Aus dem Deutschen Krebsforschungszentrum Heidelberg
(Vorstand: Prof. Dr. med. Michael Baumann, Ursula Weyrich)

Genome-wide epigenomic analyses of cell-free DNA from *anaplastic lymphoma kinase*-rearranged non-small cell lung cancer patients

Inauguraldissertation
zur Erlangung des Doctor scientiarum humanarum (Dr. sc. hum.)
an der
Medizinischen Fakultät Heidelberg
der
Ruprecht-Karls-Universität

vorgelegt von
Florian Janke

aus
Mannheim

2021

Dekan: Prof. Dr. Hans-Georg Kräusslich
Doktorvater: Prof. Dr. Holger Sültmann

Table of Contents

List of Figures

Supplementary Figures

List of Tables

Supplementary Tables

Abbreviations

5C	unmodified cytosine
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
ALK	anaplastic lymphoma kinase
ALK$^+$	ALK-driven
BER	base excision repair
bp	base pairs
BRAF	v-raf murine sarcoma viral oncogene homolog B
BRCA	breast carcinoma
CAPP-seq	Cancer Personalized Profiling by deep Sequencing
CDKN2A	cyclin-dependent kinase inhibitor 2A
cfDNA	cell-free DNA
cfMeDIP	cell-free methylation DNA immunoprecipitation
cfRNA	cell-free RNA
CGI	CpG island
ChIP-seq	chromatin immunoprecipitation DNA-sequencing
CNA	copy number alteration
COAD	colon adenocarcinoma
CpG	cytosine-guanine dinucleotide
CPM	counts per million
CT	computed tomography
CTC	circulating tumor cell, circulating tumor cell
ctDNA	circulating tumor DNA
CUP	cancer of unknown primary
DAPK	death-associated protein kinase
dATP	deoxyadenosine triphosphate
DBCO-PEG$_4$-biotin	dibenzylcyclooctyne-polyethylenglycol-4-distearylether-biotin
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
DhMR	differentially hydroxymethylated region
DMR	differentially methylated region
DNMT	DNA methyltransferase
dNTP	deoxyribonucleotide triphosphate
dsDNA	double-stranded DNA
dTTP	deoxythymidine triphosphate
EDTA	ethylenediamine tetraacetic acid
EGFR	epidermal growth factor receptor
EML4	echinoderm microtubule-associated protein-like 4
EPO	erythropoietin
FPKM	fragments per kilobase per million reads

Abbreviations

PR ....*partial response*
PRAD...*prostate adenocarcinoma*
PRKAR1A ....*protein kinase CAMP-dependent type I regulatory subunit alpha*
qPCR ....*quantitative PCR*
RASSF1A ....*Ras associated domain family member 1A*
RET..*RET proto-oncogene*
ROS1 ....*ROS proto-oncogene*
RT ....*room temperature*
RTK ....*receptor tyrosine kinase*
SCF...*stem cell factor*
SCLC ....*small cell lung cancer*
SD..*stable disease*
SNV ....*single nucleotide variant*
STK11...*serine/threonine kinase 11*
sWGS ....*shallow whole-genome sequencing*
T4-BGT...*T4 phage ß-glucosyltransferase*
T790M ....*threonine-to-methionine substitution in codon 790*
TCGA ....*The Cancer Genome Atlas*
TD ....*trimerization domain*
TDG...*thymine DNA glycosylase*
TES ....*transcription end sites*
TET ....*ten-eleven translocation dioxygenase*
TF...*transcription factor*
TKI..*tyrosine kinase inhibitor*
t-MAD ....*trimmed median absolute deviation from copy number neutrality*
TMB ....*tumor mutational burden*
TMM ....*trimmed mean of M-values*
TP53...*tumor protein p53*
TPR ....*translocated promoter region, nuclear basket protein*
TSS ....*transcription start site*
UDP-glucose ....*uridine diphosphoglucose*
UTR ....*untranslated region*
V1...*variant 1*
V2...*variant 2*
V3...*variant 3*
VAF ....*variant allele frequency*
VAF_{max} ....*maximal variant allele frequency*
voom ....*variance modeling on the observation level*
WD ....*tryptophan-aspartic acid repeats*
WES ....*whole-exome sequencing*
WGS..*whole-genome sequencing*
ZNF ....*zinc finger nuclease*

X

1 Introduction

1.1 Lung cancer

Lung cancer is the leading cause of cancer-related mortality worldwide. With an estimated 1.8 million fatalities each year, it is responsible for nearly 1 in 5 (18.4%) cancer deaths (Bray et al. 2018). Tobacco usage is the most frequent cause of the disease, accounting for >80% of cases in countries were smoking is common (Alberg et al. 2013). Strategies for eradicating tobacco usage, like nicotine-replacement therapy, counselling or taxation are key to the global fight against lung cancer (Hays and Ebbert 2008). However, besides prevention, increasingly sophisticated measures for early disease detection as well as advancements in therapy are required to lower the number of fatalities. The need for a timely detection of lung cancer becomes evident when comparing the estimated 5-year survival of lung cancer patients at different pathological stages. Approximately two thirds (65.7%) of lung cancer patients present with locally advanced (stage III) or metastatic disease (stage IV) (Morgensztern et al. 2010). At these stages, the 5-year survival rate is only 25% and 5%, respectively, whereas earlier diagnoses have much more favorable outcomes (stage I: 80% and stage II: 57% 5-year survival rate) (Goldstraw et al. 2016). Benefits of diagnosis when the disease is still localized have been demonstrated in several studies performing low-dose computed tomography (LDCT) in individuals with high-risk to develop lung cancer. These studies resulted in a mortality reduction of up to 20% in LDCT-screened participants, however, at the expense of high rates of overdiagnosis and potential harm due to radiation (de Koning et al. 2020; Field et al. 2016; National Lung Screening Trial Research et al. 2011).

Therapy decision in lung cancer is influenced by several factors, including disease stage, general medical condition of the patient, as well as histology of the tumor and its molecular profile. Surgery, often in combination with perioperative chemo- and/or radiotherapy, is the recommended treatment for early-stage and some locally advanced lung cancers (National Institute for Health and Care Excellence 2019; Vansteenkiste et al. 2014). In advanced stages, histological assessment of the tumor is pivotal for therapy decision (Vansteenkiste et al. 2014). Histopathologically, lung cancer can be divided into small cell lung cancer (SCLC), representing ~15% of patients, and non-small cell lung cancer (NSCLC; ~85% of patients). The latter is further subdivided into adenocarcinoma (LUAD; 50.4%), squamous cell carcinoma (LUSC; 22.6%) and large cell carcinoma (LCC; 1.3%) (Figure 1A) (Howlader et al. 2017). Each histological subtype is associated with discrete treatment guidelines. For advanced-stage SCLC treatment options are limited to platinum-based chemotherapy and radiotherapy, NSCLC patients can profit from a wider range of therapies (National Institute for Health and Care Excellence 2019). For example, a subset of NSCLC patients with a programmed death-ligand 1 (PD-L1) tumor proportion score ≥50% can be treated with immune checkpoint inhibitors while molecular alterations, such as a mutation in the epidermal growth factor receptor (EGFR), may render the patient susceptible to targeted

therapy using tyrosine kinase inhibitors (TKIs). Both treatment options were demonstrated to result in substantially prolonged overall (OS) and progression-free survival (PFS) compared to platinum-based cytotoxic therapy alone, highlighting the importance of therapeutic advancements in lung cancer (Brahmer et al. 2015; Excellence; 2019; Langer et al. 2016; Mok et al. 2009; Reck et al. 2016; Shaw and Solomon 2015; Solomon et al. 2014).
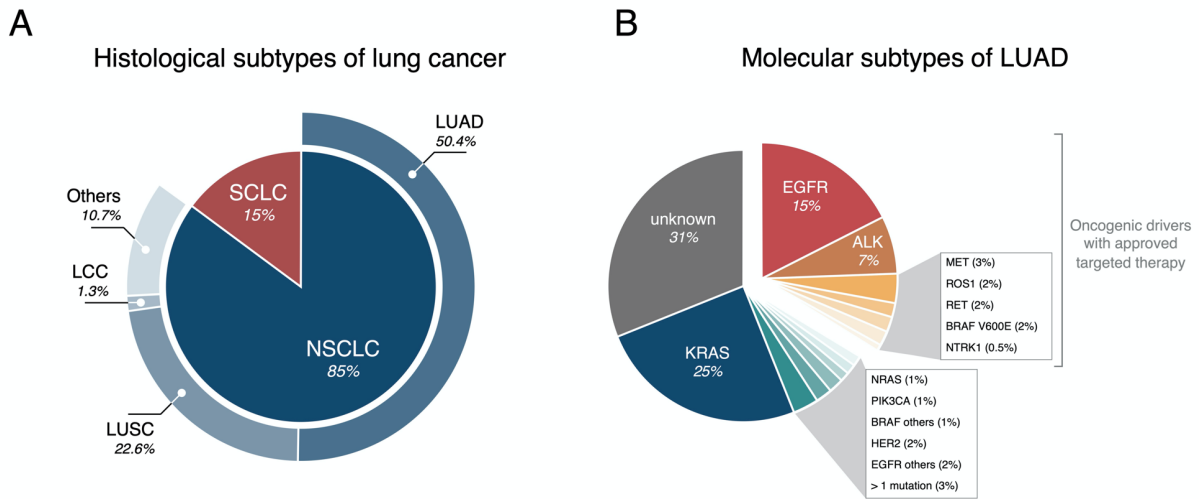
Figure 1: Histological and molecular subtypes of lung cancer. (A) Lung cancer is histologically classified into small cell (SCLC) and non-small cell lung cancer (NSCLC). NSCLC can be further subdivided into lung adenocarcinoma (LUAD), lung squamous cell carcinoma (SCC) and large cell carcinoma (LCC). Data source: (Howlader et al. 2017). (B) Frequency of driver mutations in LUAD. Oncogenic drivers with approved targeted therapy are highlighted. Adapted from Hirsch *et al.* 2016. ALK, anaplastic lymphoma kinase; BRAF, v-raf murine sarcoma viral oncogene homolog B; *EGFR*, epidermal growth factor receptor; *HER2*, Erb-B2 receptor tyrosine kinase 2; *KRAS*, v-Kir-Ras2 Kirsten rat sarcoma viral oncogene homolog; *MET*, MET proto-oncogene; *NRAS*, neuroblastoma RAS viral oncogene homolog; *NTRK1*, neurotrophic tyrosine kinase receptor type 1; *PIK3CA*, phosphatidylinositol-4,5-bisphosphate 3-kinase; *RET*, RET proto-oncogene; *ROS1*, ROS proto-oncogene 1.

1.2 Molecular subtypes and precision therapy in lung adenocarcinoma

The emergence of next-generation sequencing (NGS) approaches expanded the histology-based classification of LUAD towards considering the molecular profile of the tumor. Large cooperative studies identified a variety of genomic alterations in LUAD that led to a better understanding of the molecular causes of the disease and allowed a more detailed classification based on the presence of mutually exclusive oncogenic drivers (Figure 1B) (Campbell et al. 2016; Collisson et al. 2014). The mutational landscape of LUAD is very heterogeneous. With a median frequency of 8.7 somatic mutations per megabase (Mb), LUAD ranked third among cancers with the highest tumor mutational burden (TMB) (Alexandrov et al. 2013; Lawrence et al. 2014). However, mutation frequencies can vary considerably between patients, ranging from 0.1 to 100 mutations per Mb (Lawrence et al. 2013b). Recurrent molecular alterations in LUAD include silencing mutations and deletions of tumor suppressor genes, such as tumor protein p53 (*TP53*; 46% of LUADs), kelch-like ECH-associated protein 1 (*KEAP1*; 17%) or serine/threonine kinase 11 (*STK11*; 19%) (Collisson et al. 2014). Additionally, pathways regulating cell proliferation and survival often become constitutively activated through genomic aberrations. The receptor tyrosine kinase (RTK)/Ras/Raf signaling pathway harbors carcinogenic alterations in 75 to 85% of LUADs (Campbell et al. 2016). Activating mutations in the v-Kir-Ras2 Kirsten rat sarcoma viral oncogene homolog (*KRAS)* and *EGFR* are the most common alterations in this pathway, occurring in 25 and 15% of cases, respectively (Hirsch et al. 2016). Moreover, oncogenic rearrangements of the anaplastic lymphoma kinase (*ALK* (Soda et al. 2007)), ROS proto-oncogene 1 (*ROS1* (Bergethon et al. 2012)), RET proto-oncogene (*RET* (Wang et al. 2012a)), and neurotrophic tyrosine kinase receptor type (*NTRK* (Vaishnavi et al. 2013)) as well as amplifications of the MET proto-oncogene (*MET* (Awad et al. 2016)) and Erb-B2 receptor tyrosine kinase 2 (*HER2* (Pillai et al. 2017)) have been described to promote tumorigenesis.

In the early 2000s, it was discovered that LUAD patients with activating mutations in the tyrosine kinase domain of *EGFR* exhibit drastic clinical response to the *EGFR*-specific TKI, gefitinib (Lynch et al. 2004; Paez et al. 2004). Since then, lung cancer – and in particular LUAD – has emerged as a cancer entity in which precision medicine has been transformative. As of today, targeted treatment using TKIs is approved for seven oncogenic driver genes, comprising *EGFR* (Maemondo et al. 2010; Soria et al. 2018; Yang et al. 2015), *ALK* (Shaw et al. 2017; Solomon et al. 2014), *MET* (Frampton et al. 2015; Wolf et al. 2020), *ROS1* (Shaw et al. 2017; Shaw et al. 2014), *RET* (Ackermann et al. 2019), *BRAF* (Planchard et al. 2017), and *NTRK1* (Vaishnavi et al. 2013). As a result, almost one third of LUAD patients can potentially be treated with TKI-based therapy (Figure 1B). A multitude of studies has proven the superior outcome of patients treated with TKIs when compared to conventional chemotherapy. Although targeted therapies in LUAD can be very effective, objective

response and duration of response differ substantially between patients (Camidge et al. 2020; Mok et al. 2009; Shaw et al. 2014; Soria et al. 2018). Here, inherent and acquired resistance mechanisms are major confounding factors, limiting the efficacy of TKI therapy. Those resistance mechanisms include alterations in the targeted driver gene (i.e. secondary mutations or amplification of the tumor driver), mutation or amplification of other oncogenes (e.g. *MET* or *BRAF* mutations in *EGFR*-driven tumors), and/or activation of bypass signaling pathways (e.g. *EGFR*, *HER2* or *KIT* signaling) (Camidge et al. 2014; Gainor et al. 2016). As a consequence, targeted agents are constantly developed and refined to cope with the emergence of resistances. For instance, the most common acquired resistance under first generation *EGFR* inhibitor treatment is the secondary mutation T790M (threonine-to-methionine substitution in codon 790). Due to the development of the third generation TKI, osimertinib, T790M mutated patients can still benefit from a targeted therapy (Mok et al. 2017). Moreover, osimertinib demonstrated superior outcome in therapy naïve patients and is now considered the preferred option as first-line therapy for *EGFR*-mutated NSCLC (Ramalingam et al. 2019).

1.3 *ALK* rearrangements in lung adenocarcinoma

1.3.1 Biology of ALK⁺ tumors

The ALK tyrosine kinase receptor was first described as a contributor to malignant transformation in anaplastic large-cell non-Hodgkin´s lymphomas (Morris et al. 1994). While its role under normal physiological conditions is not yet fully elucidated, *ALK* rearrangements have been extensively characterized in the context of LUAD.

