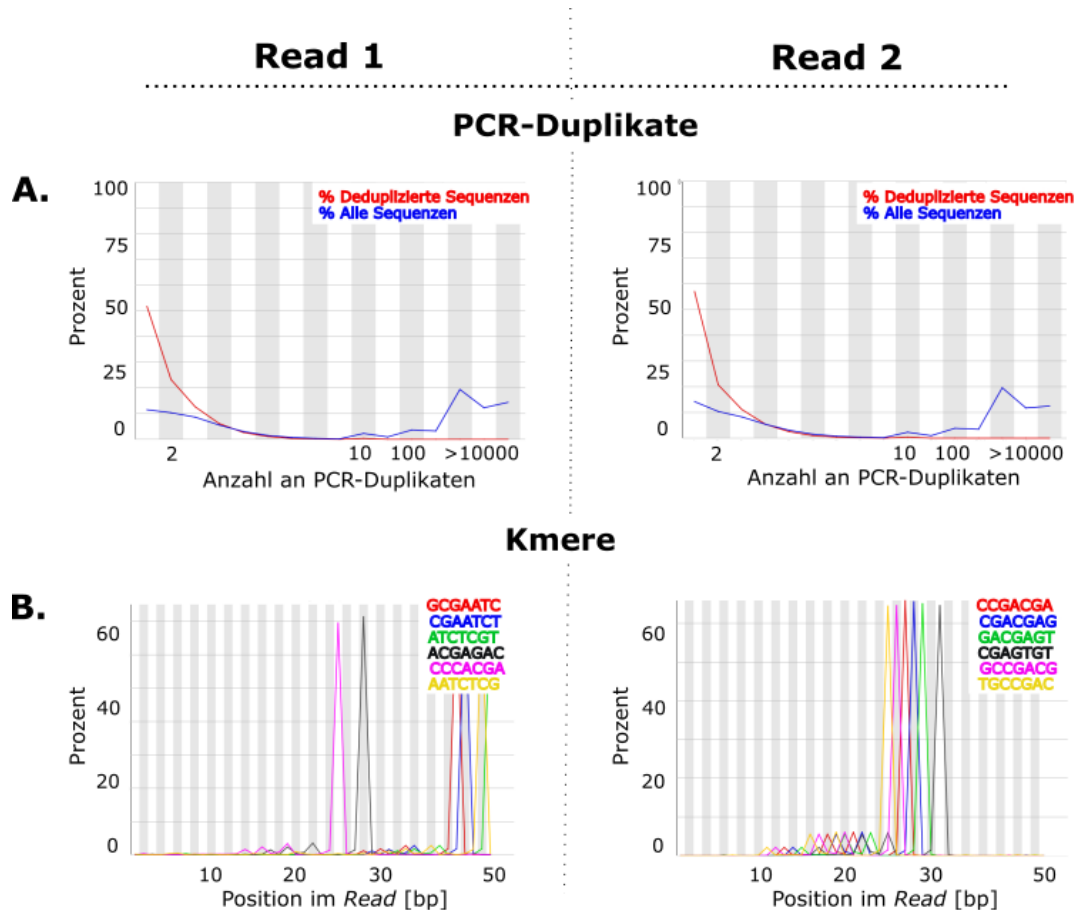


Abb. 48: Ergebnisse der mit „FastQC“ Software durchgeführten Qualitätskontrolle der ATAC-Seq. Rohdaten **A.** Die Anzahl an PCR-Duplikaten war hoch und musste deshalb im weiteren Verlauf der Auswertung unter Verwendung von „Picard“ eliminiert werden. **B.** Innerhalb der *Reads* waren nur wenige Kmere enthalten (farbig dargestellte Sequenzen sowie entsprechend gefärbte *Peaks*).

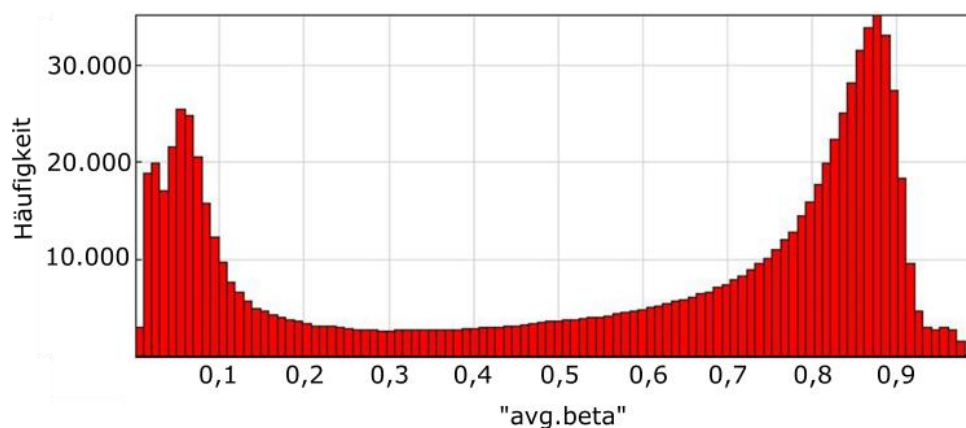


Abb. 49: Die Qualitätskontrolle der EPIC BC Rohdaten wurde unter Verwendung der „Illumina® GenomeStudio“ Software durchgeführt. Die Ergebnisse sind als Histogramm-Plot der „avg.beta“-Werte dargestellt. Die Werte zeigten eine bimodale Dichteverteilung mit der überwiegenden Zahl an Cytosinen entweder stark oder wenig methyliert.

4.3.2.4 Die integrative Analyse von RNA-Seq., ATAC-Seq. und EPIC BC Daten

Die integrative Analyse von RNA-Seq., ATAC-Seq. und EPIC BC Daten, sollte eine Erklärung liefern, auf welche Art und Weise die für die zelluläre Apoptose, Viabilität, Proliferation und Morphologie zuständigen Gene von klassischen Zigaretten sowie modernen E-Zigaretten beeinflusst werden. Dabei wurden die drei wichtigsten Ebenen der Genregulation abgedeckt: Transkription, Chromatinstatus und DNA-Methylierung.

In diesem Abschnitt wird noch einmal in Kürze auf die funktionellen und morphologischen Befunde eingegangen. Die damit zusammenhängenden Veränderungen der Genregulation fanden ausschließlich auf Ebenen des Transkriptom sowie Chromatinstatus statt. Es wurden keine Veränderungen des DNA-Methylierungsmusters identifiziert, die mit den Beobachtungen zusammenhängen könnten.

4.3.2.4.1 Induktion der Apoptose nach der Behandlung mit nikotinhaltigen CSE und eCSE

Die RNA-Sequenzierdaten identifizierten bei beiden Zelllinien fünf apoptoserelevanten Gene, die nach der Behandlung mit mittel konzentrierten, nikotinhaltigen Zigaretten- sowie E-Zigarettenextrakten wie folgt dysreguliert waren [253-261].

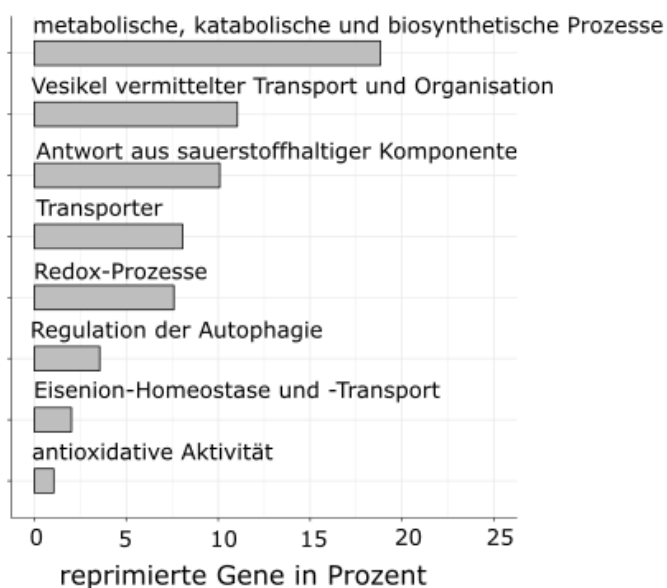
- *Brain Derived Neurotrophic Factor (BDNF)*: reprimiert.
- *Egl-9 Family Hypoxia Inducible Factor 3 (EGLN3)*: reprimiert, in AECII Zellen wurden große Bereiche des Gens in Histone verpackt
- *Heme Oxygenase 1 (HMOX1)*: überexprimiert.
- *Homeodomain Interacting Protein Kinase 2 (HIPK2)*: überexprimiert.
- *Phosphodiesterase 3A (PDE3A)*: reprimiert, in AECII Zellen wurden große Bereiche des Gens in Histone verpackt

4.3.2.4.2 Erhöhte Zellproliferation und Aneuploidie nach Behandlung der A549 Tumorzelllinie mit dem nikotinfreiem eCSE

Die Behandlung der A549 Zelllinie mit hoch konzentriertem, nikotinfreiem E-Zigarettenextrakt führte zur Fehlregulation von 1.607 Genen (Abb. 50).

Laut der GO Analyse spielten viele überexprimierte Gene eine wichtige Rolle innerhalb des TGF- β Signalwegs, der unter anderem Zellproliferation beeinflusst [262]. Des Weiteren waren viele Gene hochreguliert, die während der Aufrechterhaltung der Telomerlängen, DNA sowie RNA-Synthese entscheidend sind. Unter den reprimierten Genen waren viele für metabolische, katabolische und biosynthetische Prozesse (wie z.B. Synthese der Plasmamembran) zuständig. Des Weiteren wurden Gene reprimiert, die bei Redoxprozessen sowie Autophagie eine Rolle spielen.

A. A549 GO Analyse: 1041 reprimierte Gene



B. A549 GO Analyse: 566 überexprimierte Gene

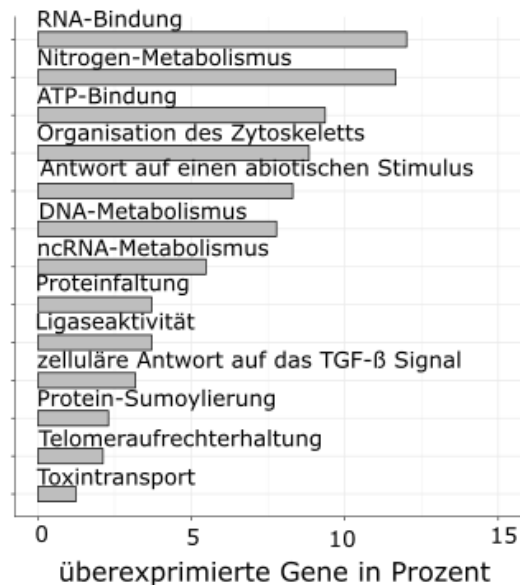


Abb. 50: Nach der Behandlung der A549 Zelllinie mit dem hoch konzentriertem, nikotinfreiem eCSE, kam es zu transkriptionellen **A.** Repression von 1041 Genen und **B.** Überexpression von 566 Genen. Die Ergebnisse der GO Analyse sind als Balkendiagramme dargestellt.

4.3.2.4.3 Weitere Ergebnisse der integrativen Datenanalyse

Nach der Behandlung der Lungenzellen mit dem klassischen Zigarettenextrakt wurden 87 differentiell exprimierte Gene identifiziert (Abb. 51A). Davon zeigten in primären AECII Zellen 14% sowie in A549 31% Veränderungen des Chromatinstatus (Abb. 52A). Die GO Analyse machte deutlich, dass die meisten aberrant exprimierten Gene in die zellulären Prozesse involviert waren, welche häufig im Zusammenhang mit Rauchen beschrieben werden, wie z.B. programmierter Zelltod, Morphogenese, Zellproliferation, Organisation der extrazellulären Matrix (inkl. Kollagensynthese), Eisen-Stoffwechsel, Fortbewegung, sensorische Organentwicklung und Antwort auf externe Stimuli.

Weil die E-Zigaretten vergleichsweise neu auf dem globalen Markt sind, existieren dazu nur wenige ernstzunehmende biologische sowie klinische Studien, die sich mit den gesundheitlichen Gefahren der E-Zigarette auseinandersetzen. In der hier durchgeführten multi-OMICS Studie wurden jedoch nach der Behandlung der Zellen mit nikotinhaltigem E-Zigarettenextrakt 241 aberrant exprimierte Gene detektiert (Abb. 51B). Davon zeigten 18% Veränderungen des Chromatinstatus in AECII Zellen sowie 38% in A549 Zelllinie (Abb. 52A). Ähnlich wie bei der klassischen Zigarette, waren auch in diesem Fall einige der detektierten Gene in Prozesse involviert, die beim programmierten Zelltod, Morphogenese sowie Entwicklungsprozessen eine Rolle spielen. Des Weiteren konnten auch Gene identifiziert werden, die für den Zellzyklus, Proteinbiogenese, Lokalisation und Bindung entscheidend waren.

Die Behandlung der Zellen mit nikotinfreiem E-Zigarettenextrakt identifizierte sogar 384 dysregulierte Gene (Abb. 51C). Davon zeigten bei beiden Zelllinien 42% der identifizierten Gene Veränderungen des Chromatinstatus (Abb. 52A). Bei den 31 differentiell exprimierten Genen handelte es sich um membrangebundene Transporter (Abb. 52B), inkl. SLC-, ABC- und ATP- Transporterfamilien.

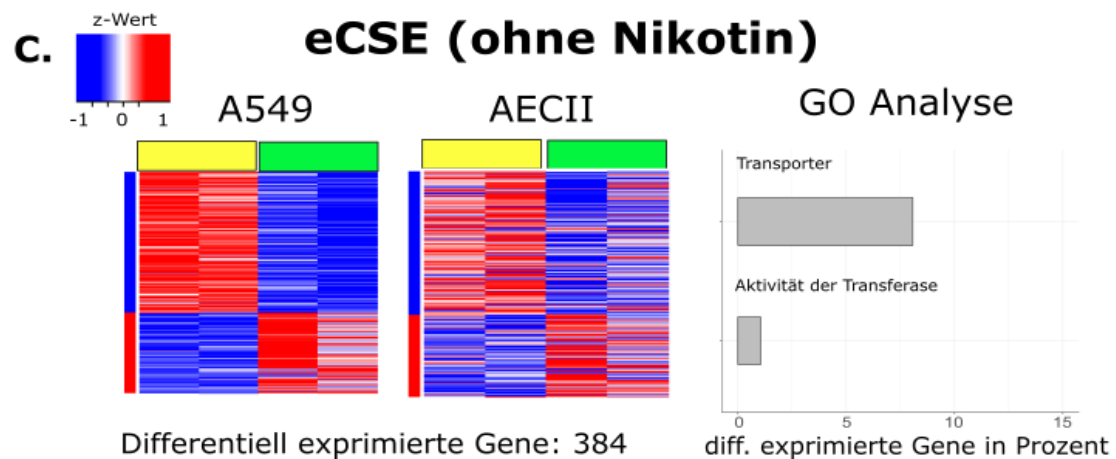
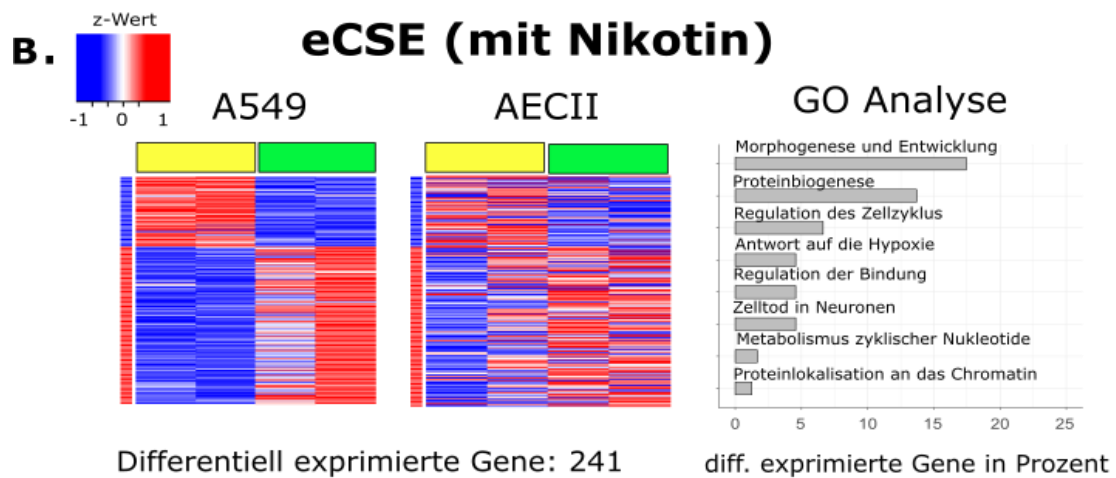
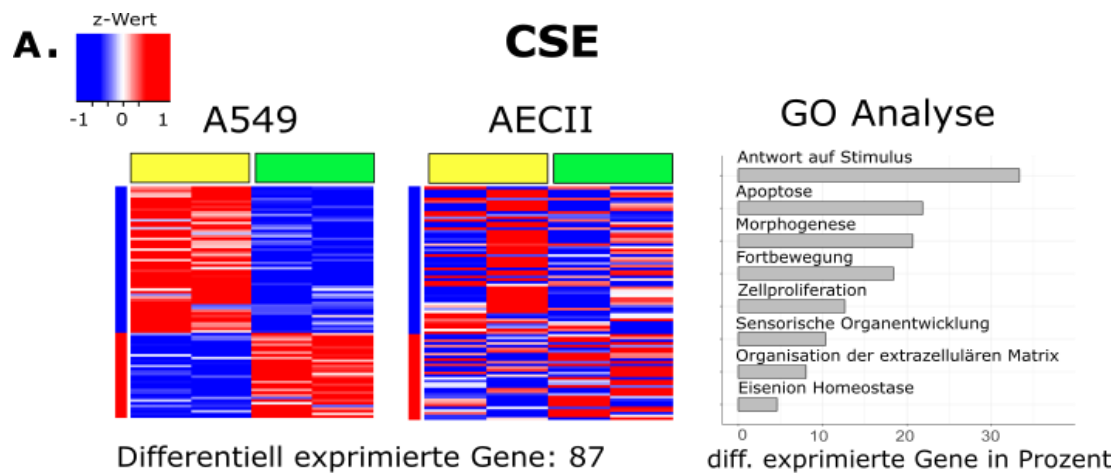


Abb. 51: Nach 48 Stunden langer Behandlung von A549 und AECII Zellen mit Zigaretten- (CSE) sowie E-Zigarettenextrakten (eCSE) wurde die RNA-Sequenzierung durchgeführt. Die differentiell exprimierten Gene wurden unter Verwendung von „edgeR“ identifiziert. Für die GO Analyse (Balkendiagramme) wurden nur die Gene ausgewählt, die bei beiden Zelllinien vorkamen und einen „log₂ fold-change“-Wert von mindesten 0,25 aufwiesen. **A.** Die Behandlung mit CSE identifizierte 87 differentiell exprimierte Gene. **B.** Nach der Behandlung mit nikotinhaltigem eCSE waren 241 Gene dysreguliert. **C.** Stimulation der Zellen mit nikotinfreiem eCSE führte zur differentiellen Expression von 384 Genen (Kontrolle: gelbe Markierung, Behandlung: grüne Markierung).

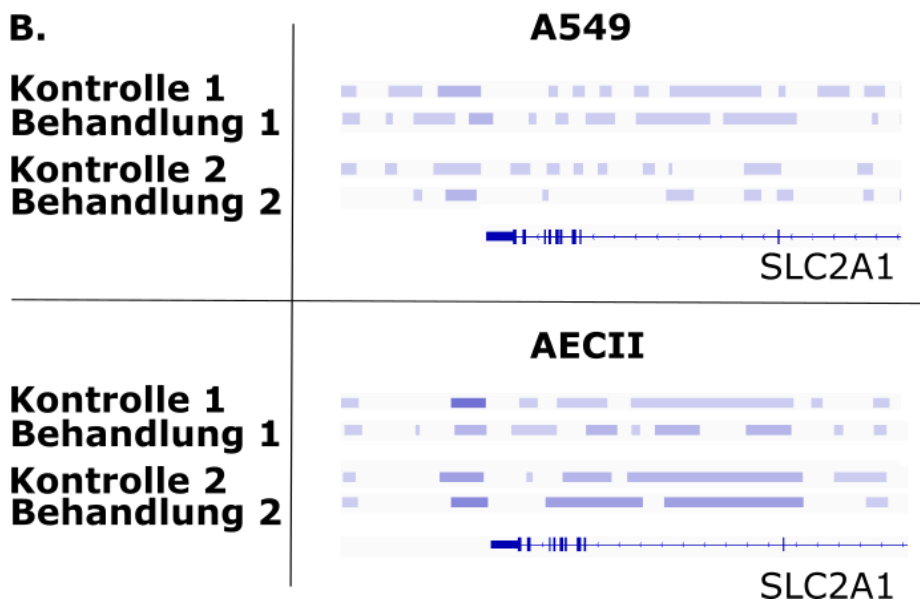
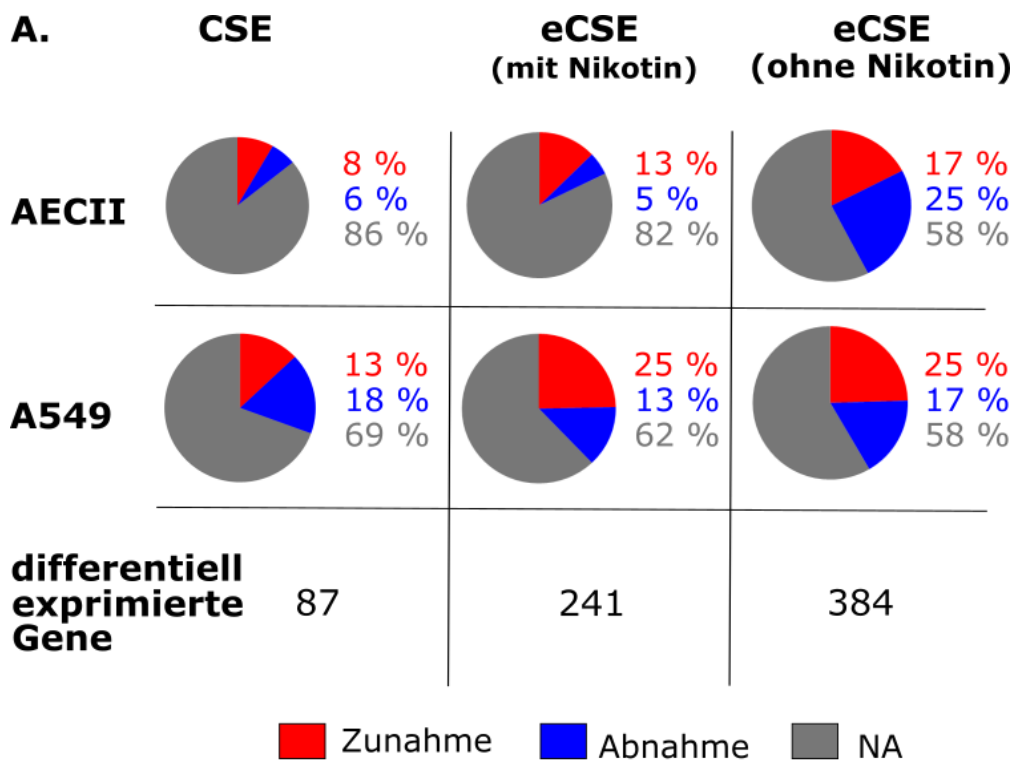


Abb. 52: Beide Zelllinien wurden 48 Stunden lang mit Zigaretten- (CSE) sowie E-Zigarettenextrakten (eCSE) behandelt. anschließend wurden gleichzeitig RNA-Seq. und ATAC-Seq. durchgeführt. **A.** Die Behandlung mit CSE identifizierte 87 differentiell exprimierte Gene. Davon zeigten in AECII Zellen 14% und in A549 Zellen 31% Veränderungen des Chromatinstatus. Nach der Behandlung mit nikotinhaltigem eCSE waren 241 Gene dysreguliert. 18% davon hatten in AECII Zellen und 38% in A549 Zellen Veränderungen des Chromatinstatus zur Folge. Stimulation der Zellen mit nikotinfreiem eCSE führte zu differentieller Expression von 384 Genen. In 42% davon detektierte das ATAC-Seq. Veränderungen des Chromatinstatus, inkl. **B.** membrangebundener Transporter wie z.B. SLC2A1.

5. Diskussion

5.1 Entwicklung und Validierung des Plasma *Panel*s

Lungenkrebs ist die zweithäufigste Krebsart bei Männern und Frauen [263]. Dass die Erkrankung meist im fortgeschrittenen Stadium diagnostiziert wird, spiegelt sich in einer schlechten Prognose wieder [99]. Bisher gibt es keine Möglichkeiten der Lungenkrebsfrüherkennung. Deshalb wurde im Rahmen der vorliegenden Doktorarbeit erstmalig ein diagnostisches *Panel* auf Basis genomweiter epigenetischer Alterationen im Blutplasma entwickelt und unter Verwendung von *Liquid Biopsies* einer kleinen Patientenkohorte (n=12) validiert. Während der Pilot-Studie differenzierte das Plasma *Panel* unter Verwendung von differentiell methylierten CpG *Loci* innerhalb der *SERPINB5*, *DOCK10*, *PCDHB2*, *HIF3A*, *FGD5*, *RCAN2*, *HOXD12*, *OCA2*, *SLC22A20*, *FADL-1*, *NRXN1*, *ACOXL*, *FAM53A*, *UBE3D* und *AUTS2* Gene mit 100%iger Richtigkeit maligne Lungentumore bereits ab Stadium I und identifizierte die häufigsten NSCLC-Subtypen. Anhand von 523 weiteren Positionen detektierte das *Panel* späte Stadien der Erkrankung (III und IV) und lieferte somit für das *Staging* relevante Informationen. Somit qualifizierte sich das entwickelte Plasma *Panel* für Lungenkrebsfrüherkennung. Die Ergebnisse der in dieser Arbeit durchgeführten Pilot-Studie wurden für die erfolgreiche Beantragung der deutschlandweiten *Clinical Trial* Studie „EmoLung“ des DZL verwendet.

Das entwickelte Plasma *Panel* nutzte *Liquid Biopsies*, um die zfDNA-Methylierungsmuster zu analysieren. Im Vergleich zu klassischen Gewebebiopsien haben *Liquid Biopsies* viele Vorteile, die sie für Krebsdiagnostik und -therapie interessant machen. Aufgrund der geringen Invasivität können *Liquid Biopsies* z.B. viel häufiger und regelmäßiger als klassische Gewebebiopsien entnommen und analysiert werden [264]. Operative Komplikationen, die durch die Entnahme von Gewebebiopsien entstehen können, werden so vermieden [265]. Über die Effizienz der verwendeten Therapie, auftretende Resistenzen oder Prognose kann rechtzeitig entschieden werden [264]. Ein potentieller Rückfall kann schneller erkannt werden [266, 267]. Des Weiteren soll nicht außer Acht gelassen werden, dass Tumore eine hohe Heterogenität besitzen [268]. Das bedeutet, dass eine Gewebebiopsie nur einen Ausschnitt aus dem Tumor darstellt und womöglich nicht die genetischen bzw. epigenetischen Eigenschaften des ganzen Tumors widerspiegelt [268]. *Liquid Biopsies* zeigen ein Gesamtbild des Tumors, sodass therapeutisch und prognostisch gesehen bessere Entscheidungen getroffen werden können [269].

Was die Sensitivität angeht, so haben *Liquid Biopsies* auch gegenüber den bildgebenden, diagnostischen Verfahren einen entscheidenden Vorteil [270]. Mit der Methode können

Tumore schon frühzeitig ab einer Größe von 50 Millionen Zellen identifiziert und analysiert werden [271].

Bei der Entwicklung des auf *Liquid Biopsies* basierten Plasma *Panels* wurden viele Aspekte berücksichtigt und Probleme gelöst. Die erste Herausforderung war es, ein optimales Analyt zu finden. Das nicht invasive Prinzip der *Liquid Biopsies* beruht darauf, dass ein wachsender Tumor permanent DNA, RNA, Tumorzellen sowie Exosomen in den Blutstrom freisetzt [272-277]. Diese enthalten alle therapeutisch und diagnostisch relevanten Informationen über den Tumor und können anschließend z.B. aus dem Vollblut bzw. Plasma extrahiert und analysiert werden [278, 279]. Im Vergleich zu anderen Analyten hat zfDNA mehrere Vorteile. So ist sie zum einen deutlich stabiler und einfacher zu handhaben als RNA [280]. Zum anderen kann sie einfacher extrahiert sowie besser und genauer quantifiziert werden als zirkulierende Tumorzellen oder Exosomen [269]. Deswegen wurde die zfDNA als Analyt für die hier beschriebenen Experimente ausgewählt.

Da die zfDNA während apoptotischer sowie nekrotischer Prozesse von Tumoren oder auch z.B. benignen Gewebe freigesetzt wird [272], ist sie stark fragmentiert. Im Durchschnitt sind die einzel- oder doppelsträngigen zfDNA-Fragmente ca. 167bp lang. Die zirkulierende DNA bestimmter Tumorentitäten kann deutlich kürzer sein und bei ca. 134-144bp liegen [121]. Die Menge der vom Tumor stammenden, zirkulierenden DNA im Blut von Krebspatienten kann stark variieren und je nach Stadium, Größe sowie Entität des Tumors 0,05% bis 93% der zfDNA-Gesamtmenge ausmachen [273, 281].

Mittlerweile ist eine Vielzahl an zfDNA-Extraktionsmethoden verfügbar [282, 283]. Generell kann die zfDNA aus einer Vielzahl von Körperflüssigkeiten extrahiert werden. So wurde z.B. bereits eine erfolgreiche Extraktion aus Blutplasma und -serum, Pleuraerguss oder Urin in der Literatur beschrieben [284]. Für die Entwicklung des Plasma *Panels* wurde Blutplasma für die zfDNA-Extraktion ausgewählt. Da Pleuraerguss erst im Laufe der Erkrankung auftritt, ist dieses Material für die Früherkennung ungeeignet [285]. Die zfDNA-Extraktion aus dem Plasma oder Serum ist deutlich schneller und kostengünstiger als aus dem Urin, was diese Materialien für *Screening* interessanter macht [172]. Schließlich war die zfDNA-Stabilität ausschlaggebend. Denn laut Literatur ist die zfDNA im Plasma länger stabil als im Serum [286].

Nachdem nun feststand, dass zfDNA aus dem Plasma ein geeigneter Analyt sein könnte, war die nächste Herausforderung zu überlegen, welche Methode für die Detektion von Biomarkern am geeignetsten wäre. In den letzten Jahren wurden viele Protokolle optimiert und weiterentwickelt, um die zfDNA-Mutationen sowie -Epimutationen analysieren zu können

[269]. Für die genomweiten Biomarker-Analysen im Plasma wurde eine NGS basierte Methode, *whole genome bisulfite sequencing*, gewählt.

Das Lungenkarzinom weist, genauso wie die meisten anderen Tumore, eine hohe genomische Heterogenität auf [287]. So können z.B. Mutationen innerhalb von *KRAS*, *EGFR*, *BRAF*, *MEK1*, *MET*, *HER2*, *ALK*, *ROS1*, *RET*, *FGFR1*, *DDR2*, *PTEN*, *LKB1*, *RB1*, *CDKN2A* oder *TP53* Genen die Entstehung eines primären Lungenkarzinoms induzieren [100]. Zusätzlich akkumulieren im Laufe der Tumor-Evolution die sogenannten *Passenger*-Mutationen, die zu verschiedenen Subklonen führen können [101]. Diese Tatsache macht die Entwicklung eines zuverlässigen, nur auf molekulargenetischen Mutationsanalysen basierten Früherkennungstests sehr schwierig [288], was an vielen Beispielen in der Literatur sichtbar wird.

So haben z.B. Uchida *et al.* versucht ein Lungenkarzinom-*Screening* basierend auf typischen Mutationen des EGFR-Gens durchzuführen. Die durchschnittliche Sensitivität des Tests betrug dabei 54,4% und fiel bei frühen Stadien IA-IIIa auf 22,2% [289]. Couraud *et al.* entwickelten einen NGS-basierten Test, bei dem die bekanntesten Mutationen innerhalb der *EGFR*, *BRAF*, *KRAS*, *HER2* und *PIK3CA* Gene im Plasma analysiert wurden. Die Sensitivität des Tests betrug 58%. Auch hier stellte die Erkennung von Tumoren in frühen Stadien ein Problem dar [290]. Newmann *et al.* entwickelten 2014 das CAPP-Seq. Hierbei handelte es sich um ein optimiertes NGS-Protokoll mit einer dazugehörigen bioinformatischen Auswertepipeline. Beim CAPP-Seq. wurden die bekanntesten NSCLC-Mutationen im Plasma sequenziert und analysiert. Dabei wurden 100% der Lungenkrebspatienten Stadien II bis IV identifiziert. Die Identifikation von Tumoren im Stadium I stellte wieder ein Problem dar, die entsprechende Sensitivität betrug nur 50% [291].

Diese Beispiele zeigen deutlich die Problematik bei der Entwicklung eines zuverlässigen, nur auf genomischen Analysen basierten Lungenkarzinomfrüherkennungstests. Im Rahmen der vorliegenden Arbeit wurde die DNA-Methylierung als Biomarker für Krebsfrüherkennung verwendet. Diese epigenetische, kovalente Modifikation ist stabil, gewebespezifisch und wird bereits sehr früh in der Tumor-Evolution verändert. [8, 38, 41, 292]. Und genau darin liegt der Vorteil gegenüber den oben beschriebenen, auf genomischen Analysen basierenden Früherkennungsverfahren.

Wie bereits im Abschnitt 1.2.2 beschrieben konnte das diagnostische Potential der DNA-Methylierung anhand der *SOX17* [126], *BRMS* [129] sowie *CST6*-Studien [127] gezeigt werden. Des Weiteren konnten viele weitere Studien aberrante DNA-Methylierungsmuster innerhalb der *SHOX2*, *RASSF1A*, *RARB2*, *LINE-1*, *P16*, *MGMT*, *DAPK*, *APC* und *DLEC1*

Promotorbereiche detektieren [288]. Schmidt *et al.* zeigten, dass der methylierte *SHOX2*-Promotor mit einer besseren Prognose der NSCLC- sowie SCLC-Patienten assoziiert [293].