Oncogenic *ALK* in LUAD arises from its fusion to the echinoderm microtubule-associated protein-like 4 (*EML4*) gene. EML4-ALK fusions occur due to a paracentric inversion on chromosome 2. Here, the *ALK* gene is truncated close to the 5´-end of exon 20, excluding the extracellular domain and the transmembrane helix, while preserving the entire *ALK* kinase-coding domain in the fusion gene. The breakpoint in *EML4* is variable and thereby responsible for the existence of multiple EML4-ALK fusion variants. The most common breakpoints are proximal to exon 13 (variant 1; V1), exon 20 (variant 2; V2), and exon 6 (variant 3; V3). All variants retain the trimerization domain of *EML4*, which confers ALK autophosphorylation, leading to its constitutive activation (Figure 2A) (Sanders et al. 2011; Soda et al. 2007). Notably, *EML4* is not the only fusion partner of *ALK*. Other fusion genes include *TPR* (translocated promoter region, nuclear basket protein (Choi et al. 2014)), *KIF5B* (kinesin family member 5B (Takeuchi et al. 2009)) and *PRKAR1A* (protein kinase *CAMP*-dependent type I regulatory subunit alpha (Ali et al. 2016)). *In vitro* studies showed that EML4-ALK fusion proteins interact with a complex network of proteins and affect several downstream signaling pathways, including RTK/Ras/Raf, JAK/STAT and PI3K/AKT. The activation of these pathways promotes tumorigenesis, driving aberrant cell proliferation,

survival and angiogenesis (Figure 2B) (Chiarle et al. 2008; Zhang et al. 2016). A profound understanding of the downstream effects of EML4-ALK fusions is crucial for the development and improvement of ALK-directed therapies. Additionally, the dysregulation of pathways during chronic TKI exposure could advance the understanding of therapy resistance.
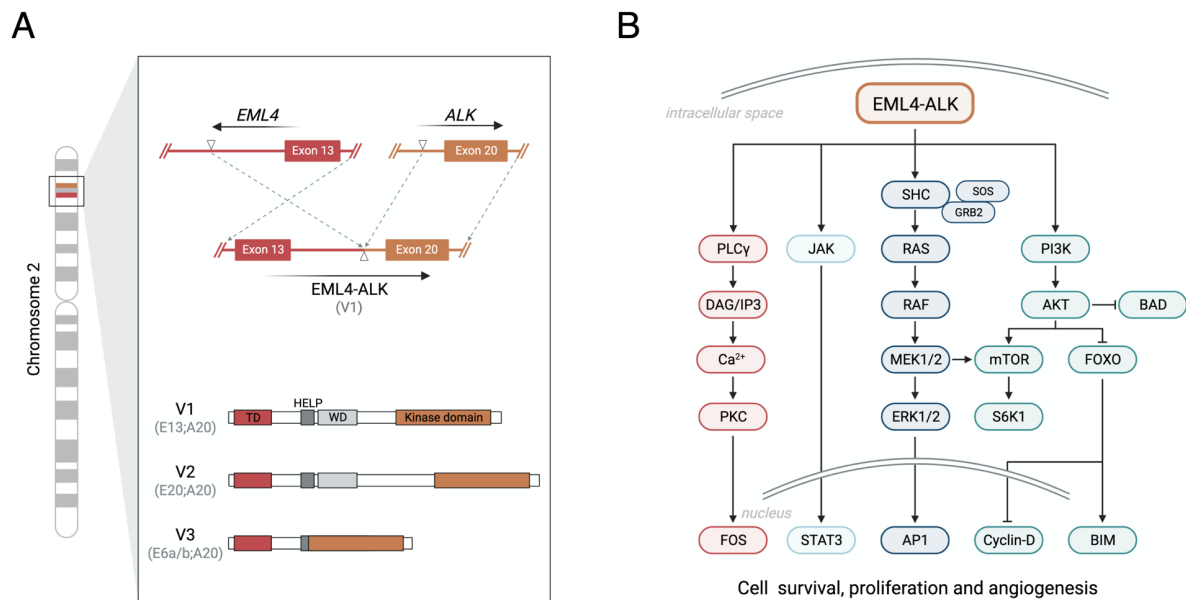
Figure 2: Oncogenic EML4-ALK fusion and the activated downstream signaling. (A) Paracentric inversion within chromosome 2p, fusing the echinoderm microtubule-associated protein-like 4 (*EML4*) gene to the tyrosine kinase domain of the anaplastic lymphoma kinase (*ALK*). EML4-ALK fusion proteins variant 1 (V1), variant 2 (V2) and variant 3 (V3) are displayed below. Adapted from Manicone *et al.* 2017. (B) Constitutive activation of the EML4-ALK fusion protein results in constant signaling of PLCγ, JAK-STAT, RTK/Ras/Raf, and PI3K-AKT pathways. This leads to aberrant regulation of a number of genes ultimately driving cellular survival, proliferation, and angiogenesis. Adapted from Lin *et al.* 2017. HELP, hydrophobic EML protein; TD, trimerization domain; WD, tryptophan-aspartic acid repeats.

1.3.2 ALK-directed therapy and therapy resistance

With a prevalence of 4 to 7%, ALK-driven (ALK$^+$) tumors represent the second most abundant molecular subtype in advanced LUAD with approved targeted treatment regimens (Hirsch et al. 2016; Skoulidis and Heymach 2019). Prior to the availability of ALK inhibitors, metastatic ALK$^+$ LUAD was associated with poor patient outcomes (median OS <1.5 years) (Pilkington et al. 2015). Today, the median life expectancy of those patients exceeds five years under sequential TKI administration (Duruisseaux et al. 2017). In 2011, the first ALK inhibitor, crizotinib, was approved by the U.S. Food and Drug Administration for the treatment of advanced ALK$^+$ LUAD. In the following years, second (i.e. ceritinib, alectinib, and brigatinib) and third generation ALK-TKIs (i.e. lorlatinib) have been developed and approved for therapy in metastatic ALK$^+$ LUAD. Compared to crizotinib, all second and third generation TKIs are more potent and exhibit improved efficiency in the brain – a common site of metastases in ALK$^+$ patients. Despite the elongated PFS and OS in ALK-TKI-*vs.* chemotherapy-receiving patients, tumors inevitably relapse due to the development of drug resistances (Figure 3) (Shaw et al. 2013; Solomon et al. 2014). Some of the drug resistances developed under crizotinib treatment can be overcome by subsequent administration of the more potent TKIs. For example, the most prevalent on-target resistance mechanism following crizotinib treatment is a L1196M (leucine-to-methionine substitution in codon 1196) mutation in the exon coding for the *ALK* tyrosine kinase domain. L1196M-mutated patients can still benefit from ceritinib, brigatinib or lorlatinib therapy. In contrast, treatment options for G1202R-mutated tumors are limited. Here, only lorlatinib therapy can overcome the acquired resistance (Figure 3) (Gainor et al. 2016). To administer the available TKIs in the most effective manner – and thereby maximize the patient´s duration of response – it is essential to identify resistance mechanisms and select therapy lines based on this information. Ideally, the checkup for therapy resistance should be reiterated at every instance of tumor progression. In this way, a second and possibly third disease remission could be achieved in patients failing first line ALK-TKI therapy (Duruisseaux et al. 2017). Besides acquired drug resistances, there are other molecular features of ALK$^+$ tumors that need to be considered during the selection of targeted therapy regimens. Tumors carrying the EML4-ALK fusion variant V3 are less sensitive towards second and third generation TKIs, when compared to V1 and V2 tumors, leading to earlier patient relapse (Woo et al. 2017). Additionally, V3 tumors present a higher incidence of metastases and have an unfavorable prognosis independent of therapy type (Christopoulos et al. 2018; Noh et al. 2017). The *ALK* fusion partner is another molecular factor worth considering during therapy selection. For instance, PRKAR1A-ALK fusions were repeatedly associated with a low sensitivity towards first, second and third generation ALK-TKIs (Childress et al. 2018).

The genetic complexity of ALK$^+$ LUAD is relatively low for a tumor of the lungs (mean TMB = 2.0 mutations/Mb (Cerami et al. 2012; Gao et al. 2013; Jordan et al. 2017)). Yet, co-mutations can severely influence the clinical phenotype of a patient. Twenty to 25% of ALK$^+$ tumors carry an additional *TP53* mutation that is associated with increased metastatic

dissemination and short survival under TKI therapy. This is especially pronounced in EML4-ALK V3 patients with a *TP53* co-mutation (Christopoulos et al. 2019; Jordan et al. 2017; Kron et al. 2018).

Thus, it is crucial to characterize molecular risk factors (i.e. fusion variant, co-mutations, and fusion partner) at diagnosis but also during the course of a patient´s therapy (i.e. resistance mutations and occurrence of new co-mutations).
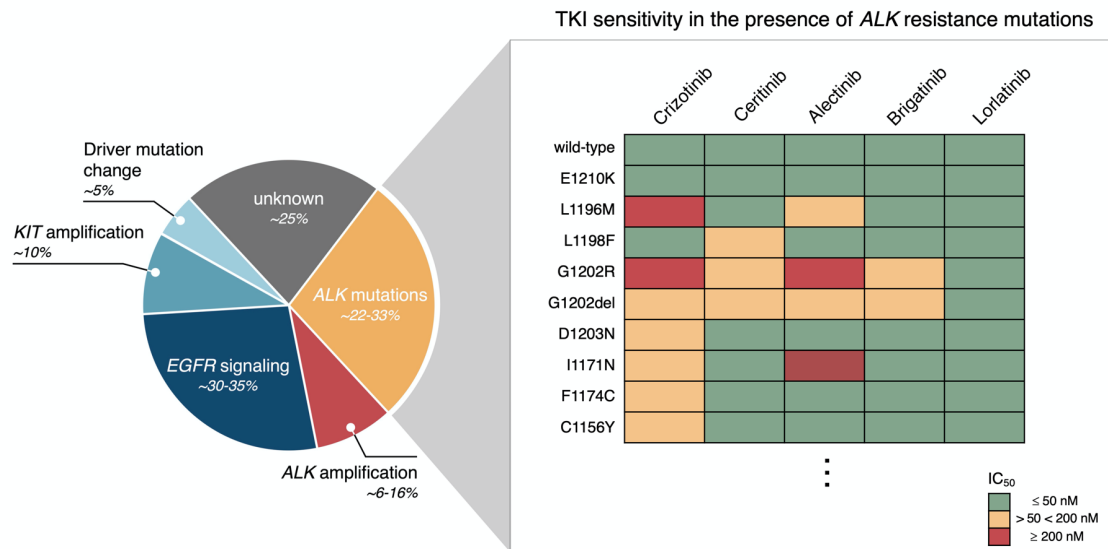
Figure 3: Resistance mechanisms and TKI sensitivity in *ALK* rearranged tumors. The pie chart illustrates acquired resistance mechanisms occurring in ALK+ LUAD patients after relapse from crizotinib treatment. Recurrent resistance mutations in the *ALK* tyrosine kinase domain and their effect on tyrosine kinase inhibitors (TKIs) are highlighted in the table. Green cells indicate potent and yellow cells intermediate inhibition by the given drug. Mutations conferring drug resistance are indicated by red cells. Inhibitory concentrations were determined *in vitro* based on phosphorylation of ALK in Ba/F3 cells. Adapted from Camidge *et al.* 2014 and Gainor *et al.* 2016. ALK, anaplastic lymphoma kinase; EGFR, epidermal growth factor receptor; IC$_{50}$, inhibitory concentration of 50%; KIT, KIT proto-oncogene; nM, nanomolar.

1.4 Epigenetics of lung adenocarcinoma

Epigenetic processes modify gene expression profiles and genomic stability without altering the coding DNA sequence. These changes in the genome´s function can occur on multiple levels, including chromatin remodeling, histone and DNA modifications, as well as expression alterations through non-coding RNAs. Under normal physiological conditions, epigenetic modifications occur in a highly controlled manner and regulate processes such as DNA imprinting, X chromosome inactivation and transcriptional (in-)activation of genes and repetitive elements (Greenberg and Bourc'his 2019). Genome-wide studies on large patient cohorts have provided important insights into epigenetic alterations occurring in various cancer entities (Collisson et al. 2014; Joehanes et al. 2016; Zhang et al. 2020b). In LUAD, epigenetic alterations have been identified during tumor initiation, progression and metastatic dissemination (Duruisseaux and Esteller 2018). Further investigations on those alterations are important for a better understanding of the disease and hold promise as reliable biomarkers for early diagnosis, molecular classification and prediction of therapy efficiency.

The modification of DNA represents a pivotal element in the epigenetic profiles of human cells. It occurs primarily at cytosine residues in the context of cytosine-guanine dinucleotides (CpGs). The most abundant DNA modification is 5-methylcytosine (5mC), occurring at 70 to 80% of CpG sites (Azzi et al. 2014). DNA methyltransferases (DNMTs) catalyze the covalent attachment of methyl groups to the carbon-5 position of cytosines (DNMT3a/b (Okano et al. 1999; Okano et al. 1998)) and faithfully maintain them during replication (DNMT1 (Bostick et al. 2007)). More recently, it was discovered that 5mC can be reverted to its unmodified state through an iterative oxidation process. Active DNA demethylation is mediated by members of the ten-eleven translocation (TET) enzyme family. TET enzymes oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Ito et al. 2011; Tahiliani et al. 2009). Those intermediates are either reverted to unmodified cytosines during replication or actively removed by thymine DNA glycosylase (TDG)-mediated base excision repair (BER; Figure 4A) (He et al. 2011; Zhang et al. 2012). Despite its substantially lower abundance compared to 5mC (5mC: 3-4% *vs.* 5hmC: 0.1-1% of all cytosines (Brazauskas and Kriaucionis 2014)), 5hmC has been recognized to be more than an intermediate state during DNA demethylation. It demonstrates a stable relative abundance over several cell divisions and was shown to have distinct effects on gene regulation (Bachman et al. 2014; Globisch et al. 2010). While high levels of 5mC at gene regulatory elements (i.e. promoters and enhancers) are commonly associated with a repressed transcriptional state, 5hmC frequently correlates with active gene expression. These effects are particularly pronounced at regions of elevated CpG density, so called CpG islands (CGIs). Two-thirds of genes within the human genome contain a CGI in their promoter region. Actively transcribed genes usually have a hypomethylated promoter CGI. 5hmC

prevents 5mC from spreading into the hypomethylated CGI by accumulating at its borders (Figure 4B; dark red and dark blue lines). Additionally, both 5mC and 5hmC levels within gene bodies were shown to correlate with gene expression. However, this correlation appears to be more distinct for 5hmC (Greenberg and Bourc'his 2019; Hansen et al. 2011; Mellén et al. 2012; Thomson et al. 2012). As regulators of transcription, 5mC and 5hmC profiles differ in accordance to their tissue/cell type of origin. Hence, both are used as biomarkers for cellular origin and identity (Brandeis et al. 1993; Eden and Cedar 1994; Forloni et al. 2016).
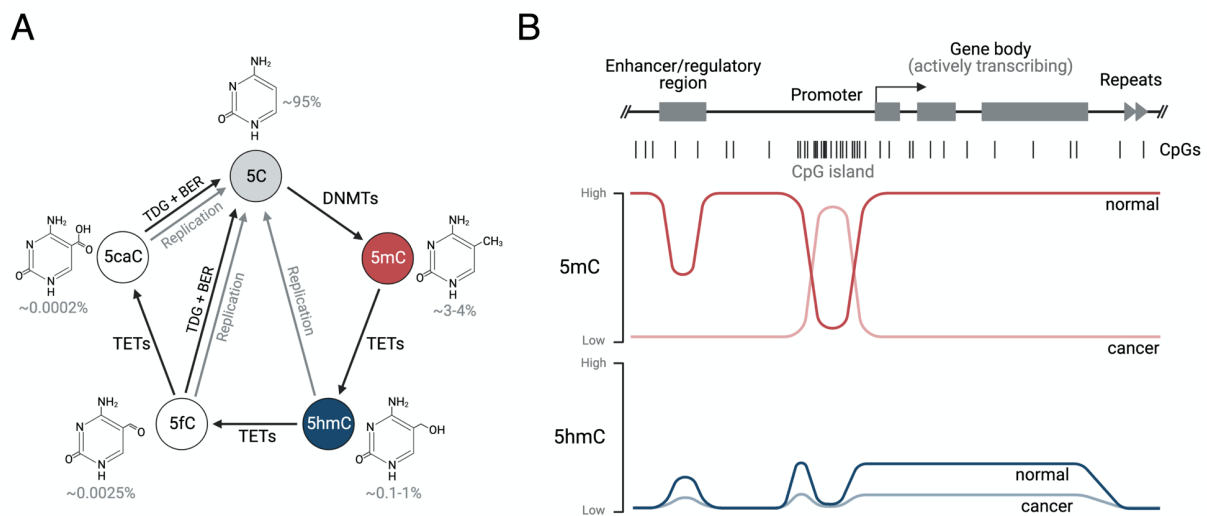
Figure 4: Active DNA demethylation and 5mC/5hmC abundances at actively transcribed genes. (A) DNA methyltransferases (DNMTs) catalyze the attachment of methyl groups to unmodified cytosines (5C). 5-methylcytosine (5mC) can be actively removed through ten-eleven translocation (TET) enzymes, iteratively oxidizing 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC). The intermediate oxidation states can be reverted to 5C during replication or enzymatically removed (thymine DNA glycosylase [TDG]-mediated base excision repair [BER]). Chemical structures and relative abundances of each cytosine modification are illustrated alongside the active DNA demethylation cycle. Adapted from Brazauskas *et al.* 2014 and Wu *et al.* 2017. (B) Conceptual illustration of 5mC (red) and 5hmC (blue) levels at an actively transcribed gene locus. Dark red and blue lines indicate the 5mC and 5hmC abundance under normal physiological conditions. Light colors highlight altered 5mC/5hmC abundances in cancer. Adapted from Skvortsova *et al.* 2019b and Thomson *et al.* 2017. CpG, cytosine-guanine dinucleotide.

Alterations of 5mC and 5hmC profiles have been identified in various cancer entities, including LUAD (Baylin and Jones 2011; Baylin and Ohm 2006; Song et al. 2017; Zhang et al. 2018b). On a global scale, 5mC abundance is reduced in tumors compared to adjacent normal tissue. Genome-wide hypomethylation coincides with high chromosomal instability and increases aneuploidy, both common characteristics of cancer genomes (Feinberg and Vogelstein 1983; Jones and Baylin 2007; Skvortsova et al. 2019b). Additionally, tumors frequently exhibit focal alterations in their 5mC profile. These changes predominantly occur at promoter CGIs and confer the activation (hypomethylation) or inactivation (hypermethylation) of associated genes (Figure 4B; light red line). In LUAD, promoter hypomethylation and consequential transcriptional upregulation was shown for oncogenes, such as *RAB25* and *MUC4*. The activation of these genes correlated with poor patient outcome (Selamat et al. 2012; Yokoyama et al. 2017). Hypermethylation of promoters is

commonly observed in the context of tumor suppressor silencing and often occurs early during carcinogenesis. Genes affected by aberrant hypermethylation are involved in a variety of cellular processes during the progression from normal lung cells to adenocarcinomas: cell cycle control (*CDKN2A*), regulation of apoptosis (*DAPK*), DNA repair (*MGMT*), Ras signaling (*RASSF1A*), and immortalization (*hTERT*) (Licchesi et al. 2008a; Licchesi et al. 2008b; Selamat et al. 2012; Selamat et al. 2011; Tsou et al. 2007). In proof-of-concept studies, *CDKN2A* promoter hypermethylation, as detected in bronchoalveolar lavage, was used as a diagnostic epigenetic biomarker for early LUAD detection (Ahrendt et al. 1999; Kim et al. 2004).

Furthermore, *CDKN2A* methylation levels were demonstrated to increase with tumor progression, suggesting its applicability for the prediction of tumor stage or disease progression (Belinsky et al. 1998). Aberrations in the hydroxymethylome have been observed in all human cancer entities. Here, the widespread reduction of 5hmC is a common feature during carcinogenesis (Figure 4B; light blue line) (Haffner et al. 2011; Lian et al. 2012a). In cancers of the lung, global 5hmC abundance was shown to decrease in a stage-dependent manner. Therefore, genome-wide 5hmC abundance could indicate disease progression in these tumors (Song et al. 2017). The lower 5hmC levels in cancer can partially be explained by inactivating mutations in genes coding for TET enzymes or lack of co-factors required for TET activity. Reduced activity of TET1 has been reported in many cancer entities, including LUAD (Lian et al. 2012a; Yang et al. 2013; Yang et al. 2012). Recently, a number of studies identified alterations in the 5hmC levels of individual genes or gene panels and successfully used them for early cancer detection, diagnosis and staging in plasma DNA derived from cancer patients and healthy individuals (Guler et al. 2020; Song et al. 2017; Zhang et al. 2018b).

1.5 Liquid Biopsy

Molecular tumor profiling is essential for selecting the appropriate therapy in clinical practice (Vogelstein et al. 2013). Currently, tissue biopsies are employed for cancer genotyping at diagnosis. However, the invasive procedure of obtaining tissue specimens often presents an inherent risk to the patient (Overman et al. 2013). Furthermore, needle biopsies often fail to capture the intratumor heterogeneity of the primary tumor tissue and cannot provide information about metastases (Gerlinger et al. 2012; McGranahan and Swanton 2015; Vogelstein et al. 2013). Selective pressure during therapy and/or disease progression can alter the tumor genome (Vogelstein et al. 2013). Therefore, repeat biopsies are vital to provide a real-time representation of the evolving tumor. Especially in advanced stages, re-biopsies are usually not feasible due to the poor medical conditions of the patient. Liquid biopsies represent a promising alternative to the conventional analysis of cancerous tissue. Liquid biopsy analysis comprises the study of body fluids, which carry material originating from tumor cells (Wan et al. 2017). The most widely used body fluid for liquid biopsy approaches is blood (i.e. plasma or serum), however, tumor-derived material was also identified in urine (Botezatu et al. 2000; Smith et al. 2020), cerebrospinal fluid (Pan et al. 2015; Wang et al. 2015), saliva (Mithani et al. 2007), pleural fluid (Sriram et al. 2012), bile (Shen et al. 2019a), and exhaled breath condensate (Koc et al. 2019). The sampling of such material presents little risk to the patient and therefore allows repeated tumor assessment over time (Pantel and Alix-Panabières 2013). In addition, tumor material within a liquid biopsy sample may originate from any growth site of the tumor (i.e. primary tumor and metastases) and might better reflect its heterogeneity compared to tissue biopsies (Murtaza et al. 2015). Today, a plethora of tumor-informative circulating analytes has been discovered. These include various types of nucleic acids, such as cell-free DNA (cfDNA), mitochondrial DNA, viral DNA, and microRNA (miRNA) (González-Masiá et al. 2013), but also proteins, extracellular vesicles (Garcia-Romero et al. 2018), circulating tumor cells (CTCs) (Alix-Panabières and Pantel 2021), as well as tumor-educated platelets (Best et al. 2015; Best et al. 2018). Among these potential biomarkers, cfDNA and its subfraction of tumor-derived DNA fragments (ctDNA; circulating tumor DNA) is a promising tool for precision medicine. For example, Diehl and colleagues (Diehl et al. 2008) were the first to show that detectable ctDNA in colorectal cancer patients following surgical tumor resection can predict early relapse. Other studies made similar observations in additional cancer entities, including lung cancer (Abbosh et al. 2017; Beaver et al. 2014; Tie et al. 2015). Furthermore, a variety of studies used mutation profiling from plasma samples to monitor therapy response and to estimate overall disease burden (Dawson et al. 2013; Dietz et al. 2020; Mok et al. 2015; Riediger et al. 2016).

1.5.1 Biology and characteristics of cfDNA

Apart from the potential clinical applications of cfDNA, the biology behind its release into the bloodstream has moved into the focus of research. In 2001, Jahr *et al.* (Jahr et al. 2001) were the first to show that the distinct size profile of cfDNA might contain information about its mechanisms of release. The characteristic fragment length of cfDNA, centered around 167 bp (and multiples thereof), reflects the length of a DNA molecule wrapped around the nucleosomal core unit (~147 bp) plus a 20 bp linker DNA fragment (Figure 5A). This size profile is reminiscent of DNA subjected to caspase-dependent cleavage and suggests that a large fraction of cfDNA is released from cells undergoing apoptosis. However, other forms of cell death (e.g. necrosis), active secretion through extracellular vesicles, and cellular proliferation might also contribute to the repertoire of circulating DNA (Galluzzi et al. 2018; Jeppesen et al. 2019). Notably, tumor-derived cfDNA fragments were reported to exhibit shorter size profiles (~145 bp) compared to non-tumor fragments, suggesting differences in the release mechanisms between ctDNA and cfDNA (Jiang et al. 2015; Mouliere et al. 2018b). Recently, studies on the cell-free methylome and histone modifications (i.e. H3K4me1/2/3) in plasma provided important insights into the different cell and tissue types contributing to the cfDNA pool. By comparing the profiles of these epigenetic marks to publicly available reference data, hematopoietic cells were identified as the major contributor to cfDNA in plasma (Guo et al. 2017; Lehmann-Werman et al. 2016; Moss et al. 2018; Serpas et al. 2019; Sun et al. 2015). The most comprehensive study in this regard used reference methylation data from 25 cell and tissue types to allow a detailed deconvolution of the cfDNA composition in healthy subjects. Their data suggests that ~85% of cfDNA in plasma originates from blood cells. Furthermore, vascular endothelial cells (8.6%) and hepatocytes (1.2%) were identified as frequent contributors to plasma cfDNA (Figure 5B) (Moss et al. 2018). In cancer, the cfDNA composition changes with the pathological condition of the patient. For example, lung cancer patients were reported to exhibit higher proportions of methylation patterns specific for lung cells (Guo et al. 2017; Moss et al. 2018).

A limitation of cfDNA-based liquid biopsies is the low quantity of tumor-derived DNA fragments detectable in the circulation. While total cfDNA concentrations consistently increase in the presence of a tumor (even at early stages (van der Pol and Mouliere 2019)), the ctDNA fraction varies considerably (<0.01 to >60% (Bettegowda et al. 2014; Diehl et al. 2008; Thierry et al. 2014)) and is affected by a multitude of parameters, including the cancer entity, tumor burden and disease stage, vascularization, cell turnover rate, histology, and proliferation capacity of the tumor (Bettegowda et al. 2014; Diehl et al. 2008; Thierry et al. 2014). Moreover, cfDNA is rapidly cleared from the bloodstream (half-life: 4 min to 2 h) (Khier and Lohan 2018). Therefore, highly sensitive analytical approaches are needed to accurately detect the minute amounts of tumor-derived DNA fragments in circulation.

A

B

Plasma DNA composition
in healthy donors
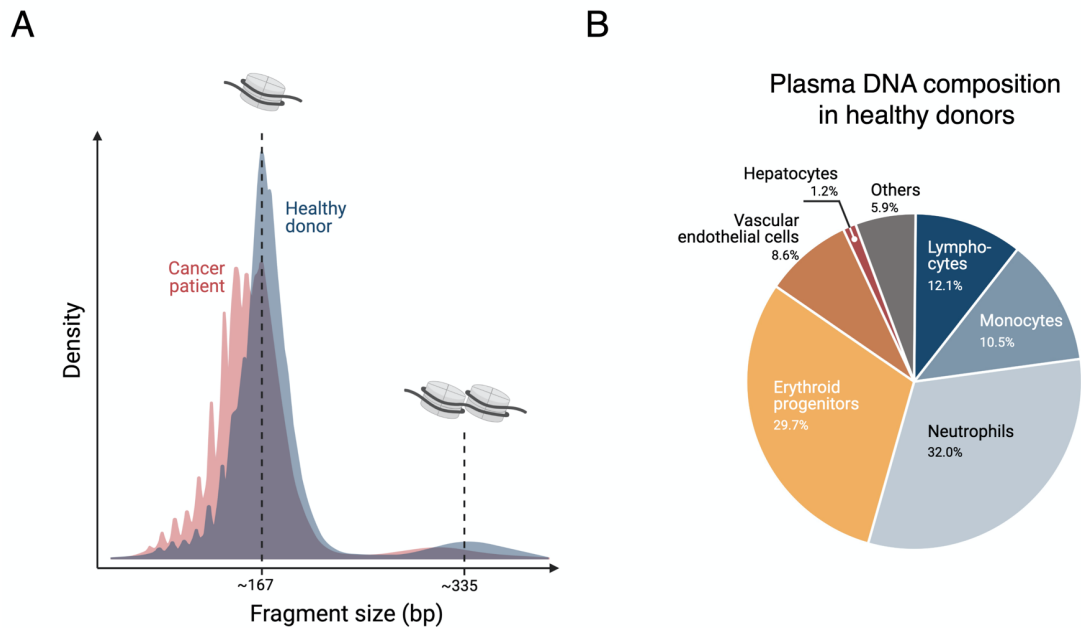
Figure 5: Fragment size distribution and cellular contributors to plasma cfDNA. (A) Schematic representation of the size profile of cfDNA isolated from plasma of a healthy donor (blue) and an advanced stage cancer patient (red). Fragment lengths of ~167 and ~335 bp reflect DNA wrapped around mono- and dinucleosomal core units. Adapted from Chan *et al.* 2016. (B) Cellular contributors to the plasma cfDNA composition in healthy individuals, as determined by methylation-based cell type deconvolution. Adapted from Moss *et al.* 2018.

1.5.2 Cancer specific genetic variants in cfDNA

The detection of ctDNA in plasma of cancer patients holds promise for clinical applications, ranging from early disease detection to therapy monitoring at advanced stages (Figure 6) (Wan et al. 2017). To cope with the low amounts of ctDNA in the circulation, methods had to be developed that identify those tumor-derived DNA fragments with high sensitivity and specificity. Today, the most widely used ctDNA-based assays focus on the detection of cancer-specific genomic alterations, including single nucleotide variants (SNVs), gene rearrangements, and copy number alterations (CNAs). These technologies can be divided into targeted and untargeted approaches (Siravegna et al. 2019). The targeted strategies require *a priori* knowledge of the cancer-associated feature to be analyzed. For instance, Cancer Personalized Profiling by deep Sequencing (CAPP-seq) uses biotinylated oligonucleotides that specifically enrich for recurrently mutated regions in the cancer of interest, followed by deep sequencing of the enriched fraction. It is designed to detect multiple classes of tumor-specific alterations (i.e. indels [insertions or deletions], rearrangements, and CNAs) and allows highly specific mutation detection in large gene panels down to variant allele frequencies (VAFs) of 0.01% (Newman et al. 2014; Newman et al. 2016). Whole-genome (WGS) and whole-exome sequencing (WES) are untargeted methods for ctDNA detection. They provide insights into cancer-associated alterations present in cfDNA, yet, at the expense of high sequencing costs and low analytic sensitivities (1 to 5% VAF) (Murtaza et al. 2013; Ulz et al. 2016). In contrast, shallow WGS (sWGS; <0.5x genome coverage) provides an alternative method for the assessment of CNAs. The majority

of metastatic cancers harbor somatic CNAs, making sWGS a cost-effective application for ctDNA detection in advanced staged patients, especially when mutations remain undetectable (Adalsteinsson et al. 2017; Dietz et al. 2020; Mouliere et al. 2018b; Smith et al. 2020).

One of the biggest advantages of liquid biopsies over conventional tissue evaluation is the feasibility of repeated sampling. This allows narrow monitoring of a patient´s disease both after surgical tumor resection and during systemic therapy (e.g. chemotherapy or TKI treatment). A number of studies demonstrated the practicality of ctDNA detection by targeted SNV profiling for the monitoring of minimal residual disease (MRD) (Abbosh et al. 2017; Chaudhuri et al. 2017; Chen et al. 2017; Murtaza et al. 2015; Ng et al. 2017). A study conducted by Chaudhuri *et al.* (Chaudhuri et al. 2017) used CAPP-seq for the surveillance of early stage lung cancer patients previously treated with curative intent. They could show that detectable ctDNA levels in the first post-treatment sample coincided with earlier relapse. Additionally, ctDNA detection preceded radiographic progression in 72% of patients with a mean lead time of 5.2 months. Another promising application of ctDNA detection is the monitoring of therapy resistance. TKI treatment is associated with inevitable relapse due to the development of drug resistance, and the early identification of these processes is crucial to maximize the clinical benefit for a patient. Since ALK[+] patients can benefit from the sequential administration of several targeted agents (Duruisseaux et al. 2017), serial plasma genotyping by NGS was demonstrated to be a reliable method for the detection of *ALK* fusions and resistance mutations. During sequential TKI therapy, *ALK* mutations emerged and cleared in accordance to the administered drug. In addition, the abundance of the *ALK* fusion, as measured in cfDNA, frequently increases during progression (Dagogo-Jack et al. 2018; Dietz et al. 2020; Li et al. 2021; Zhang et al. 2020a). This highlights the potential of plasma-based analyses to monitor and guide the selection of ALK-TKI therapies. We (Dietz et al. 2020) have recently integrated targeted NGS (based on CAPP-seq) with sWGS-based profiling to monitor the therapy of metastatic ALK[+] patients. A previously developed metric to infer global copy number changes, termed "trimmed median absolute deviation from copy number neutrality" (t-MAD) (Mouliere et al. 2018b), was derived from the sWGS data. The t-MAD score has been demonstrated to correlate with the tumor fraction in cfDNA samples and was used as a surrogate to evaluate the tumor burden at a given time point. In this study, t-MAD scores allowed therapy monitoring in patients without detectable *ALK* fusions or mutations.

Despite the advancement of ctDNA detection technologies, sensitivity remains a limiting factor, especially in cases where the tumor fraction in plasma is low (e.g. early disease, MRD detection, or sampling time points at therapy response). Moreover, important information about a patient´s tumor, such as its localization, cannot be derived from genomic analyses of ctDNA and require alternative methodologies.
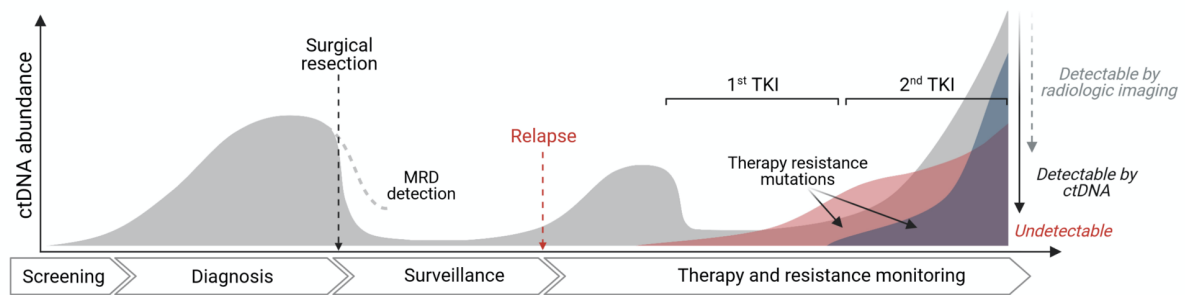
Figure 6: Potential clinical applications of ctDNA detection throughout a patient´s disease. Schematic time course of a hypothetical patient undergoing surgical tumor resection and systemic therapy. Potential applications of circulating tumor DNA (ctDNA) detection are indicated and include: cancer screening, (early) diagnosis, surveillance, minimal residual disease (MRD) detection, as well as therapy and resistance monitoring. The divergent detection limits of radiologic imaging and ctDNA detection assays are indicated on the right side. Adapted from Wan *et al.* 2017. TKI, tyrosine kinase inhibitor.

1.5.3 Epigenomic analysis of cfDNA in cancer

Profiling the cell-free epigenome of cancer patients represents an alternative to the classical evaluation of genomic aberrations that might advance the clinical utility of liquid biopsies. Nowadays, a variety of epigenomic features can be assessed through the analysis of cfDNA. These include epigenetic marks such as histone modifications (Serpas et al. 2019), DNA methylation (Liu et al. 2020; Moss et al. 2018; Shen et al. 2018) and hydroxymethylation (Bergamaschi et al. 2020; Guler et al. 2020; Song et al. 2017), as well as fragmentation profiles of cfDNA (e.g. shorter fragment length of tumor-derived cfDNA (Jiang et al. 2015; Mouliere et al. 2018b), inference of nucleosomal positioning through cfDNA coverage (Snyder et al. 2016; Ulz et al. 2016), and recurrent sequence motifs at the ends of cfDNA molecules (Jiang et al. 2020)).