Ein weiteres Beispiel für das diagnostische Potential der DNA-Methylierung stellt die „Epi proLung“-Studie dar, bei der das zfDNA-Methylierungsprofil innerhalb der *SHOX2* und *PTGER4* Gene analysiert wurde. Bei einer Spezifität von 90%, betrug die Sensitivität 67% [131]. Wie man auf den ersten Blick erkennt, ist das Ergebnis deutlich besser als bei den oben erwähnten genomischen, mutationsbasierten Ansätzen. Nichtsdestotrotz, reicht die Sensitivität von 67% noch nicht für ein zuverlässiges Krebs-Screening aus. Hier lag das Problem vermutlich daran, dass die verwendeten *SHOX2* und *PTGER4* Biomarker aus Analysen primärer Tumorgewebe stammten [294, 295]. Die vorliegende Doktorarbeit zeigt jedoch deutlich (siehe Abschnitt 4.1.3), dass die DNA-Methylierungsmuster zwischen der zfDNA aus dem Plasma und der gDNA aus einem primären Tumor nur bedingt korrelieren. Denn in der Gesamtmenge der zfDNA ist nicht nur aus der Lunge bzw. Tumor stammende DNA sondern auch DNA aus weiteren Geweben und Organen enthalten [272].

Das bedeutet, dass die im primären Tumorgewebe stark aberrant methylierte DNA-Regionen im Plasma gar keine differentielle Methylierung aufweisen könnten [272]. Deswegen reicht es nicht aus, für die Entwicklung eines nicht invasiven, zfDNA-basierten Früherkennungstests aus den primären Tumoren bekannte Biomarker zu nutzen. Vielmehr ist es notwendig, nach neuen zfDNA-spezifischen, starken und eindeutigen Methylierungssignaturen im Plasma der betroffenen Patienten zu suchen. Dies wurde in der vorliegenden Doktorarbeit durchgeführt. Die aus dem Plasma der NSCLC-Patienten extrahierte zfDNA wurde dem *whole genome bisulfite sequencing* unterzogen. Die identifizierten aberrant methylierten *SERPIN5*, *DOCK10*, *PCDHB2*, *HIF3A*, *FGD5*, *RCAN2*, *HOXD12*, *OCA2*, *SLC22A20*, *FADL-1*, *NRXN1*, *ACOXL*, *FAM53A*, *UBE3D* und *AUTS2* Gene wurden bisher noch nie speziell im Zusammenhang mit Lungenkarzinomen oder bestimmten NSCLC-Entitäten beschrieben. Deren Rolle bei der Tumor-Evolution und Prognose ist jedoch bei vielen anderen Krebsarten bekannt.

Bei *SERPIN5* handelt es sich z.B. um ein bekanntes Onkogen [296]. *HOX*-Gene werden in vielen Krebsarten aberrant exprimiert [297, 298]. Fehlregulation von *RCAN2* führt zu Proliferation der Tumorzellen [299]. Veränderte Expression von *DOCK10* hatte bei einigen Studien Migration von Melanomzellen zur Folge [300]. Auch einige *OCA2*-Mutationen sind mit erhöhtem Melanom-Risiko assoziiert [301]. Desweiteren sind *HIF3A* und *FGD5* wichtige Angiogenese-Regulatoren und spielen somit eine entscheidende Rolle während der Tumor-Evolution [302, 303]. Die DNA-Methylierung einiger *PCDHB2*-CpG *Loci* ist mit einer schlechten Prognose der Neuroblastom-Patienten assoziiert [304]. Veränderter Metabolismus ist z.B. ein Kennzeichen maligner Tumore, dabei können der *FADL-1*

Fettsäuren-Transporter sowie einige *SLC*-Transporter eine wichtige Rolle spielen [305-307]. *UBE3D* ist eine Ubiquitin-Protein-Ligase. Mehrere Studien haben gezeigt, dass einige Ubiquitin-Protein-Ligasen während der Tumor-Evolution eine wichtige Rolle spielen können [308, 309]. Bei *AUTS2* und *NRXN1* handelt es sich um neuronale Gene [310, 311]. *AUTS2*-Überexpression wurde in Lebermetastasen nachgewiesen [311]. *NRXN1* könnte für die Nikotinsucht verantwortlich sein [310]. Erhöhte Expression von *ACOXL* wurde bereits bei Prostatakarzinomen beschrieben [312]. Einige Studien beschreiben *FAM53A* als einen prognostischen und therapeutischen Mammakarzinom-Marker [313].

Um die stärksten und stabilsten Biomarker zu detektieren, wurden zunächst nur Patienten mit fortgeschrittener Erkrankung analysiert (Lungenkarzinom im Stadium III oder IV). Da die DNA-Methylierung sich bereits sehr früh in der Tumor-Evolution ändert [292], wurde angenommen, dass ein Teil der in das Plasma *Panel* aufgenommenen differentiell methylierten DNA-Regionen sich anhand des Probenmaterials der Lungenkarzinompatienten früher Stadien (I oder II) validieren lässt. Diese Annahme konnte im weiteren Verlauf bestätigt werden, was den Nutzen des Plasma *Panel*s für die Lungenkrebsfrüherkennung deutlich macht. Aus DZL-internen, organisatorischen Gründen stand zum Zeitpunkt der *Panel*-Entwicklung das Plasma einer komplett gesunden Patientenkohorte nicht zur Verfügung. Daher diente die zfDNA von Asthmapatienten als Kontrollgruppe. Die Patienten hatten zum Zeitpunkt der Blutentnahme keine malignen Erkrankungen. Sicherlich bestand das Risiko, dass ein Teil der detektierten Biomarker asthmaspezifischen Signaturen enthielt [314]. Daher wurde das entwickelte *Panel* im weiteren Verlauf der Studie an einer weiteren kleinen, unabhängigen asthma- und tumorfreen Kontrollkohorte validiert (n=3). Zukünftig wird das *Panel* im Rahmen einer deutschlandweiten *Clinical Trial* Studie „EmoLung“ des DZL validiert.

Bei den Regionen, die sich unter Verwendung der frühen Stadien nicht validieren ließen, könnte es sich z.B. um für Metastasen spezifische Signaturen handeln. Diese Regionen wurden daher für die Berechnung des *Staging*-Parameters genutzt. Bisher kann das in dieser Arbeit beschriebene *Staging*-Parameter die späten Lungenkarzinomstadien mit 80%iger Richtigkeit von frühen unterscheiden und soll in zukünftigen DZL-Projekten optimiert werden. Generell sollte der *Staging*-Parameter nur als Hinweis genutzt werden. Falls das entwickelte *Panel* ein Lungenkarzinom detektiert, wäre es weiterhin ratsam therapeutisch relevante Informationen z.B. bezüglich der Größe oder Lage des Tumors durch bildgebende Verfahren, wie z.B. wie MRT, CT oder PET CT zu generieren [315-317].

Ein weiteres Problem, das im Laufe der Biomarker-Identifikation gelöst wurde, stellte die Heterogenität der individuellen DNA-Methylierungsmuster dar, die je nach Lebensweise von Individuum zu Individuum stark variieren kann [318, 319]. Diese individuellen

Methylierungsunterschiede müssen herausgefiltert werden, um universelle Lungenkrebs-biomarker zu identifizieren.

Hierfür kann nach Gallego-Fabrega *et al.* die Heterogenität der Proben unter Verwendung einer der folgenden Ansätze minimiert werden. Der erste Ansatz besteht darin möglichst viele Einzelproben zu analysieren [320]. So berechneten z.B. Rakan *et al.*, dass innerhalb einer Methylomstudie mindestens 400 Kontrollen und 400 erkrankte Probanden analysiert werden sollten, um auch schwächere Effekte erfassen zu können [321]. In der Realität wäre dieser Ansatz viel zu teuer und daher schwer umzusetzen [320]. Die meisten publizierten Studien verwenden deutlich kleinere Kohorten. So nutzten z.B. Eberth *et al.* 40 Lymphomzelllinien und 50 Tumorgewebeproben, um *CD44* als einen potentiellen Biomarker für Lymphome zu identifizieren [322]. Bediaga *et al.* analysierten 28 Brusttumore sowie deren korrespondierende Kontrollen, um verschiedene Mammakarzinomentitäten anhand der DNA-Methylierungssignatur zu klassifizieren [323]. Lugthart *et al.* analysierten CD34 positive Zellen von 26 Patienten, die an akuter myeloischer Leukämie litten, sowie die Zellen acht gesunder Probanden und schlossen auf eine für die Erkrankung typische Methylomsignatur [324]. Der zweite von Gallego-Fabrega *et al.* beschriebene Ansatz ist viel kostengünstiger und besteht darin, mehrere Patientenproben vor der Methylierungsanalyse zu *poolen*. Dabei werden die allgemeinen, für die Erkrankung spezifischen Veränderungen stärker hervorgehoben, während die individuellen DNA-Methylierungssignaturen verloren gehen [320]. Als Beispiel analysierten die Autoren 20 einzelne sowie 4 *gepoolte* Proben bezüglich des genomweiten DNA-Methylierungsmusters. Die Korrelation beider Methoden betrug dabei 99% ($p < 10^{-16}$) [320]. Doch das *Poolen* kann auch einige Nachteile mit sich bringen. So kann im Nachhinein nicht nach individuellen Unterschieden innerhalb der Kohorte geschaut werden. Des Weiteren kann die Fehlerrate der falschpositiv detektierten Signaturen etwas höher als beim ersten, individuellen Ansatz ausfallen [325]. Aufgrund der beeindruckenden Ergebnisse der Gallego-Fabrega *et al.* Studie sowie der im Rahmen der Pilotstudie limitierten Materialverfügbarkeit, wurde der zweite, *Pool*-basierte Ansatz gewählt. Dabei wurde sehr auf präzise Vorgehensweise geachtet [326, 327], alle verwendeten zfDNA-Proben wurden äquimolar *gepoolt*.

Nach der Sequenzierung stellte die bioinformatische Auswertepipeline ein weiteres Problem dar. Die herkömmlichen, gDNA-WGBS-Libraries werden nach dem Prozessieren meist mit dem „Bismarck“ Algorithmus *aligned* [199, 328]. Die Ergebnisse des *Alignments* können dann anschließend von zahlreichen Auswertepipelines analysiert werden, wobei genomweite DNA-Methylierungssignaturen extrahiert werden [328]. Das in dieser Arbeit durchgeführte WGBS-Experiment der zirkulierenden DNA war das erste seiner Art. Dabei stellte sich heraus, dass die zfDNA-Libraries eine andere Komplexität sowie Fragmentverteilung als

herkömmliche gDNA-*Libraries* besaßen (siehe Abschnitt 3.1.2.5). Das könnte der Grund dafür sein, dass der am häufigsten verwendete „Bismarck“ Algorithmus eine unzufriedenstellende *Mapping* Effizienz von nur 70% lieferte. Aus diesem Grund wurden weitere Algorithmen ausgetestet. Die besten Ergebnisse mit einer *Mapping* Effizienz von mindestens 98% lieferte der „Segemehl“ Algorithmus (siehe Abschnitt 3.1.2.5). Das Format des „Segemehl“-*Output*-Files sah anders als das typische „Bismarck“-Format aus [200, 209]. Daher konnte im nächsten Schritt nicht die gängige, für „Bismarck“ angepasste *Peak Calling* Software verwendet werden. Vielmehr musste eine geeignete, mit „Segemehl“ kompatible Analysepipeline gefunden, etabliert und optimiert werden. Nach weiterer Literaturrecherche fiel die Entscheidung auf das „Bisulfite Analysis Toolkit“. Die modular aufgebaute Software kann auf zahlreichen Rechenclustern verwendet werden und durch weitere Software sowie eigene Skripte erweitert werden [201]. Vor allem letztere Eigenschaft war wichtig. Denn für die Identifikation von für Lungenkrebsfrüherkennung geeigneten, differentiell methylierten Regionen musste die Analysepipeline mit eigenen bioinformatischen Skripten ergänzt werden können. Die Skripte sind im Anhang zusammengestellt und wurden in Kooperation mit Dr. Helene Kretzmer entworfen.

Im nächsten Schritt wurden die identifizierten Biomarker in Zusammenarbeit mit „Roche“ für die Entwicklung eines für DZL maßgefertigten „SeqCap Epi Enrichment“ *Panels* verwendet. Aufgrund der repetitiven Sequenzen konnten ca. 100 der identifizierten, differentiell methylierten Regionen nicht für das *Panel*-Design verwendet werden. Das endgültige und für die Validierungsversuche verwendete *Panel* bestand daher aus 638 differentiell methylierten Regionen.

Während der darauffolgenden Validierung trat ein weiteres technisches Problem auf, das in zukünftigen umfangreicheren Studien behoben werden sollte. Nach der Sequenzierung wurde eine erhöhte *off-target* Rate detektiert. Dieses Problem ist auch für viele kommerziell verfügbare NGS-*Panels* typisch [329]. Dabei werden unter anderem unspezifische DNA-Fragmente von *Capture Probes* gebunden. Diese unspezifischen Fragmente enthalten keine für das *Panel* relevanten Positionen, werden aber trotzdem sequenziert [329]. Nach der Sequenzierung resultiert es in einer schlechteren *Coverage* der untersuchten, im *Panel* enthaltenen, differentiell methylierten Positionen. Die erreichte *Coverage* ist somit deutlich niedriger als die theoretisch berechnete [330]. Beim Minimieren der *off-target* Rate sollte allerdings die PCR-Duplikat-Problematik nicht außer Acht gelassen werden. Denn beim Reduzieren der *off-target* Rate sinkt auch die Komplexität der *Library*. Somit wäre es denkbar, dass die Menge an PCR-Duplikaten stark ansteigen könnte [331].

Zusammenfassend lässt sich sagen, dass es im Rahmen der vorliegenden Doktorarbeit erstmalig gelungen ist, ein NGS-*Panel* zu entwickeln, das auf genomweiten zfDNA-

Methylierungssignaturen aus dem Plasma basiert. Das Plasma *Panel* konnte unter Verwendung von *Liquid Biopsies* einer kleinen Patientenkohorte (n=12) erfolgreich validiert werden. Während der Pilot-Studie differenzierte das Plasma *Panel* mit 100%iger Richtigkeit maligne Lungentumore bereits ab Stadium I, identifizierte die häufigsten NSCLC-Subtypen und lieferte weitere Informationen bezüglich des *Staging*. Die Ergebnisse dieser Arbeit zeigten deutlich das erhebliche Potential der epigenetischen Alterationen in der Lungenkrebsfrüherkennung und stellten damit eine wesentliche Grundlage für die erfolgreiche Beantragung der deutschlandweiten *Clinical Trial* Studie „EmoLung“ des DZL dar, die voraussichtlich 2018 beginnt. Im Rahmen der klinischen Studie werden weitere Validierung und Optimierung des Plasma *Panels* anhand einer großen Patientenkohorte erfolgen. Ferner wird geprüft werden, ob es möglich sei, die Anzahl an CpG *Loci*, die für eine 100%ig genaue Aussage relevant sind, zu reduzieren sowie die Aussagekraft des *Staging*-Parameters zu optimieren.

5.2 multi-OMICS Studie zur Untersuchung von Risiken der E-Zigaretten

Es ist schon seit langer Zeit bekannt, dass Rauchen Lungenkrebs verursachen kann [137]. Deshalb wechseln heutzutage immer mehr Raucher zu E-Zigaretten, die als gesündere Alternative suggeriert werden [149-151]. In den letzten 10 Jahren ist die Beliebtheit der E-Zigaretten stark gestiegen. Mittlerweile ist ein großes Angebot an Modellen sowie dazu passenden *Liquids* erhältlich [153].

Kaum ein Produkt auf dem Markt ist so umstritten wie die E-Zigarette. Einige Forscher und Ärzte behaupten, dass die E-Zigarette viel harmloser als Tabak sei. Sie gehen davon aus, dass der E-Zigarettenkonsum zukünftige Lungenkrebsrate erheblich senken wird [149-151]. Andere Studien identifizierten dagegen hohe Mengen an giftigen Substanzen, wie z.B. Acrolein und Formaldehyd im E-Zigaretten-Dampf. Sie raten daher vom E-Zigarettenkonsum ab und fordern striktere Kontrollen der *Liquid*-Zutaten [154-157]. Das spiegelt sich auch in den weltweiten Gesetzen wieder. Während in den USA und den meisten EU-Ländern, inkl. Deutschland, der Verkauf jeglicher E-Zigaretten ab 18 legal ist [332-334], dürfen z.B. in Japan nur nikotinfreie E-Zigaretten verkauft werden [335]. In Bahrain, Singapur, Thailand und Arabischen Emiraten ist der Verkauf von E-Zigaretten sogar komplett verboten [159]. Auch die WHO warnt vom E-Zigarettenkonsum [158].

Da die E-Zigarette ein vergleichsweise neues Produkt ist, fehlen Langzeitstudienresultate, mit denen potentielle Risiken und Nebenwirkungen zuverlässig ermittelt werden können. Des Weiteren sind kaum *in vitro* bzw. *in vivo* Studien vorhanden, welche die Wirkung der E-

Zigarette auf Genom bzw. Epigenom untersuchen [155, 157, 158, 160]. Um also die ersten Hinweise zumindest über die Kurzzeitwirkung der E-Zigaretten auf die Genregulation der Lunge zu erhalten, wurde im Rahmen der vorliegenden Doktorarbeit eine multi-OMICS Studie entwickelt und etabliert.

Die multi-OMICS Studie zeigte erstmalig, dass bereits geringe Mengen nikotinfreier E-Zigarettenextrakts Aneuploidie und Zellproliferation induzierten, die Apoptose wurde dabei inhibiert. Des Weiteren zeigten molekulare, epigenetische Untersuchungen zum ersten Mal eine Fehlregulation von SLC- sowie ABC- Transportern. Zusätzlich induzierten geringe Mengen des nikotinhaltigen E-Zigarettenextrakts bereits nach drei Stunden Apoptose der Lungenzellen. Genomweite, molekulare Analysen detektierten erstmalig im weiteren Verlauf der Studie ähnlich wie bei den klassischen Zigaretten Fehlregulation wichtiger zellulärer Prozesse, wie z.B. Morphogenese, Entwicklung, Zellzyklus, Proteinbiogenese, -lokalisierung und -bindung.

Während der Entwicklung der multi-OMICS Studie wurde im ersten Schritt ein passendes biologisches Modell gewählt. Laut Literatur wurden in den vergangenen Jahren häufig Mäuse mit Zigarettenrauch über mehrere Tage behandelt, um die Veränderungen der Genregulation detektieren zu können [336-338]. Da sich die humane Genregulation jedoch stark von der murinen unterscheidet, sind die Ergebnisse solcher Studien mit Vorsicht zu betrachten [339]. Um die humane, lungenspezifische Genregulation zu untersuchen, verwenden viele Wissenschaftler primäre AECII-Zellen [340-342]. Da die AECII-Zellen direkt aus menschlicher Lunge isoliert werden [228], besitzen sie alle genetischen und epigenetischen Charakteristika einer benignen Lungenzelle.

Neben den erwähnten Vorteilen, gibt es eine Reihe von Nachteilen, die in Verbindung mit primären Zellen einhergehen. So ist z.B. das Extraktionsverfahren aus der Lunge lang, aufwendig und kann aufgrund der Materialverfügbarkeit nur von bestimmten Laboratorien durchgeführt werden (siehe Abschnitt 3.3.1.2) [228, 343]. Nach der Isolierung können die AECII-Zellen nur wenige Tage in Kultur gehalten werden, da sie nach längerer Kultivierung dedifferenzieren [344, 345]. Des Weiteren können die isolierten primären Zellen je nach Patient in ihren Eigenschaften sehr heterogen sein [346]. So kommt es vor, dass die Zellen des einen Spenders stark auf eine bestimmte Behandlung reagieren, während die Zellen des anderen kaum oder gar keine Reaktion zeigen [347]. Daher kann auch das Genexpressionsprofil von Spendern nach der Behandlung sehr unterschiedlich ausfallen [348]. Das liegt vor allem an der Flexibilität des Epigenoms, das von verschiedenen Umweltfaktoren sowie persönlicher Lebensweise permanent beeinflusst wird [318]. Die Heterogenität kann für die Aussagekraft und Reproduzierbarkeit molekulargenetischer Experimente im Nachhinein ein nicht zu vernachlässigendes Problem darstellen [349]. Aus

diesem Grund wurde zusätzlich zu den primären AECII Zellen im Rahmen der vorliegenden Arbeit die A549-Zelllinie untersucht. A549 Zellen sind gut charakterisiert und können über längere Zeiträume in Kultur gehalten werden. Da es sich hierbei um eine artifizielle Zelllinie handelt, ist die Heterogenität innerhalb der Zellpopulation gering [343].

Als nächstes wurde das eigentliche Studiendesign konzipiert. Laut Civelek *et al.* vernetzt eine gut designte multi-OMICS Studie Ursache, Wirkung und funktionelle Konsequenz miteinander [350]. In diesem Fall war die Ursache der E-Zigarettenrauch. Für die Identifikation funktioneller Konsequenzen erfolgte die Etablierung verschiedener Mikroskopie-, Apoptose- sowie Zellviabilitätsprotokolle (siehe Abschnitt 3.3). Um die Wirkung des *Liquids* auf die Genregulation zu erfassen, wurden genomweite Untersuchungen des Chromatinstatus, Methyloms und Transkriptoms durchgeführt. Bei jedem Experiment wurden biologische Duplikate analysiert. Insgesamt wurden 24 Proben jeweils einer genomweiten EPIC BC, ATAC-Seq.- und RNA-Seq.-Analyse unterzogen. Davor wurden diese Methoden sowie die dazugehörigen bioinformatischen Auswertepipelines im Rahmen der vorliegenden Doktorarbeit etabliert. Um die *Batch* Effekte zu minimieren, wurden die hergestellten *Libraries* mit *Indices* markiert, *gepoolt* und auf gleichen *Lanes* sequenziert [351, 352].

Aufgrund der bereits erwähnten individuellen, epigenetischen Unterschiede, wurden pro Replikat DNA-, RNA-, sowie Chromatin-Extraktionen aus Lungenzellen des gleichen Spenders durchgeführt [318]. Die Spender Sp.7 und Sp.9 bildeten eine Ausnahme, da aus deren Lungengewebe im Vergleich zu anderen Probanden weniger AECII-Zellen extrahiert werden konnten.

Da primäre Zellen nur in einer geringen Menge gewonnen werden können [228], wurden im Rahmen dieser Arbeit nur solche molekulargenetischen Methoden etabliert, die mit einer geringen Zellmenge auskommen. So reichten z.B. bei den verwendeten RNA-Seq. Protokollen *Input*-Mengen von 100ng aus, um komplexe *Libraries* herzustellen. Die Methylom-Analysen erfolgten unter Verwendung der EPIC BC Arrays, hierfür waren 500ng DNA ausreichend [246]. Für die Analyse des Chromatin-Status wurde im Rahmen dieser Arbeit die ATAC-Sequenzierung etabliert. Die Methode benötigte für *Library* Herstellung nur 50.000 Zellen [241].

Sicherlich hätten die *Input*-Mengen unter Verwendung der Einzelzell-Sequenzierung (engl. *single cell sequencing*) noch weiter reduziert werden können [353]. Jedoch hat diese immer beliebter werdende Methode einen deutlichen Nachteil [353], da Zellen innerhalb einer Population starke individuelle Unterschiede aufweisen [354]. So können sie z.B. verschiedene molekulargenetische Mechanismen verwenden, um gleiche zelluläre Prozesse zu aktivieren [355]. Ohne diese Heterogenität wäre ein Organismus deutlich anfälliger für

äußere Einwirkungen [356, 357]. Das bedeutet, dass die Ergebnisse der Einzelzell-Analysen für die individuelle Zelle zwar zutreffen mögen, für die gesamte Zellpopulation könnten sie jedoch nicht repräsentativ sein. Daher ist es ratsam, eine Zellmenge zu analysieren, um die Effekte zu identifizieren, die für die Zellpopulation im Durchschnitt zutreffen und somit reproduzierbar bleiben [353].

Für die bioinformatische Auswertung der Array- sowie NGS-Ergebnisse wurden im nächsten Schritt die dazugehörigen bioinformatischen Pipelines etabliert. Die Auswertung der Daten stellte aufgrund der Heterogenität der AECII-Zellen und einem $n=2$ eine große Herausforderung dar [358-360]. Durch die Verwendung höherer Anzahl an biologischen Replikaten hätte die Heterogenität der Proben stärker reduziert werden können [361]. Aufgrund der hohen Kosten der genomweiten Analysen war es jedoch nicht möglich. Um also das Problem zu lösen, wurden die RNA-Seq. Daten der AECII- und A549-Zellen zunächst getrennt voneinander unter Verwendung von „DoDas“ und „edgeR“ wie im Abschnitt 3.3.6.3.2 beschrieben ausgewertet. Durch die hohe Sensitivität der Software konnten auch sehr schwach exprimierte Transkripte detektiert werden [237]. Nachfolgend wurden unter Verwendung von „R“ nur solche Transkripte ausgewählt, die in allen Replikaten beider Lungenzelllinien differentiell exprimiert waren ($p < 0,05$). Der „log₂ fold-change“-Wert betrug dabei mindestens 0,25 [362, 363]. Im nächsten Schritt wurden dann die ATAC-Seq. Daten verwendet, um gezielt nach Veränderungen des Chromatinstatus innerhalb dieser differentiell exprimierter Genen zu suchen.

Nach der Behandlung der Lungenzellen mit Zigaretten- sowie E-Zigarettenextrakten konnten keine signifikanten Veränderungen des DNA-Methylierungsmusters nachgewiesen werden. Obwohl die DNA-Methylierung die stabilste epigenetische Modifikation ist [364], kann sie je nach Stimulus nach Stunden oder Jahren modifiziert werden. Während die Gene des zirkadianen Rhythmus im Laufe des Tages methyliert und anschließend demethyliert werden [365], braucht die Lunge eines ehemaligen Rauchers laut Ambatipudi *et al.* bis zu 22 Jahre, um ihre ursprüngliche Methylierungssignatur wiederzuerlangen [366]. Es war nicht möglich, die AECII-Zellen länger als 48h zu stimulieren, da sie bei längerer Kultivierung dedifferenzierten.

Insgesamt wurden nach der Behandlung mit CSE 87, eCSE (mit Nikotin) 241 sowie eCSE (ohne Nikotin) 384 dysregulierte Gene identifiziert. Die Anzahl an fehlregulierten Genen bei der nikotinfreien E-Zigarette war deutlich höher als bei den nikotinhaltigen Produkten. Dieses Ergebnis könnte wie folgt erklärt werden: bei den Spendern, aus deren Lunge die AECII-Zellen extrahiert wurden, handelte es sich um Patienten, die aufgrund von COPD-Verdacht untersucht wurden. Alle Spender haben entweder zum Zeitpunkt der Biopsie oder in der

Vergangenheit klassische Zigaretten konsumiert, sodass die verwendeten Lungenzellen bereits chronisch dem Zigarettenrauch und somit Nikotin ausgesetzt wurden.

Spira *et al.* konnten zeigen, dass die aus Rauchern isolierten Lungenzellen andere Genexpressionsmuster als die Lungenzellen der Nicht-Raucher aufweisen [367]. Man kann daher davon ausgehen, dass auch im Rahmen dieser Arbeit verwendeten AECII-Zellen bereits Genexpressionsmuster zeigten, die durch Zigarettenrauch beeinflusst wurden. Also wurden während der vorliegenden multi-OMICS Studie nicht alle Veränderungen des Transkriptoms und Epigenoms, sondern nur besonders starke erfasst, die im Zusammenhang mit CSE und Nikotin auftraten. Die wichtigsten Bestandteile des *Liquids* sind Glycerin sowie Propylenglykol und kommen in der klassischen Zigarette nicht vor [157]. Im Gegensatz zum nikotinhaltigen enthält der nikotinfreie E-Zigarettenextrakt weder das Nikotin noch das Tabakaroma. [368, 369]. Das bedeutet, dass die verwendeten Lungenzellen vor den Experimenten den Zutaten des nikotinfreien *Liquids* nicht ausgesetzt wurden, daher konnten vermutlich sowohl starke als auch schwächere Veränderungen des Transkriptoms und Epigenoms erfasst werden, die im Zusammenhang mit nikotinfreiem eCSE auftreten. Das wiederum resultierte im Vergleich zu nikotinhaltigen Produkten in Detektion einer höheren Anzahl differentiell exprimierter Transkripte.

Die Tatsache, dass die verwendeten AECII-Zellen eine verminderte Empfindlichkeit gegenüber Bestandteilen der klassischen Zigaretten zeigten, stellte für die designte multi-OMICS Studie ein geringes Problem dar. Denn wie bereits erwähnt, wurde hier die Wirkung der E-Zigaretten untersucht. Die molekulargenetische Wirkung der klassischen Zigaretten auf den menschlichen Organismus ist in den vergangenen Jahrzehnten bereits sehr detailliert untersucht worden [100, 107, 370], sodass die CSE-Stimulation nur als Positivkontrolle für die Plausibilität des Studiendesigns diente. Wie im Abschnitt 4.3.2.4.3 beschrieben, detektierte die integrative Analyse der RNA-Seq.- sowie ATAC-Seq.-Daten nach der CSE-Stimulation überwiegend bekannte und in der Literatur mehrfach beschriebene Effekte des Rauchens. Eins der bekanntesten Beispiele ist sicherlich die frühzeitige Hautalterung [371, 372] sowie schlechte Wundheilung [373, 374] bei Rauchern. Hierfür ist die verminderte Synthese von Kollagenen Typ I und III verantwortlich [375, 376]. Die Fehlregulation vieler mit extrazellulärer Matrix verbundener Signalwege konnte auch bereits nach 48 Stunden im Rahmen dieser Arbeit detektiert werden. Auch der Effekt des Rauchens auf die sensorische Prozesse (wie z.B. den Geschmack) [377], die negative Auswirkung auf embryonale Entwicklung [378, 379], Eisen-Stoffwechsels [380, 381] sowie Stütz- und Bewegungsapparat [382, 383] eines Menschen werden häufig in der Fachliteratur betont.

Überraschend war jedoch die Tatsache, dass die Behandlung der Lungenzellen mit nikotinhaltigem CSE sowie eCSE in funktionellen Experimenten zu Apoptose führte. Die molekulargenetischen Untersuchungen identifizierten Fehlregulationen von fünf apoptoserelevanten Genen: *Brain Derived Neurotrophic Factor (BDNF)*, *Homeodomain Interacting Protein Kinase 2 (HIPK2)*, *Heme Oxygenase 1 (HMOX1)*, *Egl-9 Family Hypoxia Inducible Factor 3 (EGLN3)*, *Phosphodiesterase 3A (PDE3A)*.

BDNF, *EGLN3* und *PDE3A* waren nach der Behandlung mit nikotinhaltigen Extrakten reprimiert. *BDNF* ist als ein neuronaler Wachstumsfaktor bekannt geworden, kann aber zusätzliche Apoptose inhibieren [253, 259]. *EGLN3* ist ein wichtiger Sensor des Sauerstoffmangels, auch Hypoxie genannt. Als Folge der Hypoxie kommt es zu einer verminderten Aktivität des Enzyms. Daraufhin wird die Expression der sogenannten Hypoxie-induzierten Faktoren und somit die Apoptose induziert [258, 261]. *PDE3A* ist für die cAMP-Hydrolyse verantwortlich. Einige Studien mit Kardiomyozyten haben gezeigt, dass *PDE3A*-Depletion Apoptose initiieren kann [254, 255].

HIPK2 und *HMOX1* waren nach der Behandlung überexprimiert. *HIPK2* ist eine Serin/Threonin-Kinase, die unter anderem mit P53 interagieren und somit Apoptose aktivieren kann [256, 257]. *HMOX1* wirkt normalerweise antiapoptotisch. Falls jedoch bestimmte karzinogene Substanzen, wie z.B. Benzene, Toluene und O-Xylene, auf die Zelle einwirken, kommt es zum Anstieg an *HMOX1* und reaktiven Sauerstoffspezies. Kurze Zeit später wird Apoptose induziert, der genaue Mechanismus ist bisher nicht geklärt [260].

Somit lässt sich sagen, dass sowohl funktionelle als auch genomweite Untersuchungen im Rahmen der multi-OMICS Studie auf eine Apoptose-induzierende Wirkung der nikotinhaltigen Zigaretten- bzw. E-Zigarettenextrakten hindeuten. Im Fall von CSE trifft das zu. Denn mehrere Studien bestätigen, dass Zigarettenextrakt in einigen Zelltypen zu einer akuten Apoptose-Induktion führen kann [384-386]. Nach Aoshiba *et al.* ist die CSE-induzierte Apoptose mit erhöhtem oxidativem Stress, mitochondrialer Fehlregulation sowie Freisetzung von Cytochrom C assoziiert [387]. Der Zigarettenrauch ist ein komplexes Gemisch, das aus mehr als 5.000 Komponenten besteht. Davon werden 98 als gesundheitsgefährlich eingestuft [388]. Daher ist es schwierig zu sagen, welche chemische Verbindung oder Kombination davon genau für Apoptose verantwortlich ist.

Nikotin an sich wirkt jedoch laut zahlreichen Studien Apoptose-hemmend [389-392]. Die Substanz ist in der Lage durch Opiode, Etoposide, Cisplatin, Gemcitabine, Taxol sowie UV-Strahlung induzierte Apoptose effizient zu hemmen [393-395]. Dabei interagiert das Nikotin mit dem nikotinischen Acetylcholinrezeptor und aktiviert darauffolgend über Phosphorylierung unterschiedlichste molekulare Mechanismen der Apoptose-Inhibierung. So

können z.B. die antiapoptotisch wirkenden BCL2 und AKT Proteine aktiviert werden, während die proapoptotisch wirkenden BAX oder BAD Proteine gehemmt werden [393, 396-398]. Des Weiteren assoziiert die protektive Nikotinwirkung mit dem intrazellulären XIAP-Anstieg [394].

Laut Hersteller bestehen die „Dipse“-*Liquids* aus Monopropylenglykol, Glycerin, Wasser, Alkohol und für die jeweilige Sorte charakteristische Aromen. Bei nikotinhaltigen *Liquids* kommen noch 24mg/ml Nikotin hinzu [369]. Wie bereits beschrieben, ist es unwahrscheinlich, dass Nikotin für die zytotoxische Wirkung des nikotinhaltigen CSE verantwortlich war. Auch konnten Monopropylenglykol, Glycerin, Wasser sowie Alkohol ausgeschlossen werden, da sie sowohl in nikotinhaltigen als auch nikotinfreien E-Zigarettenextrakten in gleicher Konzentration enthalten sind. Daher ist es durchaus möglich, dass das in der nikotinhaltigen *Liquid*-Sorte enthaltene Tabakaroma für die Apoptose mitverantwortlich war. Der Begriff Aroma ist breit definiert und bezeichnet einen Geruch oder Geschmack, der durch einzelne chemische Verbindungen oder Stoffgemische verursacht wird. Bei den einzelnen Komponenten kann es sich um verschiedene Aromaten, Ester, Terpene, Alkylpyrazine, Aldehyde oder Ketone handeln [399]. Daher sind an dieser Stelle zukünftig umfassende spektrometrische und funktionelle Untersuchungen notwendig, um die genaue zytotoxische Verbindung zu identifizieren. Für die schädliche Wirkung einiger Aromen würde auch die Studie von Rowell *et al.* sprechen, bei der 13 *Liquids* mit verschiedenen Geschmackrichtungen analysiert wurden. Alle 13 untersuchten *Liquids* führten im MTT-Test zu einer dosenabhängigen Abnahme zellulären Stoffwechsels. Vier davon zeigten sogar eine starke zytotoxische Wirkung. Darauffolgend empfehlen die Autoren der Studie, die handelsüblichen E-Zigaretten viel gründlicher auf ihre gesundheitsschädliche Wirkung zu prüfen [368].