The detection of these epigenetic alterations – in particular DNA methylation and hydroxymethylation changes – has several advantages over the assessment of genomic aberrations. Aberrant 5(h)mC events occur early during carcinogenesis and might present better biomarkers for early disease detection as well as cancer screening (Dor and Cedar 2018). The tissue specificity of 5(h)mC patterns allows to not only detect a tumor but also localize its tissue-of-origin. Here, a prime example is the usage of cfDNA-based methylation profiling to localize the primary growth site in cancers of unknown primary (CUPs) (Moss et al. 2018). Additionally, cancer-associated alterations of the (hydroxy-)methylome are both more prevalent and more pervasive when compared to mutations. Therefore, 5(h)mC-based assays for ctDNA detection are more likely to identify cancer-specific alterations and can be applied to a wider range of patients. This translates into a higher sensitivity of those assays (Kandoth et al. 2013; Li and Zhou 2020). Despite the advantages of harnessing the cell-free (hydroxy-)methylomes for minimal invasive cancer detection, several conceptual and technical limitations remain. The high sensitivity of epigenetic assays comes at the expense of a reduced tumor specificity of the detected alterations. Changes in the 5(h)mC profile do not only occur in a cancer- or tissue-specific manner but also as stochastic events

15

accumulating with age (Dor and Cedar 2018). Bisulfite conversion is an essential step in many assays for 5(h)mC profiling. It is commonly associated with extensive DNA degradation and results in low cfDNA recovery rates (22% to 66%) (Worm Ørntoft et al. 2017). This presents a limiting factor considering the low amounts of cfDNA available in most clinical settings. New approaches circumvent the necessity for bisulfite conversion in using prior enrichment of the DNA modification of interest. Shen *et al.* (Shen et al. 2018) developed an immunoprecipitation-based protocol for the enrichment of methylated cfDNA fragments followed by sequencing. This method, termed cell-free methylation DNA immunoprecipitation (cfMeDIP), interrogates 5mC events on a genome-wide scale, while requiring only minute amounts of DNA (i.e. 1 to 10 ng). Thus, cfMeDIP-seq is applicable in most clinical settings, even when cfDNA quantities are scarce. This was demonstrated by two studies that used cfMeDIP-seq on patients with intracranial malignancies and renal cell carcinomas. Both cancer types usually shed little amount of ctDNA and are difficult to detect in liquid biopsies. cfMeDIP-seq accurately identified both tumor types and could even distinguish different primary brain tumors that are usually difficult to discern by imaging (Nassiri et al. 2020; Nuzzo et al. 2020). Another enrichment-based approach found widespread application for the analysis of 5hmC from plasma samples (Bergamaschi et al. 2020; Guler et al. 2020; Song et al. 2017; Tian et al. 2018; Zhang et al. 2018b). Hydroxymethylation-selective chemical labeling (hMeSEAL) enzymatically modifies the hydroxyl-groups on cfDNA fragments in a two-step process, finally resulting in its biotinylation. Subsequently, hydroxymethylated DNA fragments are recovered using streptavidin-coated magnetic beads, followed by sequencing of the enriched fraction (Song et al. 2011; Song et al. 2017). This method has been applied to plasma samples of multiple different cancer entities (e.g. lung, liver, colorectal, pancreatic, and breast cancer) and demonstrated promising results for the (early) detection and classification of cancers (Bergamaschi et al. 2020; Guler et al. 2020; Song et al. 2017; Tian et al. 2018; Zhang et al. 2018b). Notably, 5hmC levels within gene bodies were reported to positively correlate with the gene's transcriptional status, encouraging the concept of 5hmC as a surrogate for minimal invasive expression inference (Song et al. 2017).

1.6 Aim of this study

NSCLC patients with ALK[+] tumors demonstrate significant clinical benefit from ALK-directed TKIs, however, eventually relapse due to the development of drug resistances. Timely detection of therapy resistance and disease progression is crucial to guide subsequent therapy lines. We and others have demonstrated the utility of mutation and CNA profiling from cfDNA for the monitoring of ALK[+] patients under sequential TKI treatment. Genomic alterations in cfDNA are often tumor-specific, yet suffer from a limited sensitivity due to their low relative abundance throughout the genome. Epigenetic alterations occur much more frequently in the cancer genome and might present a more sensitive alternative for the detection of tumor-derived DNA from blood samples.

In this study, genome-wide 5mC and 5hmC profiles were generated from longitudinally taken cfDNA samples of ALK[+] NSCLC patients receiving TKI therapy. The study's first objective was to enrich for tumor-derived alterations in the (hydroxy-)methylome by excluding genomic regions with high 5(h)mC levels in cell types abundant in non-tumor cfDNA. For this purpose, 5(h)mC profiles of primary monocytes, neutrophils and erythroid progenitor cells were generated and used to filter the patient-derived 5(h)mC datasets. This step intended to facilitate the identification of tumor 5(h)mC alterations and thereby allow cancer assessment in samples with low ctDNA burden. The second objective was to utilize the tumor-enriched 5(h)mC signals to infer cancer-specific gene expression from cfDNA. This would provide information about the tumor which cannot be obtained from the analysis of genomic alterations in cfDNA and might enable the identification of additional drug resistance mechanisms (i.e. bypass signaling pathway activation). Finally, the study aimed to compare the utility of tumor-specific 5(h)mC signatures to genomic alterations in cfDNA for minimal-invasive therapy monitoring in ALK[+] patients receiving TKI therapy. Hereby, the high prevalence of perturbations in the cancer (hydroxy-)methylome might translate into sensitive 5(h)mC markers and allow for therapy surveillance in patients without detectable genomic cfDNA alterations. The epigenetic biomarkers were correlated to mutations and CNAs detected in the same cfDNA samples. Dynamic changes in the marker abundance were examined in serial plasma samples.

2 Materials and Methods

2.1 Materials

Table 1: Equipment

Name	Manufacturer
Bioruptor® Pico sonication device	Diagenode, Seraing, Belgium
Capillary electrophoresis system, Bioanalyzer 2100	Agilent Technologies, Santa Clara, USA
Centrifuge Heraeus Megafuge 16 R	ThermoFisher Scientific, Waltham MA, USA
Fluorometer, Qubit 2.0	ThermoFisher Scientific, Waltham MA, USA
Heating bock, Thermomixer comfort	Eppendorf, Hamburg, Germany
LightCycler® 480	Roche, Mannheim, Germany
Magnetic separator, 0.2 mL PCR Strip	Permagen Labware, Peabody, USA
Magnetic stand, 1.5 mL tubes	ThermoFisher Scientific, Waltham MA, USA
Microcentrifuge, Heraeus™ Fresco™	ThermoFisher Scientific, Waltham MA, USA
Micropipettes, Pipetman (10, 20, 100, 200, 1000 μL)	Gilson, Middleton, USA
Multichannel pipet, Rainin pipet lite L-10 (1-10 μL)	Mettler-Toledo, Columbus, USA
Multichannel pipet, Rainin pipet lite L-200 (20-200 μL)	Mettler-Toledo, Columbus, USA
Thermocycler, T100™	Bio-Rad Laboratories, Hercules, USA
Tube rotator, Rotobot	Benchmark Scientific, Edison NJ, USA
Vortexer, Vortex-Genie 2	Scientific Industries, Bohemia, USA

Table 2: Consumables

Name	Manufacturer
Bioruptor® microtubes (1.5 mL)	Diagenode, Seraing, Belgium
Conical tubes (15 mL, 50 mL)	BD Bioscience, Bedford, USA
Filter tips (10 μL, 20 μL, 100 μL, 200 μL, 1000 μL)	Neptune, San Diego, USA
Gloves, Microflex XCEED	Microflex, Reno, USA
LightCycler® 480 Multiwell Plate 384, white	Roche, Mannheim, Germany
LightCycler® 480 Sealing Foil	Roche, Mannheim, Germany
Microcentrifuge tubes (0.5 mL, 1.5 mL, 2.0 mL)	Eppendorf, Hamburg, Germany
PCR plate 96, semi-skirted	Eppendorf, Hamburg, Germany
PCR strips (0.2 mL)	Steinbrenner, Wiesenbach, Germany
Reagent Reservoirs	ThermoFisher Scientific, Waltham MA, USA
Serological pipettes (5 mL, 10 mL, 25 mL, 50mL)	BD Bioscience, Bedford, USA
S-Monovette 7.5 mL K3 EDTA tubes	Sarstedt, Nürnberg, Germany
QIAshredder	Qiagen, Hilden, Germany

Table 3: Molecular biology kits

Kit	Manufacturer
Agilent High Sensitivity DNA Kit	Agilent Technologies, Santa Clara, USA
AllPrep DNA/RNA/miRNA Universal Kit	Qiagen, Hilden, Germany
Ambion® Buffer Kit	ThermoFisher Scientific, Waltham MA, USA
AmpliSeq RNA Lung Cancer Fusion Panel	ThermoFisher Scientific, Waltham MA, USA
CpG methyltransferase (M.SssI)	ThermoFisher Scientific, Waltham MA, USA
HpyCH4IV	New England Biolabs, Ipswich, USA
iPure Kit v2	Diagenode, Seraing, Belgium
KAPA Dual-Indexed Adapter Kit (15 µM)	Roche, Mannheim, Germany
KAPA HyperPrep Kit	Roche, Mannheim, Germany
MagMeDIP qPCR Kit	Diagenode, Seraing, Belgium
Phusion® High-Fidelity PCR Kit	New England Biolabs, Ipswich, USA
Platinum™ SuperFi™ DNA Polymerase	ThermoFisher Scientific, Waltham MA, USA
QIAamp MinElute ccfDNA Kit	Qiagen, Hilden, Germany
Qubit dsDNA HS Assay Kit	ThermoFisher Scientific, Waltham MA, USA

Table 4: Reagents

Reagent	Manufacturer
5-hydroxymethyl-dCTP	Jena Bioscience, Jena, Germany
5-methyl-dCTP	Jena Bioscience, Jena, Germany
Ambion uclease-free water	ThermoFisher Scientific, Waltham MA, USA
D5F3 clone	Roche, Mannheim, Germany
DBCO-PEG4-biotin	Jena Bioscience, Jena, Germany
Dynabeads™ M270	ThermoFisher Scientific, Waltham MA, USA
Ethanol absolute	Sigma-Aldrich, St. Louis, USA
Gibco™ Dulbecco´s Phosphate-Buffered Saline (no calcium, no magnesium)	ThermoFisher Scientific, Waltham MA, USA
Isopropyl alcohol	Sigma-Aldrich, St. Louis, USA
Lambda-DNA	ThermoFisher Scientific, Waltham MA, USA
PrimaQuant CYBR qPCR master mix	Steinbrenner, Wiesenbach, Germany
SPRI AMPure XP magnetic beads	Beckman Coulter, Brea, USA
T4 phage ß-glucosyltransferase enzyme	New England Biolabs, Ipswich, USA
Tween-20	ThermoFisher Scientific, Waltham MA, USA
UDP-azide-glucose	Active Motif, Carlsbad, USA
UltraPure™ salmon sperm DNA	ThermoFisher Scientific, Waltham MA, USA
ZytoLight SPEC *ALK* probe	ZytoVision, Bremerhaven, Germany

Table 5: Primary cells

Cell type	Number of cells	Supplier
Erythroid progenitor cells	1×10^6	StemExpress, Folsom CA, USA
Monocytes	1×10^7	Stemcell Technologies, Vancouver, Canada
Neutrophils	1×10^7	HemaCare, Los Angeles CA, USA

Table 6: Primer pairs for spike-in amplicon generation

Spike-in name	Forward primer (5´-3´)	Reverse primer (5´-3´)	Amplicon size (bp)
5C spike-in	CGTTTCCGTTCTTCTTCGTC	TACTCGCACCGAAAATGTCA	184
5mC spike-in	GTGGCGGGTTATGATGAACT	CATAAAATGCGGGGATTCAC	183
5hmC spike-in	TGAAAACGAAAGGGGATACG	GTCCAGCTGGGAGTCGATAC	185

Primers were synthesized and purchased at Eurofins Genomics (Ebersberg, Germany)

Table 7: Fragments and primer pairs for lambda filler DNA generation

Fragment name	Forward primer (5´-3´)	Reverse primer (5´-3´)	Amplicon size (bp)
1CpG	GAGGTGATAAAATTAACTGC	GGCTCTACCATATCTCCTA	196
5CpG	CATGTCCAGAGCTCATTC	GTTTAAAATCACTAGGCGA	269
10CpG	CTGACCATTTCCATCATTC	GTAACTAAACAGGAGCCG	359
15CpG	ATGTATCCATTGAGCATTGCC	CACGAATCAGCGGTAAAGGT	461
20LCpG	GAGATATGGTAGAGCCGCAGA	TTTCAGCAGCTACAGTCAGAATTT	495
20SCpG	CGATGGGTTAATTCGCTCGTTGTGG	GCACAACGGAAAGAGCACTG	274

Primers were synthesized and purchased at Eurofins Genomics (Ebersberg, Germany)

20

Table 8: Software

Software	Company
Bioanalyzer 2100 Expert Software	Agilent Technologies, Santa Clara, USA
LightCycler® 480 Software	Roche, Mannheim, Germany
Microsoft Excel 2010	Microsoft, Redmond, USA
R Studio	R Studio, Boston, USA

Table 9: Bioinformatic tools and R packages

Tool	Reference
Bedtools v2.29.2	Quinlan *et al.*, 2010
Bowtie2 v2.2.6.2	Langmead *et al.*, 2012
clusterprofiler v4.0.0	Yu *et al.*, 2012
CNAclinic v0.2.0	Mouliere *et al.*, 2018
chromHMM v1.22	Ernst *et al.*, 2012
Cutadapt v.1.16	Martin *et al.*, 2011
DKFZ Galaxy instance	Giardine *et al.*, 2005
edgeR v3.34.0	Robinson *et.al.*, 2010
FastQC v0.11.2	Andrews *et al.*, 2010
GenomicRanges v1.44.0	Lawrence *et al.*, 2013
ggplot2 v3.3.4	Wickham *et al.*, 2016
HOMER v4.11	Heinz *et al.*, 2010
HMMcopy Suite v1.34.0	Lai *et al.*, 2021
ichorCNA	Adalsteinsson *et al.*, 2017
limma v3.48.0	Ritchie *et al.*, 2015
MACS2 v2.1.1	Zhang *et al.*, 2008
MEDIPS v1.44.0	Lienhard *et al.*, 2014
minfi v1.38.0	Aryee *et al.*, 2014
pheatmap v1.0.12	Kolde *et al.*, 2015
regioneR v1.24.0	Gel *et al.*, 2015
SAMtools v1.3.1	Lie *et al.*, 2009
Subread v1.5.3	Liao *et al.*, 2014
VennDiagram v1.6.20	Chen *et al.*, 2011

2.2 Methods

2.2.1 Plasma samples

All individuals provided informed consent and the study was approved by the ethics committees at Heidelberg (S-270/2001, S-445/2015, S-296/2016) and Lübeck Universities (AZ 12-238). Peripheral blood was drawn from 31 metastatic ALK+ NSCLC patients (Table S1) and 14 healthy donors at the Thoraxklinik Heidelberg / LungenClinic Grosshansdorf, Germany, and provided *via* the Lung Biobank Heidelberg. Eleven patients provided multiple samples throughout TKI therapy (range: 2 – 14 samples per patient), resulting in a total of 93 blood samples. Plasma isolation from whole blood was performed within one hour of blood draw at the site of collection. In brief, whole blood was collected in S-Monovette 7.5 mL K3 EDTA tubes, followed by centrifugation (Centrifuge Heraeus Megafuge 16 R) at 1,600 x g for 10 min (no brakes). Subsequently, plasma was aspirated and re-centrifuged at 3,400 x g for 10 min (no brakes) to remove cell debris and stored at -80 °C until use.

Molecular tumor characterization was performed at the Institute of Pathology Heidelberg, Heidelberg University Hospital. Diagnosed ALK rearrangements were based on positivity of at least two of the following assays: ALK immunohistochemistry (D5F3 clone), ALK fluorescent *in situ* hybridization (ZytoLight SPEC *ALK* probe), and RNA-based next-generation sequencing (NGS, AmpliSeq RNA Lung Cancer Fusion Panel). Clinical data and radiographic assessments by chest/abdominal computed tomography (CT) and brain magnetic resonance imaging (MRI) were collected based on patient record reviews with a cut-off on May 30, 2020.

2.2.2 Primary blood cell samples

Cryopreserved primary monocytes, neutrophils and erythroid progenitor cells were purchased for 5mC and 5hmC profiling. Extraction procedures are described by the commercial providers as follows: Monocytes (Stemcell Technologies) and neutrophils (HemaCare) were isolated from peripheral blood using immunomagnetic separation of CD14+ mononuclear cells (monocytes) and red cell lysis followed by density gradient centrifugation (neutrophils). Erythroid progenitors (StemExpress) were isolated from bone marrow mononuclear cells. Immunomagnetic anti-CD34 beads were used to enrich for CD34+ cells. CD34+ cells were then cultured for 10 days in serum-free expansion media supplemented with SCF, EPO and IL-3. Thereby, the differentiation of CD34+ to CD36+ erythroid progenitor cells was promoted. The providers confirmed a purity >90% of the primary cell types by flow cytometry.

2.2.3 Isolation of cfDNA from plasma

Plasma samples were thawed on ice and centrifuged at 10,000 x g for 1 min at room temperature (RT) to remove residual blood cells. CfDNA isolation was performed from 0.5 to 3 mL of plasma using the QIAamp MinElute ccfDNA Kit. In brief, cfDNA was bound to magnetic beads by combining 30 μL of beads, 55 μL proteinase K and 150 μL bead binding buffer, followed by a 10 min incubation step at RT. Subsequently, samples were placed on a magnetic rack to separate the beads from the cfDNA-free plasma. To detach cfDNA from the beads, 200 μL of bead elution buffer were added and incubated for 5 min at RT. CfDNA was then mixed with 1.5 volumes of ACB buffer and bound to a QIAamp mini spin column by centrifugation. Following one wash step (500 μL ACW2 buffer), the silica membrane of the spin column was dried at 56 °C for 3 min (RT). Afterwards, cfDNA was eluted in 30 μL nuclease-free water by centrifugation at 20,000 x g for 1 min and stored at -20 °C until further use. Wash steps and cfDNA binding to the membrane was performed by centrifugation at 6,000 x g for 1 min at RT.

2.2.4 Isolation of genomic DNA from blood cells

Genomic DNA (gDNA) of primary blood cells was isolated using the AllPrep DNA/RNA/miRNA Universal Kit, following the DNA isolation steps of the protocol "Simultaneous purification of genomic DNA and total RNA, including miRNA, from cells". DNA isolation was performed from $5x10^5$ to $5x10^6$ cells. Briefly, cell pellets were lysed by adding 350 μL RLT plus buffer, followed by homogenization using QIAshredder spin columns. The homogenized lysate was then bound to AllPrep DNA Mini spin columns by centrifugation at 20,000 x g for 30 sec at RT. Subsequently, DNA was washed with 350 μL AW1 buffer and subjected to proteinase K digestion (5 min at RT). The digestion step was followed by two wash cycles (350 μL AW1 and 500 μL AW2 buffer) and DNA elution in 150 μL EB buffer. Wash steps and DNA elution was performed by centrifugation at 20,000 x g for 30 sec at RT.

2.2.5 Qubit fluorometric DNA quantification

Quantification of DNA and NGS libraries was performed with the Qubit dsDNA HS Assay Kit reagents. To prepare a working solution, Qubit dsDNA HS dye was diluted 1:200 in Qubit dsDNA HS buffer. For the quantification, 1 μL of sample was added to 199 μL of Qubit dsDNA HS buffer-dye-mixture, briefly vortexed and incubated for 2 min at RT. The resulting fluorescence intensity was measured using the Qubit 2.0 Fluorometer and concentrations were inferred based on an external standard provided in the Qubit dsDNA HS Assay Kit.

2.2.6 DNA quality assessment by capillary electrophoresis

Fragment size distribution and integrity of DNA samples and NGS libraries was assessed by capillary gel electrophoresis using the Bioanalyzer 2100 instrument with the High Sensitivity DNA Kit reagents. Sample preparation and instrument loading was performed according to the manufacturer´s instructions, applying 1 µL of undiluted DNA sample or NGS library.

2.2.7 Genomic DNA shearing

Genomic DNA was sheared to a mode fragment length of 180 bp using the Bioruptor® Pico sonication device. Per sample, 500 ng of gDNA (in 100 µL nuclease-free water) were transferred to a 1.5 mL Bioruptor® microtube and sonicated in the Bioruptor® Pico device at 4 °C for 12 cycles applying the 30-second on/off settings. DNA fragment length was confirmed by capillary electrophoresis using the Bioanalyzer 2100 instrument.

2.2.8 Global methylation and hydroxymethylation profiling

Genome-wide 5mC and 5hmC profiles were prepared by cfMeDIP-seq and hMeSEAL-seq using the protocols established by Shen *et al.* 2019 (Shen et al. 2019b) and Song *et al.* 2017 (Song et al. 2017). Both methods are based on selective labeling of the respective epigenetic DNA modification followed by pull-down-based enrichment and sequencing of the enriched fraction (Figure 7). The following sections describe (i) the preparation of spike-in amplicons for the assessment of 5(h)mC enrichment efficiency, (ii) lambda filler DNA generation to adapt cfMeDIP to low cfDNA input quantities, and (iii) 5mC enrichment by cfMeDIP-seq as well as (iv) 5hmC enrichment by hMeSEAL-seq. All steps were carried out in 0.2 mL PCR stripes if not stated otherwise.
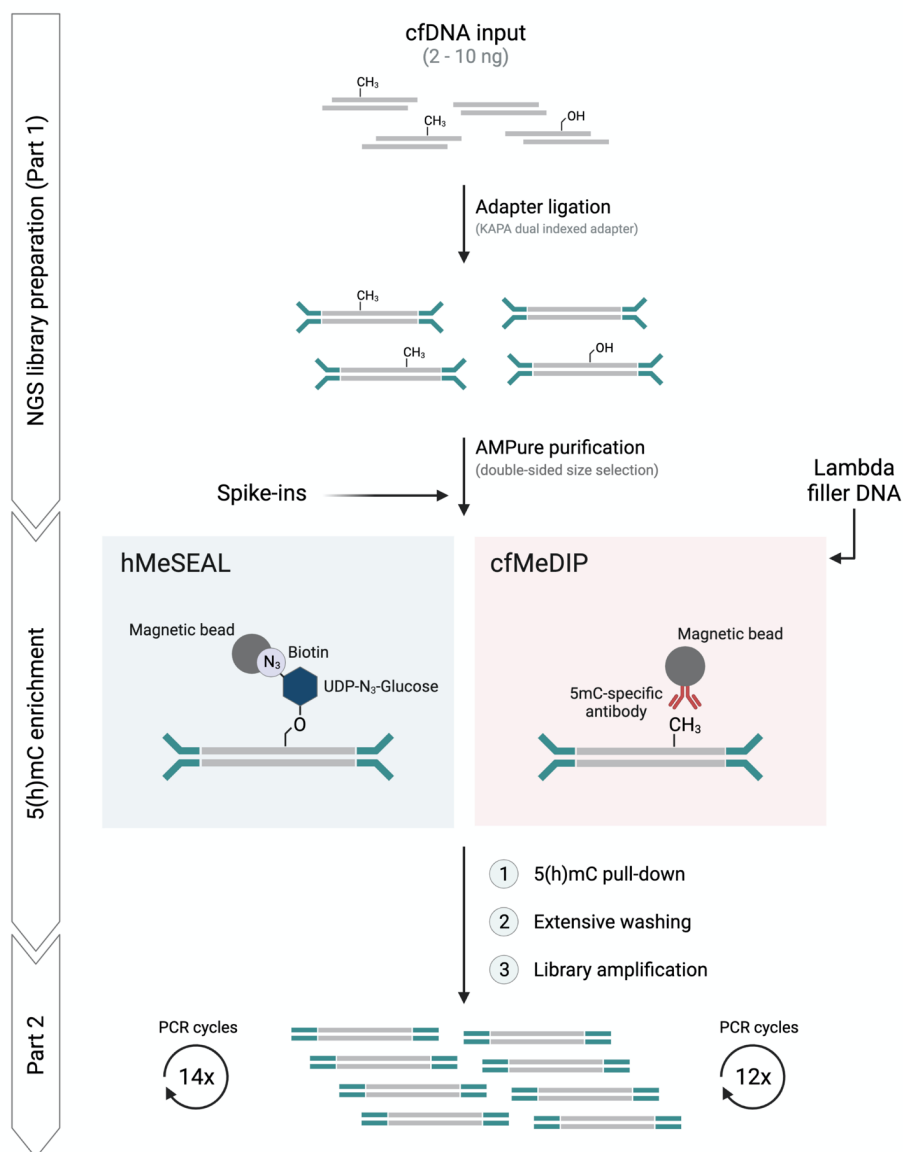
Figure 7: Schematic workflow of the library preparation and 5(h)mC enrichment protocol.

Spike-in amplicon preparation

Spike-in controls were generated by PCR amplification of lambda-DNA using the Phusion® High-Fidelity PCR Kit. Three types of amplicons were generated, containing either unmodified (5C spike-in), methylated (5mC spike-in) or hydroxymethylated CpG sites (5hmC spike-in). 5C spike-ins were prepared using a standard dNTP mix (i.e. dATP, dGTP, dTTP, and dCTP). For 5mC and 5hmC spike-ins, dCTPs were substituted by a mixture of 90% dCTPs and 10% 5-methyl-dCTPs or 10% 5-hydroxymethyl-dCTPs, respectively. The PCR reaction and temperature program was adapted from the Phusion® High-Fidelity PCR Kit protocol using 35 cycles and the primer sequences listed in Table 6. Resulting amplicons were purified by immobilization to 2 volumes of SPRI AMPure XP magnetic beads for 10 min at RT, washed twice using 200 µL freshly prepared 80% ethanol and eluted in 50 µL nuclease-free water. Spike-in controls were diluted to a working concentration of 5 pg/µL.

Generation of lambda filler DNA

Lambda filler DNA acts as a carrier for the immunoprecipitation reaction of cfMeDIP, thereby enhancing its specificity to allow the use of very low cfDNA input quantities. Filler DNA consists of a mixture of *in vitro* methylated and unmethylated lambda-DNA amplicons of fragment sizes resembling the length of sequencing adapter ligated cfDNA (Table 7). In a first step, lambda-DNA was diluted to 0.1 ng/µL and PCR amplified using 10 µL 5x SuperFi Buffer, 1 µL dNTPs (10 mM), 0.5 µL Platinum™ SuperFi™ DNA Polymerase (2 U/µL) as well as 3 µL of the appropriate primer pairs (10 µM) listed in Table 7. Samples were amplified at 98 °C for 30 sec, 30 cycles of 98 °C for 10 sec, 57 °C for 10 sec, 72 °C for 15 sec, followed by 5 min at 72 °C. Resulting PCR products were purified by immobilization to 1 volume of SPRI AMPure XP magnetic beads (for 10 min at RT), followed by two wash cycles (200 µL 80% ethanol) and elution in 100 µL nuclease-free water. Amplicon sizes were verified using the Bioanalyzer 2100 and quantified with the Qubit 2.0 Fluorometer. *In vitro* methylation of amplicons 1CpG, 5CpG, 10CpG, 15CpG, and 20LCpG was carried out using the CpG methyltransferase (M.SssI) by adding 2 µL M.SssI buffer, 0.4 µL S-adenosylmethionine (SAM) and 1 µL M.SssI enzyme per 1 µg of amplicon. Subsequently, samples were incubated for 15 min at 37 °C, followed by 20 min at 65 °C. Methylated amplicons were again purified and quantified as described previously. Efficiency of the methylation reaction was examined by subjecting aliquots of both methylated and unmethylated amplicons to restriction enzyme digest by the methylation non-sensitive HpyCH4IV enzyme. Two µL of CutSmart buffer, 1 µL HpyCH4IV enzyme and 100 ng amplicon were combined in a total volume of 20 µL and incubated for one hour at 37 °C, followed by 20 min at 65 °C. Fragment sizes of the restriction enzyme treated amplicons were checked using the Bioanalyzer 2100, verifying the digestion only in unmethylated fragments. Methylated fragments were pooled at equal concentrations and mixed in a 50:50 ratio with the unmethylated 20SCpG amplicon to generate the final filler DNA pool. Filler DNA was adjusted to a concentration of 5 ng/µL.

Cell-free methylation DNA immunoprecipitation (cfMeDIP)

NGS libraries for genome-wide 5mC profiling were prepared from 87 plasma samples of 29 ALK[+] NSCLC patients and 14 healthy donors as well as sheared gDNA from primary monocytes, neutrophils and erythroid progenitor cells using the cfMeDIP protocol. Six of the complete set of 93 plasma samples were excluded due to insufficient amount of cfDNA. On average, 6.9 ng cfDNA (range: 2 – 10 ng) and 250 ng gDNA served as starting material. The cfMeDIP protocol comprises (1) ligation of sequencing adapters, (2) enrichment of methylated cfDNA fragments, and (3) library amplification.

Ligation of sequencing adapters

Initially, sequencing adapters were ligated to the cfDNA using the KAPA HyperPrep Kit with KAPA Dual-Indexed Adapters for Illumina platforms. To achieve optimal ligation efficiency, cfDNA was end-repaired and A-tailed by adding 7 µL End-repair/A-tailing buffer and 3 µL End-repair/A-tailing enzyme mix to 50 µL of sample. Upon incubation at 20 °C for 30 min and 65 °C for 30 min, 30 µL of adapter ligation buffer, 10 µL DNA ligase and 10 µL 750 nM dual-indexed KAPA adapters were added for adapter ligation. Thereby, differently barcoded sequencing adapters were used for sample indexing to allow library pooling and multiplexed sequencing. Following overnight incubation at 16 °C, the adapter ligated libraries were purified using SPRI AMPure XP magnetic beads. In a first step, remaining, unbound sequencing adapters and adapter dimers were removed by adding 0.8 volumes of SPRI AMPure XP magnetic beads for 10 min at RT. After two consecutive washing cycles using 200 µL 80% ethanol, libraries were eluted from the beads with 50 µL nuclease-free water. Next, SPRI AMPure XP magnetic beads were used for a double-sided size selection to further exclude short DNA fragments and remaining long fragments (>800 bp) potentially derived from blood cell contamination. To this end, long fragments were immobilized by adding 0.25 volumes of SPRI AMPure XP magnetic beads to the sample. After 10 min incubation at RT, the supernatant, containing all but long DNA fragments, was transferred to a fresh tube and mixed with 0.8 volumes of SPRI AMPure XP magnetic beads to remove short fragments. Libraries were incubated and washed as described previously, following their elution in nuclease-free water. At this point, lambda filler DNA (5 ng/µL) was added to the libraries to increase the final amount of DNA (adapter ligated cfDNA and lambda filler DNA) to 100 ng in a total volume of 51 µL. In case DNA input quantities were >100 ng, no filler DNA was added.

Enrichment of methylated DNA fragments

Adapter ligated libraries were enriched for DNA fragments containing methylated CpG sites using the MagMeDIP qPCR Kit. First, 24 µL 1x MagBuffer A, 6 µL 1x MagBuffer B and 15 pg of each spike-in control (5C, 5mC and 5hmC) were added to the libraries. During 10 min incubation at 95 °C, the libraries were denatured into single strands, facilitating the subsequent binding of the 5mC antibody to CpG sites. All following steps were performed

on ice to prohibit re-hybridization of DNA fragments. Per sample, 11 µL of MagBeads were washed twice with 30 µL 1x MagBuffer A and eluted in 22 µL of the same buffer. Afterwards, the 5mC antibody was diluted 1:15 in nuclease-free water. Five µL of the antibody dilution alongside 20 µL washed MagBeads were added to the sample. Antibody binding to methylated DNA was carried out on a rotator (10 rpm) at 4 °C for 18 hours. After the incubation, the samples underwent four consecutive washing cycles: (i) bead separation and supernatant removal on a magnetic stand, (ii) bead resuspension in 100 µL MagWash Buffer-1, and (iii) incubation on a rotator (15 rpm) at 4 °C for 4 min. For the last washing step, MagWash Buffer-1 was replaced by MagWash Buffer-2. Separation from the 5mC antibody and purification of methylated DNA was performed using the iPure Kit v2. Fifthy µL iPure elution buffer were added per sample and incubated on a rotator (15 rpm) for 15 min at RT. The supernatant was separated from the MagBeads on a magnetic stand and transferred into a fresh 0.2 mL PCR strip. The previous DNA elution steps were repeated with another 50 µL of iPure elution buffer to optimize DNA recovery from the beads. Libraries were purified by adding 2 µL iPure carrier DNA, 100 µL 100% isopropyl alcohol and 10 µL iPure magnetic beads, followed by incubation on a rotator (15 rpm) for 10 min at RT. Samples were washed two times (5 min per washing step) on a rotator using first 100 µL iPure wash buffer 1 and then 100 µL iPure wash buffer 2. Afterwards, the supernatant was aspirated and the beads were air-dried for 2 min on a magnetic stand. To elute the libraries from the beads, 25 µL iPure buffer C were added and samples were incubated for 30 min on a rotator (15 rpm) at RT. The supernatant containing the 5mC-enriched libraries was then separated from the beads on a magnetic stand and transferred to a fresh 0.2 mL PCR strip. Two µL per 5mC-enriched library were taken for the assessment of enrichment efficiency and specificity, further described in section 2.2.9.

Library amplification

To obtain sufficient material for the consecutive sequencing analysis, libraries were PCR amplified using 22 µL of 5mC-enriched library, 25 µL of KAPA HiFi HotStart Ready mix and 5 µL of KAPA Library Amplification primer mix per sample. Amplification was carried out at 98 °C for 45 sec, 12 cycles of 98 °C for 15 sec, 60 °C for 30 sec, 72 °C for 30 sec, followed by 1 min at 72 °C. Amplified libraries were purified using 0.8 volumes SPRI AMPure XP magnetic beads as described previously and eluted in 15 µL nuclease-free water. Fragment sizes, library integrity and quantity were assessed using the Bioanalyzer 2100 and Qubit 2.0 Fluorometer, respectively.

Hydroxymethylation-selective chemical labeling of cfDNA (hMeSEAL)

5hmC-enriched NGS libraries were generated from all 93 plasma samples of 31 ALK$^+$ NSCLC patients and 14 healthy donors as well as sheared gDNA from primary monocytes, neutrophils and erythroid progenitor cells using the hMeSEAL protocol. On average 7.5 ng cfDNA (range: 2.5 – 10 ng) and 250 ng gDNA were used as starting material. Ligation of sequencing adapters and library amplification was performed analogous to the descriptions of the cfMeDIP method with three exceptions: (1) spike-in controls were added directly after the double-sided size selection of NGS libraries, (2) lambda filler DNA was not included for the preparation of hMeSEAL libraries and (3) libraries were amplified while still attached to magnetic streptavidin beads using 14 PCR cycles.