Andere fehlerhaft regulierte Gene, die nach der Behandlung der Lungenzellen mit dem nikotinhaltigen E-Zigarettenextrakt identifiziert wurden, sind in Morphogenese, Entwicklung, Zellzyklus, Proteinbiogenese, -lokalisation und -bindung involviert. Die fehlregulierten Prozesse hängen vermutlich mit hohem Nikotingehalt des verwendeten *Liquids* zusammen [400-405].

Ein weiteres ernstzunehmendes Ergebnis der multi-OMICS Studie war, dass der nikotinfreie E-Zigarettenextrakt zwar keine Apoptose jedoch Proliferation der Lungenzellen induzierte. Auch wenn Krebs eine komplexe und äußerst heterogene Erkrankung ist, haben alle malignen Krebszellen zwei Eigenschaften gemeinsam: fehlregulierte, erhöhte Zellproliferation sowie supprimierte Apoptose. Diese zwei Punkte sind sozusagen eine Mindestanforderung für die neoplastische Progression und Tumorentstehung [406].

Wie bereits im Abschnitt 1.2.3 beschrieben, sind Propylenglykol und Glycerin die Hauptkomponenten aller handelsüblichen *Liquids*. Beim Verdampfen entstehen große Mengen an Formaldehyd und Acrolein [157]. Auf zellulärer Ebene ist die Wirkung des Formaldehyds stark von der Dosis abhängig. Während sehr hohe Formaldehyd-Konzentrationen von über 10mM zu Nekrose führen, kommt es bei sehr geringen Konzentrationen unter 0,5mM zu erhöhten Proliferations- sowie verminderten Apoptoseraten [407].

Obwohl zahlreiche Studien vor gesundheitlichen Gefahren des Passivrauchens warnen, gibt es bisher keinerlei Publikationen, die sich mit dem Passivrauchen von E-Zigaretten auseinandersetzen. So sind z.B. die Formaldehydkonzentrationen im ausgeatmeten E-Zigarettenrauch 20 mal höher als in der umgebenden Luft [408]. Das könnte auch für Schwangere problematisch sein. Derzeit gibt es zwar keine Studien, die gezielt Auswirkungen des E-Zigarettenrauchs auf Embryonalentwicklung untersucht haben. Allerdings raten zahlreiche Publikationen vor Formaldehydexposition während der Schwangerschaft ab, da es ein perinatales Risiko für den Fetus bedeutet [409, 410]. Besonders beunruhigend in diesem Zusammenhang ist die Studie von Wagner *et al.* Demnach denken aufgrund zahlreicher Werbekampagnen viele Schwangere, dass E-Zigaretten weniger ungesund als klassische Zigaretten sind. Das führt dazu, dass sie während der Schwangerschaft zur E-Zigarette wechseln oder es für unnötig halten, mit dem *Liquid*-Konsum aufzuhören [411].

Zusätzlich zu erhöhten Proliferationsraten wurde in dieser Arbeit erstmals eine erhöhte Anzahl an multinuklearen bzw. aneuploiden A549-Zellen nach der Behandlung mit nikotinfreien eCSE beschrieben. Die genomweiten, molekulargenetischen Untersuchungen identifizierten 566 überexprimierte Gene, die für folgende Prozesse von Bedeutung sind: TGF- β Signalweg, Aufrechterhaltung der Telomerlängen sowie DNA- und RNA-Synthese. Des Weiteren wurden 1041 herunterregulierte Gene identifiziert, die für Redoxprozesse, Autophagie, Metabolismus, Katabolismus und Biosynthese entscheidend sind. Somit wird klar, dass der nikotinfreie E-Zigarettenextrakt nicht nur über den TGF- β Signalweg die Zellproliferation induziert. Zusätzlich werden Autophagie sowie zahlreiche biosynthetische und metabolische Prozesse so gehemmt, dass die Zellteilung offenbar nicht vollständig ablaufen kann. Das Ergebnis sind multinukleare, aneuploide Zellen. Dieser Hinweis wird von mehreren Publikationen untermauert, die besagen, dass vor allem fehlerhafte Sphingolipid-Biosynthese sowie Autophagie zur Entstehung multinuklearer Zellen führen können [412, 413].

Multinukleare bzw. aneuploide Zellen kommen in einigen NSCLC-Subtypen vor. Besonders häufig sind die Zugewinne von Chromosomen 7, 8 und 12 sowie Verluste an Chromosom 9

[414, 415]. Multinukleare Zellen sind vor allem für eine seltene und sehr aggressive Form des NSCLC, dem Riesenzellkarzinom, typisch [416]. Es wurde aber auch von einem Plattenepithelkarzinompatienten berichtet, dessen Primärtumor multinukleare Zellen aufwies [417]. Bei einem weiteren Patienten mit Adenokarzinom wurden multinukleare Zellen zwar nicht im primären Tumor, aber in verschiedenen Metastasen nachgewiesen [418]. Der Studie von Weihua et.al. zufolge, kann eine einzelne mononukleare Zelle in Mäusen zur Entstehung solider Tumore führen [419].

Des Weiteren konnten die genomweiten Analysen nach der Behandlung der Lungenzellen mit nikotinfreiem eCSE Fehlregulation von 31 membrangebundenen Transportern detektieren, inkl. SLC-, ABC- und ATP-Transporterfamilien. Viele der betroffenen Transporter sind unter anderem im Zusammenhang mit Chemoresistenz sowie Metabolismus von Lungenkarzinomzellen beschrieben worden [420-422].

Für die zelluläre Aufnahme der Chemotherapeutika sind membrangebundene Transporter, wie z.B. organische Kationentransporter, entscheidend. Interessanterweise wurde nach der Behandlung mit nikotinfreiem eCSE die zelluläre Expression von SLC22A3 reprimiert [422]. Gleichzeitig kam es zu Überexpression einiger ABC-Efflux-Transporter, wie z.B. ABCA1. Die Mitglieder der ABC-Transporterfamilie sind für das Herauspumpen zytotoxischer Agentien aus der Zelle zuständig und wurden häufig im Zusammenhang mit Chemoresistenz beschrieben [420, 423, 424]. Auch der SLC46A3-Transporter, der für den Transport einiger Antikörper-Wirkstoff-Konjugate aus dem Lysosom ins Zytoplasma verantwortlich ist, wurde nach der Stimulation reprimiert [425].

Des Weiteren konnte Überexpression von mehreren für den zellulären Stoffwechsel relevanten Transportern detektiert werden. So wurden z.B. SLC1A2 oder SLC6A9 im Zusammenhang mit erhöhtem Metabolismus maligner Tumore in der Literatur beschrieben [305, 426-429].

Ein weiteres interessantes Beispiel stellt SLC25A23 dar, der die intrazelluläre Kalzium-Aufnahme verstärken und somit durch den oxidativen Stress vermittelte Apoptose induzieren kann [430]. Nach der Behandlung mit nikotinfreiem eCSE war der SLC25A23-Transporter herunterreguliert.

Zusammenfassend lässt sich sagen, dass die Ergebnisse der im Rahmen der vorliegenden Doktorarbeit durchgeführten multi-OMICS Studie Hinweise auf erstzunehmende gesundheitliche Gefahren der E-Zigaretten liefern und zahlreiche wissenschaftliche Arbeiten untermauern, die vor E-Zigarettenkonsum warnen.

Bereits geringe Mengen an nikotinfreiem E-Zigarettenextrakt induzierten Aneuploidie, Zellproliferation und supprimierten Apoptose. Der tägliche *Liquid*-Konsum eines durchschnittlichen Rauchers ist jedoch deutlich höher [154], was vor allem langfristig zur Tumorentstehung führen könnte. Laut Berechnungen von Jensen *et al.* haben E-Zigarettenkonsumenten im Vergleich zu Tabakkonsumenten ein 15fach höheres Risiko im Laufe des Lebens am Lungenkrebs zu erkranken [154]. Die dabei entstehenden Tumore könnten aufgrund der Fehlregulation von SLC- sowie ABC-Transportern chemoresistent und aufgrund des erhöhten Metabolismus besonders aggressiv sein [420-422].

Des Weiteren induzierte das im nikotinhaltigen *Liquid* enthaltene Tabakaroma Apoptose. Die Apoptose des Lungenepithels hat Entzündung zur Folge [431, 432], was nicht nur das Risiko an Bronchialkarzinom zu erkranken noch weiter erhöht [433] sondern auch andere schwerwiegende Erkrankungen der Lunge, wie z.B. Asthma, COPD oder ARDS auslösen könnte [432, 434, 435].

Die im Rahmen dieser Arbeit durchgeführte multi-OMICS Studie zeigt deutlich, dass die Zusammensetzung der auf dem Markt erhältlichen *Liquids* viel strikter kontrolliert, dokumentiert und reguliert werden sollte. Es ist notwendig, die Öffentlichkeit genauso wie bei Tabakprodukten über die gesundheitlichen Risiken des E-Zigarettenkonsums aufzuklären und die verharmlosenden Werbekampagnen zu unterbinden.

6. Zusammenfassung

Das nicht-kleinzellige Lungenkarzinom (engl. *non-small cell lung cancer*; NSCLC) ist die zweithäufigste Krebsart bei Männern und Frauen, zu den häufigsten Entitäten zählen das Adeno- und das Plattenepithelkarzinom. Der NSCLC wird häufig im fortgeschrittenen Stadium diagnostiziert und ist mit einer schlechten Prognose assoziiert. Bisher gibt es keine Möglichkeiten der routinemäßigen Lungenkrebsfrüherkennung.

Rauchen zählt zu den häufigsten Ursachen für die Entstehung von Bronchialkarzinomen, weswegen heutzutage viele Raucher zu der vermeintlich gesünderen Alternative, der E-Zigarette, wechseln. Da beim Verdampfen des *Liquids* u.a. auch Formaldehyd und Acrolein entstehen, raten einige Studien vom E-Zigarettenkonsum ab. Um die tatsächlichen Risiken der E-Zigarette abzuschätzen, fehlen Ergebnisse aus Langzeitstudien. Desweiteren liegen keine biologischen Studien vor, welche die Wirkung des E-Zigarettdampfes auf molekulargenetischer Ebene genomweit untersucht haben.

Die vorliegende Doktorarbeit wurde im Rahmen des Deutschen Zentrums für Lungenkrankheiten (DZL) durchgeführt. Ziel war es, einerseits einen zuverlässigen Früherkennungstest anhand von bestimmten, für das Lungenkarzinom spezifischen Epimutationen zu entwickeln. Andererseits wurde die Wirkung des E-Zigarettdampfes auf die Entstehung von Epimutationen untersucht.

Bei dem entwickelten Test für Lungenkrebsfrüherkennung handelte es sich um ein auf *Liquid Biopsies* basiertes NGS-*Panel*. Die Entwicklung des Plasma *Panel*s erfolgte in drei voneinander unabhängigen Ansätzen. Im ersten Ansatz wurden für den primären Tumor spezifische DNA-Methylierungssignaturen detektiert. Hierfür wurden „Illumina Infinium Human Methylation450K BeadChip“ (HM 450K)-Datensätze, bestehend aus 40 Lungenkarzinomen sowie deren korrespondierenden Kontrollen, analysiert. Die *in silico* Analysen ergaben 224 CpG *Loci*, die am besten für die Differenzierung maligner Tumorgewebe geeignet waren ($FDR < 1 \times 10^{-23}$, $\sigma/\sigma_{\max} > 0,4$). 63 weitere Positionen zeigten eine starke, entitätsspezifische DNA-Methylierung ($FDR < 1 \times 10^{-4}$). Im zweiten Ansatz wurden für das Blutplasma der Lungenkrebspatienten spezifische DNA-Methylierungssignaturen detektiert. Hierfür erfolgte *Whole Genome Bisulfite Sequencing* (WGBS) zirkulierender, zellfreier DNA (zfdNA) aus dem Plasma von Adeno- und Plattenepithelkarzinompatienten. Als Kontrolle diente Plasma einer tumorfreen Kohorte. WGBS identifizierte erstmalig 29.877 tumor- sowie 76.374 entitätsspezifisch methylierte CpG *Loci* im Blutplasma. Für das *Panel* wurden 292 Regionen ausgewählt, die maligne Lungentumore sowie deren Entitäten am zuverlässigsten differenzierten. Im dritten Ansatz wurde das Plasma *Panel* mit 59 tumorspezifischen sowie prognostisch relevanten CpG *Loci*

(FDR < 0,05) aus Datensätzen weiterer Studien ergänzt. Anschließend wurde das Plasma *Panel* synthetisiert und anhand einer kleinen Patientenkohorte (n=12) validiert. Im Rahmen der Pilot-Studie differenzierte das entwickelte Plasma *Panel* mit einer 100%igen Richtigkeit maligne Lungentumore (Stadium: I bis IV) und lieferte weitere Informationen bezüglich deren Entität und *Staging*.

Um die Wirkung des E-Zigarettdampfes auf Lungenzellen zu untersuchen, wurde eine multi-OMICS Studie designt. Im Rahmen der Studie wurden primäre Lungenepithelzellen sowie die A549 Lungenzelllinie parallel analysiert. Nach Etablierung zahlreicher zellphysiologischer, molekularbiologischer sowie epigenetischer Assays inkl. „Infinium MethylationEPIC BeadChip“, ATAC-Sequenzierung und RNA-Sequenzierung wurden die Daten ausgewertet.

Die Ergebnisse der multi-OMICS Studie zeigten, dass bereits geringe Mengen an nikotinfreiem E-Zigarettenextrakt Aneuploidie und Zellproliferation induzierten, die Apoptose wurde dabei inhibiert. Diese Beobachtungen wurden bisher in der Literatur noch nie im Zusammenhang mit E-Zigaretten beschrieben und untermauern andere Studien, die vor dem erhöhten Lungenkrebsrisiko der E-Zigarettenkonsumenten warnen. Desweiteren zeigten molekulargenetischen Untersuchungen zum ersten Mal eine Fehlregulation von SLC- sowie ABC-Transportern. Das lässt den Rückschluss zu, dass die entstehenden Tumore eventuell chemoresistent und aufgrund des erhöhten Metabolismus besonders aggressiv sein könnten.

Geringe Mengen des nikotinhaltigen E-Zigarettenextrakts induzierten bereits nach drei Stunden Apoptose der Lungenzellen. Laut mehreren internationalen Studien hat die Apoptose des Lungenepithels Entzündungen zur Folge, was das Risiko an Bronchialkarzinom zu erkranken noch weiter erhöhen könnte. Desweiteren detektierten genomweite, molekulare Analysen erstmalig ähnlich wie bei den klassischen Zigaretten die Fehlregulation wichtiger zellulärer Prozesse, wie z.B. Morphogenese, Entwicklung, Zellzyklus, Proteinbiogenese, -lokalisation und -bindung.

In dieser Arbeit konnte somit ein diagnostisch-prognostisches *Panel* auf Basis epigenetischer Alterationen entwickelt werden. Die Ergebnisse dieser Arbeit stellen damit eine wesentliche Grundlage für die erfolgreiche Beantragung der deutschlandweiten *Clinical Trial* Studie „EmoLung“ des DZL dar, die voraussichtlich 2018 beginnt. Darüber hinaus konnte mit einer Kombination epigenetischer, molekularbiologischer und zellphysiologischer Untersuchungen erstmals eine Proliferations- und Aneuploidie fördernde Wirkung durch E-Zigarettenextrakt gezeigt werden.

7. Summary

Non-small cell lung cancer (NSCLC) is worldwide the second most common cancer in both men and women. The most common entities include adeno- and squamous cell carcinoma. As there are no proper tools for lung cancer early detection, NSCLC is often diagnosed in advanced stages and is therefore associated with poor prognosis.

Smoking is one of the most frequent causes of bronchial carcinomas. As a consequence many smokers switch to e-cigarettes, which are supposed to be a healthier alternative to conventional cigarettes. However, during vapouring big amounts of formaldehyde and acrolein develop. For this reason several studies dis advise e-cigarette consumption. Unfortunately, there is a lack of long term as well as short term molecular studies to estimate the risk of e-cigarettes.

This work was performed in the framework of The German Center for Lung Research (DZL). The aims were to design a reliable, epimutation-based test for lung cancer early detection and to analyze the effect of e-cigarette vapour on the appearance of those epimutations.

The developed lung cancer early detection test was a liquid biopsy-based NGS-panel. The development occurred in several approaches. First, tumor-specific DNA methylation signatures were detected. To this aim „Illumina Infinium Human Methylation450K BeadChip“-data of 40 lung cancer surgical specimen as well as their corresponding controls were analysed. *In silico* data analysis identified 224 CpG loci, whose methylation was significantly different in malignant tumor tissue ($FDR < 1 \times 10^{-23}$, $\sigma/\sigma_{max} > 0.4$). 63 other positions showed entity specific DNA methylation patterns ($FDR < 1 \times 10^{-4}$). The second approach was performed to identify tumor specific DNA methylation profiles in blood plasma. Therefore, whole genome bisulfite sequencing (WGBS) of circulating, cell-free DNA (cfDNA), isolated from plasma of adeno- and squamous cell carcinoma patients, was performed. Tumor-free patients served as a control group. For the first time, WGBS identified 29.877 tumor and 76.374 entity specific, differentially methylated CpG loci in blood plasma. Therefrom 292 targets were selected, depending on their ability to differentiate malignant lung tumors and their entities. During the third and the last approach the panel was complemented with 59 tumor specific and prognostic markers ($FDR < 0.05$) from other studies. After a successful development, the plasma panel was synthesized and validated by using a small patient cohort (n=12). The developed panel was able to analyze liquid biopsy samples and identify lung cancer patients (stage I to IV) with an accuracy of 100%. Moreover, it provided information about the entity and the tumor stage.

For the second aim of this work, a multi-OMICS study was designed to analyse the effect of e-cigarette vapour on lung cells. In this study primary epithelial lung cells as well as A549 cells were used. After establishing numerous physiological, molecular and epigenetic protocols including the "Infinium Methylation EPIC BeadChip", ATAC-Sequencing and RNA-Sequencing, data analysis was performed.

For the first time, it was shown, that even small amounts of nicotine free e-cigarette extract induced aneuploidy as well as cell proliferation. At the same time apoptosis was inhibited. These observations are in line with the results of other international studies indicating the potential harm of e-cigarettes. Moreover, for the first time the multi-OMICS study identified aberrant transcriptional regulation of SLC- and ABC-transporters, leading to potential chemoresistance and aggressiveness of arising tumors.

Small amounts of nicotine containing e-cigarette extract induced after three hours apoptosis in lung cells. According to several international studies enhanced apoptosis leads to lung epithelia inflammation, enhancing the risk of tumor development. Here, genome-wide molecular assays detected for the first time deregulation of important cellular processes including morphogenesis, development, cell cycle, protein biogenesis, protein localisation and protein binding.

During this work epigenetic alterations in lung cancer specimen were determined in order to develop a prognostic and diagnostic lung cancer panel. The results of this study provided the basis for a successful application of the German Clinical Trial DZL-Study „EmoLung“, which will be initiated in 2018. Moreover, a proliferative and aneuploidy promoting effect of e-cigarettes was shown for the first time using epigenetic, molecular and physiological approaches.

8. Literaturverzeichnis

1. Hansen, T., *Type 2 diabetes mellitus-a multifactorial disease*. Ann Univ Mariae Curie Sklodowska Med, 2002. **57**(1): p. 544-9.
2. Basile, J.N., *A multifactorial approach to reduce cardiovascular disease in type 2 diabetes mellitus: now more than ever*. Hosp Pract (1995), 2016. **44**(1): p. 9-20.
3. Lafeber, M., et al., *Multifactorial prevention of cardiovascular disease in patients with hypertension: the cardiovascular polypill*. Curr Hypertens Rep, 2016. **18**(5): p. 40.
4. Stine, N. and D. Chokshi, *Multifactorial risk assessment for atherosclerotic cardiovascular disease-reply*. JAMA, 2015. **313**(9): p. 972.
5. Shahzad, N., et al., *Phytosterols as a natural anticancer agent: current status and future perspective*. Biomed Pharmacother, 2017. **88**: p. 786-794.
6. Koch, C.A., *How can environmental factors contribute to the incidence of thyroid cancer?* Horm Metab Res, 2017. **49**(3): p. 229-231.
7. Blumenthal, M.N., *Genetic, epigenetic, and environmental factors in asthma and allergy*. Ann Allergy Asthma Immunol, 2012. **108**(2): p. 69-73.
8. Gapp, K., et al., *Early life epigenetic programming and transmission of stress-induced traits in mammals: how and when can environmental factors influence traits and their transgenerational inheritance?* Bioessays, 2014. **36**(5): p. 491-502.
9. Herceg, Z. and A. Paliwal, *Epigenetic mechanisms in hepatocellular carcinoma: how environmental factors influence the epigenome*. Mutat Res, 2011. **727**(3): p. 55-61.
10. Jadiya, P. and A. Nazir, *Environmental toxicants as extrinsic epigenetic factors for parkinsonism: studies employing transgenic C. elegans model*. CNS Neurol Disord Drug Targets, 2012. **11**(8): p. 976-83.
11. Kargul, J., I. Irminger-Finger, and G.J. Laurent, *Epigenetics regulation of disease: There is more to a gene than its sequence*. Int J Biochem Cell Biol, 2015. **67**: p. 43.
12. Portha, B., et al., *Early environmental factors, alteration of epigenetic marks and metabolic disease susceptibility*. Biochimie, 2014. **97**: p. 1-15.
13. Warburton, D. and B.E. Olver, *Coordination of genetic, epigenetic, and environmental factors in lung development, injury, and repair*. Chest, 1997. **111**(6 Suppl): p. 119S-122S.
14. Wissenschaft, S.d. *Epigenetik*. Available from: <http://www.spektrum.de/thema/epigenetik/1191602> (abgerufen am 24.10.2017).
15. Morgan, H.D., et al., *Epigenetic inheritance at the agouti locus in the mouse*. Nat Genet, 1999. **23**(3): p. 314-8.

16. Guerrero-Bosagna, C. and M.K. Skinner, *Environmentally induced epigenetic transgenerational inheritance of male infertility*. *Curr Opin Genet Dev*, 2014. **26**: p. 79-88.
17. Srinageshwar, B., et al., *Role of epigenetics in stem cell proliferation and differentiation: implications for treating neurodegenerative diseases*. *Int J Mol Sci*, 2016. **17**(2).
18. Xu, W., et al., *Epigenetics and cellular metabolism*. *Genet Epigenet*, 2016. **8**: p. 43-51.
19. Wahlin, K.J., et al., *Epigenetics and cell death: DNA hypermethylation in programmed retinal cell death*. *PLoS One*, 2013. **8**(11): p. e79140.
20. Doerfler, W., *In pursuit of the first recognized epigenetic signal-DNA methylation: a 1976 to 2008 synopsis*. *Epigenetics*, 2008. **3**(3): p. 125-33.
21. Hattori, M. and T.D. Taylor, *The human intestinal microbiome: a new frontier of human biology*. *DNA Res*, 2009. **16**(1): p. 1-12.
22. Murrell, A., V.K. Rakyan, and S. Beck, *From genome to epigenome*. *Hum Mol Genet*, 2005. **14 Spec No 1**: p. R3-R10.
23. Armstrong, L., *Epigenetic control of embryonic stem cell differentiation*. *Stem Cell Rev*, 2012. **8**(1): p. 67-77.
24. Wu, H. and Y.E. Sun, *Epigenetic regulation of stem cell differentiation*. *Pediatr Res*, 2006. **59**(4 Pt 2): p. 21R-5R.
25. Bemmer, M., *Unraveling the complex epigenetic mechanisms that regulate gene activity*. *Methods Mol Biol*, 2018. **1675**: p. 205-231.
26. Gan, Q., et al., *Concise review: epigenetic mechanisms contribute to pluripotency and cell lineage determination of embryonic stem cells*. *Stem Cells*, 2007. **25**(1): p. 2-9.
27. Yan, M.S., C.C. Matouk, and P.A. Marsden, *Epigenetics of the vascular endothelium*. *J Appl Physiol (1985)*, 2010. **109**(3): p. 916-26.
28. Colot, V., L. Maloisel, and J.L. Rossignol, *DNA repeats and homologous recombination: a probable role for DNA methylation in genome stability of eukaryotic cells*. *J Soc Biol*, 1999. **193**(1): p. 29-34.
29. Harbers, K., et al., *DNA methylation and gene expression: endogenous retroviral genome becomes infectious after molecular cloning*. *Proc Natl Acad Sci U S A*, 1981. **78**(12): p. 7609-13.
30. Razin, A. and B. Kantor, *DNA methylation in epigenetic control of gene expression*. *Prog Mol Subcell Biol*, 2005. **38**: p. 151-67.
31. Shalginikh, N., et al., *Retroviral DNA methylation and epigenetic repression are mediated by the antiviral host protein Daxx*. *J Virol*, 2013. **87**(4): p. 2137-50.

32. Wagener, M., *Molekulare Onkologie: Entstehung, Progression, klinische Aspekte* G.T. Verlag, Editor. 2010.
33. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development*. Cell, 1999. **99**(3): p. 247-57.
34. Robert, M.F., et al., *DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells*. Nat Genet, 2003. **33**(1): p. 61-5.
35. Bashtrykov, P., et al., *Specificity of Dnmt1 for methylation of hemimethylated CpG sites resides in its catalytic domain*. Chem Biol, 2012. **19**(5): p. 572-8.
36. Bashtrykov, P. and A. Jeltsch, *DNMT1-associated DNA methylation changes in cancer*. Cell Cycle, 2015. **14**(1): p. 5.
37. Shaffer, B., et al., *The DNMT1 intrinsically disordered domain regulates genomic methylation during development*. Genetics, 2015. **199**(2): p. 533-41.
38. Ehrlich, M., et al., *Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells*. Nucleic Acids Res, 1982. **10**(8): p. 2709-21.
39. Razin, A. and H. Cedar, *DNA methylation in eukaryotic cells*. Int Rev Cytol, 1984. **92**: p. 159-85.
40. Caiafa, P. and M. Zampieri, *DNA methylation and chromatin structure: the puzzling CpG islands*. J Cell Biochem, 2005. **94**(2): p. 257-65.
41. Cheong, J., et al., *Diverse DNA methylation statuses at alternative promoters of human genes in various tissues*. DNA Res, 2006. **13**(4): p. 155-67.
42. Comb, M. and H.M. Goodman, *CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2*. Nucleic Acids Res, 1990. **18**(13): p. 3975-82.
43. Newell-Price, J., A.J. Clark, and P. King, *DNA methylation and silencing of gene expression*. Trends Endocrinol Metab, 2000. **11**(4): p. 142-8.
44. Fatemi, M. and P.A. Wade, *MBD family proteins: reading the epigenetic code*. J Cell Sci, 2006. **119**(Pt 15): p. 3033-7.
45. Kohli, R.M. and Y. Zhang, *TET enzymes, TDG and the dynamics of DNA demethylation*. Nature, 2013. **502**(7472): p. 472-9.
46. Patil, V.S., R. Zhou, and T.M. Rana, *Gene regulation by non-coding RNAs*. Crit Rev Biochem Mol Biol, 2014. **49**(1): p. 16-32.
47. Brimacombe, R. and W. Stiege, *Structure and function of ribosomal RNA*. Biochem J, 1985. **229**(1): p. 1-17.
48. Price, S.R., P.R. Evans, and K. Nagai, *Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA*. Nature, 1998. **394**(6694): p. 645-50.

49. Kiss, T., *Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions*. Cell, 2002. **109**(2): p. 145-8.
50. Schmitz, K.M., et al., *Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes*. Genes Dev, 2010. **24**(20): p. 2264-9.
51. Vance, K.W. and C.P. Ponting, *Transcriptional regulatory functions of nuclear long noncoding RNAs*. Trends Genet, 2014. **30**(8): p. 348-55.
52. Agrawal, N., et al., *RNA interference: biology, mechanism, and applications*. Microbiol Mol Biol Rev, 2003. **67**(4): p. 657-85.
53. Khraiwesh, B., et al., *Transcriptional control of gene expression by microRNAs*. Cell, 2010. **140**(1): p. 111-22.
54. Grewal, S.I. and D. Moazed, *Heterochromatin and epigenetic control of gene expression*. Science, 2003. **301**(5634): p. 798-802.
55. Gilbert, N., et al., *Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers*. Cell, 2004. **118**(5): p. 555-66.
56. Watson, B., *Molecular Biology of the Gene*. Pearson, Editor. 2004.
57. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 Å resolution*. Nature, 1997. **389**(6648): p. 251-60.
58. McGinty, R.K. and S. Tan, *Nucleosome structure and function*. Chem Rev, 2015. **115**(6): p. 2255-73.
59. Thoma, F., T. Koller, and A. Klug, *Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin*. J Cell Biol, 1979. **83**(2 Pt 1): p. 403-27.
60. Nemeth, A. and G. Langst, *Chromatin higher order structure: opening up chromatin for transcription*. Brief Funct Genomic Proteomic, 2004. **2**(4): p. 334-43.
61. Woodcock, C.L. and R.P. Ghosh, *Chromatin higher-order structure and dynamics*. Cold Spring Harb Perspect Biol, 2010. **2**(5): p. a000596.
62. Li, G. and D. Reinberg, *Chromatin higher-order structures and gene regulation*. Curr Opin Genet Dev, 2011. **21**(2): p. 175-86.
63. Maeshima, K., et al., *Chromatin as dynamic 10-nm fibers*. Chromosoma, 2014. **123**(3): p. 225-37.
64. Cairns, B.R., *The logic of chromatin architecture and remodelling at promoters*. Nature, 2009. **461**(7261): p. 193-8.
65. Ettig, R., et al., *Dissecting DNA-histone interactions in the nucleosome by molecular dynamics simulations of DNA unwrapping*. Biophys J, 2011. **101**(8): p. 1999-2008.
66. Erler, J., et al., *The role of histone tails in the nucleosome: a computational study*. Biophys J, 2014. **107**(12): p. 2911-22.

67. Yun, M., et al., *Readers of histone modifications*. Cell Res, 2011. **21**(4): p. 564-78.
68. Peterson, C.L. and M.A. Laniel, *Histones and histone modifications*. Curr Biol, 2004. **14**(14): p. R546-51.
69. Kothapalli, N., et al., *Biological functions of biotinylated histones*. J Nutr Biochem, 2005. **16**(7): p. 446-8.
70. Dong, X. and Z. Weng, *The correlation between histone modifications and gene expression*. Epigenomics, 2013. **5**(2): p. 113-6.
71. Kouzarides, T., *Chromatin modifications and their function*. Cell, 2007. **128**(4): p. 693-705.
72. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications*. Cell Res, 2011. **21**(3): p. 381-95.
73. Handy, D.E., R. Castro, and J. Loscalzo, *Epigenetic modifications: basic mechanisms and role in cardiovascular disease*. Circulation, 2011. **123**(19): p. 2145-56.
74. Kuroishi, T., et al., *Biotinylation is a natural, albeit rare, modification of human histones*. Mol Genet Metab, 2011. **104**(4): p. 537-45.
75. Martin, C. and Y. Zhang, *The diverse functions of histone lysine methylation*. Nat Rev Mol Cell Biol, 2005. **6**(11): p. 838-49.
76. Tang, L., E. Nogales, and C. Ciferri, *Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription*. Prog Biophys Mol Biol, 2010. **102**(2-3): p. 122-8.
77. Gangaraju, V.K. and B. Bartholomew, *Mechanisms of ATP dependent chromatin remodeling*. Mutat Res, 2007. **618**(1-2): p. 3-17.
78. Nigg, E.A., *Assembly and cell cycle dynamics of the nuclear lamina*. Semin Cell Biol, 1992. **3**(4): p. 245-53.
79. Goldman, R.D., et al., *Nuclear lamins: building blocks of nuclear architecture*. Genes Dev, 2002. **16**(5): p. 533-47.
80. Gruenbaum, Y., et al., *The nuclear lamina and its functions in the nucleus*. Int Rev Cytol, 2003. **226**: p. 1-62.
81. Shimi, T., et al., *The A- and B-type nuclear lamin networks: microdomains involved in chromatin organization and transcription*. Genes Dev, 2008. **22**(24): p. 3409-21.
82. Kind, J. and B. van Steensel, *Genome-nuclear lamina interactions and gene regulation*. Curr Opin Cell Biol, 2010. **22**(3): p. 320-5.
83. Collas, P., E.G. Lund, and A.R. Oldenburg, *Closing the (nuclear) envelope on the genome: how nuclear lamins interact with promoters and modulate gene expression*. Bioessays, 2014. **36**(1): p. 75-83.

84. van Steensel, B. and A.S. Belmont, *Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression*. Cell, 2017. **169**(5): p. 780-791.
85. Pombo, A. and N. Dillon, *Three-dimensional genome architecture: players and mechanisms*. Nat Rev Mol Cell Biol, 2015. **16**(4): p. 245-57.
86. Pontvianne, F., et al., *Identification of nucleolus-associated chromatin domains reveals a role for the nucleolus in 3D organization of the A. thaliana Genome*. Cell Rep, 2016. **16**(6): p. 1574-1587.
87. van Koningsbruggen, S., et al., *High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli*. Mol Biol Cell, 2010. **21**(21): p. 3735-48.
88. Dekker, J., M.A. Marti-Renom, and L.A. Mirny, *Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data*. Nat Rev Genet, 2013. **14**(6): p. 390-403.
89. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions*. Nature, 2012. **485**(7398): p. 376-80.
90. Acemel, R.D., I. Maeso, and J.L. Gomez-Skarmeta, *Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals*. Wiley Interdiscip Rev Dev Biol, 2017. **6**(3).
91. Bolzer, A., et al., *Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes*. PLoS Biol, 2005. **3**(5): p. e157.
92. Cremer, T. and M. Cremer, *Chromosome territories*. Cold Spring Harb Perspect Biol, 2010. **2**(3): p. a003889.
93. iGEM. Available from: https://2014.igem.org/Team:Heidelberg/Project/PCR_2.0 (abgerufen am 26.10.2017).
94. EPIGENTEK. *DNA Demethylation*. Available from: <https://www.epigentek.com/catalog/dna-demethylation.php> (abgerufen am 26.10.2017).
95. NIH, N.I.o.H. *epigenetische Mechanismen*. Available from: <https://commonfund.nih.gov/epigenomics/figure> (abgerufen am 26.10.2017).
96. Krebsforschungszentrums, D.K.d.D. *Lungenkrebs*. Available from: <https://www.krebsinformationsdienst.de/> (abgerufen am 26.10.2017).
97. Mabry, M., *The biology of lung cancer: an overview*. Cancer Control, 1994. **1**(5): p. 528-536.
98. Gridelli, C., et al., *Non-small-cell lung cancer*. Nat Rev Dis Primers, 2015. **1**: p. 15009.
99. Collins, L.G., et al., *Lung cancer: diagnosis and management*. Am Fam Physician, 2007. **75**(1): p. 56-63.

100. Cooper, W.A., et al., *Molecular biology of lung cancer*. J Thorac Dis, 2013. **5 Suppl 5**: p. S479-90.
101. Pon, J.R. and M.A. Marra, *Driver and passenger mutations in cancer*. Annu Rev Pathol, 2015. **10**: p. 25-50.
102. Brzezianska, E., A. Dutkowska, and A. Antczak, *The significance of epigenetic alterations in lung carcinogenesis*. Mol Biol Rep, 2013. **40**(1): p. 309-25.
103. Belinsky, S.A., et al., *Gene promoter methylation in plasma and sputum increases with lung cancer risk*. Clin Cancer Res, 2005. **11**(18): p. 6505-11.
104. Zochbauer-Muller, S., J.D. Minna, and A.F. Gazdar, *Aberrant DNA methylation in lung cancer: biological and clinical implications*. Oncologist, 2002. **7**(5): p. 451-7.
105. Kim, H., et al., *Elevated mRNA levels of DNA methyltransferase-1 as an independent prognostic factor in primary nonsmall cell lung cancer*. Cancer, 2006. **107**(5): p. 1042-9.
106. Lin, R.K., et al., *Alteration of DNA methyltransferases contributes to 5'CpG methylation and poor prognosis in lung cancer*. Lung Cancer, 2007. **55**(2): p. 205-13.
107. Langevin, S.M., R.A. Kratzke, and K.T. Kelsey, *Epigenetics of lung cancer*. Transl Res, 2015. **165**(1): p. 74-90.
108. Kim, S.H., et al., *Expression of cancer-testis antigens MAGE-A3/6 and NY-ESO-1 in non-small-cell lung carcinomas and their relationship with immune cell infiltration*. Lung, 2009. **187**(6): p. 401-11.
109. Guan, P., et al., *Meta-analysis of human lung cancer microRNA expression profiling studies comparing cancer tissues with normal tissues*. J Exp Clin Cancer Res, 2012. **31**: p. 54.
110. Liu, X.H., et al., *The long non-coding RNA HOTAIR indicates a poor prognosis and promotes metastasis in non-small cell lung cancer*. BMC Cancer, 2013. **13**: p. 464.
111. Schmidt, L.H., et al., *The long noncoding MALAT-1 RNA indicates a poor prognosis in non-small cell lung cancer and induces migration and tumor growth*. J Thorac Oncol, 2011. **6**(12): p. 1984-92.
112. Rivera, M.P., A.C. Mehta, and M.M. Wahidi, *Establishing the diagnosis of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines*. Chest, 2013. **143**(5 Suppl): p. e142S-e165S.
113. Hammerschmidt, S. and H. Wirtz, *Lung cancer: current diagnosis and treatment*. Dtsch Arztebl Int, 2009. **106**(49): p. 809-18; quiz 819-20.
114. Arroliga, A.C. and R.A. Matthay, *The role of bronchoscopy in lung cancer*. Clin Chest Med, 1993. **14**(1): p. 87-98.

115. Zaric, B., et al., *Advanced bronchoscopic techniques in diagnosis and staging of lung cancer*. J Thorac Dis, 2013. **5 Suppl 4**: p. S359-70.
116. Ho, C., et al., *Histologic classification of non-small-cell lung cancer over time: reducing the rates of not-otherwise-specified*. Curr Oncol, 2015. **22**(3): p. e164-70.
117. Thomas, A., et al., *Refining the treatment of NSCLC according to histological and molecular subtypes*. Nat Rev Clin Oncol, 2015. **12**(9): p. 511-26.
118. Carson, J. and D.J. Finley, *Lung cancer staging: an overview of the new staging system and implications for radiographic clinical staging*. Semin Roentgenol, 2011. **46**(3): p. 187-93.
119. Francis, G. and S. Stein, *Circulating cell-free tumour DNA in the management of cancer*. Int J Mol Sci, 2015. **16**(6): p. 14122-42.
120. Stroun, M., et al., *About the possible origin and mechanism of circulating DNA apoptosis and active DNA release*. Clin Chim Acta, 2001. **313**(1-2): p. 139-42.
121. Underhill, H.R., et al., *Fragment length of circulating tumor DNA*. PLoS Genet, 2016. **12**(7): p. e1006162.
122. Heitzer, E., P. Ulz, and J.B. Geigl, *Circulating tumor DNA as a liquid biopsy for cancer*. Clin Chem, 2015. **61**(1): p. 112-23.
123. Qin, Z., et al., *Cell-free circulating tumor DNA in cancer*. Chin J Cancer, 2016. **35**: p. 36.
124. Han, X., J. Wang, and Y. Sun, *Circulating tumor DNA as biomarkers for cancer detection*. Genomics Proteomics Bioinformatics, 2017. **15**(2): p. 59-72.
125. Puszyk, W.M., et al., *Methylation of the imprinted GNAS1 gene in cell-free plasma DNA: equal steady-state quantities of methylated and unmethylated DNA in plasma*. Clin Chim Acta, 2009. **400**(1-2): p. 107-10.
126. Balgouranidou, I., et al., *Assessment of SOX17 DNA methylation in cell free DNA from patients with operable gastric cancer. Association with prognostic variables and survival*. Clin Chem Lab Med, 2013. **51**(7): p. 1505-10.
127. Chimonidou, M., et al., *CST6 promoter methylation in circulating cell-free DNA of breast cancer patients*. Clin Biochem, 2013. **46**(3): p. 235-40.
128. Liggett, T., et al., *Differential methylation of cell-free circulating DNA among patients with pancreatic cancer versus chronic pancreatitis*. Cancer, 2010. **116**(7): p. 1674-80.
129. Balgouranidou, I., et al., *Breast cancer metastasis suppressor-1 promoter methylation in cell-free DNA provides prognostic information in non-small cell lung cancer*. Br J Cancer, 2014. **110**(8): p. 2054-62.
130. Marwitz, S., et al., *Epigenetic modifications of the immune-checkpoint genes CTLA4 and PDCD1 in non-small cell lung cancer results in increased expression*. Clin Epigenetics, 2017. **9**: p. 51.

131. Weiss, G., et al., *Validation of the SHOX2/PTGER4 DNA methylation marker panel for plasma-based discrimination between patients with malignant and nonmalignant lung disease*. J Thorac Oncol, 2017. **12**(1): p. 77-84.
132. Loeb, L.A., et al., *Smoking and lung cancer: an overview*. Cancer Res, 1984. **44**(12 Pt 1): p. 5940-58.
133. Du, Y.X., B.S. Zhou, and J.M. Wu, *Lifestyle factors and human lung cancer: an overview of recent advances (review)*. Int J Oncol, 1998. **13**(3): p. 471-9.
134. Heintz, N.H., Y.M. Janssen-Heininger, and B.T. Mossman, *Asbestos, lung cancers, and mesotheliomas: from molecular approaches to targeting tumor survival pathways*. Am J Respir Cell Mol Biol, 2010. **42**(2): p. 133-9.
135. Hoek, G. and O. Raaschou-Nielsen, *Impact of fine particles in ambient air on lung cancer*. Chin J Cancer, 2014. **33**(4): p. 197-203.
136. Silverman, D.T., *Diesel exhaust causes lung cancer: now what?* Occup Environ Med, 2017. **74**(4): p. 233-234.
137. Islami, F., L.A. Torre, and A. Jemal, *Global trends of lung cancer mortality and smoking prevalence*. Transl Lung Cancer Res, 2015. **4**(4): p. 327-38.
138. Cooper, C.S., *Smoking, lung cancers and their TP53 mutations*. Mutagenesis, 2002. **17**(4): p. 279-80.
139. Anna, L., et al., *Relationship between TP53 tumour suppressor gene mutations and smoking-related bulky DNA adducts in a lung cancer study population from Hungary*. Mutagenesis, 2009. **24**(6): p. 475-80.
140. Porta, M., et al., *Cigarette smoking and K-ras mutations in pancreas, lung and colorectal adenocarcinomas: etiopathogenic similarities, differences and paradoxes*. Mutat Res, 2009. **682**(2-3): p. 83-93.
141. *New research shows genetic chaos caused by smoking. Results provide a "cancer map" that catalogues mutations; findings could revolutionize cancer treatment*. Duke Med Health News, 2010. **16**(3): p. 5-6.
142. Dogan, S., et al., *Molecular epidemiology of EGFR and KRAS mutations in 3,026 lung adenocarcinomas: higher susceptibility of women to smoking-related KRAS-mutant cancers*. Clin Cancer Res, 2012. **18**(22): p. 6169-77.
143. Baykara, O., et al., *Association of epidermal growth factor receptor and K-Ras mutations with smoking history in non-small cell lung cancer patients*. Exp Ther Med, 2013. **5**(2): p. 495-498.
144. Kim, M.H., et al., *Impact of cigarette smoking on response to epidermal growth factor receptor (EGFR)-tyrosine kinase inhibitors in lung adenocarcinoma with activating EGFR mutations*. Lung Cancer, 2014. **84**(2): p. 196-202.

145. Risch, A. and C. Plass, *Lung cancer epigenetics and genetics*. Int J Cancer, 2008. **123**(1): p. 1-7.
146. Harjes, U., *Non-small cell lung cancer: Where there's smoke*. Nat Rev Cancer, 2017. **17**(11): p. 634-635.
147. Fasanelli, F., et al., *Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts*. Nat Commun, 2015. **6**: p. 10192.
148. Zhang, Y., et al., *Smoking-associated DNA methylation markers predict lung cancer incidence*. Clin Epigenetics, 2016. **8**: p. 127.
149. Farsalinos, K.E. and R. Polosa, *Safety evaluation and risk assessment of electronic cigarettes as tobacco cigarette substitutes: a systematic review*. Ther Adv Drug Saf, 2014. **5**(2): p. 67-86.
150. Rodu, B., *The scientific foundation for tobacco harm reduction, 2006-2011*. Harm Reduct J, 2011. **8**: p. 19.
151. Goniewicz, M.L., et al., *Levels of selected carcinogens and toxicants in vapour from electronic cigarettes*. Tob Control, 2014. **23**(2): p. 133-9.
152. Brown, C.J. and J.M. Cheng, *Electronic cigarettes: product characterisation and design considerations*. Tob Control, 2014. **23 Suppl 2**: p. ii4-10.
153. Goniewicz, M.L., E.O. Lingas, and P. Hajek, *Patterns of electronic cigarette use and user beliefs about their safety and benefits: an internet survey*. Drug Alcohol Rev, 2013. **32**(2): p. 133-40.
154. Jensen, R.P., et al., *Hidden formaldehyde in e-cigarette aerosols*. N Engl J Med, 2015. **372**(4): p. 392-4.
155. Cai, H. and C. Wang, *Graphical review: The redox dark side of e-cigarettes; exposure to oxidants and public health concerns*. Redox Biol, 2017. **13**: p. 402-406.
156. Canistro, D., et al., *E-cigarettes induce toxicological effects that can raise the cancer risk*. Sci Rep, 2017. **7**(1): p. 2028.
157. Bekki, K., et al., *Carbonyl compounds generated from electronic cigarettes*. Int J Environ Res Public Health, 2014. **11**(11): p. 11192-200.
158. Kaleta, D., *Recommendations of the World Health Organization Tobacco Control Research Team regarding electronic nicotine delivery devices*. Przegl Lek, 2010. **67**(10): p. 1088-90.
159. Health, J.H.B.S.o.P. 2017; Available from: <http://globaltobaccocontrol.org/e-cigarette/country-laws-regulating-e-cigarettes> (abgerufen am 15.01.2018).
160. Glasser, A.M., et al., *Overview of electronic nicotine delivery systems: a systematic review*. Am J Prev Med, 2017. **52**(2): p. e33-e66.
161. Vergleichsportal, E.-Z.-I.n.d.I.-u. *Aufbau der E-Zigarette*. Available from: <http://e-zigarette-info.net/e-zigarette-aufbau/> (abgerufen am 23.10.2017).

162. DSMZ. *A549*. Available from: https://www.dsmz.de/catalogues/details/culture/ACC-107.html?tx_dsmzresources_pi5%5BreturnPid%5D=192 (abgerufen am 23.10.2017).
163. Duden, *In Silico Datenanalyse*. Available from: https://www.duden.de/rechtschreibung/PC_Personal_Computer (abgerufen am 18.12.2018)
164. Extension, L. *Lungenkrebs*. Available from: <http://www.lifeextension.com/Protocols/Cancer/Lung-Cancer/Page-01> (abgerufen am 10.02.2018).
165. Utah, U.o. *Detecting circulating tumor DNA*. Available from: <http://uofuhealth.utah.edu/huntsman/labs/varley/research/detecting-circulating.php> (abgerufen am 10.02.2018).
166. Illumina. *HM 450K*. Available from: https://support.illumina.com/array/array_kits/infinium_humanmethylation450_beadchip_kit.html (abgerufen am 10.02.2018).
167. Illumina. *NextSeq 500*. Available from: <https://www.illumina.com/configuration/presscenter/multimedia/images/nextseq-500-front.html> (abgerufen am 10.02.2018).
168. Qlucore. *Qlucore Omics Explorer*. Available from: <https://www.qlucore.com/> (abgerufen am 14.11.2017).
169. Kindlon, A.R.Q.a.N. *Bedtools*. Available from: <http://bedtools.readthedocs.io/en/latest/> (abgerufen am 14.11.2017).
170. Warton, K., K.L. Mahon, and G. Samimi, *Methylated circulating tumor DNA in blood: power in cancer prognosis and response*. *Endocr Relat Cancer*, 2016. **23**(3): p. R157-71.
171. Malentacchi, F., et al., *Influence of storage conditions and extraction methods on the quantity and quality of circulating cell-free DNA (ccfDNA): the SPIDIA-DNAplas External Quality Assessment experience*. *Clin Chem Lab Med*, 2015. **53**(12): p. 1935-42.
172. AG, A.J. *PME free circulating DNA extraction kit*. Available from: https://www.analytik-jena.de/fileadmin/content/pdf_life_science/Manual/Manual_PME_free_circulating_DNA_Extraction_Kit.pdf (abgerufen am 26.10.2017).
173. Gallagher, S.R. and P.R. Desjardins, *Quantitation of DNA and RNA with absorption and fluorescence spectroscopy*. *Curr Protoc Hum Genet*, 2007. **Appendix 3**: p. Appendix 3D.
174. „Thermo Fisher Scientific“, U., *Qubit*. Available from: <https://www.thermofisher.com/de/de/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit.html> (abgerufen am 18.02.2018)
175. Panaro, N.J., et al., *Evaluation of DNA fragment sizing and quantification by the agilent 2100 bioanalyzer*. *Clin Chem*, 2000. **46**(11): p. 1851-3.
176. Patterson, K., et al., *DNA methylation: bisulphite modification and analysis*. *J Vis Exp*, 2011(56).

177. Frommer, M., et al., *A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands*. Proc Natl Acad Sci U S A, 1992. **89**(5): p. 1827-31.
178. Li, Y. and T.O. Tollefsbol, *DNA methylation detection: bisulfite genomic sequencing analysis*. Methods Mol Biol, 2011. **791**: p. 11-21.
179. NEB, N.E.B. *Bisulfitkonvertierung*. Available from: http://www.neb-online.de/wp-content/uploads/2015/04/NEB_epigenetik_bisulfit3.jpg (abgerufen an 26.10.2017).
180. Clark, S.J., et al., *High sensitivity mapping of methylated cytosines*. Nucleic Acids Res, 1994. **22**(15): p. 2990-7.
181. Niedringhaus, T.P., et al., *Landscape of next-generation sequencing technologies*. Anal Chem, 2011. **83**(12): p. 4327-41.
182. van Dijk, E.L., et al., *Ten years of next-generation sequencing technology*. Trends Genet, 2014. **30**(9): p. 418-26.
183. Voelkerding, K.V., S.A. Dames, and J.D. Durtschi, *Next-generation sequencing: from basic research to diagnostics*. Clin Chem, 2009. **55**(4): p. 641-58.
184. Illumina. *Sequencing Platforms*. Available from: <https://www.illumina.com/systems/sequencing-platforms/comparison-tool.html> (abgerufen am 26.10.2017).
185. Yuan Lu, Y.S., Wesley Warren and Ronald Walter, *Next generation sequencing in aquatic models*. Available from: <https://www.intechopen.com/books/next-generation-sequencing-advances-applications-and-challenges/next-generation-sequencing-in-aquatic-models> (abgerufen am 26.10.2017).
186. Tanaka, K. and A. Okamoto, *Degradation of DNA by bisulfite treatment*. Bioorg Med Chem Lett, 2007. **17**(7): p. 1912-5.
187. Kurdyukov, S. and M. Bullock, *DNA methylation analysis: choosing the right method*. Biology (Basel), 2016. **5**(1).
188. Board, R.E., et al., *Isolation and extraction of circulating tumor DNA from patients with small cell lung cancer*. Ann N Y Acad Sci, 2008. **1137**: p. 98-107.
189. Biosciences, S. *Accel-NGS Methyl-Seq DNA Library Kit*. Available from: <https://swiftbiosci.com/products/accel-ngs-methyl-seq-dna-library-kit/accel-ngs-methyl-seq-dna-library-kit/#exp-workflow> (abgerufen am 27.10.2017).
190. K, S.C.K.L.C., *NGS library preparation for balanced, comprehensive methylome coverage from low-input quantities*. Nature Methods | Application Note 2015. **12**.
191. Westburg. *Swift Biosciences NGS Kits*. Available from: <https://www.westburg.eu/swift-biosciences-ngs-kits> (abgerufen am 26.10.2017).
192. Trivedi, U.H., et al., *Quality control of next-generation sequencing data without a reference*. Front Genet, 2014. **5**: p. 111.
193. S., A., *A quality control tool for high throughput sequence data*. January 2014.

194. Xiaolei, W., et al., *Kmer-indexer: A Fast K-mer Indexing Program*. Stud Health Technol Inform, 2015. **216**: p. 1083.
195. Yang, X., S. Aluru, and K.S. Dorman, *Repeat-aware modeling and correction of short read errors*. BMC Bioinformatics, 2011. **12 Suppl 1**: p. S52.
196. Del Fabbro, C., et al., *An extensive evaluation of read trimming effects on Illumina NGS data analysis*. PLoS One, 2013. **8**(12): p. e85024.
197. Li, H. and N. Homer, *A survey of sequence alignment algorithms for next-generation sequencing*. Brief Bioinform, 2010. **11**(5): p. 473-83.
198. Chatterjee, A., et al., *Comparison of alignment software for genome-wide bisulphite sequence data*. Nucleic Acids Res, 2012. **40**(10): p. e79.
199. Krueger, F. and S.R. Andrews, *Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications*. Bioinformatics, 2011. **27**(11): p. 1571-2.
200. Otto, C., P.F. Stadler, and S. Hoffmann, *Fast and sensitive mapping of bisulfite-treated sequencing data*. Bioinformatics, 2012. **28**(13): p. 1698-704.
201. Kretzmer, H., C. Otto, and S. Hoffmann, *BAT: Bisulfite Analysis Toolkit: BAT is a toolkit to analyze DNA methylation sequencing data accurately and reproducibly. It covers standard processing and analysis steps from raw read mapping up to annotation data integration and calculation of correlating DMRs*. F1000Res, 2017. **6**: p. 1490.
202. Sims, D., et al., *Sequencing depth and coverage: key considerations in genomic analyses*. Nat Rev Genet, 2014. **15**(2): p. 121-32.
203. R Development Core Team, R.F.f.S.C., *R: A language and environment for statistical computing*.
204. Foundation, P.S. *Python*. Available from: <https://www.python.org/> (abgerufen am 14.11.2017).
205. Foundation, T.P. *Perl*. Available from: <https://www.perl.org/> (abgerufen am 14.11.2017).
206. Cock, P.J., et al., *The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants*. Nucleic Acids Res, 2010. **38**(6): p. 1767-71.
207. Bioinformatics, B. *FastQC*. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (abgerufen am 26.10.2017).
208. Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing reads*. EMBnet.journal, 2011. **17**(1).
209. Hoffmann, S., et al., *Fast mapping of short sequences with mismatches, insertions and deletions using index structures*. PLoS Comput Biol, 2009. **5**(9): p. e1000502.
210. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. **25**(16): p. 2078-9.

211. Github. *BamUtil*. Available from: <https://github.com/statgen/bamUtil> (abgerufen am 14.11.2017).
212. Leipzig, I.f.I. *BAT*. Available from: <http://www.bioinf.uni-leipzig.de/Software/BAT/> (abgerufen am 26.10.2017).
213. Quinlan, A.R., *BEDTools: The Swiss-Army Tool for Genome Feature Analysis*. *Curr Protoc Bioinformatics*, 2014. **47**: p. 11 12 1-34.
214. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features*. *Bioinformatics*, 2010. **26**(6): p. 841-2.
215. Kent, W.J., et al., *The human genome browser at UCSC*. *Genome Res*, 2002. **12**(6): p. 996-1006.
216. Bernhart, S.H., et al., *Changes of bivalent chromatin coincide with increased expression of developmental genes in cancer*. *Sci Rep*, 2016. **6**: p. 37393.
217. Roche. *SeqCap Epi Enrichment Kit*. Available from: <https://roche-biochem.jp/products/NimbleGen/10-learn/20-technology/40-automation.html> (abgerufen am 14.11.2017).
218. Jiang, H., et al., *Classification, predictive modelling, and statistical analysis of cancer data (a)*. *Cancer Inform*, 2014. **13**(Suppl 2): p. 1-3.
219. Gomez, I., et al., *Supervised discretization can discover risk groups in cancer survival analysis*. *Comput Methods Programs Biomed*, 2016. **136**: p. 11-9.
220. Sarkar, M. and T.Y. Leong, *Application of K-nearest neighbors algorithm on breast cancer diagnosis problem*. *Proc AMIA Symp*, 2000: p. 759-63.
221. Wang, X., *A fast exact k-nearest neighbors algorithm for high dimensional search using k-means clustering and triangle inequality*. *Proc Int Jt Conf Neural Netw*, 2012. **43**(6): p. 2351-2358.
222. Pontil, M. and A. Verri, *Properties of support vector machines*. *Neural Comput*, 1998. **10**(4): p. 955-74.
223. Raudys, S., *How good are support vector machines?* *Neural Netw*, 2000. **13**(1): p. 17-9.
224. Clark, D.E. and M. el-Taha, *Generation of correlated logistic-normal random variates for medical decision trees*. *Methods Inf Med*, 1998. **37**(3): p. 235-8.
225. McCabe, G.P., Jr. and M.L. Samuels, *Random censoring and dendritic trees*. *Biometrics*, 1977. **33**(1): p. 69-84.
226. Farsalinos, K.E., et al., *Characteristics, perceived side effects and benefits of electronic cigarette use: a worldwide survey of more than 19,000 consumers*. *Int J Environ Res Public Health*, 2014. **11**(4): p. 4356-73.

227. Lucchiari, C., et al., *Benefits of e-cigarettes among heavy smokers undergoing a lung cancer screening program: randomized controlled trial protocol*. JMIR Res Protoc, 2016. **5**(1): p. e21.
228. Marwitz, S., et al., *Downregulation of the TGFbeta pseudoreceptor BAMBI in non-small cell lung cancer enhances TGFbeta signaling and invasion*. Cancer Res, 2016. **76**(13): p. 3785-801.
229. Kreindler, J.L., et al., *Inhibition of chloride secretion in human bronchial epithelial cells by cigarette smoke extract*. Am J Physiol Lung Cell Mol Physiol, 2005. **288**(5): p. L894-902.
230. Promega. *ApoONE® Homogeneous Caspase-3/7 Assay*. Available from: <https://www.promega.com/-/media/files/resources/protocols/technical-bulletins/0/apoone-homogeneous-caspase-3-7-assay-protocol.pdf> (abgerufen am 14.11.2017).
231. Promega. *CellTiter-Blue® Cell Viability Assay*. Available from: <https://www.promega.com/-/media/files/resources/protocols/technical-bulletins/101/celltiter-blue-cell-viability-assay-protocol.pdf?la=en> (abgerufen am 14.11.2017).
232. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics*. Nat Rev Genet, 2009. **10**(1): p. 57-63.
233. O'Neil, D., H. Glowatz, and M. Schlumpberger, *Ribosomal RNA depletion for efficient use of RNA-seq capacity*. Curr Protoc Mol Biol, 2013. **Chapter 4**: p. Unit 4 19.
234. Abernathy, J. and K. Overturf, *Comparison of ribosomal RNA removal methods for transcriptome sequencing workflows in teleost fish*. Anim Biotechnol, 2016. **27**(1): p. 60-5.
235. Benes V., B.J., Doyle K, *Ribo-Zero Gold Kit: improved RNA-seq results after removal of cytoplasmic and mitochondrial ribosomal RNA*. Nature Methods, 2011.
236. Inc., A.B.M.a. *RNA-Seq.*; Available from: https://www.abmgood.com/marketing/knowledge_base/next_generation_sequencing_experimental_design.php (abgerufen am 23.10.2017).
237. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data*. Bioinformatics, 2010. **26**(1): p. 139-40.
238. Zhang, B., S. Kirov, and J. Snoddy, *WebGestalt: an integrated system for exploring gene sets in various biological contexts*. Nucleic Acids Res, 2005. **33**(Web Server issue): p. W741-8.
239. Kirov, S., et al., *Functional annotation of differentially regulated gene set using WebGestalt: a gene set predictive of response to ipilimumab in tumor biopsies*. Methods Mol Biol, 2014. **1101**: p. 31-42.

240. Wang, J., et al., *WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit*. Nucleic Acids Res, 2017.
241. Buenrostro, J.D., et al., *Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position*. Nat Methods, 2013. **10**(12): p. 1213-8.
242. Buenrostro, J.D., et al., *ATAC-seq: a method for assaying chromatin accessibility genome-wide*. Curr Protoc Mol Biol, 2015. **109**: p. 21 29 1-9.
243. Institute, B. *Picard*. Available from: <http://broadinstitute.github.io/picard/faq.html> (abgerufen am 23.10.2017).
244. Ebbert, M.T., et al., *Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches*. BMC Bioinformatics, 2016. **17 Suppl 7**: p. 239.
245. Foundation, P.S. *MACS2*. Available from: <https://pypi.python.org/pypi/MACS2> (abgerufen am 14.11.2017).
246. Illumina. *EPIC BC*. Available from: <https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/humanmethylationepic-data-sheet-1070-2015-008.pdf> (abgerufen am 26.10.2017).
247. Bibikova, M., et al., *High density DNA methylation array with single CpG site resolution*. Genomics, 2011. **98**(4): p. 288-95.
248. Du, P., et al., *Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis*. BMC Bioinformatics, 2010. **11**: p. 587.
249. Pidsley, R., et al., *A data-driven approach to preprocessing Illumina 450K methylation array data*. BMC Genomics, 2013. **14**: p. 293.
250. Carless, M.A., *Determination of DNA methylation levels using Illumina HumanMethylation450 BeadChips*. Methods Mol Biol, 2015. **1288**: p. 143-92.
251. Illumina. *GenomeStudio*. Available from: <https://www.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html> (abgerufen am 14.11.2017).
252. Meyer, C.A. and X.S. Liu, *Identifying and mitigating bias in next-generation sequencing methods for chromatin biology*. Nat Rev Genet, 2014. **15**(11): p. 709-21.
253. Almeida, R.D., et al., *Neuroprotection by BDNF against glutamate-induced apoptotic cell death is mediated by ERK and PI3-kinase pathways*. Cell Death Differ, 2005. **12**(10): p. 1329-43.
254. Ding, B., et al., *A positive feedback loop of phosphodiesterase 3 (PDE3) and inducible cAMP early repressor (ICER) leads to cardiomyocyte apoptosis*. Proc Natl Acad Sci U S A, 2005. **102**(41): p. 14771-6.

255. Ding, B., et al., *Functional role of phosphodiesterase 3 in cardiomyocyte apoptosis: implication in heart failure*. *Circulation*, 2005. **111**(19): p. 2469-2476.
256. Feng, Y., et al., *Homeodomain-interacting protein kinase 2 (HIPK2): a promising target for anti-cancer therapies*. *Oncotarget*, 2017. **8**(12): p. 20452-20461.
257. Lee, M.G., et al., *XAF1 directs apoptotic switch of p53 signaling through activation of HIPK2 and ZNF313*. *Proc Natl Acad Sci U S A*, 2014. **111**(43): p. 15532-7.
258. Pescador, N., et al., *Identification of a functional hypoxia-responsive element that regulates the expression of the egl nine homologue 3 (egln3/phd3) gene*. *Biochem J*, 2005. **390**(Pt 1): p. 189-97.
259. Petersen, A.A., et al., *Brain-derived neurotrophic factor inhibits apoptosis and dopamine-induced free radical production in striatal neurons but does not prevent cell death*. *Brain Res Bull*, 2001. **56**(3-4): p. 331-5.
260. Sarma, S.N., et al., *Induction of apoptosis in human leukemia cells through the production of reactive oxygen species and activation of HMOX1 and Noxa by benzene, toluene, and o-xylene*. *Toxicology*, 2011. **280**(3): p. 109-17.
261. Xie, L., et al., *Oxygen-regulated beta(2)-adrenergic receptor hydroxylation by EGLN3 and ubiquitylation by pVHL*. *Sci Signal*, 2009. **2**(78): p. ra33.
262. Huang, S.S. and J.S. Huang, *TGF-beta control of cell proliferation*. *J Cell Biochem*, 2005. **96**(3): p. 447-62.
263. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2016*. *CA Cancer J Clin*, 2016. **66**(1): p. 7-30.
264. Crowley, E., et al., *Liquid biopsy: monitoring cancer-genetics in the blood*. *Nat Rev Clin Oncol*, 2013. **10**(8): p. 472-84.
265. Stahl, D.L., K.M. Richard, and T.J. Papadimos, *Complications of bronchoscopy: A concise synopsis*. *Int J Crit Illn Inj Sci*, 2015. **5**(3): p. 189-95.
266. Birkenkamp-Demtroder, K., et al., *Monitoring, treatment, response and metastatic relapse in advanced bladder cancer by liquid biopsy analysis*. *Eur Urol*, 2017.
267. Dawson, S.J., et al., *Analysis of circulating tumor DNA to monitor metastatic breast cancer*. *N Engl J Med*, 2013. **368**(13): p. 1199-209.
268. Ilie, M., et al., *Current challenges for detection of circulating tumor cells and cell-free circulating nucleic acids, and their characterization in non-small cell lung carcinoma patients. What is the best blood substrate for personalized medicine?* *Ann Transl Med*, 2014. **2**(11): p. 107.
269. Perez-Callejo, D., et al., *Liquid biopsy based biomarkers in non-small cell lung cancer for diagnosis and treatment monitoring*. *Transl Lung Cancer Res*, 2016. **5**(5): p. 455-465.

270. Perez-Ramirez, C., et al., *Liquid biopsy in early stage lung cancer*. *Transl Lung Cancer Res*, 2016. **5**(5): p. 517-524.
271. Ilie, M. and P. Hofman, *Pros: Can tissue biopsy be replaced by liquid biopsy?* *Transl Lung Cancer Res*, 2016. **5**(4): p. 420-3.
272. Cheng, F., L. Su, and C. Qian, *Circulating tumor DNA: a promising biomarker in the liquid biopsy of cancer*. *Oncotarget*, 2016. **7**(30): p. 48832-48841.
273. Jahr, S., et al., *DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells*. *Cancer Res*, 2001. **61**(4): p. 1659-65.
274. Koproski, M.S., F.A. Benko, and C.D. Gocke, *Circulating RNA as a tumor marker: detection of 5T4 mRNA in breast and lung cancer patient serum*. *Ann N Y Acad Sci*, 2001. **945**: p. 172-8.
275. Shaw, J.A., et al., *Circulating tumor cells and plasma DNA analysis in patients with indeterminate early or metastatic breast cancer*. *Biomark Med*, 2011. **5**(1): p. 87-91.
276. Thakur, B.K., et al., *Double-stranded DNA in exosomes: a novel biomarker in cancer detection*. *Cell Res*, 2014. **24**(6): p. 766-9.
277. Pantel, K. and M.R. Speicher, *The biology of circulating tumor cells*. *Oncogene*, 2016. **35**(10): p. 1216-24.
278. Cai, X., et al., *Accessing genetic information with liquid biopsies*. *Trends Genet*, 2015. **31**(10): p. 564-75.
279. Lavon, I., et al., *Serum DNA can define tumor-specific genetic and epigenetic markers in gliomas of various grades*. *Neuro Oncol*, 2010. **12**(2): p. 173-80.
280. Wurtmann, E.J. and S.L. Wolin, *RNA under attack: cellular handling of RNA damage*. *Crit Rev Biochem Mol Biol*, 2009. **44**(1): p. 34-49.
281. Diehl, F., et al., *Circulating mutant DNA to assess tumor dynamics*. *Nat Med*, 2008. **14**(9): p. 985-90.
282. Devonshire, A.S., et al., *Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification*. *Anal Bioanal Chem*, 2014. **406**(26): p. 6499-512.
283. Perez-Barrios, C., et al., *Comparison of methods for circulating cell-free DNA isolation using blood from cancer patients: impact on biomarker testing*. *Transl Lung Cancer Res*, 2016. **5**(6): p. 665-672.
284. Rolfo, C., et al., *Liquid biopsies in lung cancer: the new ambrosia of researchers*. *Biochim Biophys Acta*, 2014. **1846**(2): p. 539-46.
285. Canto, A., et al., *Lung cancer and pleural effusion. Clinical significance and study of pleural metastatic locations*. *Chest*, 1985. **87**(5): p. 649-52.