Enrichment of hydroxymethylated DNA fragments

cfDNA fragments containing hydroxymethylated CpG sites were first enzymatically coupled to azide-modified uridine diphosphoglucose (UDP-glucose), using the T4 phage ß-glucosyltransferase enzyme (T4-BGT). Three µL 10x NEBuffer 4, 3 µL T4-BGT (10 U/µL), and 1 µL UDP-glucose (3 mM) were added to the adapter ligated libraries (20 µL) and filled up to 30 µL with nuclease-free water, following one hour of incubation at 37 °C. To minimize carry over of unbound UDP-glucose, libraries were purified using 1 volume SPRI AMPure XP magnetic beads, as described previously, and eluted in 20 µL nuclease-free water. Utilizing Huisgen cycloaddition (click-) chemistry (Huisgen et al. 1963), dibenzylcyclooctyne-polyethylenglycol-4-distearylether-biotin conjugate (DBCO-PEG$_4$-biotin) was attached to the azide-residue of the UDP-glucose. To this end, 5 µL DBCO-PEG$_4$-biotin (1 mM) was added per sample, incubated for 1.5 hours at 37 °C, purified using 1 volume SPRI AMPure XP magnetic beads and eluted in 30 µL nuclease-free water. Next, biotin-tagged DNA fragments were immobilized on streptavidin-coupled magnetic Dynabeads™ M270. Per sample, 5 µL Dynabeads™ M270, pre-blocked with 50 µg UltraPure™ salmon sperm DNA, were added and filled up to 150 µL, using Buffer 1 (5 mM Tris-HCl (pH 7), 1 M NaCl, 0.5 mM EDTA, and 0.2% Tween-20) (Ambion® Buffer Kit). After 30 min of rotation (15 rpm) at RT, beads underwent four 5-min washes each with Buffer 1, Buffer 2 (Buffer 1 without NaCl), Buffer 3 (Buffer 1 with Tris-HCl (pH 8)), and Buffer 4 (Buffer 3 without NaCl). In brief, samples were placed on a magnetic stand, supernatant was removed without disturbing the beads, beads were resuspended in 200 µL of buffer and rotated (15 rpm) for 5 min at RT. After the last washing step, the bead-coupled 5hmC-enriched libraries were resuspended in 25 µL nuclease-free water. Two µL per library were taken for the assessment of enrichment efficiency and specificity, further described in section 2.2.9.

2.2.9 Evaluation of 5mC and 5hmC enrichment efficiency and specificity

For each 5mC- and 5hmC-enriched NGS library, enrichment specificity and efficiency was evaluated before subjecting samples to sequencing analysis. To this end, 2 μL per library were taken before library amplification and diluted in 28 μL nuclease-free water. Spike-in controls were quantified in triplicates by quantitative PCR (qPCR) using the LigthCycler® 480 instrument. Per reaction, 2 μL of diluted library were mixed with 5.5 μL 2x PrimaQuant CYBR qPCR master mix, 0.31 μL of the primer pairs (2.5 μM) used for spike-in amplicon generation (Table 6) and filled up to 11 μL with nuclease-free water. The qPCR was carried out in a LightCycler® 480 Multiwell Plate 384 using the following temperature program: 3 min at 98 °C, followed by 50 cycles of 10 sec at 98 °C and 30 sec at 60 °C. Concentrations of 5C, 5mC and 5hmC amplicons were determined based on external DNA standard curves.

2.2.10 Next-generation sequencing

Next-generation sequencing of 5mC- and 5hmC-enriched libraries was performed at the High Throughput Sequencing Unit of the DKFZ Genomic and Proteomics Core Facility. Libraries were pooled in equimolar concentrations to a final of 10 μM for multiplexed sequencing. For 5mC-enriched libraries, eight samples were pooled per sequencing run. 5hmC-enriched libraries were 23-plexed. All libraries were subjected to 75 bp paired-end sequencing on the Illumina NextSeq 550 system (high throughput). Flow-cell loading, cluster formation and NGS were performed according to the manufacturer´s recommendations.

2.2.11 NGS data processing and alignment

Initial NGS data processing was composed of the following steps: (i) raw sequence data quality assessment, (ii) removal of contaminating adapter sequences, (iii) read alignment to the human reference genome (hg19/GRCh38.p2), followed by (iv) duplicate collapsing and quality filtering. All processing steps were performed on the Galaxy instance of the DKFZ (Giardine et al. 2005). Raw sequence quality was determined using FastQC v0.11.2 (Andrews et al. 2012), followed by removal of Illumina universal sequencing adapters by Cutadapt v1.16 (Martin 2011). Next, adapter-trimmed sequencing reads in fastq format were mapped to the human reference genome hg19/GRCh38.p2 (obtained from the GENCODE release 22 (Frankish et al. 2019)) using Bowtie2 v2.2.6.2 (Langmead and Salzberg 2012). Paired-end options were enabled, removing paired reads with insert sizes <30 bp and >700 bp. Duplicate reads were marked using the RmDup function of SAMtools v1.3.1 (Li et al. 2009) and collapsed to allow only one read per alignment position. Finally, reads were filtered according to the following criteria (SAMtools v1.3.1): keep reads with a mapping quality score (MAPQ) > 10, keep properly paired reads and discard unmapped reads. Reads mapping to chromosomes X and Y as well as the mitochondrial genome were removed.

2.2.12 Quality control of 5mC- and 5hmC-enriched sequencing data

The R package MEDIPS v1.44.0 (Lienhard et al. 2014) was used to generate quality metrics of the sequenced libraries. This comprises the assessment of library saturation, CpG enrichment and CpG coverage. Saturation analysis was performed to assess whether the given set of paired reads is sufficient to create saturated – and therefore reproducible – coverage profiles of the human genome. First, sequencing reads were artificially doubled and then randomly divided into two sets. Afterwards, read coverage at 300-bp genomic windows was calculated using both sets and compared by Pearson correlation. Read coverages resulting in Pearson correlation coefficients >0.8 were considered to yield reproducible results. Both 5mC- and 5hmC-enriched libraries were expected to contain high CpG densities. The CpG enrichment score compares the relative frequency of CpG dinucleotides in the reference genome (genome.CpG$_{rel}$) to the CpG frequency in the sequenced fragments (fragments.CpG$_{rel}$). The number of CpGs within the human genome (genome.cg) and the sequencing data (fragments.cg) were first counted and then divided by the total number of bases in the reference genome (m) and sequencing reads (n). Subsequently, CpG enrichment scores were calculated as:

$$CpG\ enrichment\ score = \frac{fragments.CpG_{rel}}{genome.CpG_{rel}}$$

where

$$genome.CpG_{rel} = \frac{genome.cg}{m}$$

and

$$fragments.CpG_{rel} = \frac{fragments.cg}{n}$$

Afterwards, the number of CpGs covered at least once by the given set of reads was calculated. Additionally, the depth of coverage per CpG sites was determined. Fragments containing no CpG dinucleotide were counted as a measure for unspecific DNA pull-down.

2.2.13 5mC and 5hmC peak calling

Model-based analysis of ChIP-seq 2 (MACS2) v2.1.1 (Zhang et al. 2008) was used to identify genomic regions containing 5(h)mC modifications. For each sample, peak calling was performed using the "callpeak" command with parameters set to: "--format BAMPE --genome hs --keep-dup all --nolambda --p 1e-05 --call-summits". Peak summits were extended by ±100 bp to a total width of 201 bp. Fixed-width peaks were chosen to reduce the bias of differing peak widths during the comparison between 5mC- and 5hmC-enriched regions. To minimize the effect of varying sequencing depths on the peak calling algorithm, 5mC and 5hmC datasets were downsampled to a common coverage. When patient samples were compared to healthy controls, 5mC data was downsampled to a maximum of 15 million and 5hmC to 10 million paired reads per sample. 5mC- and 5hmC-enriched loci were compared in samples with both datasets available. Here, read coverage was adjusted to the sample with the lower sequencing depth. Downsampling was performed using the "view" command of SAMtools v1.3.3 (Li et al. 2009).

2.2.14 Peak annotation to genomic features and chromatin states

5(h)mC-enriched peaks were associated with genomic features using the "annotatePeaks.pl" command of the HOMER software v4.11 (Heinz et al. 2010) with default parameters. The following features were considered: promoter regions, 5′- and 3′-untranslated regions (UTRs), first exons and other exons, introns, transcription end sites (TESs), and intergenic regions. The core 15-state model containing coordinates of genomic regions associated with different functional chromatin states was downloaded from the ROADMAP epigenomics project (https://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/Chmm Models/coreMarks/jointModel/final/E116_15_coreMarks_segments.bed). The chromatin states were inferred from chromatin immunoprecipitation DNA-sequencing data (ChIP-seq; H3K4me3, H3K4me1, H3K36me3, H3K9me3, and H3K27me3) of GM12878 cells using the chromHMM software v1.22 (Ernst and Kellis 2012). Inferred states included active transcription start sites (TSSs), flanking active TSSs, sites of strong transcription, transcription at 5′ and 3′ of genes, sites of weak transcription, genic enhancers, enhancers, zinc finger nuclease (ZNF) genes and repeats, heterochromatin, bivalent/poised TSSs, flanking bivalent TSSs/enhancers, bivalent enhancers, repressed polycomb, weak repressed polycomb, and quiescent regions. 5(h)mC-enriched peaks were counted at each chromatin state using the "countOverlaps" function of the GenomicRanges R package v1.44.0 (Lawrence et al. 2013a).

Enrichment of 5(h)mC peaks at genomic features and chromatin states was determined by comparing the observed number of peaks to the peak number following random permutation of the peak set throughout the human genome. This was performed for each sample individually. Bedtools v.2.29.2 (Quinlan and Hall 2010) was employed for peak set permutation using the "shuffle" command with default parameters.

2.2.15 Gene ontology analysis

Gene ontology (GO) analysis of genes containing 5(h)mC-enriched peaks was performed using the clusterprofiler R package v4.0.0 (Yu et al. 2012) with the included "enricher" function. Gene sets were obtained from the Molecular Signature Database (MSigDB) v7.4.

2.2.16 Processing of primary blood cell 5mC and 5hmC data

Monocytes, neutrophils and erythroid progenitor cells are major contributors to the cfDNA pool in healthy individuals (Moss et al. 2018). Genomic regions with high 5(h)mC levels in these cell types are unlikely to be tumor-informative in plasma samples of cancer patients. The "featureCounts" function of Subread v1.5.3 (Liao et al. 2014) was used to obtain read coverages at 300-bp genomic windows from 5mC- and 5hmC-enriched NGS libraries of primary monocytes, neutrophils and erythroid progenitors. Paired-end options were enabled for read counting and fragments were assigned to the genomic window with the largest overlap. Afterwards, count data was normalized to the sample´s respective library size and by trimmed mean of M-values (TMM) normalization using the edgeR R package v3.34.0 (Robinson et al. 2009). The cell type´s 5mC and 5hmC signals at 300-bp windows were then multiplied by its relative contribution to cfDNA in plasma of healthy individuals (Figure 5B) (Moss et al. 2018). As the profiled cell types account for 72.2% of the predicted cfDNA composition, relative contributions were scaled to 100% (i.e. monocytes: 14.5%, neutrophils: 44.3%, and erythroid progenitors: 41.1%). Subsequently, count data of the three blood cell types were combined by summation to yield profiles approximating the 5(h)mC levels in healthy plasma.

2.2.17 Differential methylation and hydroxymethylation analysis

Differentially methylated (DMRs) and hydroxymethylated regions (DhMRs) between cfDNA samples from ALK[+] NSCLC patients and healthy individuals were determined using the R packages limma v3.48.0 (Ritchie et al. 2015) and edgeR v3.34.0. For each sample, read coverages at 300-bp genomic windows were determined and TMM normalized as described in section 2.2.16. Genomic windows with counts >20% of the total number of samples across both groups were considered for differential analysis. In addition, genomic windows covering no CpG dinucleotide were excluded. Normalized counts were subjected to the "voom" function of limma. Variance modeling on the observation level (voom) estimates the mean-variance relationship of $\log_2$-counts, generating a precision weight for each observation (Law et al. 2014). Next, a linear model using weighted least squares was fit for each gene and standard errors were smoothed by empirical bayes smoothing. DMRs and DhMRs at $p < 0.01$ and $|\log_2(\text{fold-change})| > 1$ were extracted and associated to genomic features as well as their nearest TSS using the "annotatePeaks.pl" function of HOMER. For heatmap and principle component analysis (PCA) visualization, z-scores were calculated

from TMM normalized counts of DMRs and DhMRs. Heatmaps were generated using the pheatmap R package v1.0.12 (Kolde 2013) with Euclidean distance and Ward clustering. Principle components were calculated using the "prcomp" function of the stats R package. 5mC and 5hmC signals displayed in boxplots and plots showing disease kinetics refer to TMM normalized counts per million (CPM).

2.2.18 Acquisition and processing of reference datasets

Reference tissue methylation and expression datasets used in this study were analyzed as part of The Cancer Genome Atlas (TCGA) project and downloaded from the GDC data portal as IDAT files (methylation data) or raw counts (expression data) (Grossman et al. 2016; Weisenberger 2014). Illumina 450k methylation array data of cfDNA from healthy individuals was obtained from GSE122126 (Moss et al. 2018) and whole blood expression data from the Genotype-Tissue Expression (GTEx) project v6 (Aguet et al. 2017).

Illumina 450k methylation array data

5mC profiles from primary tumor tissue of LUAD (n = 469), breast carcinoma (BRCA; n = 789), colon adenocarcinoma (COAD; n = 295), prostate adenocarcinoma (PRAD; n = 502), kidney renal clear cell carcinoma (KIRC; n = 322), and stomach adenocarcinoma (STAD; n = 395) patients were downloaded as raw IDAT files. Additionally, 5mC data of non-malignant adjacent lung, breast, colon, prostate, as well as kidney tissue (n = 29 per tissue type) and cfDNA of 4 healthy individuals was obtained. Normal stomach tissue was excluded due to only two samples with methylation data available. IDAT files were uniformly processed using the minfi R package v1.38.0 (Aryee et al. 2014). CpG sites represented by less than three beads or mapping to chromosomes X/Y or the mitochondrial genome were excluded. The detection p-value was set to <0.01. To adapt the base resolution 450k array data to the 5mC signals at 300-bp genomic windows determined from cfMeDIP-seq, β-values of CpGs mapping to the same window were averaged, resulting in one β-value per genomic window. Differential analysis between LUAD and all non-malignant tissue samples (n = 145) was performed as described in section 2.2.17. LUAD-specific DMRs were extracted at $p < 0.001$ and $|\Delta\beta| > 0.15$, with $\Delta\beta$ referring to the mean β-value difference between LUAD and non-malignant tissue samples. Next, regions uniquely hyper- and hypomethylated in each of the six considered cancer entities were determined adapting the feature selection procedure described by Moss and colleagues (Moss et al. 2018). In a first step, regions with β-value variances <0.1% across all samples were excluded. Afterwards, mean β-values were calculated across samples for each cancer entity. A methylation matrix X with N rows (genomic regions) and d columns (cancer entities) was built and each value within a row was divided by the sum of the same row to obtain relative 5mC levels comparing the six cancer types (X_i').

$$X_i{'} = \frac{X_i}{\sum_j^d X_{i,j}}$$

The top 100 hypermethylated regions with the highest $X_{i,j}{'}$ were determined for each cancer entity j. Uniquely hypomethylated regions were identified using the same strategy with an inversed methylation matrix $(1 - X)$.

RNA sequencing data

RNA sequencing data of primary tumor tissues from LUAD patients (n = 533) and non-malignant tissues (i.e. lung, breast, colon, prostate, kidney, and stomach; n = 29 samples per tissue type) were downloaded as raw counts. LUAD-specific gene expression was determined by differential analysis between LUAD and normal tissues (n = 174) using the procedure described in section 2.2.17. Differentially expressed genes were extracted at $p < 0.001$ and $|\log_2(\text{fold-change})| > 1$. Whole blood cell gene expression data from 755 healthy individuals were downloaded as fragments per kilobase per million reads (FPKM).

2.2.19 Copy number inference from 5mC and 5hmC data

Genome-wide copy number profiles were estimated from 5mC- and 5hmC-enriched sequencing data of patient and healthy control samples using the ichorCNA R package v0.2.0 (Adalsteinsson et al. 2017). In a first step, reads were counted at 1 Mb genomic windows and corrected for GC content and mappability bias by the HMMcopy Suite v1.34.0 (Lai et al. 2021). CNAs were then predicted by a Bayesian statistical framework of the hidden Markov model and an expectation-maximization algorithm. 5mC and 5hmC profiles of healthy individuals were used as a copy number neutral reference.

The CNAclinic R package (Mouliere et al. 2018a) was used to calculate t-MAD scores from 5mC- and 5hmC-enriched sequencing data. First, sequencing reads were randomly downsampled to 5 million paired reads per sample using SAMtools v1.3.3, as described in 2.2.13. Subsequently, reads were counted at 1 Mb genomic windows and corrected for GC content and mappability. Samples from healthy controls were processed identically and normalized by their median genome-wide count. $\log_2$-transformed copy number ratios were calculated comparing patient to control samples. The median absolute deviation from the $\log_2$-transformed copy number ratio was determined and denotes the t-MAD score.

2.2.20 Determination of molecular alterations from plasma samples

As part of another project, molecular alterations (i.e. SNVs and the *EML4-ALK* fusion) and CNAs were determined from the same ALK+ NSCLC patient plasma samples by CAPP-seq and sWGS, respectively. Experimental procedures and data analysis were described previously (Dietz et al. 2020).

2.2.21 Statistical analyses

Quantitative data was represented as boxplots with center lines indicating the median and boxes showing the interquartile range with Tukey whiskers. Unpaired two-sided Wilcoxon tests were performed to accept or reject the null hypothesis that there is no difference between two sample groups. Permutation tests were used to estimate the significance of overlap between two sets of genomic regions. Random permutations were generated using the regioneR R package v1.24.0 (Gel et al. 2015) and p-values were computed using a normal standard distribution. P-values < 0.01 were considered significant.

2.2.22 Data visualization

Visualization of data was performed using R packages ggplot2 v3.3.4 (Wickham 2016), VennDiagram v1.6.20 (Chen and Boutros 2011) and pheatmap v1.0.12 (Kolde 2013). Infographics were prepared in BioRender (BioRender, Toronto, Canada).