286. Warton, K., et al., *Methylation-capture and Next-Generation Sequencing of free circulating DNA from human plasma*. BMC Genomics, 2014. **15**: p. 476.
287. de Bruin, E.C., N. McGranahan, and C. Swanton, *Analysis of intratumor heterogeneity unravels lung cancer evolution*. Mol Cell Oncol, 2015. **2**(3): p. e985549.
288. Lu, Y., et al., *Methylated DNA/RNA in body fluids as biomarkers for lung cancer*. Biol Proced Online, 2017. **19**: p. 2.
289. Uchida, J., et al., *Diagnostic accuracy of noninvasive genotyping of EGFR in lung cancer patients by deep sequencing of plasma cell-free DNA*. Clin Chem, 2015. **61**(9): p. 1191-6.
290. Couraud, S., et al., *Noninvasive diagnosis of actionable mutations by deep sequencing of circulating free DNA in lung cancer from never-smokers: a proof-of-concept study from BioCAST/IFCT-1002*. Clin Cancer Res, 2014. **20**(17): p. 4613-24.
291. Newman, A.M., et al., *An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage*. Nat Med, 2014. **20**(5): p. 548-54.
292. Sharma, S., T.K. Kelly, and P.A. Jones, *Epigenetics in cancer*. Carcinogenesis, 2010. **31**(1): p. 27-36.
293. Schmidt, B., et al., *Quantification of cell-free mSHOX2 Plasma DNA for therapy monitoring in advanced stage non-small cell (NSCLC) and small-cell lung cancer (SCLC) patients*. PLoS One, 2015. **10**(2): p. e0118195.
294. Murn, J., et al., *Prostaglandin E2 regulates B cell proliferation through a candidate tumor suppressor, Ptger4*. J Exp Med, 2008. **205**(13): p. 3091-103.
295. Schneider, K.U., et al., *Correlation of SHOX2 gene amplification and DNA methylation in lung cancer tumors*. BMC Cancer, 2011. **11**: p. 102.
296. Lei, K.F., et al., *Serpina5 interacts with KHDRBS3 and FBXO32 in gastric cancer cells*. Oncol Rep, 2011. **26**(5): p. 1115-20.
297. Abe, M., et al., *Disordered expression of HOX genes in human non-small cell lung cancer*. Oncol Rep, 2006. **15**(4): p. 797-802.
298. Bhatlekar, S., J.Z. Fields, and B.M. Boman, *HOX genes and their role in the development of human cancers*. J Mol Med (Berl), 2014. **92**(8): p. 811-23.
299. Niitsu, H., et al., *KRAS mutation leads to decreased expression of regulator of calcineurin 2, resulting in tumor proliferation in colorectal cancer*. Oncogenesis, 2016. **5**(8): p. e253.
300. Gadea, G., et al., *DOCK10-mediated Cdc42 activation is necessary for amoeboid invasion of melanoma cells*. Curr Biol, 2008. **18**(19): p. 1456-65.
301. Hawkes, J.E., et al., *Report of a novel OCA2 gene mutation and an investigation of OCA2 variants on melanoma risk in a familial melanoma pedigree*. J Dermatol Sci, 2013. **69**(1): p. 30-7.

302. Jackson, A.L., B. Zhou, and W.Y. Kim, *HIF, hypoxia and the role of angiogenesis in non-small cell lung cancer*. *Expert Opin Ther Targets*, 2010. **14**(10): p. 1047-57.
303. Kurogane, Y., et al., *FGD5 mediates proangiogenic action of vascular endothelial growth factor in human vascular endothelial cells*. *Arterioscler Thromb Vasc Biol*, 2012. **32**(4): p. 988-96.
304. Abe, M., et al., *CpG island methylator phenotype is a strong determinant of poor prognosis in neuroblastomas*. *Cancer Res*, 2005. **65**(3): p. 828-34.
305. Lin, L., et al., *SLC transporters as therapeutic targets: emerging opportunities*. *Nat Rev Drug Discov*, 2015. **14**(8): p. 543-60.
306. Black, P.N., *Primary sequence of the Escherichia coli fadL gene encoding an outer membrane protein required for long-chain fatty acid transport*. *J Bacteriol*, 1991. **173**(2): p. 435-42.
307. DeBerardinis, R.J. and N.S. Chandel, *Fundamentals of cancer metabolism*. *Sci Adv*, 2016. **2**(5): p. e1600200.
308. Lisztwan, J., et al., *The von Hippel-Lindau tumor suppressor protein is a component of an E3 ubiquitin-protein ligase activity*. *Genes Dev*, 1999. **13**(14): p. 1822-33.
309. Scheffner, M., et al., *The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53*. *Cell*, 1993. **75**(3): p. 495-505.
310. Ching, M.S., et al., *Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders*. *Am J Med Genet B Neuropsychiatr Genet*, 2010. **153B**(4): p. 937-47.
311. Oksenberg, N. and N. Ahituv, *The role of AUTS2 in neurodevelopment and human evolution*. *Trends Genet*, 2013. **29**(10): p. 600-8.
312. O'Hurley, G., et al., *Analysis of the human prostate-specific proteome defined by transcriptomics and antibody-based profiling identifies TMEM79 and ACOXL as two putative, diagnostic markers in prostate cancer*. *PLoS One*, 2015. **10**(8): p. e0133449.
313. Fagerholm, R., et al., *TP53-based interaction analysis identifies cis-eQTL variants for TP53BP2, FBXO28, and FAM53A that associate with survival and treatment outcome in breast cancer*. *Oncotarget*, 2017. **8**(11): p. 18381-18398.
314. Nicodemus-Johnson, J., et al., *DNA methylation in lung cells is associated with asthma endotypes and genetic risk*. *JCI Insight*, 2016. **1**(20): p. e90151.
315. Harwood, K.V., *Non-small cell lung cancer: an overview of diagnosis, staging, and treatment*. *Semin Oncol Nurs*, 1996. **12**(4): p. 285-94.
316. Sommer, G. and B. Stieltjes, *Magnetic resonance imaging for staging of non-small-cell lung cancer-technical advances and unmet needs*. *J Thorac Dis*, 2015. **7**(7): p. 1098-102.

317. Yousefi-Koma, A., M. Panah-Moghaddam, and V. Kalff, *The utility of metabolic imaging by 18F-FDG PET/CT in lung cancer: impact on diagnosis and staging*. Tanaffos, 2013. **12**(1): p. 16-25.
318. Alegria-Torres, J.A., A. Baccarelli, and V. Bollati, *Epigenetics and lifestyle*. Epigenomics, 2011. **3**(3): p. 267-77.
319. Lim, U. and M.A. Song, *Dietary and lifestyle factors of DNA methylation*. Methods Mol Biol, 2012. **863**: p. 359-76.
320. Gallego-Fabrega, C., et al., *DNA methylation levels are highly correlated between pooled samples and averaged values when analysed using the Infinium HumanMethylation450 BeadChip array*. Clin Epigenetics, 2015. **7**: p. 78.
321. Rakyan, V.K., et al., *Epigenome-wide association studies for common human diseases*. Nat Rev Genet, 2011. **12**(8): p. 529-41.
322. Eberth, S., et al., *Epigenetic regulation of CD44 in Hodgkin and non-Hodgkin lymphoma*. BMC Cancer, 2010. **10**: p. 517.
323. Bediaga, N.G., et al., *DNA methylation epigenotypes in breast cancer molecular subtypes*. Breast Cancer Res, 2010. **12**(5): p. R77.
324. Lugthart, S., et al., *Aberrant DNA hypermethylation signature in acute myeloid leukemia directed by EVI1*. Blood, 2011. **117**(1): p. 234-41.
325. Teumer, A., et al., *Comparison of genotyping using pooled DNA samples (allelotyping) and individual genotyping using the affymetrix genome-wide human SNP array 6.0*. BMC Genomics, 2013. **14**: p. 506.
326. Sham, P., et al., *DNA Pooling: a tool for large-scale association studies*. Nat Rev Genet, 2002. **3**(11): p. 862-71.
327. Norton, N., et al., *DNA pooling as a tool for large-scale association studies in complex traits*. Ann Med, 2004. **36**(2): p. 146-52.
328. Tsuji, J. and Z. Weng, *Evaluation of preprocessing, mapping and postprocessing algorithms for analyzing whole genome bisulfite sequencing data*. Brief Bioinform, 2016. **17**(6): p. 938-952.
329. Bodi, K., et al., *Comparison of commercially available target enrichment methods for next-generation sequencing*. J Biomol Tech, 2013. **24**(2): p. 73-86.
330. Meldrum, C., M.A. Doyle, and R.W. Tothill, *Next-generation sequencing for cancer diagnostics: a practical perspective*. Clin Biochem Rev, 2011. **32**(4): p. 177-95.
331. Garcia-Garcia, G., et al., *Assessment of the latest NGS enrichment capture methods in clinical context*. Sci Rep, 2016. **6**: p. 20948.
332. Bundesministerium für Familie, S., Frauen und Jugend, *Protection of Young Persons Act § 10*. 2016. Available from: <https://www.bmfsfj.de/blob/90278/7a45fcfc6a8f06a8fcca28df3274e947/juschg-englisch-2016-data.pdf> (abgerufen am 18.02.2018)

333. Commission, E.; Available from: http://europa.eu/rapid/press-release_MEMO-14-134_en.htm (abgerufen am 15.01.2018).
334. Barraza, L.F., et al., *Regulations and policies regarding e-cigarettes*. *Cancer*, 2017. **123**(16): p. 3007-3014.
335. Tabuchi, T., et al., *Awareness and use of electronic cigarettes and heat-not-burn tobacco products in Japan*. *Addiction*, 2016. **111**(4): p. 706-13.
336. Izzotti, A., et al., *Alterations of gene expression in skin and lung of mice exposed to light and cigarette smoke*. *FASEB J*, 2004. **18**(13): p. 1559-61.
337. Meng, Q.R., et al., *Gene expression profiling in lung tissues from mice exposed to cigarette smoke, lipopolysaccharide, or smoke plus lipopolysaccharide by inhalation*. *Inhal Toxicol*, 2006. **18**(8): p. 555-68.
338. Morissette, M.C., et al., *Impact of cigarette smoke on the human and mouse lungs: a gene-expression comparison study*. *PLoS One*, 2014. **9**(3): p. e92498.
339. Lin, S., et al., *Comparison of the transcriptional landscapes between human and mouse tissues*. *Proc Natl Acad Sci U S A*, 2014. **111**(48): p. 17224-9.
340. Fujino, N., et al., *Gene expression profiles of alveolar type II cells of chronic obstructive pulmonary disease: a case-control study*. *BMJ Open*, 2012. **2**(6).
341. Kobayashi, K., et al., *Identification of genes whose expression is upregulated in lung adenocarcinoma cells in comparison with type II alveolar cells and bronchiolar epithelial cells in vivo*. *Oncogene*, 2004. **23**(17): p. 3089-96.
342. Stegemann-Koniszewski, S., et al., *Alveolar type II epithelial cells contribute to the anti-influenza A virus response in the lung by integrating pathogen- and microenvironment-derived signals*. *MBio*, 2016. **7**(3).
343. Cooper, J.R., et al., *Long term culture of the A549 cancer cell line promotes multilamellar body formation and differentiation towards an alveolar type II pneumocyte phenotype*. *PLoS One*, 2016. **11**(10): p. e0164438.
344. Mao, P., et al., *Human alveolar epithelial type II cells in primary culture*. *Physiol Rep*, 2015. **3**(2).
345. Fuchs, S., et al., *Differentiation of human alveolar epithelial cells in primary culture: morphological characterization and synthesis of caveolin-1 and surfactant protein-C*. *Cell Tissue Res*, 2003. **311**(1): p. 31-45.
346. Wang, F., et al., *Heterogeneity of claudin expression by alveolar epithelial cells*. *Am J Respir Cell Mol Biol*, 2003. **29**(1): p. 62-70.
347. Beck, G.C., et al., *Heterogeneity in lipopolysaccharide responsiveness of endothelial cells identified by gene expression profiling: role of transcription factors*. *Clin Exp Immunol*, 2006. **143**(3): p. 523-33.

348. Peixoto, A., et al., *Quantification of multiple gene expression in individual cells*. Genome Res, 2004. **14**(10A): p. 1938-47.
349. Leek, J.T. and J.D. Storey, *Capturing heterogeneity in gene expression studies by surrogate variable analysis*. PLoS Genet, 2007. **3**(9): p. 1724-35.
350. Civelek, M. and A.J. Lusis, *Systems genetics approaches to understand complex traits*. Nat Rev Genet, 2014. **15**(1): p. 34-48.
351. Consortium, S.M.-I., *A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium*. Nat Biotechnol, 2014. **32**(9): p. 903-14.
352. Peixoto, L., et al., *How data analysis affects power, reproducibility and biological insight of RNA-seq studies in complex datasets*. Nucleic Acids Res, 2015. **43**(16): p. 7664-74.
353. Yuan, G.C., et al., *Challenges and emerging directions in single-cell analysis*. Genome Biol, 2017. **18**(1): p. 84.
354. Pelkmans, L., *Cell Biology. Using cell-to-cell variability - a new era in molecular biology*. Science, 2012. **336**(6080): p. 425-6.
355. Stevens, J.B., et al., *Diverse system stresses: common mechanisms of chromosome fragmentation*. Cell Death Dis, 2011. **2**: p. e178.
356. Abdallah, B.Y., et al., *Single cell heterogeneity: why unstable genomes are incompatible with average profiles*. Cell Cycle, 2013. **12**(23): p. 3640-9.
357. Altschuler, S.J. and L.F. Wu, *Cellular heterogeneity: do differences make a difference?* Cell, 2010. **141**(4): p. 559-63.
358. Ching, T., S. Huang, and L.X. Garmire, *Power analysis and sample size estimation for RNA-Seq differential expression*. RNA, 2014. **20**(11): p. 1684-96.
359. Hart, S.N., et al., *Calculating sample size estimates for RNA sequencing data*. J Comput Biol, 2013. **20**(12): p. 970-8.
360. Li, C.I., et al., *Sample size calculation for differential expression analysis of RNA-seq data under Poisson distribution*. Int J Comput Biol Drug Des, 2013. **6**(4): p. 358-75.
361. Hasin, Y., M. Seldin, and A. Lusis, *Multi-omics approaches to disease*. Genome Biol, 2017. **18**(1): p. 83.
362. Chen, Y., A.T. Lun, and G.K. Smyth, *From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline*. F1000Res, 2016. **5**: p. 1438.
363. Vaes, E., M. Khan, and P. Mombaerts, *Statistical analysis of differential gene expression relative to a fold change threshold on NanoString data of mouse odorant receptor genes*. BMC Bioinformatics, 2014. **15**: p. 39.

364. Jin, B., Y. Li, and K.D. Robertson, *DNA methylation: superior or subordinate in the epigenetic hierarchy?* *Genes Cancer*, 2011. **2**(6): p. 607-17.
365. Lim, A.S., et al., *24-hour rhythms of DNA methylation and their relation with rhythms of RNA expression in the human dorsolateral prefrontal cortex.* *PLoS Genet*, 2014. **10**(11): p. e1004792.
366. Ambatipudi, S., et al., *Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study.* *Epigenomics*, 2016. **8**(5): p. 599-618.
367. Spira, A., et al., *Effects of cigarette smoke on the human airway epithelial cell transcriptome.* *Proc Natl Acad Sci U S A*, 2004. **101**(27): p. 10143-8.
368. Rowell, T.R., et al., *Flavored e-cigarette liquids reduce proliferation and viability in the CALU3 airway epithelial cell line.* *Am J Physiol Lung Cell Mol Physiol*, 2017. **313**(1): p. L52-L66.
369. "Dipse". *Liquid-Zusammensetzung.* Available from: https://www.amazon.de/Dipse-Premium-elektrische-elektronische-Nikotinfrei/dp/B005YZBLE0/ref=sr_1_2?s=drugstore&ie=UTF8&qid=1511382574&sr=1-2&keywords=dipse+liquid (abgerufen am 22.11.2017).
370. El-Telbany, A. and P.C. Ma, *Cancer genes in lung cancer: racial disparities: are there any?* *Genes Cancer*, 2012. **3**(7-8): p. 467-80.
371. Knuutinen, A., et al., *Smoking and skin: a study of the physical qualities and histology of skin in smokers and non-smokers.* *Acta Derm Venereol*, 2002. **82**(1): p. 36-40.
372. Urbanska, M., G. Nowak, and E. Florek, *Cigarette smoking and its influence on skin aging.* *Przegl Lek*, 2012. **69**(10): p. 1111-4.
373. Silverstein, P., *Smoking and wound healing.* *Am J Med*, 1992. **93**(1A): p. 22S-24S.
374. McDaniel, J.C. and K.K. Browning, *Smoking, chronic wound healing, and implications for evidence-based practice.* *J Wound Ostomy Continence Nurs*, 2014. **41**(5): p. 415-23; quiz E1-2.
375. Knuutinen, A., et al., *Smoking affects collagen synthesis and extracellular matrix turnover in human skin.* *Br J Dermatol*, 2002. **146**(4): p. 588-94.
376. Jorgensen, L.N., et al., *Less collagen production in smokers.* *Surgery*, 1998. **123**(4): p. 450-5.
377. Cheruel, F., M. Jarlier, and H. Sancho-Garnier, *Effect of cigarette smoke on gustatory sensitivity, evaluation of the deficit and of the recovery time-course after smoking cessation.* *Tob Induc Dis*, 2017. **15**: p. 15.
378. Jaakkola, J.J. and M. Gissler, *Maternal smoking in pregnancy, fetal development, and childhood asthma.* *Am J Public Health*, 2004. **94**(1): p. 136-40.
379. Anblagan, D., et al., *Maternal smoking during pregnancy and fetal organ growth: a magnetic resonance imaging study.* *PLoS One*, 2013. **8**(7): p. e67223.

380. Ghio, A.J., et al., *Particulate matter in cigarette smoke alters iron homeostasis to produce a biological effect*. Am J Respir Crit Care Med, 2008. **178**(11): p. 1130-8.
381. Ghio, A.J. and E.D. Hilborn, *Indices of iron homeostasis correlate with airway obstruction in an NHANES III cohort*. Int J Chron Obstruct Pulmon Dis, 2017. **12**: p. 2075-2084.
382. Beyth, S., et al., *Cigarette smoking is associated with a lower concentration of CD105(+) bone marrow progenitor cells*. Bone Marrow Res, 2015. **2015**: p. 914935.
383. Abate, M., et al., *Cigarette smoking and musculoskeletal disorders*. Muscles Ligaments Tendons J, 2013. **3**(2): p. 63-9.
384. Ma, L., et al., *Cigarette smoke increases apoptosis in the gastric mucosa: role of epidermal growth factor*. Digestion, 1999. **60**(5): p. 461-8.
385. Carnevali, S., et al., *Cigarette smoke extract induces oxidative stress and apoptosis in human lung fibroblasts*. Am J Physiol Lung Cell Mol Physiol, 2003. **284**(6): p. L955-63.
386. Chen, H., et al., *Cigarette smoke extract induces apoptosis of rat alveolar Type II cells via the PLTP/TGF-beta1/Smad2 pathway*. Int Immunopharmacol, 2015. **28**(1): p. 707-14.
387. Aoshiba, K., J. Tamaoki, and A. Nagai, *Acute cigarette smoke exposure induces apoptosis of alveolar macrophages*. Am J Physiol Lung Cell Mol Physiol, 2001. **281**(6): p. L1392-401.
388. Talhout, R., et al., *Hazardous compounds in tobacco smoke*. Int J Environ Res Public Health, 2011. **8**(2): p. 613-28.
389. Wright, S.C., et al., *Nicotine inhibition of apoptosis suggests a role in tumor promotion*. FASEB J, 1993. **7**(11): p. 1045-51.
390. Hakki, A., et al., *Nicotine inhibition of apoptosis in murine immune cells*. Exp Biol Med (Maywood), 2001. **226**(10): p. 947-53.
391. Hakki, A., H. Friedman, and S. Pross, *Nicotine modulation of apoptosis in human coronary artery endothelial cells*. Int Immunopharmacol, 2002. **2**(10): p. 1403-9.
392. Wang, C., et al., *Nicotine suppresses apoptosis by regulating alpha7nAChR/Prx1 axis in oral precancerous lesions*. Oncotarget, 2017. **8**(43): p. 75065-75075.
393. West, K.A., et al., *Rapid Akt activation by nicotine and a tobacco carcinogen modulates the phenotype of normal human airway epithelial cells*. J Clin Invest, 2003. **111**(1): p. 81-90.
394. Dasgupta, P., et al., *Nicotine inhibits apoptosis induced by chemotherapeutic drugs by up-regulating XIAP and survivin*. Proc Natl Acad Sci U S A, 2006. **103**(16): p. 6332-7.

395. Maneckjee, R. and J.D. Minna, *Opioid and nicotine receptors affect growth regulation of human lung cancer cell lines*. Proc Natl Acad Sci U S A, 1990. **87**(9): p. 3294-8.
396. Mai, H., et al., *A functional role for nicotine in Bcl2 phosphorylation and suppression of apoptosis*. J Biol Chem, 2003. **278**(3): p. 1886-91.
397. Jin, Z., et al., *Nicotine induces multi-site phosphorylation of Bad in association with suppression of apoptosis*. J Biol Chem, 2004. **279**(22): p. 23837-44.
398. Xin, M. and X. Deng, *Nicotine inactivation of the proapoptotic function of Bax through phosphorylation*. J Biol Chem, 2005. **280**(11): p. 10781-9.
399. Lexikon, C. *Aroma*. Available from: <http://www.chemie.de/lexikon/Aroma.html> (abgerufen am 13.02.2018).
400. Yang, L., et al., *Nicotine alters morphology and function of retinal pigment epithelial cells in mice*. Toxicol Pathol, 2010. **38**(4): p. 560-7.
401. Fan, J., et al., *Prenatal and lactation nicotine exposure affects morphology and function of brown adipose tissue in male rat offspring*. Ultrastruct Pathol, 2016. **40**(5): p. 288-95.
402. Lee, H.J., et al., *Effects of nicotine on proliferation, cell cycle, and differentiation in immortalized and malignant oral keratinocytes*. J Oral Pathol Med, 2005. **34**(7): p. 436-43.
403. Dwyer, J.B., S.C. McQuown, and F.M. Leslie, *The dynamic effects of nicotine on the developing brain*. Pharmacol Ther, 2009. **122**(2): p. 125-39.
404. Smith, A.M., L.P. Dwoskin, and J.R. Pauly, *Early exposure to nicotine during critical periods of brain development: Mechanisms and consequences*. J Pediatr Biochem, 2010. **1**(2): p. 125-141.
405. Colombo, S.F., et al., *Biogenesis, trafficking and up-regulation of nicotinic ACh receptors*. Biochem Pharmacol, 2013. **86**(8): p. 1063-73.
406. Evan, G.I. and K.H. Vousden, *Proliferation, cell cycle and apoptosis in cancer*. Nature, 2001. **411**(6835): p. 342-8.
407. Szende, B. and E. Tyihak, *Effect of formaldehyde on cell proliferation and death*. Cell Biol Int, 2010. **34**(12): p. 1273-82.
408. Wilson, N., et al., *Should e-cigarette use be included in indoor smoking bans?* Bull World Health Organ, 2017. **95**(7): p. 540-541.
409. Amiri, A., et al., *Formaldehyde exposure during pregnancy*. MCN Am J Matern Child Nurs, 2015. **40**(3): p. 180-5.
410. Haffner, M.J., et al., *Formaldehyde exposure and its effects during pregnancy: recommendations for laboratory attendance based on available data*. Clin Anat, 2015. **28**(8): p. 972-9.

411. Wagner, N.J., M. Camerota, and C. Propper, *Prevalence and perceptions of electronic cigarette use during pregnancy*. *Matern Child Health J*, 2017. **21**(8): p. 1655-1661.
412. Belaid, A., et al., *Autophagy plays a critical role in the degradation of active RHOA, the control of cell cytokinesis, and genomic stability*. *Cancer Res*, 2013. **73**(14): p. 4311-22.
413. Kozutsumi, Y., et al., *Sphingolipids involved in the induction of multinuclear cell formation*. *Biochim Biophys Acta*, 2002. **1582**(1-3): p. 138-43.
414. Gao, S.G., et al., *The study on the chromosome aneuploidy in human lung cancer*. *Zhonghua Yi Xue Za Zhi*, 2007. **87**(24): p. 1701-3.
415. Romeo, M.S., et al., *Chromosomal abnormalities in non-small cell lung carcinomas and in bronchial epithelia of high-risk smokers detected by multi-target interphase fluorescence in situ hybridization*. *J Mol Diagn*, 2003. **5**(2): p. 103-12.
416. Ginsberg, S.S., et al., *Giant cell carcinoma of the lung*. *Cancer*, 1992. **70**(3): p. 606-10.
417. Kong, L., et al., *Squamous cell carcinoma of lung associated with osteoclast-like giant cells: report of a case*. *Int J Clin Exp Pathol*, 2015. **8**(9): p. 11823-5.
418. Matsukuma, S., et al., *Numerous osteoclast-like giant cells in metastases from lung adenocarcinoma, but absent from primary tumor*. *Thorac Cancer*, 2014. **5**(4): p. 354-7.
419. Weihua, Z., et al., *Formation of solid tumors by a single multinucleated cancer cell*. *Cancer*, 2011. **117**(17): p. 4092-9.
420. Decleves, X., et al., *Role of ABC transporters in the chemoresistance of human gliomas*. *Curr Cancer Drug Targets*, 2006. **6**(5): p. 433-45.
421. Huang, Y. and W. Sadee, *Membrane transporters and channels in chemoresistance and -sensitivity of tumor cells*. *Cancer Lett*, 2006. **239**(2): p. 168-82.
422. Li, Q. and Y. Shu, *Role of solute carriers in response to anticancer drugs*. *Mol Cell Ther*, 2014. **2**: p. 15.
423. Rama, A.R., et al., *ABC transporters as differentiation markers in glioblastoma cells*. *Mol Biol Rep*, 2014. **41**(8): p. 4847-51.
424. Ihlefeld, K., et al., *Upregulation of ABC transporters contributes to chemoresistance of sphingosine 1-phosphate lyase-deficient fibroblasts*. *J Lipid Res*, 2015. **56**(1): p. 60-9.
425. Hamblett, K.J., et al., *SLC46A3 Is Required to Transport Catabolites of Noncleavable Antibody Maytansine Conjugates from the Lysosome to the Cytoplasm*. *Cancer Res*, 2015. **75**(24): p. 5329-40.

426. Li, X. and U. Francke, *Assignment of the gene SLC1A2 coding for the human glutamate transporter EAAT2 to human chromosome 11 bands p13-p12*. Cytogenet Cell Genet, 1995. **71**(3): p. 212-3.
427. Deng, X., et al., *Association study of polymorphisms in the neutral amino acid transporter genes SLC1A4, SLC1A5 and the glycine transporter genes SLC6A5, SLC6A9 with schizophrenia*. BMC Psychiatry, 2008. **8**: p. 58.
428. Shinmura, K., et al., *CD44-SLC1A2 fusion transcripts in primary colorectal cancer*. Pathol Oncol Res, 2015. **21**(3): p. 759-64.
429. Xiang, C., et al., *LncRNA-SLC6A9-5:2: A potent sensitizer in 131I-resistant papillary thyroid carcinoma with PARP-1 induction*. Oncotarget, 2017. **8**(14): p. 22954-22967.
430. Hoffman, N.E., et al., *SLC25A23 augments mitochondrial Ca(2)(+) uptake, interacts with MCU, and induces oxidative stress-mediated cell death*. Mol Biol Cell, 2014. **25**(6): p. 936-47.
431. Kuwano, K., *Epithelial cell apoptosis and lung remodeling*. Cell Mol Immunol, 2007. **4**(6): p. 419-29.
432. Martin, T.R., et al., *Apoptosis and epithelial injury in the lungs*. Proc Am Thorac Soc, 2005. **2**(3): p. 214-20.
433. Gomes, M., et al., *The role of inflammation in lung cancer*. Adv Exp Med Biol, 2014. **816**: p. 1-23.
434. King, P.T., *Inflammation in chronic obstructive pulmonary disease and its role in cardiovascular disease and lung cancer*. Clin Transl Med, 2015. **4**(1): p. 68.
435. Murdoch, J.R. and C.M. Lloyd, *Chronic inflammation and asthma*. Mutat Res, 2010. **690**(1-2): p. 24-39.

9. Anhang

9.1 Entwicklung und Validierung des Plasma *Panel*s

Auswählen der CpG *Loci* für das Plasma *Panel*

a. Filtern nach Chromosom

Die Chromosome M, X und Y wurden verworfen, die Befehle lauteten:

```
grep -v "chrM" <Name>.bedgraph | grep -v "chrX" | grep -v "chrY" > <Name>.ohneMXY.bedgraph  
cut -f1 <Name>ohneMXY.bedgraph | sort | uniq
```

b. Annotation mit dem „HG19“ Referenzgenom

```
less gencode.v19.only.genes.bed | perl -ane 'if($F[5] eq "+"){F[1]=F[1]-1500}else{F[2] =F[2]+1500}; print "$F[0]\t$F[1]\t$F[2]\t$F[3]\t$F[4]\t$F[5]\n"' > gencode.v19.only.genes.TSS_1500nt.bed
```

```
bedtools intersect -wa -wb -a <Name>ohneMXY.bedgraph -b gencode.v19.only.genes.TSS_1500nt.bed
```

c. Auswählen der CpG *Loci*, die von WGBS und HM 450K detektiert wurden

```
bedtools intersect -wa -wb -a <WGBS_data>.bedgraph -b <450K_BeadChip_data>.bed | perl -ane 'if(($F[3]>0 && $F[7]>0) || ($F[3]<0 && $F[7]<0)){print $_}' > overlap_WGBS_450K_BeadChip.bed
```

```
bedtools intersect -wa -wb -a overlap_WGBS_450K_BeadChip.bed -b gencode.v19.only.genes.TSS_1500nt.bed | cut -f1-4,8,9,13 > overlap_WGBS_450K_BeadChip_gencode.v19.bed
```

d. Auswählen der differentiell methylierten CpG Cluster

Hierbei wurden CpG *Loci* ausgewählt, die innerhalb eines Clusters bestehend aus mindestens zwei weiteren differentiell methylierten CpG *Loci* lagen. Alle CpG *Loci* des Clusters waren entweder hypo- oder hypermethyliert. Der Abstand zwischen den CpG *Loci* betrug zwei bis 20 Nukleotide.

```
less <Name>ohneMXY.bedgraph | sort -k10,10 | bedtools groupby -g 7,8,9,10,11,12 -c 1,2,3,1 -o collapse,collapse,collapse,count | perl -ane 'if($F[-1]>=3){print $_}' | perl -ane '@chr=split(/,,$F[6]); @start=split(/,,$F[7]); @end=split(/,,$F[8]); for($i=0; $i<$F[-1]; $i++){print "$chr[$i]\t$start[$i]\t$end[$i]\t$F[0]\t$F[1]\t$F[2]\t$F[3]\t$F[4]\t$F[5]\n"' > <Name>ohneMXY_mind3CpG_annotation.bedgraph
```



```
perl CpG_cluster_Swetlana --min 2 --max 20 --in <Name>ohneMXY_mind3CpG_annotation.bedgraph  
| grep protein > <Name>ohneMXY_mind3CpG_3diffCpG.bedgraph
```

```
less <Name>ohneMXY_mind3CpG_3diffCpG.bedgraph | bedtools groupby -g 7,11 -c 3,1,2,3 -o  
collapse,distinct,min,max | perl -ane 'print "$F[3]\t$F[4]\t$F[5]\t$F[2];$F[0]\n"' > <Name>ohneMXY  
_mind3CpG_3diffCpG_sortiert.bedgraph
```

```
bedtools intersect -wa -wb -a <Name>ohneMXY_mind3CpG_3diffCpG_sortiert.bedgraph -b  
<Diff_fiel>.bedgraph | bedtools groupby -g 1,2,3,4 -c 8 -o mean | perl -ane '$a=abs($F[4]); chomp  
$_; print "$_ \t$a\n"' | sort -k6,6n | tail -150 > <Name>ohneMXY_mind3CpG_3diffCpG_sortiert_beste  
_150_regionen.bedgraph
```

e. Auswählen der Positionen mit der höchsten differentiellen DNA-Methylierung

```
bedtools intersect -v -a <Name>ohneMXY.bedgraph -b <Name>ohneMXY_mind3CpG_3diffCpG_  
sortiert_beste_150_regionen.bedgraph | bedtools intersect -wa -wb -a stdin -b gencode.v19.only  
.genes.TSS_1500nt_ohnechrM.bed | grep protein | perl -ane '$a=abs($F[5]); chomp $_; print  
"$_ \t$a\n"' | sort -V -k13,13n | cut -f1,2,3,10,13 | tail -100 > <Name>ohneMXY_die_besten_einzel_  
cpg.bedgraph
```

9.2 multi-OMICS Studie zur Untersuchung von Risiken der E-Zigaretten

RNA-Seq.: Identifikation differentiell exprimierter Transkripte

Die Identifikation differentiell exprimierter Transkripte erfolgte mit der „edgeR“ Software. Hierfür wurde zunächst eine „Count_table.txt“-Datei generiert.

```
for i in *.count; do n=`echo $i | sed 's/.*JDK-/' | sed 's/_R.*//'; echo $n >$n.tmp; cut -f2 $i | grep -v  
Count >>$n.tmp; done  
cut -f1 170407_SNK268_A_L_JDK-48_R1.alignment.count | sed 's/\s.*//>genes  
paste genes *.tmp >count_table.txt  
rm *.tmp
```

Die „Count_table.txt“-Datei wurde im nächsten Schritt mit der „edgeR“ Software ausgewertet. Dabei wurden „run.edgeR.sh“ und „run.edgeR.R“ Skripte verwendet.

a. run.edgeR.sh

```
ctr=`head -1 count_table.txt | sed 's/\t/\n/g' | grep -n ${1} | perl -ane '$F[0]=~s/:.*/; print "$F[0],"'`  
case=`head -1 count_table.txt | sed 's/\t/\n/g' | grep -n ${2} | perl -ane '$F[0]=~s/:.*/; print "$F[0],"'`  
d=`echo "1,"${ctr}${case} | sed 's/,,$/'`  
f=`echo ${1}"_"${2}`  
cut -f${d} count_table.txt >${f}.txt
```

```
Rscript run.edgeR.R ${f} ${1} ${2}
```

b. runedgeR.R

```
inname <- paste0(args[1], ".txt")  
  
# Input-Files lesen:  
data <- as.matrix(read.table(inname, row.names=1))  
countData <- as.matrix(read.table(inname, header=T, row.names=1))  
annotation <- read.table("/sfs/fs5/home-sh/sukkm527/RNA-Seq_auswertung/  
Diff_expressionannotation.bed", header=F, col.names=c("chr", "start", "end", "ENSG", "gene", "strand"))  
  
# Gruppen definieren:  
g <- as.vector(data[1,])  
  
# Größe der Library für die Normalisierung ermitteln:  
libSizes <- as.vector(colSums(countData))  
  
# edgeR:  
d <- DGEList(counts=countData, group=g)  
keep <- rowSums(cpm(d)>0.5) >= 1  
d <- d[keep, , keep.lib.sizes=FALSE]  
d <- calcNormFactors(d)  
d <- estimateCommonDisp(d)  
d <- estimateTagwiseDisp(d)  
de.com <- exactTest(d, pair=c(args[2],args[3]))  
results <- topTags(de.com, n=length(data[, 1]))  
  
# „diff.expr.table“ annotieren:  
out <- data.frame(as.matrix(results$table))  
out$ENSG <- rownames(out)  
rownames(out) <- NULL  
out <- merge(out, annotation, by="ENSG")  
  
# das finale File mit den Informationen über die ermittelte, differentielle Expression schreiben:  
write.table(out, file=paste0(args[1], "_edgeR.out"), sep="\t", row.names=F, quote=F)
```

Um besonders starke Effekte des Rauchens zu untersuchen, wurden nur die Transkripte ausgewählt und visualisiert, die einen „log2 fold-change“-Wert von mindestens 0,25 hatten und bei beiden Zelllinien differentiell exprimiert waren. Die Befehle in „R“ waren:

```
sig.diff <- read.table('<Name>_edgeR_p_0.05.txt', header=T)
fresh <- read.table('<Name>_edgeR.out', header=T)
data <- merge(sig.diff, fresh, by="ENSG")
max(data[data$logFC.x < 0, 2])
min(data[data$logFC.x > 0, 2])
data$color <- "black"
data$color[which(data$logFC.x < 0 & data$logFC.y < -0.25)] <- "blue"
data$color[which(data$logFC.x > 0 & data$logFC.y > 0.25)] <- "red"
data$color <- factor(data$color, levels=c("black","red","blue"))

ggplot(data, aes(x=logFC.x, y=logFC.y)) + geom_point(aes(colour=color)) + theme_classic() +
  xlab('Differentially Expressed Genes in A549 (treatment)') + ylab('Differentially Expressed Genes in
  AECII (treatment)') + coord_fixed(ratio=1) + geom_hline(yintercept=0, color="grey") +
  geom_vline(xintercept=0, color="grey") + coord_cartesian(xlim=c(-6,6), ylim=c(-6,6)) +
  scale_color_manual(name="logFC", values=c("black" = "black", "red" = "red", "blue" = "blue"), labels =
  c("NA", "Upregulated Genes", "Downregulated Genes"))

write.table(Both_cell_lines_<treatment>.txt', data[data$color!="black",], col.names=T, row.names=F,
  quote=F)
```

ATAC-Seq.: Integrative Analyse in „R“

Als erstes wurden die Ergebnisse der ATAC-Sequenzierung direkt in „R“ geladen und nach Genen gefiltert, deren Chromatinstatus nach der Behandlung mit CSE bzw. eCSE Veränderungen von mindestens 10% zeigte.

```
ctr <- read.table('<Name>.txt', header=F, col.names=c('ENSG','ctr_nt.ATAC'))
treat <- read.table('<Name>.txt', header=F, col.names=c('ENSG','treat_nt.ATAC'))
data <- merge(ctr, treat, by='ENSG')
data$change <- "NA"
data$change[which(data$treat_nt.ATAC > (data$ctr_nt.ATAC+350))] <- "gain"
data$change[which(data$treat_nt.ATAC < (data$ctr_nt.ATAC-350))] <- "loss"
data$change.perc <- (data$treat_nt.ATAC - data$ctr_nt.ATAC)/35
```

Dann wurden die Ergebnisse der RNA-Sequenzierung direkt in „R“ geladen. Nachfolgend analysierte „R“, welches der differentiell exprimierten Gene eine Veränderung des

Chromatinstatus von mindestens 10% zeigte. Die Ergebnisse wurden schließlich in einer „TXT“-Datei gespeichert.

```
expr <- read.table('<Name>.txt', header=T)
all <- merge(data, expr, by='ENSG')
write.table(all, '<Name>.txt', col.names=T, row.names=F, quote=F)
```

9.3 Tabellen

Tab. 7: Plasma Panel (638 differentiell methylierte Regionen) **A.** 352 Regionen, die einen malignen Tumor der Lunge detektieren **B.** 253 Regionen, welche die häufigsten Lungenkarzinomentitäten (Adeno- und Plattenepithelkarzinom) voneinander unterscheiden **C.** 33 prognostisch relevante CpG Loci.

| A. Lungenkarzinom- oder Lungengewebe? | | | |
|---------------------------------------|-----------|-----------|--------------------------------------|
| Chromosom | Start | Ende | Methode |
| chr1 | 57955028 | 57955174 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr1 | 193191311 | 193191476 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr10 | 85985699 | 85985859 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr10 | 110084584 | 110084739 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr10 | 130860130 | 130860266 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr11 | 57798784 | 57798925 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr11 | 57948628 | 57948769 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr11 | 58034333 | 58034464 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr11 | 59634150 | 59634282 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr11 | 59824464 | 59824610 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr11 | 131547241 | 131547390 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr12 | 7818556 | 7818707 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr12 | 111016497 | 111016626 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr12 | 128899218 | 128899363 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr14 | 58064847 | 58064987 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr14 | 88621354 | 88621491 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr15 | 63349114 | 63349271 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr15 | 87516105 | 87516269 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr16 | 20055126 | 20055299 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr16 | 34255462 | 34255596 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr17 | 46799562 | 46799708 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr2 | 2019860 | 2020000 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr2 | 66671403 | 66671543 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr2 | 118569132 | 118569281 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr2 | 155089787 | 155089940 | zfDNA (WGBS), OP-Präparate (HM 450K) |

| | | | |
|-------|-----------|-----------|--------------------------------------|
| chr20 | 29960903 | 29961067 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr21 | 31987899 | 31988061 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr3 | 159175958 | 159176133 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr4 | 77703312 | 77703460 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr5 | 5033914 | 5034062 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr5 | 5568513 | 5568662 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr5 | 141130550 | 141130698 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr6 | 5132810 | 5132954 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr6 | 20877268 | 20877408 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr6 | 27648240 | 27648385 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr6 | 55956239 | 55956395 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr7 | 149112327 | 149112464 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr8 | 54798658 | 54798811 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr1 | 2198804 | 2198961 | OP-Präparate (HM 450K) |
| chr1 | 6515521 | 6515702 | OP-Präparate (HM 450K) |
| chr1 | 6520115 | 6520257 | OP-Präparate (HM 450K) |
| chr1 | 19764609 | 19764757 | OP-Präparate (HM 450K) |
| chr1 | 34642324 | 34642455 | OP-Präparate (HM 450K) |
| chr1 | 47694840 | 47694995 | OP-Präparate (HM 450K) |
| chr1 | 50883315 | 50883461 | OP-Präparate (HM 450K) |
| chr1 | 50886707 | 50886857 | OP-Präparate (HM 450K) |
| chr1 | 50886870 | 50887021 | OP-Präparate (HM 450K) |
| chr1 | 79472375 | 79472516 | OP-Präparate (HM 450K) |
| chr1 | 110610821 | 110610964 | OP-Präparate (HM 450K) |
| chr1 | 110611386 | 110611542 | OP-Präparate (HM 450K) |
| chr1 | 110611971 | 110612108 | OP-Präparate (HM 450K) |
| chr1 | 119522559 | 119522707 | OP-Präparate (HM 450K) |
| chr1 | 150595130 | 150595282 | OP-Präparate (HM 450K) |
| chr1 | 153896523 | 153896648 | OP-Präparate (HM 450K) |
| chr1 | 155162673 | 155162808 | OP-Präparate (HM 450K) |
| chr1 | 158324396 | 158324540 | OP-Präparate (HM 450K) |
| chr1 | 158549201 | 158549351 | OP-Präparate (HM 450K) |
| chr1 | 158575697 | 158575854 | OP-Präparate (HM 450K) |
| chr1 | 158736216 | 158736378 | OP-Präparate (HM 450K) |
| chr1 | 159284004 | 159284160 | OP-Präparate (HM 450K) |
| chr1 | 159284209 | 159284363 | OP-Präparate (HM 450K) |
| chr1 | 159682419 | 159682564 | OP-Präparate (HM 450K) |
| chr1 | 160782978 | 160783141 | OP-Präparate (HM 450K) |
| chr1 | 161008634 | 161008907 | OP-Präparate (HM 450K) |
| chr1 | 166039366 | 166039510 | OP-Präparate (HM 450K) |
| chr1 | 175050401 | 175050549 | OP-Präparate (HM 450K) |
| chr1 | 182025968 | 182026117 | OP-Präparate (HM 450K) |
| chr1 | 223948836 | 223948969 | OP-Präparate (HM 450K) |
| chr1 | 248903024 | 248903175 | OP-Präparate (HM 450K) |
| chr10 | 15688934 | 15689073 | OP-Präparate (HM 450K) |

| | | | |
|-------|-----------|-----------|------------------------|
| chr10 | 34405682 | 34405834 | OP-Präparate (HM 450K) |
| chr10 | 44285786 | 44285947 | OP-Präparate (HM 450K) |
| chr10 | 98129672 | 98129823 | OP-Präparate (HM 450K) |
| chr10 | 98129826 | 98129981 | OP-Präparate (HM 450K) |
| chr10 | 102894966 | 102895098 | OP-Präparate (HM 450K) |
| chr10 | 104000754 | 104000901 | OP-Präparate (HM 450K) |
| chr10 | 118892505 | 118892640 | OP-Präparate (HM 450K) |
| chr10 | 118893055 | 118893205 | OP-Präparate (HM 450K) |
| chr10 | 121075240 | 121075380 | OP-Präparate (HM 450K) |
| chr10 | 134598276 | 134598414 | OP-Präparate (HM 450K) |
| chr11 | 627096 | 627254 | OP-Präparate (HM 450K) |
| chr11 | 31826508 | 31826642 | OP-Präparate (HM 450K) |
| chr11 | 40136733 | 40136880 | OP-Präparate (HM 450K) |
| chr11 | 40312591 | 40312717 | OP-Präparate (HM 450K) |
| chr11 | 57005866 | 57005971 | OP-Präparate (HM 450K) |
| chr11 | 57006196 | 57006350 | OP-Präparate (HM 450K) |
| chr11 | 59270333 | 59270463 | OP-Präparate (HM 450K) |
| chr11 | 68166958 | 68167099 | OP-Präparate (HM 450K) |
| chr11 | 69061832 | 69061978 | OP-Präparate (HM 450K) |
| chr11 | 75831643 | 75831777 | OP-Präparate (HM 450K) |
| chr11 | 86085859 | 86085993 | OP-Präparate (HM 450K) |
| chr11 | 123885618 | 123885776 | OP-Präparate (HM 450K) |
| chr11 | 133005846 | 133005990 | OP-Präparate (HM 450K) |
| chr12 | 5918113 | 5918249 | OP-Präparate (HM 450K) |
| chr12 | 21590167 | 21590318 | OP-Präparate (HM 450K) |
| chr12 | 50665695 | 50665835 | OP-Präparate (HM 450K) |
| chr12 | 54423481 | 54423625 | OP-Präparate (HM 450K) |
| chr12 | 54448654 | 54448816 | OP-Präparate (HM 450K) |
| chr12 | 54448836 | 54448981 | OP-Präparate (HM 450K) |
| chr12 | 56329564 | 56329709 | OP-Präparate (HM 450K) |
| chr12 | 62584958 | 62585102 | OP-Präparate (HM 450K) |
| chr12 | 75601386 | 75601538 | OP-Präparate (HM 450K) |
| chr12 | 114847503 | 114847664 | OP-Präparate (HM 450K) |
| chr12 | 126142819 | 126142966 | OP-Präparate (HM 450K) |
| chr12 | 129595318 | 129595466 | OP-Präparate (HM 450K) |
| chr13 | 41593317 | 41593485 | OP-Präparate (HM 450K) |
| chr13 | 42188553 | 42188701 | OP-Präparate (HM 450K) |
| chr13 | 58207783 | 58207923 | OP-Präparate (HM 450K) |
| chr14 | 21623728 | 21623873 | OP-Präparate (HM 450K) |
| chr14 | 37128511 | 37128658 | OP-Präparate (HM 450K) |
| chr14 | 55907221 | 55907370 | OP-Präparate (HM 450K) |
| chr14 | 57274684 | 57274828 | OP-Präparate (HM 450K) |
| chr14 | 57275089 | 57275229 | OP-Präparate (HM 450K) |
| chr14 | 57275889 | 57276137 | OP-Präparate (HM 450K) |
| chr14 | 57276179 | 57276336 | OP-Präparate (HM 450K) |

| | | | |
|-------|----------|----------|------------------------|
| chr14 | 57276449 | 57276590 | OP-Präparate (HM 450K) |
| chr14 | 57277149 | 57277295 | OP-Präparate (HM 450K) |
| chr14 | 57278109 | 57278251 | OP-Präparate (HM 450K) |
| chr14 | 57284449 | 57284596 | OP-Präparate (HM 450K) |
| chr14 | 60977778 | 60977928 | OP-Präparate (HM 450K) |
| chr14 | 60978086 | 60978221 | OP-Präparate (HM 450K) |
| chr14 | 77769608 | 77769754 | OP-Präparate (HM 450K) |
| chr15 | 42749674 | 42749956 | OP-Präparate (HM 450K) |
| chr15 | 45409243 | 45409393 | OP-Präparate (HM 450K) |
| chr15 | 72520560 | 72520691 | OP-Präparate (HM 450K) |
| chr15 | 86233150 | 86233290 | OP-Präparate (HM 450K) |
| chr15 | 89920745 | 89920964 | OP-Präparate (HM 450K) |
| chr15 | 89922266 | 89922403 | OP-Präparate (HM 450K) |
| chr16 | 23850031 | 23850175 | OP-Präparate (HM 450K) |
| chr16 | 29086204 | 29086356 | OP-Präparate (HM 450K) |
| chr16 | 31580915 | 31581053 | OP-Präparate (HM 450K) |
| chr16 | 48592619 | 48592755 | OP-Präparate (HM 450K) |
| chr16 | 59789141 | 59789301 | OP-Präparate (HM 450K) |
| chr16 | 59790110 | 59790246 | OP-Präparate (HM 450K) |
| chr16 | 66613021 | 66613174 | OP-Präparate (HM 450K) |
| chr16 | 66613201 | 66613354 | OP-Präparate (HM 450K) |
| chr16 | 76342543 | 76342697 | OP-Präparate (HM 450K) |
| chr17 | 750165 | 750314 | OP-Präparate (HM 450K) |
| chr17 | 31689711 | 31689863 | OP-Präparate (HM 450K) |
| chr17 | 32613223 | 32613361 | OP-Präparate (HM 450K) |
| chr17 | 35299524 | 35299661 | OP-Präparate (HM 450K) |
| chr17 | 55951984 | 55952129 | OP-Präparate (HM 450K) |
| chr17 | 59532229 | 59532369 | OP-Präparate (HM 450K) |
| chr17 | 67536233 | 67536383 | OP-Präparate (HM 450K) |
| chr17 | 72619477 | 72619639 | OP-Präparate (HM 450K) |
| chr18 | 20714264 | 20714392 | OP-Präparate (HM 450K) |
| chr18 | 21596836 | 21596981 | OP-Präparate (HM 450K) |
| chr18 | 61143869 | 61144219 | OP-Präparate (HM 450K) |
| chr18 | 61144261 | 61144399 | OP-Präparate (HM 450K) |
| chr19 | 9609321 | 9609462 | OP-Präparate (HM 450K) |
| chr19 | 18761488 | 18761632 | OP-Präparate (HM 450K) |
| chr19 | 19625186 | 19625348 | OP-Präparate (HM 450K) |
| chr19 | 42600201 | 42600339 | OP-Präparate (HM 450K) |
| chr19 | 48285227 | 48285396 | OP-Präparate (HM 450K) |
| chr19 | 53038895 | 53039056 | OP-Präparate (HM 450K) |
| chr2 | 2336353 | 2336494 | OP-Präparate (HM 450K) |
| chr2 | 3642551 | 3642688 | OP-Präparate (HM 450K) |
| chr2 | 43496147 | 43496286 | OP-Präparate (HM 450K) |
| chr2 | 45171739 | 45171891 | OP-Präparate (HM 450K) |
| chr2 | 45232352 | 45232491 | OP-Präparate (HM 450K) |

| | | | |
|-------|-----------|-----------|------------------------|
| chr2 | 63280990 | 63281212 | OP-Präparate (HM 450K) |
| chr2 | 63281305 | 63281462 | OP-Präparate (HM 450K) |
| chr2 | 63282625 | 63282788 | OP-Präparate (HM 450K) |
| chr2 | 63282935 | 63283081 | OP-Präparate (HM 450K) |
| chr2 | 63283888 | 63284202 | OP-Präparate (HM 450K) |
| chr2 | 73021274 | 73021424 | OP-Präparate (HM 450K) |
| chr2 | 100516804 | 100516939 | OP-Präparate (HM 450K) |
| chr2 | 105069122 | 105069275 | OP-Präparate (HM 450K) |
| chr2 | 105086941 | 105087083 | OP-Präparate (HM 450K) |
| chr2 | 124920570 | 124920719 | OP-Präparate (HM 450K) |
| chr2 | 127453687 | 127453859 | OP-Präparate (HM 450K) |
| chr2 | 162280362 | 162280605 | OP-Präparate (HM 450K) |
| chr2 | 176964058 | 176964200 | OP-Präparate (HM 450K) |
| chr2 | 176964383 | 176964599 | OP-Präparate (HM 450K) |
| chr2 | 176964651 | 176964790 | OP-Präparate (HM 450K) |
| chr2 | 176980760 | 176980908 | OP-Präparate (HM 450K) |
| chr2 | 176980985 | 176981133 | OP-Präparate (HM 450K) |
| chr2 | 176982263 | 176982420 | OP-Präparate (HM 450K) |
| chr2 | 176986185 | 176986321 | OP-Präparate (HM 450K) |
| chr2 | 176988868 | 176989016 | OP-Präparate (HM 450K) |
| chr2 | 176989280 | 176989410 | OP-Präparate (HM 450K) |
| chr2 | 177014478 | 177014625 | OP-Präparate (HM 450K) |
| chr2 | 177014869 | 177015014 | OP-Präparate (HM 450K) |
| chr2 | 177027372 | 177027510 | OP-Präparate (HM 450K) |
| chr2 | 177029509 | 177029683 | OP-Präparate (HM 450K) |
| chr2 | 192113945 | 192114079 | OP-Präparate (HM 450K) |
| chr2 | 200326645 | 200326782 | OP-Präparate (HM 450K) |
| chr2 | 208989171 | 208989315 | OP-Präparate (HM 450K) |
| chr2 | 223161815 | 223161963 | OP-Präparate (HM 450K) |
| chr2 | 223162956 | 223163101 | OP-Präparate (HM 450K) |
| chr2 | 223163250 | 223163396 | OP-Präparate (HM 450K) |
| chr20 | 5282874 | 5283037 | OP-Präparate (HM 450K) |
| chr20 | 29979773 | 29979904 | OP-Präparate (HM 450K) |
| chr20 | 60119429 | 60119586 | OP-Präparate (HM 450K) |
| chr21 | 38076799 | 38076947 | OP-Präparate (HM 450K) |
| chr21 | 38076967 | 38077102 | OP-Präparate (HM 450K) |
| chr21 | 38077182 | 38077314 | OP-Präparate (HM 450K) |
| chr21 | 38082537 | 38082677 | OP-Präparate (HM 450K) |
| chr3 | 128202420 | 128202557 | OP-Präparate (HM 450K) |
| chr3 | 147106484 | 147106639 | OP-Präparate (HM 450K) |
| chr3 | 147108444 | 147108594 | OP-Präparate (HM 450K) |
| chr3 | 147108764 | 147108915 | OP-Präparate (HM 450K) |
| chr3 | 147109715 | 147109852 | OP-Präparate (HM 450K) |
| chr3 | 147113649 | 147113806 | OP-Präparate (HM 450K) |
| chr3 | 147113839 | 147113992 | OP-Präparate (HM 450K) |

| | | | |
|------|-----------|-----------|------------------------|
| chr3 | 147127584 | 147127733 | OP-Präparate (HM 450K) |
| chr3 | 147128049 | 147128198 | OP-Präparate (HM 450K) |
| chr3 | 147131253 | 147131426 | OP-Präparate (HM 450K) |
| chr3 | 160167891 | 160168052 | OP-Präparate (HM 450K) |
| chr3 | 178907589 | 178907716 | OP-Präparate (HM 450K) |
| chr3 | 181421475 | 181421609 | OP-Präparate (HM 450K) |
| chr3 | 181421626 | 181421778 | OP-Präparate (HM 450K) |
| chr4 | 16639102 | 16639249 | OP-Präparate (HM 450K) |
| chr4 | 16773604 | 16773747 | OP-Präparate (HM 450K) |
| chr4 | 16795659 | 16795823 | OP-Präparate (HM 450K) |
| chr4 | 16862004 | 16862148 | OP-Präparate (HM 450K) |
| chr4 | 38871251 | 38871401 | OP-Präparate (HM 450K) |
| chr4 | 40336180 | 40336325 | OP-Präparate (HM 450K) |
| chr4 | 81189629 | 81189764 | OP-Präparate (HM 450K) |
| chr4 | 84469489 | 84469634 | OP-Präparate (HM 450K) |
| chr4 | 111550587 | 111550727 | OP-Präparate (HM 450K) |
| chr4 | 111550752 | 111550898 | OP-Präparate (HM 450K) |
| chr4 | 151504646 | 151504792 | OP-Präparate (HM 450K) |
| chr5 | 1879607 | 1879772 | OP-Präparate (HM 450K) |
| chr5 | 5146264 | 5146403 | OP-Präparate (HM 450K) |
| chr5 | 9782073 | 9782214 | OP-Präparate (HM 450K) |
| chr5 | 33737859 | 33738007 | OP-Präparate (HM 450K) |
| chr5 | 140174811 | 140174969 | OP-Präparate (HM 450K) |
| chr5 | 140810843 | 140810977 | OP-Präparate (HM 450K) |
| chr5 | 140811566 | 140811712 | OP-Präparate (HM 450K) |
| chr6 | 28227021 | 28227193 | OP-Präparate (HM 450K) |
| chr6 | 30130783 | 30131058 | OP-Präparate (HM 450K) |
| chr6 | 33141218 | 33141345 | OP-Präparate (HM 450K) |
| chr6 | 34984865 | 34985013 | OP-Präparate (HM 450K) |
| chr6 | 36253031 | 36253158 | OP-Präparate (HM 450K) |
| chr6 | 50791127 | 50791269 | OP-Präparate (HM 450K) |
| chr6 | 50813472 | 50813737 | OP-Präparate (HM 450K) |
| chr6 | 100905379 | 100905516 | OP-Präparate (HM 450K) |
| chr6 | 100912869 | 100913017 | OP-Präparate (HM 450K) |
| chr6 | 101846889 | 101847032 | OP-Präparate (HM 450K) |
| chr6 | 138866798 | 138866965 | OP-Präparate (HM 450K) |
| chr7 | 811109 | 811265 | OP-Präparate (HM 450K) |
| chr7 | 1596186 | 1596331 | OP-Präparate (HM 450K) |
| chr7 | 2609786 | 2609933 | OP-Präparate (HM 450K) |
| chr7 | 3988693 | 3988828 | OP-Präparate (HM 450K) |
| chr7 | 4786820 | 4787032 | OP-Präparate (HM 450K) |
| chr7 | 7759144 | 7759281 | OP-Präparate (HM 450K) |
| chr7 | 27142023 | 27142169 | OP-Präparate (HM 450K) |
| chr7 | 27204708 | 27204859 | OP-Präparate (HM 450K) |
| chr7 | 27204903 | 27205058 | OP-Präparate (HM 450K) |

| | | | |
|-------|-----------|-----------|------------------------|
| chr7 | 54612258 | 54612404 | OP-Präparate (HM 450K) |
| chr7 | 65617286 | 65617424 | OP-Präparate (HM 450K) |
| chr7 | 96621248 | 96621396 | OP-Präparate (HM 450K) |
| chr7 | 96622543 | 96622774 | OP-Präparate (HM 450K) |
| chr7 | 154087897 | 154088047 | OP-Präparate (HM 450K) |
| chr7 | 154428954 | 154429110 | OP-Präparate (HM 450K) |
| chr8 | 12236159 | 12236321 | OP-Präparate (HM 450K) |
| chr8 | 24151806 | 24151954 | OP-Präparate (HM 450K) |
| chr8 | 70981967 | 70982102 | OP-Präparate (HM 450K) |
| chr8 | 128807993 | 128808124 | OP-Präparate (HM 450K) |
| chr8 | 133072190 | 133072337 | OP-Präparate (HM 450K) |
| chr9 | 37002618 | 37002762 | OP-Präparate (HM 450K) |
| chr1 | 6165201 | 6165361 | zfDNA (WGBS) |
| chr1 | 17567892 | 17568189 | zfDNA (WGBS) |
| chr1 | 15426262 | 15426418 | zfDNA (WGBS) |
| chr1 | 15670403 | 15670539 | zfDNA (WGBS) |
| chr10 | 96279972 | 96280055 | zfDNA (WGBS) |
| chr10 | 97033594 | 97033733 | zfDNA (WGBS) |
| chr11 | 134245966 | 134246129 | zfDNA (WGBS) |
| chr12 | 8004422 | 8004573 | zfDNA (WGBS) |
| chr12 | 97140774 | 97140905 | zfDNA (WGBS) |
| chr12 | 111566555 | 111566698 | zfDNA (WGBS) |
| chr12 | 117750775 | 117750937 | zfDNA (WGBS) |
| chr13 | 36828740 | 36828902 | zfDNA (WGBS) |
| chr14 | 93214072 | 93214242 | zfDNA (WGBS) |
| chr15 | 56006471 | 56006552 | zfDNA (WGBS) |
| chr15 | 101547384 | 101547527 | zfDNA (WGBS) |
| chr16 | 4141795 | 4141956 | zfDNA (WGBS) |
| chr18 | 21857621 | 21857750 | zfDNA (WGBS) |
| chr18 | 29528340 | 29528468 | zfDNA (WGBS) |
| chr18 | 46845901 | 46846043 | zfDNA (WGBS) |
| chr19 | 874766 | 874934 | zfDNA (WGBS) |
| chr19 | 6799968 | 6800095 | zfDNA (WGBS) |
| chr2 | 1126410 | 1126557 | zfDNA (WGBS) |
| chr2 | 225642009 | 225642217 | zfDNA (WGBS) |
| chr2 | 236745514 | 236745688 | zfDNA (WGBS) |
| chr2 | 240881986 | 240882138 | zfDNA (WGBS) |
| chr2 | 2179742 | 2179886 | zfDNA (WGBS) |
| chr2 | 30747398 | 30747539 | zfDNA (WGBS) |
| chr2 | 175998270 | 175998415 | zfDNA (WGBS) |
| chr2 | 219647407 | 219647560 | zfDNA (WGBS) |
| chr20 | 20243607 | 20243747 | zfDNA (WGBS) |
| chr20 | 55079800 | 55079945 | zfDNA (WGBS) |
| chr21 | 30502729 | 30502871 | zfDNA (WGBS) |
| chr21 | 46587906 | 46588052 | zfDNA (WGBS) |

| | | | |
|-------|-----------|-----------|--------------------------------|
| chr3 | 56445240 | 56445378 | zfDNA (WGBS) |
| chr3 | 85143433 | 85143600 | zfDNA (WGBS) |
| chr3 | 146123966 | 146124095 | zfDNA (WGBS) |
| chr3 | 68947379 | 68947542 | zfDNA (WGBS) |
| chr3 | 197767819 | 197767978 | zfDNA (WGBS) |
| chr4 | 143487129 | 143487273 | zfDNA (WGBS) |
| chr4 | 26398190 | 26398329 | zfDNA (WGBS) |
| chr4 | 77647893 | 77648027 | zfDNA (WGBS) |
| chr4 | 102497551 | 102497732 | zfDNA (WGBS) |
| chr5 | 39187156 | 39187287 | zfDNA (WGBS) |
| chr5 | 56145736 | 56145896 | zfDNA (WGBS) |
| chr5 | 160171748 | 160171896 | zfDNA (WGBS) |
| chr5 | 16793080 | 16793219 | zfDNA (WGBS) |
| chr5 | 76869108 | 76869253 | zfDNA (WGBS) |
| chr6 | 169050287 | 169050447 | zfDNA (WGBS) |
| chr6 | 76773251 | 76773422 | zfDNA (WGBS) |
| chr6 | 123869831 | 123869971 | zfDNA (WGBS) |
| chr7 | 6268960 | 6269087 | zfDNA (WGBS) |
| chr7 | 38508407 | 38508486 | zfDNA (WGBS) |
| chr7 | 153743779 | 153743947 | zfDNA (WGBS) |
| chr7 | 137230794 | 137230963 | zfDNA (WGBS) |
| chr7 | 151300131 | 151300282 | zfDNA (WGBS) |
| chr8 | 3672236 | 3672387 | zfDNA (WGBS) |
| chr8 | 99510084 | 99510252 | zfDNA (WGBS) |
| chr8 | 101170822 | 101170975 | zfDNA (WGBS) |
| chr8 | 141127042 | 141127183 | zfDNA (WGBS) |
| chr9 | 2050654 | 2050804 | zfDNA (WGBS) |
| chr9 | 9227683 | 9227824 | zfDNA (WGBS) |
| chr9 | 79060522 | 79060633 | zfDNA (WGBS) |
| chr9 | 124334690 | 124334848 | zfDNA (WGBS) |
| chr9 | 126166694 | 126166828 | zfDNA (WGBS) |
| chr1 | 180202441 | 180202578 | Die bivalente Chromatin Studie |
| chr10 | 102984159 | 102984316 | Die bivalente Chromatin Studie |
| chr10 | 102986926 | 102987078 | Die bivalente Chromatin Studie |
| chr10 | 124905661 | 124905811 | Die bivalente Chromatin Studie |
| chr11 | 18416284 | 18416422 | Die bivalente Chromatin Studie |
| chr11 | 20178032 | 20178171 | Die bivalente Chromatin Studie |
| chr11 | 20181732 | 20181875 | Die bivalente Chromatin Studie |
| chr11 | 31821190 | 31821332 | Die bivalente Chromatin Studie |
| chr11 | 31831813 | 31831955 | Die bivalente Chromatin Studie |
| chr12 | 6644024 | 6644165 | Die bivalente Chromatin Studie |
| chr12 | 51158311 | 51158455 | Die bivalente Chromatin Studie |
| chr13 | 100621076 | 100621217 | Die bivalente Chromatin Studie |
| chr13 | 100624236 | 100624376 | Die bivalente Chromatin Studie |
| chr14 | 23790611 | 23790772 | Die bivalente Chromatin Studie |

| | | | |
|-------|-----------|-----------|--------------------------------|
| chr17 | 46674335 | 46674487 | Die bivalente Chromatin Studie |
| chr17 | 48048913 | 48049070 | Die bivalente Chromatin Studie |
| chr2 | 162283732 | 162283879 | Die bivalente Chromatin Studie |
| chr2 | 175199619 | 175199764 | Die bivalente Chromatin Studie |
| chr2 | 175200596 | 175200742 | Die bivalente Chromatin Studie |
| chr2 | 223163736 | 223163879 | Die bivalente Chromatin Studie |
| chr4 | 13525615 | 13525755 | Die bivalente Chromatin Studie |
| chr4 | 113432474 | 113432622 | Die bivalente Chromatin Studie |
| chr6 | 100051116 | 100051256 | Die bivalente Chromatin Studie |
| chr6 | 100054673 | 100054827 | Die bivalente Chromatin Studie |
| chr6 | 100060971 | 100061117 | Die bivalente Chromatin Studie |
| chr7 | 5569502 | 5569641 | Die bivalente Chromatin Studie |

| B. Entität: Adeno- oder Plattenepithelkarzinom? | | | |
|--|-----------|-----------|--------------------------------------|
| Chromosom | Start | Ende | Methode |
| chr1 | 52158087 | 52158220 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr1 | 61668739 | 61668922 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr1 | 64578151 | 64578293 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr1 | 77533495 | 77533671 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr1 | 171868017 | 171868187 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr1 | 214646125 | 214646279 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr11 | 1328403 | 1328548 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr11 | 4079459 | 4079623 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr11 | 71188639 | 71188789 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr11 | 104972062 | 104972193 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr11 | 105010212 | 105010354 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr12 | 52946925 | 52947067 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr12 | 88538122 | 88538272 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr12 | 109096126 | 109096269 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr14 | 90083196 | 90083338 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr16 | 58155114 | 58155256 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr17 | 29667240 | 29667387 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr2 | 9987364 | 9987518 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr2 | 25501964 | 25502121 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr2 | 172266609 | 172266746 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr2 | 178178843 | 178179018 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr2 | 179897218 | 179897356 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr20 | 31446106 | 31446254 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr3 | 4348279 | 4348416 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr3 | 38567580 | 38567725 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr3 | 111629808 | 111629952 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr3 | 114074222 | 114074369 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr3 | 122841556 | 122841705 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr3 | 150948199 | 150948350 | zfDNA (WGBS), OP-Präparate (HM 450K) |

| | | | |
|-------|-----------|-----------|--------------------------------------|
| chr3 | 164915101 | 164915268 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr5 | 122506718 | 122506853 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr6 | 63990944 | 63991095 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr6 | 64572767 | 64572911 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr7 | 20381014 | 20381160 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr7 | 21813010 | 21813162 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr7 | 98722395 | 98722537 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr7 | 102574027 | 102574188 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr7 | 102574397 | 102574549 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr7 | 111825737 | 111825894 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr7 | 116377388 | 116377530 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr7 | 122056569 | 122056711 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr8 | 38643499 | 38643633 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr8 | 42772303 | 42772478 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr8 | 145599209 | 145599355 | zfDNA (WGBS), OP-Präparate (HM 450K) |
| chr1 | 3607047 | 3607181 | OP-Präparate (HM 450K) |
| chr1 | 220101648 | 220101795 | OP-Präparate (HM 450K) |
| chr1 | 220101867 | 220102015 | OP-Präparate (HM 450K) |
| chr1 | 236849398 | 236849548 | OP-Präparate (HM 450K) |
| chr1 | 236849891 | 236850048 | OP-Präparate (HM 450K) |
| chr10 | 11206799 | 11206938 | OP-Präparate (HM 450K) |
| chr11 | 30606998 | 30607133 | OP-Präparate (HM 450K) |
| chr11 | 64992997 | 64993132 | OP-Präparate (HM 450K) |
| chr11 | 64993266 | 64993396 | OP-Präparate (HM 450K) |
| chr11 | 65360248 | 65360394 | OP-Präparate (HM 450K) |
| chr11 | 77160268 | 77160416 | OP-Präparate (HM 450K) |
| chr11 | 82444721 | 82444866 | OP-Präparate (HM 450K) |
| chr12 | 4381723 | 4381963 | OP-Präparate (HM 450K) |
| chr12 | 33592568 | 33592710 | OP-Präparate (HM 450K) |
| chr13 | 28674372 | 28674520 | OP-Präparate (HM 450K) |
| chr15 | 69087740 | 69087878 | OP-Präparate (HM 450K) |
| chr15 | 83316148 | 83316297 | OP-Präparate (HM 450K) |
| chr16 | 1202369 | 1202544 | OP-Präparate (HM 450K) |
| chr16 | 56224714 | 56224858 | OP-Präparate (HM 450K) |
| chr16 | 81564475 | 81564626 | OP-Präparate (HM 450K) |
| chr16 | 86600252 | 86600386 | OP-Präparate (HM 450K) |
| chr17 | 693067 | 693222 | OP-Präparate (HM 450K) |
| chr17 | 693313 | 693458 | OP-Präparate (HM 450K) |
| chr17 | 66292297 | 66292442 | OP-Präparate (HM 450K) |
| chr17 | 74696666 | 74696814 | OP-Präparate (HM 450K) |
| chr17 | 75196873 | 75197007 | OP-Präparate (HM 450K) |
| chr17 | 80794200 | 80794346 | OP-Präparate (HM 450K) |
| chr18 | 2847458 | 2847590 | OP-Präparate (HM 450K) |
| chr18 | 24131050 | 24131188 | OP-Präparate (HM 450K) |
| chr18 | 24131310 | 24131449 | OP-Präparate (HM 450K) |