3 Results

3.1 Patient characteristics

Seventy-nine plasma samples from 31 patients were collected at the Thoraxklinik Heidelberg and LungenClinic Großhansdorf, Germany. For 11/31 patients, serial plasma samples (range: 2 – 14 samples) taken throughout therapy were obtained (Figure 8 and Table S1). All patients were diagnosed with metastatic ALK$^+$ NSCLC as confirmed by molecular pathological tissue examination at the Institute of Pathology in Heidelberg. In addition, 14 plasma samples from healthy donors, i.e. subjects without known current disease, were collected at the Thoraxklinik Heidelberg.

Comprehensive clinical annotations were available for 24 patients, comprising 72 plasma samples (Table 10). Per patient demographics are listed in Table S1. For healthy reference samples, demographic information was unavailable. Clinical information included results of radiological assessment of therapy success and disease status (by chest CT and brain MRI), molecular information about *ALK* fusion variants and *TP53* co-mutations, and number/type of sequentially administered therapy regimens. On average, patients received 3.9 lines of sequential therapy (range: 1 – 8) since first diagnosis, including at least one line of ALK-TKI. The median OS since diagnosis of stage IV disease was 31 months, and 13 patients deceased during follow-up. Plasma-based mutation and CNA profiling was performed on the same 24 patients, as part of another project (Dietz et al. 2020). Targeted panel sequencing from cfDNA (CAPP-seq) revealed mutations in 22 out of 24 patients in at least one of the longitudinally taken plasma samples (oncoprint (Dietz et al. 2020); Figure S1). The *ALK* fusion variant was detectable in 41.7% patients and 7 patients had additional resistance mutations in the *ALK* gene (Table S1).
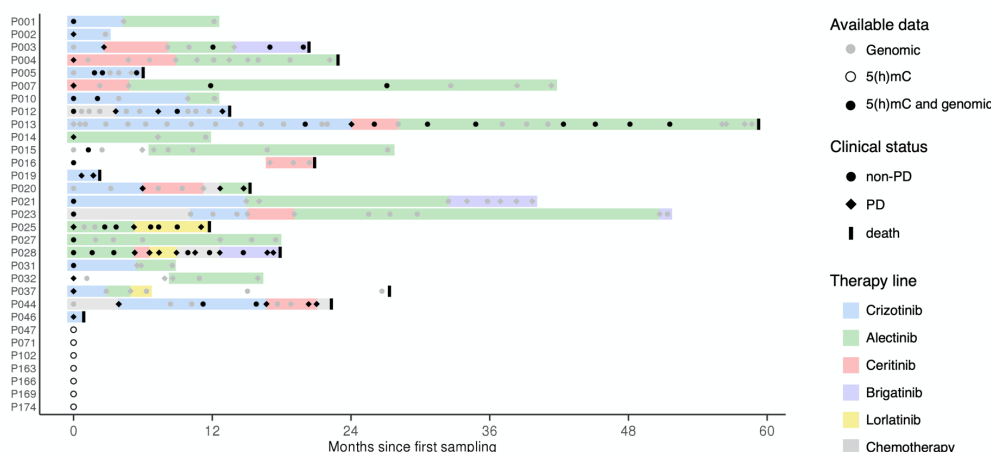
Figure 8: Cohort overview. Swimmer plot providing an overview of longitudinally taken plasma samples, survival and administered therapy lines. Dots and diamonds represent plasma sampling at non-progressive (non-PD) and progressive (PD) time points, respectively. Available datasets are indicated by black or gray colors. Genomic data includes profiling of molecular alterations by targeted panel sequencing and chromosomal instability inference from shallow whole-genome sequencing data. Administered therapy lines since first plasma sampling are indicated by the background colors.

Table 10: Patient characteristics (n = 31)

		ALK⁺ NSCLC patients
Age (mean, range)		57.5 (42 – 84)
Gender, % male		45.2%
Histology, % adenocarcinoma		100.0%
Stage	IV	29
	no data	2
ALK fusion variant	*EML4-ALK* V1	11
	EML4-ALK V2	3
	EML4-ALK V3	9
	others[1]	2
	no data	6
TP53 status, mutated	positive	6
	negative	17
	no data	8
Treatment, sample number	1st gen ALK TKI	21
	2nd gen ALK TKI	32
	3rd gen ALK TKI	6
	chemotherapy	9
	immunotherapy	1
	naive	9
	no data	1
Follow-up, months (median, range)		36.6 (3 – 130)
Number of samples per patient (mean, range)		2.5 (1 – 14)
Number of therapy lines (mean, range)		3.9 (1 – 8)
Radiological evaluation at sampling, sample number	PD	32
	SD	36
	PR	4
	no data	7

ALK, anaplastic lymphoma kinase; EML4, echinoderm microtubule-associated protein-like 4; gen, generation; SD, stable disease; PD, progressive disease; PR, partial response.

[1] one patient with a K9A20 and one with an E9A10 fusion.

3.2 Sample characteristics

cfDNA was isolated from 93 plasma samples, using on average 1.6 mL (range: 0.5 – 3.0 mL) of plasma as starting material. DNA concentrations varied considerably between samples with a range from 2.85 to 218.40 ng DNA per mL of plasma (mean = 29.68 ng/mL). Consistent with the literature (van der Pol and Mouliere 2019), significantly higher DNA concentrations were observed in plasma samples from patients when compared to healthy donors (Figure 9A; p = 3e-05). No significant difference was observed comparing cfDNA concentrations by demographic and molecular characteristics as well as radiologically assessed disease status (Figure S2A-E). The significantly increased cfDNA concentrations of samples taken during the second line of therapy is likely derived from the overrepresentation of samples taken from patient P012 (comprising 5 out of 14 plasma samples in this group) and not due to increased cfDNA shedding at this therapy line (Figure S2F). Assessment of cfDNA integrity revealed the presence of small DNA fragments with a mode size of 167 bp (range: 161 – 178 bp) in all samples (Figure 9B). Thirty-one out of 93 samples showed additional peaks at double and/or triple the length of the first peak. High molecular weight DNA (>700 bp) was observed in a minor fraction of the cfDNA samples (11/93) and comprised on average 28% of the total DNA content. Fragment length profiles were unaffected by demographic or clinicopathological parameters (data not shown). The quality and quantity of all samples was sufficient for subsequent library preparation. No sample was excluded.
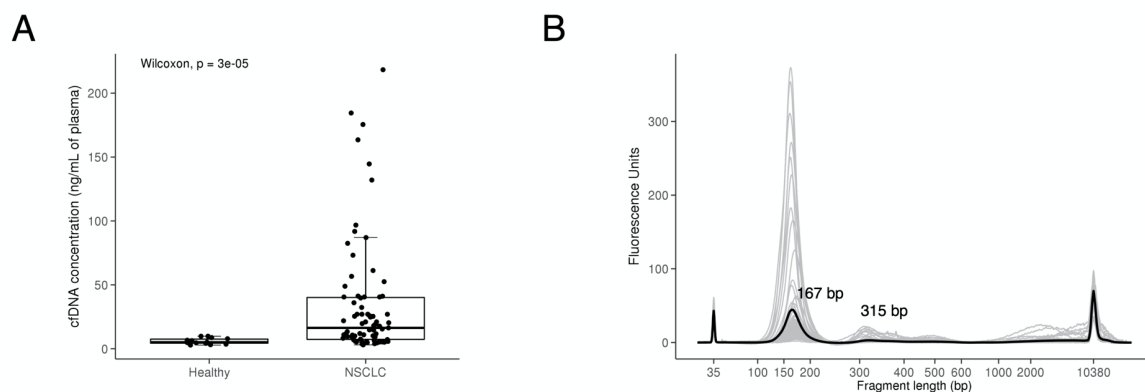
Figure 9: Concentration and integrity of cfDNA samples. (A) CfDNA concentrations comparing healthy donors (n = 14) and NSCLC patients (n = 79). Each dot represents one cfDNA sample. (B) Integrity and fragment length distribution of all cfDNA samples (n = 93). The black line indicates the median fragment length profile. Gray lines show the profiles of the individual samples. Statistical analysis was performed using a two-sided Wilcoxon test. Boxplot center lines indicate the median and boxes represent interquartile range with Tukey whiskers.

3.3 Performance evaluation of 5mC and 5hmC profiling

3.3.1 Preparation of 5mC- and 5hmC-enriched sequencing libraries

After cfDNA isolation, sequencing libraries were prepared combining the KAPA Hyper Prep library preparation kit and 5(h)mC enrichment methods (Shen et al. 2019b; Song et al. 2017). DNA input for library preparation varied depending on the available quantity of cfDNA per sample (Table 11). 5hmC-enriched libraries were prepared from all 93 samples, whereas for 5mC enrichment, six samples had to be excluded due to insufficient amounts of cfDNA starting material (<2 ng). Quantity, integrity and enrichment efficiency was evaluated in all 180 sequencing libraries. All libraries were of sufficient quantity for sequencing analysis (Table 11). Capillary gel electrophoresis revealed high library quality, resembling the DNA laddering observed in the cfDNA input (Figure S3A). The fragment sizes of the libraries are shifted to higher molecular weights due to the attachment of Illumina sequencing adapters. Prior to 5(h)mC enrichment, libraries were spiked with three amplicons carrying either methylated, hydroxymethylated or unmodified CpGs. Following enrichment, the residual spike-in quantity was measured by qPCR (external standard curves; Figure S3B). Both protocols showed high enrichment efficiencies and specificities towards their respective epigenetic mark (Figure 10). All 180 sequencing libraries passed the quality criteria and were sequenced at the DKFZ core facility.

Table 11: Quality metrics of 5(h)mC enrichment and sequencing data

		5mC	5hmC
DNA input, ng (mean, range)		6.9 (2.0 – 10.0)	7.5 (2.5 – 10.0)
NGS library quantity, ng (mean, range)		93.4 (19.8 – 264.0)	94.1 (10.6 – 356.0)
Recovered spike-in DNA, pg (mean)	5C	0.050	0.003
	5mC	1.700	0.001
	5hmC	0.090	1.600
Number of paired reads, x1e6 (mean, range)	raw	55.4 (29.2 – 72.8)	22.0 (14.0 – 27.6)
	after QC	26.4 (7.7 – 41.5)	12.9 (3.4 – 19.6)
Non-duplicate mapping rate, % (mean, range)		57.3 (52.1 – 61.9)	62.0 (25.3 – 74.7)
Saturation analysis, r (mean, range)		0.98 (0.96 – 0.98)	0.86 (0.66 – 0.92)
CpGs covered, % (mean, range)		68.8 (50.0 – 82.1)	56.5 (30.7 – 66.1)
CpGs covered >5x, % (mean, range)		32.7 (12.9 – 43.2)	10.3 (0.4 – 19.3)
CpG enrichment score (mean, range)		2.9 (2.5 – 3.3)	2.0 (1.8 – 2.3)
Reads covering no CpG site, % (mean, range)		3.1 (1.3 – 10.9)	3.1 (1.1 – 10.8)

CpG, cytosine-guanine dinucleotide; ng, nanogram; pg, picogram; QC, quality control.
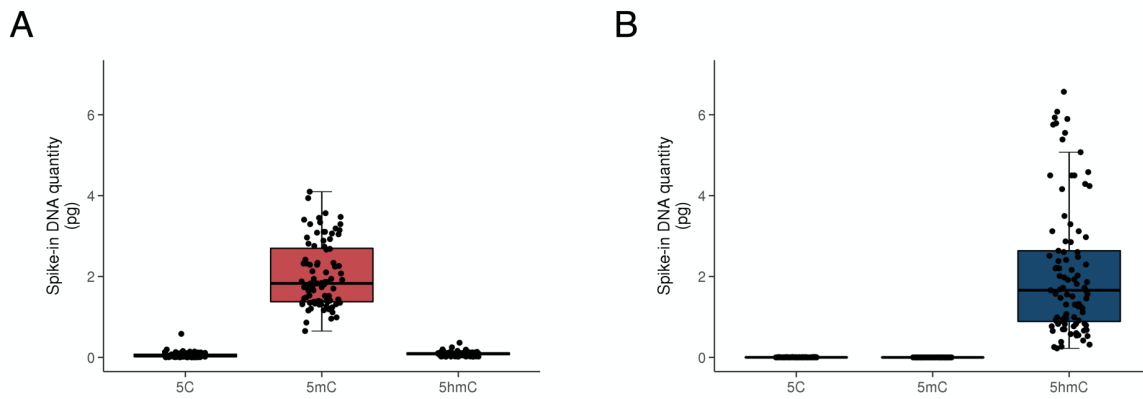
Figure 10: Spike-in recovery following 5mC and 5hmC enrichment. Boxplots represent the recovered spike-in quantities in 5mC- (A) and 5hmC-enried (B) sequencing libraries. Each dot represents one sample.

3.3.2 Evaluation of sequencing performance

The following quality parameters were considered for the evaluation of sequencing performance: (i) total read coverage after quality filtering, (ii) sequencing library saturation at the given read coverage, (iii) CpG enrichment, (iv) CpG coverage, and (v) number of reads covering no CpG site. Based on recent literature, I aimed for sequencing coverages of 20 million (5mC) and 10 million (5hmC) paired reads following sequence quality filtering and deduplication (Shen et al. 2019b; Song et al. 2017). With an average of 26.4 million (5mC) and 12.9 million (5hmC) paired reads, this number was surpassed in the majority of samples (Table 11 and Figure 11A; x-axis). Three 5mC-enriched and one 5hmC-enriched samples were excluded due to low sequence coverage (<50% of the desired number of reads). Saturation analysis was performed to evaluate the reproducibility of each library at the given read coverage. Libraries with an estimated Pearson correlation coefficient <0.8 were excluded due to insufficient reproducibility (Lienhard and Chavez 2016). 5mC-enriched samples were highly reproducible (mean Pearson, $r = 0.98$), while 5hmC-enriched samples demonstrated lower correlation coefficients (mean Pearson, $r = 0.86$; Figure 11A) at similar sequencing depths. Four additional 5hmC-enriched libraries had to be excluded due to low reproducibility. Next, the density of CpG sites within the sequenced fragments was determined and compared to the expected CpG abundance in the human genome (CpG enrichment score). The CpG enrichment score was used as a quality metric for the 5(h)mC precipitation reaction. A mean 2.9-fold enrichment of CpG sites in 5mC- and mean 2.0-fold enrichment in 5hmC-enriched samples was observed, confirming the successful pull-down of both epigenetic marks (Figure 11B). On average, 68.8% and 56.5% of CpGs were covered by the 5mC and 5hmC datasets, respectively. High sequence coverage (>5x) was observed at 32.7% (5mC) and 10.3% (5hmC) of CpG sites (Figure 11B). Only a small fraction of reads (mean = 3.1% in both datasets) covered no CpG site, suggesting low unspecific DNA pull-down in both protocols.
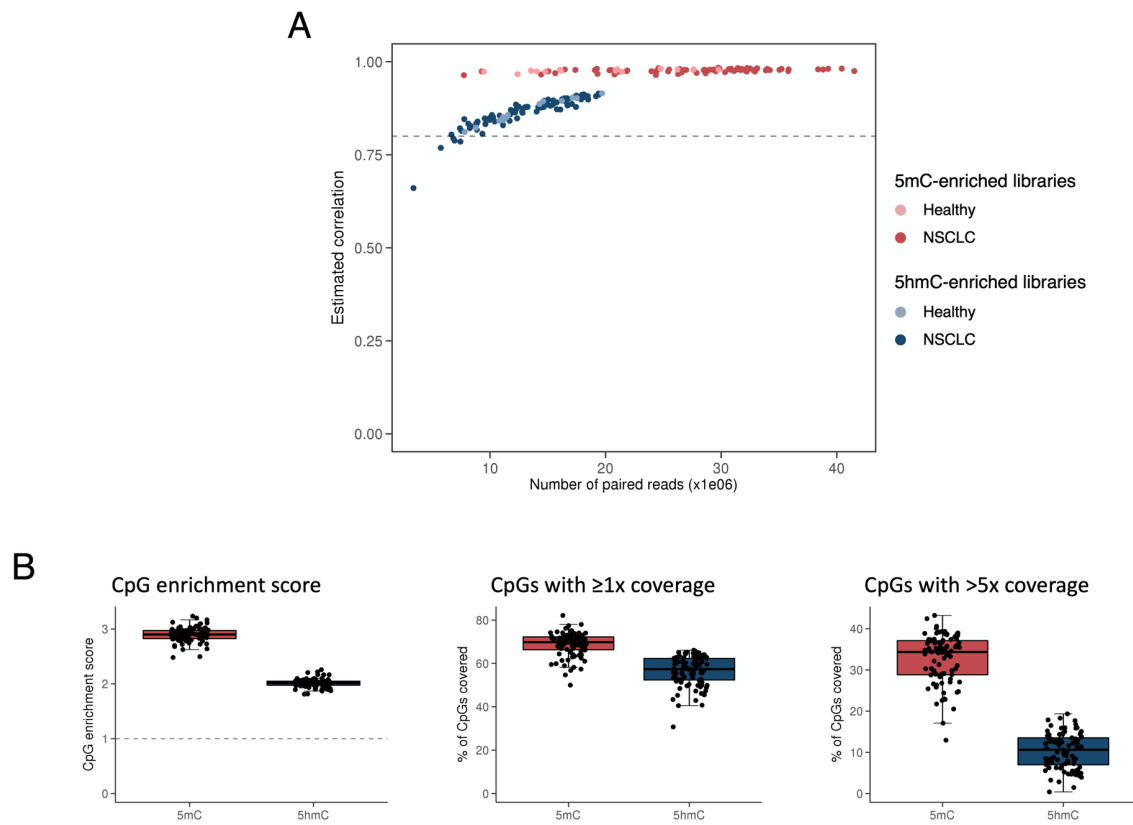
Figure 11: Quality metrics of 5(h)mC sequencing data. (A) Estimated saturation of 5mC (red) and 5hmC (blue) libraries at their respective sequencing coverage. Pearson correlation coefficients were calculated by dividing the sequencing data of each library into two random sets, followed by their artificial doubling and correlation. The dotted horizontal line represents the cut-off at a Pearson correlation coefficient of 0.8. (B) CpG enrichment scores, percent of CpGs covered ≥1x and >5x for each sample. Each dot represents one sample. CpG, cytosine-guanine dinucleotide; NSCLC, non-small cell lung cancer.