| | | | |
|-------|-----------|-----------|------------------------|
| chr19 | 10572284 | 10572428 | OP-Präparate (HM 450K) |
| chr2 | 30834597 | 30834737 | OP-Präparate (HM 450K) |
| chr2 | 50574632 | 50574774 | OP-Präparate (HM 450K) |
| chr2 | 54054275 | 54054427 | OP-Präparate (HM 450K) |
| chr2 | 63276135 | 63276276 | OP-Präparate (HM 450K) |
| chr2 | 236444206 | 236444348 | OP-Präparate (HM 450K) |
| chr20 | 20349092 | 20349238 | OP-Präparate (HM 450K) |
| chr20 | 47444494 | 47444648 | OP-Präparate (HM 450K) |
| chr20 | 47444775 | 47445083 | OP-Präparate (HM 450K) |
| chr21 | 26934497 | 26934635 | OP-Präparate (HM 450K) |
| chr3 | 141102520 | 141102668 | OP-Präparate (HM 450K) |
| chr3 | 172167531 | 172167678 | OP-Präparate (HM 450K) |
| chr3 | 172394613 | 172394766 | OP-Präparate (HM 450K) |
| chr3 | 186914643 | 186914790 | OP-Präparate (HM 450K) |
| chr3 | 196435366 | 196435518 | OP-Präparate (HM 450K) |
| chr4 | 57522417 | 57522559 | OP-Präparate (HM 450K) |
| chr4 | 57522562 | 57522846 | OP-Präparate (HM 450K) |
| chr5 | 912634 | 912893 | OP-Präparate (HM 450K) |
| chr5 | 1883876 | 1884018 | OP-Präparate (HM 450K) |
| chr5 | 16179056 | 16179202 | OP-Präparate (HM 450K) |
| chr5 | 33936177 | 33936331 | OP-Präparate (HM 450K) |
| chr5 | 36607322 | 36607477 | OP-Präparate (HM 450K) |
| chr5 | 169064355 | 169064518 | OP-Präparate (HM 450K) |
| chr7 | 653234 | 653373 | OP-Präparate (HM 450K) |
| chr7 | 1491753 | 1492006 | OP-Präparate (HM 450K) |
| chr7 | 2158351 | 2158498 | OP-Präparate (HM 450K) |
| chr7 | 4228700 | 4228842 | OP-Präparate (HM 450K) |
| chr7 | 19156542 | 19156690 | OP-Präparate (HM 450K) |
| chr7 | 19157127 | 19157340 | OP-Präparate (HM 450K) |
| chr7 | 45197376 | 45197524 | OP-Präparate (HM 450K) |
| chr8 | 41754102 | 41754249 | OP-Präparate (HM 450K) |
| chr8 | 123874964 | 123875109 | OP-Präparate (HM 450K) |
| chr8 | 123875144 | 123875280 | OP-Präparate (HM 450K) |
| chr1 | 3289010 | 3289139 | zfDNA (WGBS) |
| chr1 | 17567892 | 17568189 | zfDNA (WGBS) |
| chr1 | 23284417 | 23284507 | zfDNA (WGBS) |
| chr1 | 24277975 | 24278154 | zfDNA (WGBS) |
| chr1 | 47738990 | 47739142 | zfDNA (WGBS) |
| chr1 | 79467955 | 79468081 | zfDNA (WGBS) |
| chr1 | 108975333 | 108975476 | zfDNA (WGBS) |
| chr1 | 196682870 | 196683025 | zfDNA (WGBS) |
| chr1 | 217310510 | 217310654 | zfDNA (WGBS) |
| chr1 | 237765796 | 237765947 | zfDNA (WGBS) |
| chr1 | 240656480 | 240656649 | zfDNA (WGBS) |
| chr1 | 240746545 | 240746706 | zfDNA (WGBS) |

| | | | |
|-------|-----------|-----------|--------------|
| chr1 | 246241918 | 246242056 | zfDNA (WGBS) |
| chr10 | 12533631 | 12533768 | zfDNA (WGBS) |
| chr10 | 32647546 | 32647656 | zfDNA (WGBS) |
| chr10 | 32657588 | 32657719 | zfDNA (WGBS) |
| chr10 | 37511104 | 37511239 | zfDNA (WGBS) |
| chr10 | 62708104 | 62708269 | zfDNA (WGBS) |
| chr10 | 73207931 | 73208064 | zfDNA (WGBS) |
| chr10 | 108812804 | 108812940 | zfDNA (WGBS) |
| chr10 | 115658133 | 115658275 | zfDNA (WGBS) |
| chr10 | 123914649 | 123914808 | zfDNA (WGBS) |
| chr11 | 15025357 | 15025499 | zfDNA (WGBS) |
| chr11 | 19778770 | 19778909 | zfDNA (WGBS) |
| chr11 | 26355535 | 26355711 | zfDNA (WGBS) |
| chr11 | 26600784 | 26600925 | zfDNA (WGBS) |
| chr11 | 26626367 | 26626558 | zfDNA (WGBS) |
| chr11 | 41275397 | 41275536 | zfDNA (WGBS) |
| chr11 | 62158845 | 62158985 | zfDNA (WGBS) |
| chr11 | 70503001 | 70503139 | zfDNA (WGBS) |
| chr11 | 106592142 | 106592304 | zfDNA (WGBS) |
| chr11 | 120644150 | 120644282 | zfDNA (WGBS) |
| chr11 | 122678508 | 122678636 | zfDNA (WGBS) |
| chr11 | 128851150 | 128851286 | zfDNA (WGBS) |
| chr12 | 16028663 | 16028857 | zfDNA (WGBS) |
| chr12 | 125571801 | 125571933 | zfDNA (WGBS) |
| chr13 | 48806444 | 48806588 | zfDNA (WGBS) |
| chr13 | 113527733 | 113527876 | zfDNA (WGBS) |
| chr14 | 35030336 | 35030470 | zfDNA (WGBS) |
| chr14 | 104486171 | 104486314 | zfDNA (WGBS) |
| chr15 | 22839905 | 22840043 | zfDNA (WGBS) |
| chr15 | 26964926 | 26965065 | zfDNA (WGBS) |
| chr15 | 29246303 | 29246447 | zfDNA (WGBS) |
| chr15 | 30180680 | 30180842 | zfDNA (WGBS) |
| chr15 | 32404970 | 32405130 | zfDNA (WGBS) |
| chr15 | 64244033 | 64244215 | zfDNA (WGBS) |
| chr15 | 68530927 | 68531091 | zfDNA (WGBS) |
| chr15 | 83579367 | 83579513 | zfDNA (WGBS) |
| chr15 | 88559865 | 88560003 | zfDNA (WGBS) |
| chr16 | 6257325 | 6257474 | zfDNA (WGBS) |
| chr16 | 15665564 | 15665721 | zfDNA (WGBS) |
| chr16 | 24321180 | 24321320 | zfDNA (WGBS) |
| chr16 | 75528556 | 75528698 | zfDNA (WGBS) |
| chr16 | 88013993 | 88014135 | zfDNA (WGBS) |
| chr16 | 89713952 | 89714124 | zfDNA (WGBS) |
| chr17 | 416719 | 416865 | zfDNA (WGBS) |
| chr17 | 19809670 | 19809830 | zfDNA (WGBS) |

| | | | |
|-------|-----------|-----------|--------------|
| chr17 | 21086965 | 21087112 | zfDNA (WGBS) |
| chr17 | 33364961 | 33365040 | zfDNA (WGBS) |
| chr17 | 42960413 | 42960551 | zfDNA (WGBS) |
| chr17 | 64330485 | 64330837 | zfDNA (WGBS) |
| chr17 | 75142732 | 75142885 | zfDNA (WGBS) |
| chr19 | 11890923 | 11891074 | zfDNA (WGBS) |
| chr19 | 49016450 | 49016584 | zfDNA (WGBS) |
| chr19 | 57922060 | 57922195 | zfDNA (WGBS) |
| chr2 | 1129413 | 1129596 | zfDNA (WGBS) |
| chr2 | 1334513 | 1334640 | zfDNA (WGBS) |
| chr2 | 23917010 | 23917136 | zfDNA (WGBS) |
| chr2 | 25124037 | 25124165 | zfDNA (WGBS) |
| chr2 | 46779214 | 46779381 | zfDNA (WGBS) |
| chr2 | 111874418 | 111874571 | zfDNA (WGBS) |
| chr2 | 113534514 | 113534653 | zfDNA (WGBS) |
| chr2 | 120417931 | 120418073 | zfDNA (WGBS) |
| chr2 | 131798797 | 131798977 | zfDNA (WGBS) |
| chr2 | 198073787 | 198073950 | zfDNA (WGBS) |
| chr2 | 205889570 | 205889704 | zfDNA (WGBS) |
| chr2 | 207319476 | 207319691 | zfDNA (WGBS) |
| chr20 | 9706282 | 9706429 | zfDNA (WGBS) |
| chr20 | 33713618 | 33713757 | zfDNA (WGBS) |
| chr21 | 33340955 | 33341038 | zfDNA (WGBS) |
| chr22 | 21206849 | 21206995 | zfDNA (WGBS) |
| chr22 | 30292326 | 30292475 | zfDNA (WGBS) |
| chr22 | 35697444 | 35697606 | zfDNA (WGBS) |
| chr3 | 3755582 | 3755730 | zfDNA (WGBS) |
| chr3 | 14959981 | 14960128 | zfDNA (WGBS) |
| chr3 | 25581721 | 25581859 | zfDNA (WGBS) |
| chr3 | 75834579 | 75834736 | zfDNA (WGBS) |
| chr3 | 87031909 | 87032079 | zfDNA (WGBS) |
| chr3 | 122710736 | 122710872 | zfDNA (WGBS) |
| chr3 | 139727561 | 139727706 | zfDNA (WGBS) |
| chr3 | 145864433 | 145864574 | zfDNA (WGBS) |
| chr4 | 1665996 | 1666155 | zfDNA (WGBS) |
| chr4 | 22518120 | 22518271 | zfDNA (WGBS) |
| chr4 | 77306769 | 77306948 | zfDNA (WGBS) |
| chr4 | 82520036 | 82520212 | zfDNA (WGBS) |
| chr4 | 155413871 | 155414011 | zfDNA (WGBS) |
| chr4 | 156601279 | 156601436 | zfDNA (WGBS) |
| chr4 | 162457724 | 162457860 | zfDNA (WGBS) |
| chr4 | 176636441 | 176636580 | zfDNA (WGBS) |
| chr4 | 177654193 | 177654363 | zfDNA (WGBS) |
| chr4 | 186659464 | 186659644 | zfDNA (WGBS) |
| chr5 | 14450118 | 14450272 | zfDNA (WGBS) |

| | | | |
|------|-----------|-----------|--------------|
| chr5 | 75935318 | 75935450 | zfDNA (WGBS) |
| chr5 | 132894183 | 132894320 | zfDNA (WGBS) |
| chr5 | 140475728 | 140475872 | zfDNA (WGBS) |
| chr5 | 146345906 | 146346062 | zfDNA (WGBS) |
| chr5 | 156458027 | 156458167 | zfDNA (WGBS) |
| chr5 | 157169890 | 157170038 | zfDNA (WGBS) |
| chr6 | 20832000 | 20832349 | zfDNA (WGBS) |
| chr6 | 24420281 | 24420413 | zfDNA (WGBS) |
| chr6 | 36331071 | 36331215 | zfDNA (WGBS) |
| chr6 | 54074847 | 54075021 | zfDNA (WGBS) |
| chr6 | 71122323 | 71122483 | zfDNA (WGBS) |
| chr6 | 83604672 | 83604779 | zfDNA (WGBS) |
| chr6 | 90709859 | 90710016 | zfDNA (WGBS) |
| chr6 | 111744738 | 111744881 | zfDNA (WGBS) |
| chr6 | 148806765 | 148806922 | zfDNA (WGBS) |
| chr6 | 155574119 | 155574263 | zfDNA (WGBS) |
| chr6 | 158460178 | 158460323 | zfDNA (WGBS) |
| chr7 | 5549605 | 5549675 | zfDNA (WGBS) |
| chr7 | 40669616 | 40669796 | zfDNA (WGBS) |
| chr7 | 73799798 | 73799908 | zfDNA (WGBS) |
| chr7 | 78030021 | 78030155 | zfDNA (WGBS) |
| chr7 | 81399230 | 81399365 | zfDNA (WGBS) |
| chr7 | 134452355 | 134452524 | zfDNA (WGBS) |
| chr7 | 140335200 | 140335344 | zfDNA (WGBS) |
| chr7 | 146925646 | 146925824 | zfDNA (WGBS) |
| chr7 | 153976496 | 153976643 | zfDNA (WGBS) |
| chr7 | 157941162 | 157941344 | zfDNA (WGBS) |
| chr7 | 157980130 | 157980264 | zfDNA (WGBS) |
| chr7 | 157980485 | 157980624 | zfDNA (WGBS) |
| chr7 | 158314155 | 158314301 | zfDNA (WGBS) |
| chr8 | 6392188 | 6392336 | zfDNA (WGBS) |
| chr8 | 11724061 | 11724159 | zfDNA (WGBS) |
| chr8 | 17237496 | 17237639 | zfDNA (WGBS) |
| chr8 | 21803649 | 21803801 | zfDNA (WGBS) |
| chr8 | 52696850 | 52697008 | zfDNA (WGBS) |
| chr8 | 72183950 | 72184120 | zfDNA (WGBS) |
| chr8 | 81042553 | 81042694 | zfDNA (WGBS) |
| chr8 | 85101824 | 85101952 | zfDNA (WGBS) |
| chr8 | 110703169 | 110703320 | zfDNA (WGBS) |
| chr8 | 121727803 | 121727944 | zfDNA (WGBS) |
| chr8 | 133476418 | 133476558 | zfDNA (WGBS) |
| chr9 | 8813022 | 8813150 | zfDNA (WGBS) |
| chr9 | 90258110 | 90258253 | zfDNA (WGBS) |
| chr9 | 97061691 | 97061835 | zfDNA (WGBS) |

| C. Prognose: günstig oder ungünstig? | | | |
|---|-----------|-----------|----------------------------------|
| Chromosom | Start | Ende | Methode |
| chr1 | 226187853 | 226187854 | <i>Paired Biopsies</i> (HM 450K) |
| chr1 | 226188006 | 226188007 | <i>Paired Biopsies</i> (HM 450K) |
| chr1 | 226187877 | 226187878 | <i>Paired Biopsies</i> (HM 450K) |
| chr1 | 26699448 | 26699449 | <i>Paired Biopsies</i> (HM 450K) |
| chr1 | 115677211 | 115677212 | <i>Paired Biopsies</i> (HM 450K) |
| chr1 | 18063106 | 18063107 | <i>Paired Biopsies</i> (HM 450K) |
| chr10 | 54223606 | 54223607 | <i>Paired Biopsies</i> (HM 450K) |
| chr11 | 1673436 | 1673437 | <i>Paired Biopsies</i> (HM 450K) |
| chr14 | 99700232 | 99700233 | <i>Paired Biopsies</i> (HM 450K) |
| chr2 | 241344707 | 241344708 | <i>Paired Biopsies</i> (HM 450K) |
| chr2 | 27362421 | 27362422 | <i>Paired Biopsies</i> (HM 450K) |
| chr2 | 241314589 | 241314590 | <i>Paired Biopsies</i> (HM 450K) |
| chr20 | 43028501 | 43028502 | <i>Paired Biopsies</i> (HM 450K) |
| chr20 | 584774 | 584775 | <i>Paired Biopsies</i> (HM 450K) |
| chr20 | 4201165 | 4201166 | <i>Paired Biopsies</i> (HM 450K) |
| chr20 | 2508982 | 2508983 | <i>Paired Biopsies</i> (HM 450K) |
| chr20 | 46323482 | 46323483 | <i>Paired Biopsies</i> (HM 450K) |
| chr3 | 13914731 | 13914732 | <i>Paired Biopsies</i> (HM 450K) |
| chr5 | 176167283 | 176167284 | <i>Paired Biopsies</i> (HM 450K) |
| chr6 | 29528775 | 29528776 | <i>Paired Biopsies</i> (HM 450K) |
| chr6 | 154869910 | 154869911 | <i>Paired Biopsies</i> (HM 450K) |
| chr7 | 42195876 | 42195877 | <i>Paired Biopsies</i> (HM 450K) |
| chr8 | 49496392 | 49496393 | <i>Paired Biopsies</i> (HM 450K) |
| chr8 | 49382370 | 49382371 | <i>Paired Biopsies</i> (HM 450K) |
| chr8 | 49496370 | 49496371 | <i>Paired Biopsies</i> (HM 450K) |
| chr8 | 49533444 | 49533445 | <i>Paired Biopsies</i> (HM 450K) |
| chr8 | 49494724 | 49494725 | <i>Paired Biopsies</i> (HM 450K) |
| chr8 | 10452896 | 10452897 | <i>Paired Biopsies</i> (HM 450K) |
| chr8 | 49547127 | 49547128 | <i>Paired Biopsies</i> (HM 450K) |
| chr8 | 49823434 | 49823435 | <i>Paired Biopsies</i> (HM 450K) |
| chr8 | 11614472 | 11614473 | <i>Paired Biopsies</i> (HM 450K) |
| chr8 | 49466210 | 49466211 | <i>Paired Biopsies</i> (HM 450K) |
| chr9 | 133792985 | 133792986 | <i>Paired Biopsies</i> (HM 450K) |

Tab. 8: Der kNN Algorithmus verwendete 10 Positionen, um die Lungenkarzinompatienten von den gesunden Probanden unterscheiden zu können.

| ID | Chromosom | Position |
|-------------------|-----------------------------|------------------|
| 596 | chr11 | 57006229 |
| 1717 | chr15 | 28262724 |
| 2636 | chr18 | 61144199 |
| 2805 | chr19 | 46823441 |
| 4674 | chr2 | 176964685 |
| 4999 | chr2 | 225642035 |
| 5071 | chr3 | 14960020 |
| 5576 | chr4 | 13525705 |
| 6105 | chr5 | 140475760 |
| 6434 | chr6 | 46386723 |
| Gruppe | Richtigkeit | Anzahl an Proben |
| Maligner Tumor | 1 | 9 |
| Kontrolle | 1 | 3 |
| Mittelwert | 1 | 12 |
| Weitere Parameter | | |
| K | 5 | |
| Rangordnung | Vergleich zweier Gruppen | |
| Normalisierung | Mittelwert = 0, Varianz = 1 | |
| Fehlender Wert | Mittelwert | |

Tab. 9: Der RT Algorithmus analysierte 10 Positionen um die Entität eines Tumors zu ermitteln.

| ID | Chromosom | Position |
|-------------------------------|-----------------------------|------------------|
| 650 | chr11 | 64993331 |
| 2995 | chr1 | 17568007 |
| 4233 | chr2 | 50574690 |
| 4241 | chr2 | 50574708 |
| 4428 | chr2 | 111874494 |
| 4447 | chr2 | 121276804 |
| 5537 | chr4 | 1666074 |
| 5538 | chr4 | 1666075 |
| 6524 | chr6 | 83604790 |
| 7164 | chr7 | 69971740 |
| Gruppe | Richtigkeit | Anzahl an Proben |
| Adenokarzinom | 1 | 4 |
| Plattenepithelkarzinom | 1 | 5 |
| Mittelwert | 1 | 9 |
| Weitere Parameter | | |
| Max. Tiefe | 25 | |
| Min. Anteil der Stichprobe | 1% | |
| Max. Anzahl an Kategorien | 10 | |
| Max. Anzahl an Bäumen | 250 | |
| Richtigkeit des Waldes | 0.1 | |
| Kriterien für das Terminieren | Max. Anzahl an Bäumen | |
| Rangordnung | Vergleich zweier Gruppen | |
| Normalisierung | Mittelwert = 0, Varianz = 1 | |
| Fehlender Wert | Mittelwert | |

Tab. 10: Für das *Staging* wurden vom SVM Algorithmus 523 Positionen analysiert.

| ID | Chromosom | Position |
|-----|-----------|-----------|
| 16 | chr10 | 12533708 |
| 17 | chr10 | 12533710 |
| 20 | chr10 | 12533754 |
| 26 | chr10 | 15110983 |
| 37 | chr10 | 32657656 |
| 38 | chr10 | 32657672 |
| 79 | chr10 | 62708202 |
| 104 | chr10 | 97033706 |
| 123 | chr10 | 98129889 |
| 154 | chr10 | 102895057 |

| | | |
|-----|-------|-----------|
| 164 | chr10 | 102984248 |
| 196 | chr10 | 102987003 |
| 199 | chr10 | 102987007 |
| 269 | chr10 | 121075316 |
| 281 | chr10 | 123914718 |
| 315 | chr10 | 124905781 |
| 320 | chr10 | 126494586 |
| 327 | chr10 | 126494644 |
| 333 | chr10 | 130860205 |
| 347 | chr10 | 134598357 |
| 349 | chr10 | 134598359 |
| 364 | chr11 | 627157 |
| 382 | chr11 | 1328455 |
| 385 | chr11 | 1328485 |
| 411 | chr11 | 15025433 |
| 431 | chr11 | 19778814 |
| 473 | chr11 | 26355627 |
| 479 | chr11 | 26626371 |
| 483 | chr11 | 26626471 |
| 554 | chr11 | 31831858 |
| 568 | chr11 | 31831899 |
| 572 | chr11 | 31831908 |
| 576 | chr11 | 31831919 |
| 585 | chr11 | 40136809 |
| 677 | chr11 | 69061863 |
| 696 | chr11 | 71188761 |
| 697 | chr11 | 71188762 |
| 709 | chr11 | 82444757 |
| 712 | chr11 | 82444771 |
| 742 | chr11 | 86085932 |
| 760 | chr11 | 113629710 |
| 767 | chr11 | 113629767 |
| 768 | chr11 | 114052115 |
| 793 | chr11 | 122678513 |
| 819 | chr11 | 134246113 |
| 822 | chr12 | 1943225 |
| 823 | chr12 | 1943226 |
| 824 | chr12 | 1943232 |
| 827 | chr12 | 2526357 |
| 831 | chr12 | 2526402 |
| 849 | chr12 | 2751699 |
| 867 | chr12 | 4381792 |
| 873 | chr12 | 4381812 |
| 880 | chr12 | 4381851 |
| 958 | chr12 | 33592641 |

| | | |
|------|-------|-----------|
| 963 | chr12 | 33592661 |
| 966 | chr12 | 33592673 |
| 970 | chr12 | 33592682 |
| 976 | chr12 | 34358517 |
| 1005 | chr12 | 34503613 |
| 1041 | chr12 | 54423543 |
| 1052 | chr12 | 54423567 |
| 1171 | chr12 | 97140890 |
| 1179 | chr12 | 108051521 |
| 1196 | chr12 | 112427829 |
| 1222 | chr12 | 123203607 |
| 1223 | chr12 | 123203612 |
| 1227 | chr12 | 123203644 |
| 1235 | chr12 | 125571833 |
| 1239 | chr12 | 126142895 |
| 1244 | chr12 | 129347679 |
| 1265 | chr12 | 129886069 |
| 1272 | chr12 | 129886183 |
| 1287 | chr13 | 36828832 |
| 1305 | chr13 | 48806483 |
| 1317 | chr13 | 58207814 |
| 1319 | chr13 | 58207831 |
| 1335 | chr13 | 93325573 |
| 1336 | chr13 | 93325602 |
| 1348 | chr13 | 100621150 |
| 1354 | chr13 | 100621175 |
| 1358 | chr13 | 100621185 |
| 1362 | chr13 | 100621194 |
| 1394 | chr13 | 100624346 |
| 1415 | chr14 | 23511181 |
| 1441 | chr14 | 35030414 |
| 1442 | chr14 | 35030415 |
| 1443 | chr14 | 35231195 |
| 1461 | chr14 | 37128597 |
| 1472 | chr14 | 55907282 |
| 1478 | chr14 | 55907299 |
| 1487 | chr14 | 57274719 |
| 1512 | chr14 | 57275127 |
| 1550 | chr14 | 57275995 |
| 1610 | chr14 | 57278179 |
| 1612 | chr14 | 57278187 |
| 1619 | chr14 | 57278220 |
| 1637 | chr14 | 58064926 |
| 1701 | chr14 | 104486258 |
| 1702 | chr14 | 104486260 |

| | | |
|------|-------|----------|
| 1710 | chr15 | 26964987 |
| 1714 | chr15 | 28262702 |
| 1732 | chr15 | 32405064 |
| 1733 | chr15 | 32405065 |
| 1738 | chr15 | 41925179 |
| 1740 | chr15 | 41925187 |
| 1741 | chr15 | 41925188 |
| 1767 | chr15 | 45409283 |
| 1768 | chr15 | 45409293 |
| 1780 | chr15 | 56006548 |
| 1784 | chr15 | 63349191 |
| 1786 | chr15 | 63349194 |
| 1787 | chr15 | 63349195 |
| 1807 | chr15 | 69087782 |
| 1820 | chr15 | 72520614 |
| 1831 | chr15 | 83316217 |
| 1840 | chr15 | 83316257 |
| 1841 | chr15 | 83316258 |
| 1846 | chr15 | 83316269 |
| 1847 | chr15 | 83316270 |
| 1873 | chr15 | 89920814 |
| 1879 | chr15 | 89920855 |
| 1893 | chr15 | 89922297 |
| 1908 | chr15 | 89922344 |
| 1910 | chr15 | 89922358 |
| 1915 | chr15 | 89922387 |
| 1919 | chr15 | 98477887 |
| 1934 | chr16 | 526638 |
| 1956 | chr16 | 1202458 |
| 1973 | chr16 | 2880458 |
| 1974 | chr16 | 2880459 |
| 1991 | chr16 | 15665653 |
| 2014 | chr16 | 24822719 |
| 2056 | chr16 | 34255535 |
| 2067 | chr16 | 56224754 |
| 2071 | chr16 | 56224777 |
| 2080 | chr16 | 56224809 |
| 2106 | chr16 | 66613071 |
| 2133 | chr16 | 66613250 |
| 2157 | chr16 | 71528178 |
| 2171 | chr16 | 75528661 |
| 2187 | chr16 | 86600283 |
| 2199 | chr16 | 86600325 |
| 2205 | chr16 | 86600340 |
| 2214 | chr16 | 88014067 |

| | | |
|------|-------|----------|
| 2216 | chr16 | 88014072 |
| 2218 | chr16 | 88014083 |
| 2222 | chr16 | 89714003 |
| 2225 | chr16 | 89714008 |
| 2226 | chr16 | 89714009 |
| 2237 | chr16 | 89714063 |
| 2249 | chr17 | 416795 |
| 2274 | chr17 | 750241 |
| 2285 | chr17 | 19809748 |
| 2286 | chr17 | 19809749 |
| 2291 | chr17 | 29174301 |
| 2299 | chr17 | 29174410 |
| 2317 | chr17 | 33314935 |
| 2322 | chr17 | 33314978 |
| 2327 | chr17 | 33314988 |
| 2333 | chr17 | 33365076 |
| 2356 | chr17 | 35299620 |
| 2371 | chr17 | 42960474 |
| 2373 | chr17 | 42960488 |
| 2389 | chr17 | 46799639 |
| 2391 | chr17 | 46799644 |
| 2393 | chr17 | 46799647 |
| 2394 | chr17 | 46799648 |
| 2406 | chr17 | 48048981 |
| 2411 | chr17 | 48049008 |
| 2435 | chr17 | 59532275 |
| 2443 | chr17 | 59532314 |
| 2453 | chr17 | 64330651 |
| 2459 | chr17 | 66292378 |
| 2463 | chr17 | 67536298 |
| 2465 | chr17 | 72619555 |
| 2481 | chr17 | 75142814 |
| 2506 | chr17 | 80794324 |
| 2530 | chr18 | 3971140 |
| 2541 | chr18 | 18658118 |
| 2545 | chr18 | 20714345 |
| 2550 | chr18 | 21596915 |
| 2559 | chr18 | 24131108 |
| 2574 | chr18 | 24131391 |
| 2600 | chr18 | 61143901 |
| 2624 | chr18 | 61144121 |
| 2665 | chr19 | 5141394 |
| 2688 | chr19 | 10572317 |
| 2711 | chr19 | 11891002 |
| 2739 | chr19 | 19625286 |

| | | |
|------|-------|-----------|
| 2756 | chr19 | 29991306 |
| 2767 | chr19 | 33102573 |
| 2769 | chr19 | 42600236 |
| 2789 | chr19 | 44629761 |
| 2791 | chr19 | 45782682 |
| 2803 | chr19 | 46823435 |
| 2815 | chr19 | 49016516 |
| 2820 | chr19 | 49016533 |
| 2824 | chr19 | 49503049 |
| 2839 | chr19 | 49909923 |
| 2862 | chr1 | 2198846 |
| 2863 | chr1 | 2198847 |
| 2867 | chr1 | 2198863 |
| 2958 | chr1 | 8787209 |
| 2962 | chr1 | 8787261 |
| 2973 | chr1 | 15426297 |
| 2984 | chr1 | 15670515 |
| 2992 | chr1 | 17568000 |
| 3005 | chr1 | 17568145 |
| 3017 | chr1 | 19764666 |
| 3019 | chr1 | 19764669 |
| 3025 | chr1 | 19764722 |
| 3031 | chr1 | 23284390 |
| 3033 | chr1 | 23284423 |
| 3045 | chr1 | 27234577 |
| 3053 | chr1 | 27234623 |
| 3056 | chr1 | 27234626 |
| 3066 | chr1 | 34642399 |
| 3110 | chr1 | 50883372 |
| 3121 | chr1 | 50886745 |
| 3127 | chr1 | 50886772 |
| 3164 | chr1 | 50886995 |
| 3174 | chr1 | 61668834 |
| 3176 | chr1 | 63489100 |
| 3225 | chr1 | 108975412 |
| 3227 | chr1 | 108975445 |
| 3284 | chr1 | 115677210 |
| 3315 | chr1 | 155162705 |
| 3348 | chr1 | 160783053 |
| 3390 | chr1 | 161008828 |
| 3392 | chr1 | 161008848 |
| 3402 | chr1 | 161306175 |
| 3454 | chr1 | 180202553 |
| 3488 | chr1 | 217310587 |
| 3501 | chr1 | 220101724 |

| | | |
|------|-------|-----------|
| 3509 | chr1 | 220101774 |
| 3519 | chr1 | 220101934 |
| 3538 | chr1 | 223948910 |
| 3554 | chr1 | 236849473 |
| 3579 | chr1 | 236849941 |
| 3586 | chr1 | 236849958 |
| 3616 | chr1 | 240656582 |
| 3630 | chr20 | 584741 |
| 3631 | chr20 | 584744 |
| 3636 | chr20 | 2508981 |
| 3641 | chr20 | 5282947 |
| 3651 | chr20 | 9706361 |
| 3652 | chr20 | 9706362 |
| 3655 | chr20 | 19910263 |
| 3724 | chr20 | 46323527 |
| 3748 | chr20 | 47444816 |
| 3762 | chr20 | 47444849 |
| 3764 | chr20 | 47444851 |
| 3795 | chr20 | 47444971 |
| 3809 | chr20 | 47445025 |
| 3923 | chr21 | 38077257 |
| 3970 | chr22 | 29956486 |
| 3977 | chr22 | 30292389 |
| 3990 | chr22 | 35697550 |
| 3992 | chr22 | 35697558 |
| 3999 | chr22 | 40810375 |
| 4005 | chr22 | 45129980 |
| 4018 | chr22 | 45992027 |
| 4021 | chr22 | 45992040 |
| 4070 | chr2 | 3642648 |
| 4078 | chr2 | 9987439 |
| 4142 | chr2 | 32313668 |
| 4158 | chr2 | 45171790 |
| 4183 | chr2 | 45232432 |
| 4236 | chr2 | 50574700 |
| 4257 | chr2 | 63276180 |
| 4258 | chr2 | 63276181 |
| 4265 | chr2 | 63276215 |
| 4299 | chr2 | 63281187 |
| 4391 | chr2 | 63284132 |
| 4399 | chr2 | 63284165 |
| 4406 | chr2 | 73021347 |
| 4414 | chr2 | 100209375 |
| 4434 | chr2 | 113534594 |
| 4446 | chr2 | 120418056 |

| | | |
|------|------|-----------|
| 4472 | chr2 | 152992678 |
| 4479 | chr2 | 155089854 |
| 4550 | chr2 | 162283795 |
| 4574 | chr2 | 175199659 |
| 4597 | chr2 | 175199740 |
| 4642 | chr2 | 176964121 |
| 4683 | chr2 | 176964710 |
| 4761 | chr2 | 176988910 |
| 4768 | chr2 | 176988939 |
| 4797 | chr2 | 177014908 |
| 4800 | chr2 | 177014949 |
| 4820 | chr2 | 177027453 |
| 4845 | chr2 | 179897288 |
| 4864 | chr2 | 200326684 |
| 4873 | chr2 | 200326734 |
| 4998 | chr2 | 225642025 |
| 5005 | chr2 | 236444281 |
| 5023 | chr2 | 240882054 |
| 5030 | chr2 | 240882093 |
| 5040 | chr2 | 241314636 |
| 5054 | chr2 | 242273197 |
| 5057 | chr2 | 242273208 |
| 5058 | chr2 | 242273209 |
| 5059 | chr2 | 242273212 |
| 5072 | chr3 | 14960021 |
| 5093 | chr3 | 37544000 |
| 5097 | chr3 | 38567610 |
| 5104 | chr3 | 46904943 |
| 5108 | chr3 | 48837718 |
| 5120 | chr3 | 68947458 |
| 5122 | chr3 | 69280664 |
| 5125 | chr3 | 69280734 |
| 5146 | chr3 | 87032003 |
| 5147 | chr3 | 87032004 |
| 5153 | chr3 | 114074297 |
| 5156 | chr3 | 122710786 |
| 5157 | chr3 | 122710787 |
| 5160 | chr3 | 122710815 |
| 5162 | chr3 | 122710835 |
| 5163 | chr3 | 122710836 |
| 5164 | chr3 | 122710859 |
| 5165 | chr3 | 122710860 |
| 5167 | chr3 | 122841632 |
| 5282 | chr3 | 147108882 |
| 5303 | chr3 | 147113870 |

| | | |
|------|------|-----------|
| 5380 | chr3 | 150948274 |
| 5388 | chr3 | 160167967 |
| 5391 | chr3 | 160167976 |
| 5436 | chr3 | 178907655 |
| 5483 | chr3 | 184742720 |
| 5510 | chr3 | 196440526 |
| 5522 | chr4 | 738427 |
| 5531 | chr4 | 738469 |
| 5532 | chr4 | 738470 |
| 5536 | chr4 | 1666071 |
| 5550 | chr4 | 8017839 |
| 5556 | chr4 | 8288935 |
| 5561 | chr4 | 13525657 |
| 5567 | chr4 | 13525666 |
| 5577 | chr4 | 13525721 |
| 5596 | chr4 | 21886957 |
| 5602 | chr4 | 26398259 |
| 5613 | chr4 | 57522468 |
| 5623 | chr4 | 57522505 |
| 5624 | chr4 | 57522506 |
| 5632 | chr4 | 57522618 |
| 5634 | chr4 | 57522620 |
| 5640 | chr4 | 57522642 |
| 5652 | chr4 | 57522762 |
| 5664 | chr4 | 71520485 |
| 5670 | chr4 | 77306910 |
| 5691 | chr4 | 82520132 |
| 5692 | chr4 | 82520133 |
| 5698 | chr4 | 94616253 |
| 5773 | chr4 | 113432594 |
| 5794 | chr4 | 151504732 |
| 5816 | chr4 | 176636539 |
| 5833 | chr4 | 183696160 |
| 5834 | chr4 | 183696161 |
| 5836 | chr4 | 183696165 |
| 5837 | chr4 | 186659550 |
| 5844 | chr5 | 912688 |
| 5861 | chr5 | 912783 |
| 5863 | chr5 | 912786 |
| 5865 | chr5 | 912803 |
| 5869 | chr5 | 912820 |
| 5872 | chr5 | 912834 |
| 5875 | chr5 | 912839 |
| 5947 | chr5 | 5146384 |
| 5949 | chr5 | 5568539 |

| | | |
|------|------|-----------|
| 5955 | chr5 | 5568625 |
| 5956 | chr5 | 9782141 |
| 5975 | chr5 | 15824429 |
| 5992 | chr5 | 16179153 |
| 6040 | chr5 | 39187223 |
| 6065 | chr5 | 75935385 |
| 6081 | chr5 | 125881737 |
| 6083 | chr5 | 125881794 |
| 6115 | chr5 | 140475801 |
| 6128 | chr5 | 140810902 |
| 6130 | chr5 | 140810918 |
| 6156 | chr5 | 140811674 |
| 6163 | chr5 | 146345982 |
| 6164 | chr5 | 146345983 |
| 6167 | chr5 | 146346033 |
| 6169 | chr5 | 155481326 |
| 6174 | chr5 | 155481363 |
| 6186 | chr5 | 157169968 |
| 6191 | chr5 | 160171802 |
| 6194 | chr5 | 160171823 |
| 6206 | chr5 | 169064444 |
| 6233 | chr5 | 176167243 |
| 6237 | chr5 | 176167283 |
| 6253 | chr5 | 177966092 |
| 6255 | chr5 | 179180016 |
| 6273 | chr6 | 1656954 |
| 6274 | chr6 | 1656975 |
| 6281 | chr6 | 1657068 |
| 6283 | chr6 | 5132887 |
| 6302 | chr6 | 20832186 |
| 6309 | chr6 | 20832253 |
| 6363 | chr6 | 28227164 |
| 6365 | chr6 | 29528773 |
| 6367 | chr6 | 30130880 |
| 6368 | chr6 | 30130881 |
| 6390 | chr6 | 34984930 |
| 6394 | chr6 | 36253053 |
| 6502 | chr6 | 54074944 |
| 6511 | chr6 | 63991019 |
| 6512 | chr6 | 63991020 |
| 6531 | chr6 | 90709953 |
| 6537 | chr6 | 100051149 |
| 6565 | chr6 | 100054714 |
| 6616 | chr6 | 100061098 |
| 6625 | chr6 | 100905468 |

| | | |
|------|------|-----------|
| 6666 | chr6 | 108675552 |
| 6674 | chr6 | 111744817 |
| 6693 | chr6 | 138866887 |
| 6698 | chr6 | 139205501 |
| 6723 | chr6 | 158056087 |
| 6760 | chr6 | 169050364 |
| 6778 | chr7 | 187685 |
| 6782 | chr7 | 187715 |
| 6784 | chr7 | 653308 |
| 6805 | chr7 | 1491812 |
| 6813 | chr7 | 1491843 |
| 6815 | chr7 | 1491858 |
| 6821 | chr7 | 1491902 |
| 6825 | chr7 | 1491921 |
| 6829 | chr7 | 1491966 |
| 6890 | chr7 | 4228760 |
| 6895 | chr7 | 4228818 |
| 6896 | chr7 | 4786867 |
| 6903 | chr7 | 4786899 |
| 6906 | chr7 | 4786942 |
| 6949 | chr7 | 6269027 |
| 6958 | chr7 | 19156576 |
| 6969 | chr7 | 19156620 |
| 6999 | chr7 | 19157240 |
| 7093 | chr7 | 40100597 |
| 7114 | chr7 | 44200085 |
| 7131 | chr7 | 45197448 |
| 7157 | chr7 | 65617361 |
| 7160 | chr7 | 65617365 |
| 7165 | chr7 | 69971769 |
| 7185 | chr7 | 73981010 |
| 7242 | chr7 | 96622713 |
| 7252 | chr7 | 102574104 |
| 7253 | chr7 | 102574105 |
| 7261 | chr7 | 102574475 |
| 7265 | chr7 | 102574499 |
| 7269 | chr7 | 102574518 |
| 7277 | chr7 | 111825813 |
| 7307 | chr7 | 134452411 |
| 7312 | chr7 | 134452451 |
| 7313 | chr7 | 134452452 |
| 7324 | chr7 | 140335252 |
| 7326 | chr7 | 140335261 |
| 7333 | chr7 | 141397873 |
| 7335 | chr7 | 146925681 |

| | | |
|------|------|-----------|
| 7365 | chr7 | 151300170 |
| 7368 | chr7 | 151300199 |
| 7400 | chr7 | 154087938 |
| 7410 | chr7 | 157691297 |
| 7417 | chr7 | 157691343 |
| 7420 | chr7 | 157941187 |
| 7424 | chr7 | 157941301 |
| 7430 | chr7 | 157980193 |
| 7433 | chr7 | 157980216 |
| 7439 | chr7 | 158314202 |
| 7443 | chr7 | 158314235 |
| 7463 | chr8 | 6735007 |
| 7466 | chr8 | 10452854 |
| 7473 | chr8 | 10452918 |
| 7478 | chr8 | 11614472 |
| 7482 | chr8 | 11614502 |
| 7483 | chr8 | 11614519 |
| 7484 | chr8 | 11614520 |
| 7502 | chr8 | 14422449 |
| 7521 | chr8 | 26286645 |
| 7539 | chr8 | 31676500 |
| 7544 | chr8 | 31676524 |
| 7553 | chr8 | 31676686 |
| 7555 | chr8 | 31676697 |
| 7565 | chr8 | 41754146 |
| 7608 | chr8 | 49382365 |
| 7622 | chr8 | 49494676 |
| 7626 | chr8 | 49494699 |
| 7630 | chr8 | 49494723 |
| 7634 | chr8 | 49494773 |
| 7636 | chr8 | 49496331 |
| 7671 | chr8 | 49823388 |
| 7673 | chr8 | 49823395 |
| 7696 | chr8 | 62323244 |
| 7727 | chr8 | 72184078 |
| 7734 | chr8 | 85101900 |
| 7740 | chr8 | 110703242 |
| 7742 | chr8 | 110703245 |
| 7759 | chr8 | 123875007 |
| 7762 | chr8 | 123875033 |
| 7763 | chr8 | 123875034 |
| 7766 | chr8 | 123875051 |
| 7773 | chr8 | 123875079 |
| 7782 | chr8 | 123875253 |
| 7786 | chr8 | 126258311 |

| | | |
|-------------------------------|--|-------------------------|
| 7787 | chr8 | 126258312 |
| 7793 | chr8 | 129339255 |
| 7802 | chr8 | 141127120 |
| 7821 | chr9 | 8813092 |
| 7846 | chr9 | 37002714 |
| 7859 | chr9 | 74342704 |
| 7863 | chr9 | 74342751 |
| 7868 | chr9 | 79060545 |
| 7876 | chr9 | 90258167 |
| 7879 | chr9 | 90258174 |
| 7893 | chr9 | 97061769 |
| 7894 | chr9 | 97061770 |
| 7900 | chr9 | 115478948 |
| 7904 | chr9 | 115478954 |
| 7925 | chr9 | 126166763 |
| 7926 | chr9 | 126166764 |
| 7927 | chr9 | 133792936 |
| 7932 | chr9 | 133792984 |
| Gruppe | Richtigkeit | Anzahl an Proben |
| Frühes Stadium | 0 | 4 |
| Spätes Stadium | 0,8 | 5 |
| Mittelwert | 0,4 | 9 |
| Weitere Parameter | | |
| Typ | C_SVC | |
| Kernel-Typ | Linear | |
| Ausgabe | 0.0001 | |
| Nu | 0,5 | |
| Epsilon SVR | 0,1 | |
| Kriterien für das Terminieren | Das Epsilon-Terminieren oder max. Iterationen | |
| Das Epsilon-Terminieren | 0.001 | |
| Max. Iterationen | 1000 | |
| Rangordnung | Vergleich zweier Gruppen | |
| Filtern nach | Gruppe | |
| Bedingung | ungleich, != | |
| Normalisierung | Mittelwert = 0, Varianz = 1 | |
| Fehlender Wert | Mittelwert | |

9.4 Ethikvotum



UNIVERSITÄT ZU LÜBECK

Universität zu Lübeck · Ratzeburger Allee 160 · 23538 Lübeck

Frau
PD Dr. Gaede
Forschungszentrum Borstel
Parkallee 1-40

23845 Borstel

nachrichtlich:
Herr Prof. Zabel
Direktor FZB Borstel

Ethik-Kommission

Vorsitzender:
Herr Prof. Dr. med. Dr. phil. H. Raspe
Universität zu Lübeck
Stellv. Vorsitzender:
Herr Prof. Dr. med. F. Gieseler
Ratzeburger Allee 160
23538 Lübeck

Sachbearbeitung: Frau Janine Erdmann
Tel.: +49 451 500 4639
Fax: +49 451 500 3026
janine.erdmann@medizin.uni-luebeck.de

Aktenzeichen: 12-220
Datum: 18. Dezember 2012

Sitzung der Ethik-Kommission am 13. Dezember 2012

Antragsteller: Frau Dr. Gaede / Herr Prof. Zabel

Titel: Verwendung historischer Biomaterialproben aus sog. Überschussmaterial inkl. zugehöriger "Klinischer Basisdatensätze" für die medizinische Wissenschaft der Zukunft

Sehr geehrte Frau Dr. Gaede,

der Antrag wurde unter berufsethischen, medizinisch-wissenschaftlichen und berufsrechtlichen Gesichtspunkten geprüft.

Die Kommission hat keine Bedenken.

Über alle schwerwiegenden oder unerwarteten und unerwünschten Ereignisse, die während der Studie auftreten, muss die Kommission umgehend benachrichtigt werden.

Die ärztliche und juristische Verantwortung des Studienleiters und der an der Studie teilnehmenden Ärzte bleibt entsprechend der Beratungsfunktion der Ethikkommission durch unsere Stellungnahme unberührt.

Mit freundlichem Gruß bin ich

Ihr

Prof. Dr. med. Frank Gieseler
Stellv. Vorsitzender



UNIVERSITÄT ZU LÜBECK

Universität zu Lübeck · Ratzeburger Allee 160 · 23538 Lübeck

Herrn
PD Dr. med. N. Reinmuth
Krankenhaus Großhansdorf
Onkologischer Schwerpunkt
Wöhrendamm 80

22927 Großhansdorf

Ethik-Kommission

Vorsitzender:
Herr Prof. Dr. med. Dr. phil. H. Raspe
Stellv. Vorsitzender:
Herr Prof. Dr. med. F. Gieseler
Universität zu Lübeck
Ratzeburger Allee 160
23538 Lübeck

Sachbearbeitung: Frau Janine Erdmann
Tel.: +49 451 500 4639
Fax: +49 451 500 3026
janine.erdmann@medizin.uni-
luebeck.de

Aktenzeichen: 12-238

Datum: 10. Januar 2013

Antrag an die Ethik-Kommission vom 14. Dezember 2012

Antragsteller: Herr PD Dr. Reinmuth

RETARGET: Aufbau einer Kohorte von Patienten mit Lungenkarzinom mit wiederholter Entnahme von Gewebe und Blut zum verbesserten biologischen Verständnis der Erkrankung sowie einer verbesserten Therapiesteuerung

Sehr geehrter Herr Dr. Reinmuth,

am 10. Januar 2013 habe ich mich mit der o.g. Studie befasst.

Die Prüfung bezieht sich auf folgende Unterlagen:

- Ihr Anschreiben vom 14. Dezember 2012, Basisformular, Studienprotokoll in der Version 1.1 vom 14. Dezember 2012, Aufklärungsmaterialien in der Version 1.0 vom 12. Dezember 2012 und das zustimmende Ethikvotum der Ethik-Kommission Universität Heidelberg vom 24. Mai 2012.

Die Ethik-Kommission schließt sich dem o.g. positiven Votum an.

Bei Änderung des Studiendesigns sollte der Antrag erneut vorgelegt werden. Über alle schwerwiegenden oder unerwarteten und unerwünschten Ereignisse, die während der Studie auftreten, muß die Kommission umgehend benachrichtigt werden. Die ärztliche und juristische Verantwortung des Studienleiters und der an der Prüfung teilnehmenden Ärzte bleibt entsprechend der Beratungsfunktion der Ethikkommission durch unsere Stellungnahme unberührt.

Mit freundlichem Gruß und besten Wünschen
für den weiteren Verlauf Ihrer Forschung bin ich

Ihr

Prof. Dr. med. Dr. phil. H. Raspe
Vorsitzender



UNIVERSITÄT ZU LÜBECK

Universität zu Lübeck · Ratzeburger Allee 160 · 23538 Lübeck

Herrn
Prof. Dr. med. Martin Reck
LungenClinic Grosshansdorf GmbH
Wöhrendamm 80

22927 Großhansdorf

Ethik-Kommission

Vorsitzender:
Herr Prof. Dr. med. Alexander Katalinic
Stellv. Vorsitzender:
Herr Prof. Dr. med. Frank Gieseler

Ratzeburger Allee 160
23538 Lübeck

Sachbearbeitung: Frau Janine Erdmann
Tel.: +49 451 3101 1008
Fax: +49 451 3101 1024
ethikkommission@uni-luebeck.de

Aktenzeichen: 12-238

Datum: 29. November 2016

RETARGET: Aufbau einer Kohorte von Patienten mit Lungenkarzinom mit wiederholter Entnahme von Gewebe und Blut zum verbesserten biologischen Verständnis der Erkrankung sowie einer verbesserten Therapiesteuerung

Hier: Amendment – Ihr Schreiben vom 22. November 2016

die Ethik-Kommission nimmt das Amendment zustimmend zur Kenntnis.

Folgende Unterlagen haben vorgelegen:

- Zustimmendes Primärvotum der Ethik-Kommission der Medizinischen Fakultät in Heidelberg vom 24. Mai 2012
- Modifiziertes Studienprotokoll (gilt für beide Kliniken) vom 19. September 2016
- Modifizierte Patienteninformationen und Einwilligungserklärungen „Gewebeentnahme bei Erstdiagnose in der Version 16. vom 19. September 2016 und „Gewebeentnahme im Erkrankungsverlauf“ in der Version 1.5 vom 19. September 2016 der Thoraxklinik Heidelberg
- die entsprechend angepassten Patienteninformationen und Einwilligungserklärungen der LungenClinic Grosshansdorf in der Version 2.0 vom 10. November 2016 sowie
- das zustimmende Votum der Ethik-Kommission der Universität Heidelberg vom 17. Oktober 2016 mit den darin gegebenen Hinweisen.

Mit freundlichen Grüßen


Prof. Dr. med. Alexander Katalinic
Vorsitzender



Medizinische Fakultät Heidelberg

Ethikkommission der Med. Fak. HD | Alte Glockengießerei 11/1 | D-69115 Heidelberg

Herrn Prof. Dr. med. Michael Thomas
Abt. Internistische Onkologie der Thoraxtumoren
Thoraxklinik am Universitätsklinikum Heidelberg
Amalienstr. 5
69126 Heidelberg

24.05.2012
ts/vpf

VOTUM - Berufsrechtliche Beratung nach inhaltlicher Nachreichung
(Zweitschrift: ersetzt das Votum vom 26.04.2012)

Unser Zeichen: S-048/2012 (Bitte stets angeben)

Titel: „Aufbau einer Kohorte von Patienten mit Lungenkarzinom mit
wiederholter Entnahme von Gewebe und Blut zum verbesserten
biologischen Verständnis der Erkrankung, sowie einer verbesserten
Therapiesteuerung“

Eingereichte Unterlagen: Primäreinreichung:
Anschreiben vom 31.01.2012
Zusammenfassung
Checkliste Sonstige Studien
Formular für Erstantrag vom 31.01.2012
Patienteninformation und Einwilligungserklärung Erstbiopsie Version
1.0 vom 31.01.2012
Patienteninformation und Einwilligungserklärung wiederholte
Gewebeentnahme Version 1.0 vom 31.01.2012
Studienprotokoll Version 1.0 vom 04.01.2012
CV Univ.-Prof. Dr. med. Michael Thomas vom 31.01.2012
E-Mail vom 01.02.2012
Inhaltliche Nachreichung:
E-Mail vom 27.03.2012
Anschreiben vom 22.03.2012
Studienprotokoll Version 1.1 vom 22.03.2012
Patienteninformation und Einwilligungserklärung Erstbiopsie,
Version 1.1 vom 22.03.2012
Patienteninformation und Einwilligungserklärung wiederholte
Gewebeentnahme, Version 1.1 vom 22.03.2012

Sehr geehrter Herr Professor Thomas,

mit den Änderungen bzw. Ergänzungen in den oben näher bezeichneten Dokumenten
wurden die mit Schreiben vom 21.03.2012 genannten Empfehlungen der
Ethikkommission berücksichtigt.

Die Ethikkommission hat damit keine Bedenken gegen Ihr Vorhaben.



Alte Glockengießerei 11/1
D-69115 Heidelberg

+49 6221 33822-0 (Empfang)
+49 6221 3382222
ethikkommission-l@med.uni-heidelberg.de

www.medizinische-fakultaet-hd.uni-heidelberg.de/
ethikkommission

Vorsitz:
Prof. Dr. med. Thomas Strowitzki

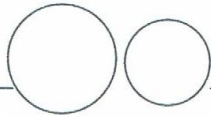
Stellv. Vorsitz:
Prof. Dr. med. Johannes Schröder
Prof. Dr. med. Klaus Herfarth

Geschäftsleitung:
Dr. med. Verena Pfeilschifter

Sonstige Studien:
Damir Jurincic, M.A.
+49 6221 33822-0
+49 6221 3382222
Damir.Jurincic@med.uni-heidelberg.de



Bankverbindung:
Baden-Württembergische Bank Stuttgart
Konto-Nr.: 7421 500 429
BLZ: 600 501 01
SWIFT/BIC Code: SOLADEST
IBAN-Nr.: DE 64600501017421500429



Ferner bestätigt wird, dass für das beantragte Forschungsvorhaben ein zwingendes Bedürfnis im Sinne des § 24 Abs. 1 Nr. 1 StrlSchV besteht, da die bisherigen Forschungsergebnisse und die die medizinischen Erkenntnisse nicht ausreichen.

Wir wünschen Ihnen bei der Durchführung der Studie viel Erfolg.

Bitte leiten Sie das Ergebnis der berufsrechtlichen Beratung und die studienrelevante Korrespondenz allen teilnehmenden Ärzten in unserem Zuständigkeitsbereich weiter.

Mit freundlichen Grüßen

Digital unterschrieben von Dr. Thomas
Strowitzki FN
DN: cn=Dr. Thomas Strowitzki FN,
pseudonym=Dr. Thomas Strowitzki FN,
serialNumber=00100000000119542300
02
Datum: 2012.05.25 13:50:03 +02'00'

Prof. Dr. med. Thomas Strowitzki
Vorsitzender

Allgemeine Hinweise:

- Änderungen in Organisation und Ablauf der Studie sind der Kommission, zusammen mit einer Bewertung der Nutzen-Risiko-Relation, umgehend mitzuteilen. Sowohl die Antragsnummer als auch die geänderten Passagen sollten in den betreffenden Unterlagen deutlich gekennzeichnet sein, da anderenfalls keine zügige Bearbeitung möglich ist.
- Die Ethikkommission der Medizinischen Fakultät Heidelberg arbeitet gemäß den nationalen gesetzlichen Bestimmungen und den ICH-GCP-Richtlinien. Ihren Beratungen liegt die Deklaration des Weltärztebundes von Helsinki in der jeweils aktuellen Fassung zugrunde.
- Unabhängig vom Beratungsergebnis macht die Ethikkommission Sie darauf aufmerksam, dass die ethische und rechtliche Verantwortung für die Durchführung einer Studie beim Leiter der Studie und bei allen teilnehmenden Ärzten liegt.



UNIVERSITÄT ZU LÜBECK

Universität zu Lübeck - Ratzeburger Allee 160 - 23538 Lübeck

Herrn
Prof. Dr. med. Kopp
Klinik für Kinder- und Jugendmedizin

im Hause

CC Prof. Zabel, PD Dr. Drömann
Med. Klinik III im Haus

Ethik-Kommission

Vorsitzender:

Herr Prof. Dr. med. Alexander Katalinic
Universität zu Lübeck
Stellv. Vorsitzender:
Herr Prof. Dr. med. Frank Gieseler
Ratzeburger Allee 160
23538 Lübeck

Sachbearbeitung: Frau Janine Erdmann
Tel.: +49 451 500 4639
Fax: +49 451 500 3026
ethikkommission@uni-luebeck.de

Aktenzeichen: 12-215
Datum: 05. Dezember 2013

Kinder Register Asthma (KIRA-Studie) - Deutsche Asthmakohorte im Deutschen Zentrum für Lungenforschung (DZL)

Hier: Amendment – Ihr Schreiben vom 02. Dezember 2013

Sehr geehrter Herr Prof. Kopp,


das Amendment vom 02. Dezember 2013 mit der beschriebenen Erweiterung des KIRA Registers um erwachsene Patienten mit bekannter Asthmaerkrankung (Erwachsenen Register Asthma ERA-Studie) nehme ich zustimmend zur Kenntnis.

Folgende Unterlagen lagen vor:

- Anlage 1 – Verantwortlicher Leiter, beteiligte Ärzte und Wissenschaftler, Prüfstellen
- Anlage 2 – Studienpopulation, Ein- und Ausschlusskriterien
- Anlage 3 – Änderung des Untersuchungsplans
- Anlage 4 – Aufklärungsmaterialien
- Anlage 5 – Anamnesefragebogen Basis und Follow Up
- Anlage 6 – Asthma Control Test, Asthma Control Questionnaire, Asthma Quality of Life Questionnaire
- Anlage 7 – Versicherungsunterlagen.

Es bedarf keiner weiteren Begutachtung durch die Kommission.

Mit freundlichem Gruß bin ich


Prof. Dr. med. Alexander Katalinic
Vorsitzender



UNIVERSITÄT ZU LÜBECK

Universität zu Lübeck · Ratzeburger Allee 160 · 23538 Lübeck

Forschungszentrum Borstel
Frau PD Dr. Karoline Gaede
Parkallee 35
23845 Borstel

nachrichtlich:

Herrn Prof. Dr. med. Torsten Goldmann
FZB Borstel, Parkallee 35 in 23845 Borstel
Herrn Prof. Dr. med. Martin Reck, E-Mail: m.reck@lungenclinik.de

Ethik-Kommission

Vorsitzender:

Herr Prof. Dr. med. Alexander Katalinic
Universität zu Lübeck

Stellv. Vorsitzender:

Herr Prof. Dr. med. Frank Gieseler
Ratzeburger Allee 160
23538 Lübeck

Sachbearbeitung: Frau Janine Erdmann

Tel.: +49 451 3101 1008

Fax: +49 451 3101 1024

ethikkommission@uni-luebeck.de

Aktenzeichen: 17-350

Datum: 01. Dezember 2017

Einfluss von Zigaretten und E-Zigaretten auf humane Lungenzellen

Hier: Ihr Schreiben vom 17.11.2017 mit Anlagen

Sehr geehrte Frau Dr. Gaede, sehr geehrter Herr Prof. Reck und Prof. Goldmann,

Ihr Antrag wurde unter berufsethischen, medizinisch-wissenschaftlichen und berufsrechtlichen Gesichtspunkten im verkürzten Verfahren geprüft.

Gegen die Abgabe von Biomaterialproben und Daten aus der BMB Nord in pseudonymisierter Form an Herrn Prof. Dr. med. Torsten Goldmann, FZB Borstel werden keine Bedenken gesehen.

Bei Änderung des Studiendesigns sollte der Antrag erneut vorgelegt werden.

Die ärztliche und juristische Verantwortung des Studienleiters und der an der Studie teilnehmenden Ärzte bleibt entsprechend der Beratungsfunktion der Ethikkommission durch unsere Stellungnahme unberührt.

Mit freundlichen Grüßen

Prof. Dr. med. Alexander Katalinic
Vorsitzender

10. Danksagung

An dieser Stelle möchte ich mich bei allen Menschen bedanken, die mich während der Doktorarbeit unterstützt haben.

Besonders möchte ich mich bei meinem Betreuer, Prof. Dr. rer. nat. Ole Ammerpohl, für das in mich gesetzte Vertrauen, die Bereitstellung spannender wissenschaftlicher Projekte, die herausragende Betreuung sowie für die viele Zeit und Geduld bedanken.

Bei Frau Prof. Dr. rer. nat. Manuela Dittmar möchte ich mich für die Übernahme der Betreuung meiner Arbeit vor der Mathematisch-Naturwissenschaftlichen Fakultät der „Christian-Albrechts-Universität zu Kiel“ bedanken.

Des Weiteren möchte ich mich bei Prof. Dr. med. Reiner Siebert und Prof. Dr. med. Almuth Caliebe für die Möglichkeit bedanken, meine Doktorarbeit am „Institut für Humangenetik, Christian-Albrechts-Universität zu Kiel“ anfertigen zu können. Allen Mitarbeitern des Instituts danke ich für die Zusammenarbeit, Hilfsbereitschaft und das angenehme Arbeitsklima. Mein besonderer Dank für die experimentelle Unterstützung während der Doktorarbeit gilt Lorena Valles, Benedikte Willert und Gabi Riesen.

Dr. rer. nat. Helene Kretzmer danke ich für die Software, Skripte und vor allem für die Vermittlung bioinformatischer Grundlagen.

Weiterhin gilt mein Dank allen DZL-Kollaborationspartnern, ohne die diese Arbeit nicht möglich gewesen wäre. Vor allem bedanke ich mich bei Prof. Dr. med. Martin Reck, Prof. Dr. rer. nat. Torsten Goldmann, Dr. rer. nat. Iris Watermann, Dr. rer. nat. Sebastian Marwitz und Dr. rer. nat. Lena Heinbockel.

Schließlich möchte ich mich bei meinem Mann, Sebastian Scheufele, Freunden und Familie bedanken, die mich während dieser Lebensphase immer unterstützt und motiviert haben.

11. Publikationsliste

11.1 Publierte Originalarbeiten

Marwitz S., Heinbockel L., Scheufele S., Nitschkowski D., Kugler C., Perner S., Reck M., Ammerpohl O., Goldmann T., *Epigenetic modifications of the VGF gene in human non-small cell lung cancer tissues pave the way towards enhanced expression*. Clin Epigenetics, 2017. **9**: p. 123.

Marwitz S., Scheufele S., Perner S., Reck M., Ammerpohl O., Goldmann T., *Epigenetic modifications of the immune-checkpoint genes CTLA4 and PDCD1 in non-small cell lung cancer results in increased expression*. Clin Epigenetics, 2017. **9**: p. 51

Ammerpohl O., Scheufele S., Siebert R., *Analysen epigenetischer Marker aus Liquid Biopsies: Informationen von jenseits des Genoms*. Medgen, 2016. **28**: p. 251

Marwitz S., Depner S., Dvornikov D., Merkle R., Szczygiel M., Muller-Decker K., Lucarelli P., Wasch M., Mairbaur H., Rabe K. F., Kugler C., Vollmer E., Reck M., Scheufele S., Kröger M., Ammerpohl O., Siebert R., Goldmann T., Klingmüller U., *Downregulation of the TGFbeta pseudoreceptor BAMBI in non-small cell lung cancer enhances TGFbeta signaling and invasion*. Cancer Res, 2016. **76**(13): p. 3785-801.

11.2 Eingereichte Originalarbeiten

Marwitz S., Heinbockel L., Scheufele S., Kugler C., Reck M., Rabe K. F., Perner S., Goldmann T., Ammerpohl O., *Fountain of youth for squamous cell carcinomas? On the epigenetic age of NSCLC and corresponding tumor free lung tissues*. eLife, 2017

Marwitz S., Heinbockel L., Scheufele S., Nitschkowski D., Kugler C., Perner S., Reck M., Ammerpohl O., Goldmann T., *VGF promotes EGFR-TKI resistance and EMT*. Cancer Res, 2017

11.3 Poster

Scheufele S., Marwitz S., Reinmuth N., Kröger M., Kugler C., Zabel P., Vollmer E., Goldmann T., Reck M., Siebert R., Ammerpohl O., *Aberrant patterns of DNA methylation in lung cancer*. Medgen, 2015. **28**: p. 120 (26. GfH-Kongress, April 2015, Graz, Österreich).

Scheufele S., Kröger M., Willert B., Seehase S., Marwitz S., Reinmuth N., Kugler C., Zabel P., Vollmer E., Goldmann T., Reck M., Siebert R., Ammerpohl O., *Aberrant chromatin status and DNA methylation in lung cancer*. DZL Annual Meeting 2016 Abstract Book, 2016. **1**: p.177 (DZL Annual Meeting, Februar 2016, Hannover, Deutschland).

Scheufele S., Kröger M., Marwitz S., Reinmuth N., Kugler C., Zabel P., Vollmer E., Goldmann T., Reck M., Siebert R., Ammerpohl O., *Aberrant DNA methylation in the diagnostics of lung cancer*. Medgen, 2016. **28**: p. 140 (27. GfH-Kongress, März 2016, Lübeck, Deutschland).

Bhat J., Scheufele S., Willert B., Siebert R., Kabelitz D., Ammerpohl O., *Mapping the accessible genome of human T-cell subsets by ATAC-sequencing*, Medgen, 2016. **28**: p. 125 (27. GfH-Kongress, März 2016, Lübeck, Deutschland).

Scheufele S., Kretzmer H., Kröger M., Marwitz S., Watermann I., Reinmuth N., Kugler C., Zabel P., Vollmer E., Perner S., Goldmann T., Reck M., Siebert R., Ammerpohl O., *Aberrant DNA methylation in the diagnostics of non-small cell lung cancer*. DZL Annual Meeting 2017 Abstract Book, 2017. **1**: p. 259 (DZL Annual Meeting, Januar 2017, München, Deutschland).

12. Eidesstattliche Erklärung

Hiermit erkläre ich, dass die vorliegende Dissertation nach Inhalt und Form eigenständig unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft angefertigt wurde. Dabei wurden abgesehen von der wissenschaftlichen Beratung meiner akademischen Betreuer nur die von mir angegebenen Quellen und Hilfsmitteln verwendet. Des Weiteren versichere ich, dass die vorliegende Arbeit weder ganz noch zum Teil an einer anderen Stelle im Rahmen eines Prüfungsverfahrens vorgelegt wurde und ich selbst zuvor noch keine Promotionsversuche unternommen habe.

Einige Teile der vorliegenden Arbeit wurden im Rahmen folgender wissenschaftlicher Kongresse veröffentlicht bzw. als Poster präsentiert:

- Scheufele S., Marwitz S., Reinmuth N., Kröger M., Kugler C., Zabel P., Vollmer E., Goldmann T., Reck M., Siebert R., Ammerpohl O., *Aberrant patterns of DNA methylation in lung cancer*. Medgen, 2015. **28**: p. 120 (26. GfH-Kongress, April 2015, Graz, Österreich).
- Scheufele S., Kröger M., Marwitz S., Reinmuth N., Kugler C., Zabel P., Vollmer E., Goldmann T., Reck M., Siebert R., Ammerpohl O., *Aberrant DNA methylation in the diagnostics of lung cancer*. Medgen, 2016. **28**: p. 140 (27. GfH-Kongress, März 2016, Lübeck, Deutschland).
- Scheufele S., Kröger M., Willert B., Seehase S., Marwitz S., Reinmuth N., Kugler C., Zabel P., Vollmer E., Goldmann T., Reck M., Siebert R., Ammerpohl O., *Aberrant chromatin status and DNA methylation in lung cancer*. DZL Annual Meeting 2016 Abstract Book, 2016. **1**: p.177 (DZL Annual Meeting, Februar 2016, Hannover, Deutschland).
- Scheufele S., Kretzmer H., Kröger M., Marwitz S., Watermann I., Reinmuth N., Kugler C., Zabel P., Vollmer E., Perner S., Goldmann T., Reck M., Siebert R., Ammerpohl O., *Aberrant DNA methylation in the diagnostics of non-small cell lung cancer*. DZL Annual Meeting 2017 Abstract Book, 2017. **1**: p. 259 (DZL Annual Meeting, Januar 2017, München, Deutschland).

Kiel, den 20.02.2018

Swetlana Scheufele