3.4 Genome-wide distribution of 5mC and 5hmC loci

To locate the genomic regions associated with 5mC and 5hmC in cfDNA, I first identified 5(h)mC-enriched loci using MACS2 peak calling (Zhang et al. 2008), defining 201-bp windows around the peak summit as 5(h)mC-enriched. Only samples with both 5mC and 5hmC data available (n = 84) were considered for this analysis, representing 28 NSCLC patients and 13 healthy donors. Comparing the total number of peaks between the 5mC and 5hmC datasets, a significantly higher number of 5mC loci was observed (Figure 12A; p < 2.2e-16), whereas the comparison between NSCLC patient samples and healthy donors showed no significant difference in both datasets (Figure S4A and B). Next, the distribution of 5mC and 5hmC peaks across eight genomic regions was assessed: promoters, 5′- and 3′-UTRs, first exons, exons, introns, TESs, and intergenic regions. The majority of 5mC and 5hmC peaks occurred within non-coding regions of the genome (introns and intergenic regions; Figure S5). However, compared to the genomic background, enrichment in both datasets was observed at (first) exons, promoters, 5′-/3′-UTRs, TESs and introns. Peaks at intergenic regions were underrepresented in both datasets (Figure 12B). The most significant differences between the two datasets were observed at introns (p < 2.2e-16), promoters (p < 2.2e-16) and 5′-UTRs (p < 2.2e-16), all showing a higher abundance of 5hmC peaks. The comparison between samples from patients and healthy donors revealed no significant difference at these genomic features (Figure S4A and B).

I then aimed to evaluate whether the differences in the mapping of 5mC and 5hmC peaks can be associated with the different functional roles of these epigenetic marks. To this end, an annotation file was downloaded which segments the genome of the human lymphoblastoid cell line GM12878 into 15 chromatin states. These chromatin states were inferred from ChIP-seq data of histone marks with well characterized functional associations (H3K4me3, H3K4me1, H3K36me3, H3K9me3, and H3K27me3), using the chromHMM software (Ernst and Kellis 2012). GM12878 data was used as reference for the comparison to cfDNA 5mC and 5hmC profiles, since the majority DNA fragments in plasma are derived from white blood cells (Lehmann-Werman et al. 2016; Moss et al. 2018; Sun et al. 2015). 5mC and 5hmC peaks were mapped to these chromatin states and relative enrichment compared to the genome background was calculated (Figure 12C). This analysis revealed higher enrichment of 5hmC peaks at chromatin states associated with active transcription and enhancer regions, while 5mC peaks were associated with repressed elements of the genome (heterochromatin, weak repressed polycomb and quiescent regions). Especially regions proximal to the TSS of actively transcribed genes showed enrichment for 5hmC and underrepresentation of 5mC peaks. This might reflect the association of 5hmC and 5mC with active and repressed gene expression, respectively.
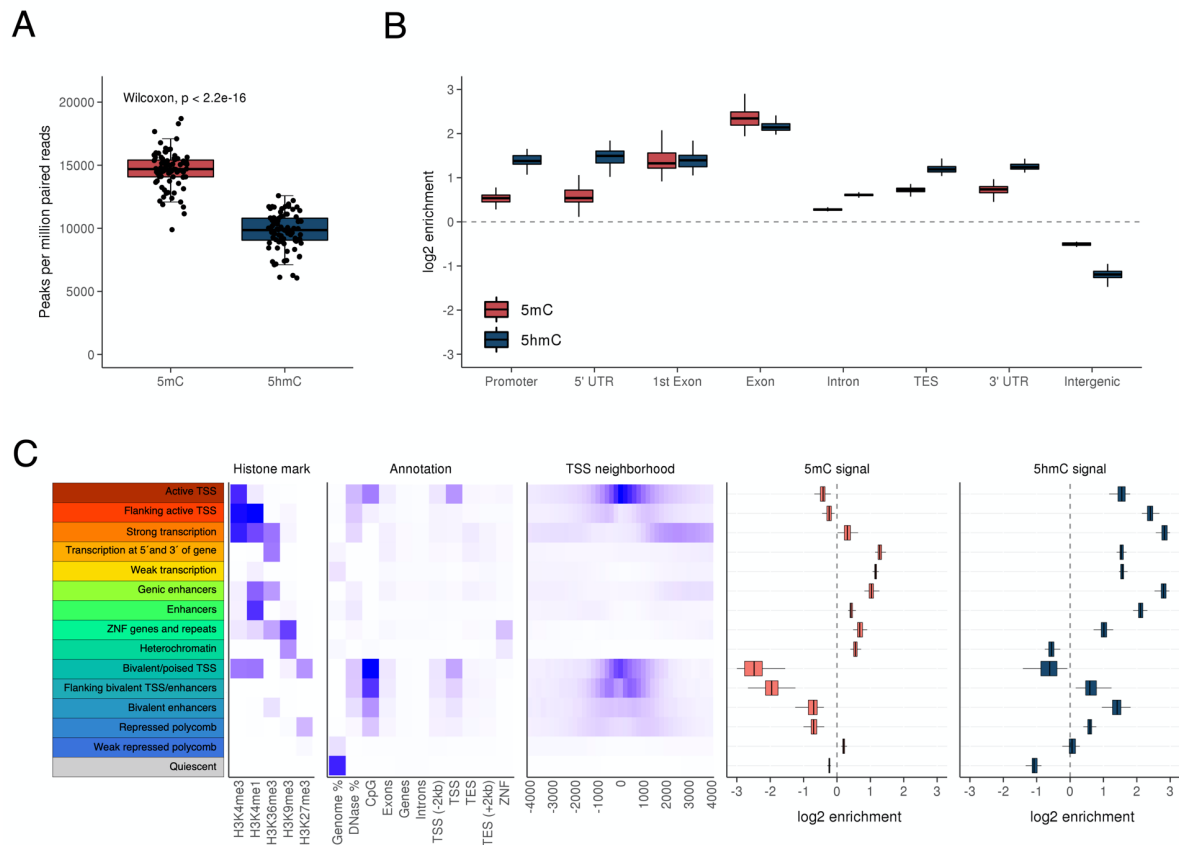
Figure 12: Differential enrichment of 5mC and 5hmC at genomic features and chromatin states. (A) Number of 5mC (red; n = 84) and 5hmC (blue; n = 84) peaks per million paired reads. Each dot depicts a plasma sample. (B) Boxplots illustrate the log₂-transformed enrichment of 5mC and 5hmC peaks at genomic features. Log₂ enrichment compares the actual number of peaks per feature to the peak number expected by random chance. Positive values represent enrichment and negative values underrepresentation of 5(h)mC peaks at the respective feature. (C) 5mC (red) and 5hmC (blue) occupancy at 15 chromatin states inferred from histone modification profiles generated from GM12878 cells (data obtained from ENCODE). For each chromatin state, histone mark, genomic elements and TSS neighborhood association is provided, with dark blue colors indicating high abundances. NSCLC patient and control samples were combined in (A), (B) and (C). Only samples with both 5mC and 5hmC enrichment data available were used for the illustrated comparisons (n = 84). CpG, cytosine-guanine dinucleotide; ENCODE, Encyclopedia of DNA Elements; kb, kilobase; NSCLC, non-small cell lung cancer; TES, transcription end site; TSS, transcription start site; UTR, untranslated region; ZNF, zinc finger nuclease.

3.5 TSS-proximal DNA hydroxymethylation is associated with active gene expression

The role of 5mC and 5hmC in the regulation of gene expression is well characterized. Especially when located within or proximal to promoter regions, 5mC is associated with repressed transcription, while 5hmC confers an active transcriptional status (Greenberg and Bourc'his 2019; Thomson et al. 2012). Earlier, high abundances of 5hmC peaks proximal to the TSS (±1,500 bp) of actively transcribing genes were observed (Figure 12C), suggesting their 5hmC-mediated transcriptional regulation. Therefore, I investigated whether 5mC and/or 5hmC abundances near promoter regions can be used to infer the expression status of genes in cfDNA.

The majority of cfDNA in healthy individuals is derived from hematopoietic cells (Lehmann-Werman et al. 2016; Moss et al. 2018; Sun et al. 2015). To test whether blood cell-associated expression profiles can be inferred from plasma, genes with at least one TSS-proximal (±1,500 bp) 5mC or 5hmC peak were extracted from the 5mC and 5hmC profiles of 13 healthy donors. Genes with 5mC or 5hmC peaks in at least 7/13 samples were considered as high confidence peaks and, therefore, used for following analysis. This resulted in 4,474 5mC- and 2,341 5hmC-associated genes as well as 2,065 genes with both 5mC and 5hmC peaks near their TSS (Figure 13A). Subsequently, gene expression data of 755 whole blood samples was obtained from the GTEx project v6 (Aguet et al. 2017). Expression levels of genes with TSS-proximal DNA modifications (i.e. 5mC, 5hmC or both) were compared to an equal number of randomly selected genes without 5(h)mC marks. The whole blood expression level of 5hmC-associated genes was significantly higher compared to genes without 5hmC peaks near their TSS (Figure 13B; $p < 2.2e{-}16$). Unexpectedly, also significantly higher expression of 5mC-associated genes was observed ($p < 2.2e{-}16$). However, the mean whole blood expression of 5hmC genes was 3.1-fold greater compared to genes with 5mC peaks proximal to the TSS, suggesting a closer association of 5hmC to active transcription. Genes with co-modified TSSs had a 2.2-fold higher mean expression (mean = 6.34 FPKM) when compared to genes carrying only 5mC marks. Similar associations with transcriptional activity were observed when comparing the 5mC and 5hmC related gene sets to various other tissue types (data not shown). Next, MSigDB was inspected to identify biological functions (C5.Ontology gene sets) associated with genes with 5mC- and 5hmC-specific modifications (Figure 13C). Genes with 5mC peaks near their TSS were mainly associated with processes involved in cell division (meiosis I cell cycle process, meiotic cell cycle, and meiotic cell cycle process). 5hmC genes were enriched in gene sets with functional associations with cytokine production as well as various GO terms regarding the migration of different blood cell types. These results suggested that especially 5hmC profiling from plasma samples can capture blood cell-derived biological signals.

A

B

C
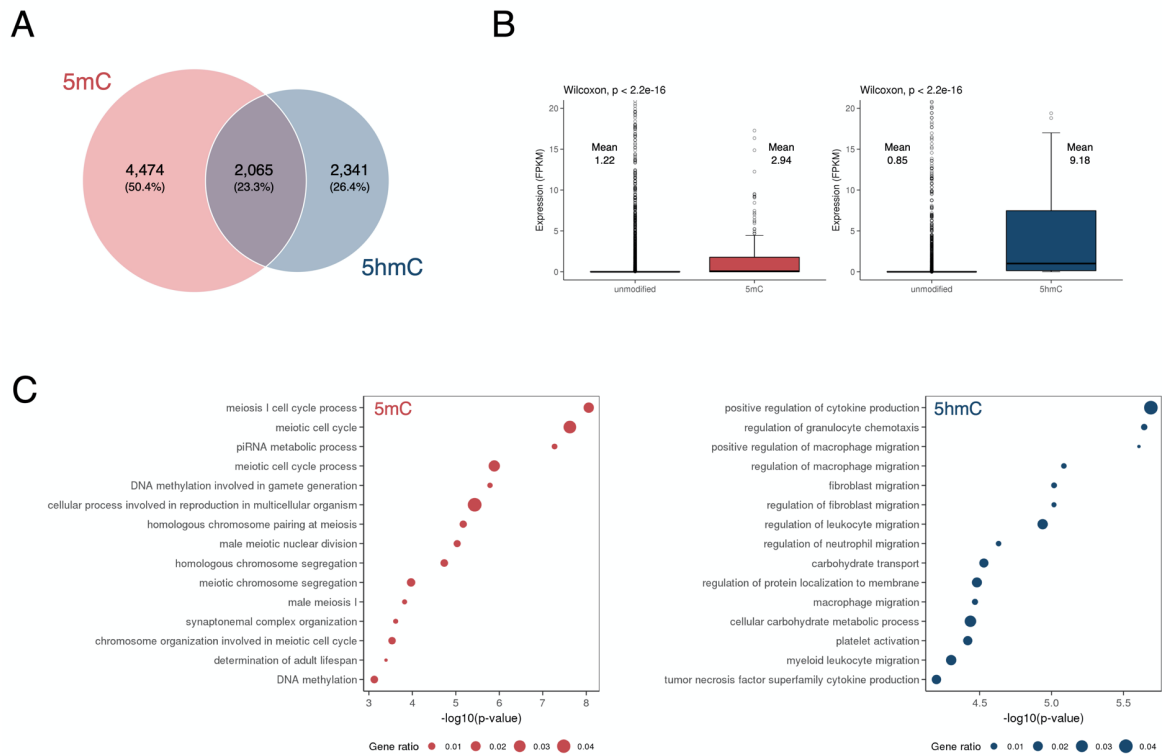
Figure 13: Association of plasma 5(h)mC levels of healthy individuals with gene expression in whole blood and gene ontology analysis . (A) Venn diagram illustrating the overlap between TSS-proximal (±1,500 bp) 5mC and 5hmC peaks. (B) Boxplots showing the difference in FPKM between genes with TSS near 5mC (red; n = 4,474 genes) and 5hmC peaks (blue; n = 2,341 genes) compared to an equal number of genes without TSS associated DNA modification. Gene expression was measured in whole blood of 755 individuals (obtained from GTEx). (C) Dots illustrate the 15 most significantly enriched gene sets with TSS-proximal 5mC (left) and 5hmC (right) peaks (MSigDB: C5.Ontology gene sets). FPKM, fragments per kilobase per million reads; GTEx, Genotype-Tissue Expression project; MSigDB, molecular signature database; piRNA, piwi-interacting RNA; TSS, transcription start site.

3.6 Identification of NSCLC-specific 5mC biomarkers by cfMeDIP-seq

To identify NSCLC-specific 5mC signatures from plasma, a bioinformatic workflow that aims to enrich tumor-derived signals from genome-wide methylation data was established. The ctDNA fraction in plasma can vary considerably between patients and is commonly lower than the non-tumor fraction (Bettegowda et al. 2014). With the majority of non-tumor cfDNA originating from blood cells (Moss et al. 2018), I first intended to exclude genomic regions exhibiting high methylation signals in blood. I reasoned that these regions are likely masked by the blood cell methylation pattern and are therefore not informative about the patient´s tumor. After the exclusion of these sites, DMRs were identified comparing samples from healthy donors to NSCLC patient samples. The tumor specificity of the identified DMRs was then validated using TCGA methylation data generated from LUAD tissue (Figure 14). In a last step, 5mC signals of selected DMRs were correlated to genomic tumor biomarkers quantified in the same plasma samples and their utility for therapy monitoring was examined.
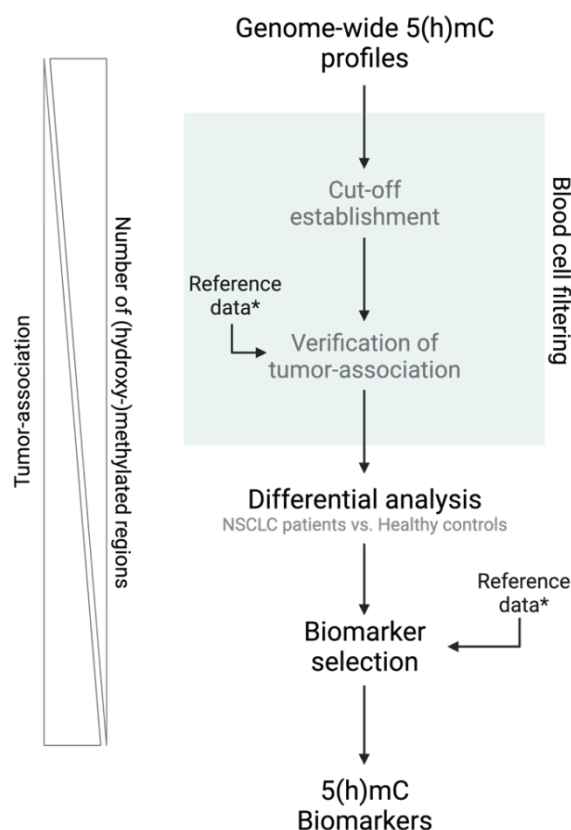
Figure 14: 5(h)mC biomarker identification workflow.

3.6.1 Exclusion of genomic regions hypermethylated in blood cells unmasks tumor-derived 5mC signals

Moss and colleagues (Moss et al. 2018) used genome-wide methylation profiles to quantify the major cell types contributing to cfDNA in healthy individuals (Figure 5B). This data was used as a basis to identify genomic regions hypermethylated in non-tumor cfDNA. By excluding these regions, I aimed to reduce the uninformative background signals to reveal tumor-derived 5mC signatures. To this end, genome-scale 5mC data from primary monocytes, neutrophils and erythroid progenitor cells was generated by cfMeDIP-seq. These cell types account for the majority of the non-tumor cfDNA fraction (72.2%). On average, 48.1 million deduplicated paired reads were obtained from blood cell cfMeDIP-seq and CpG enrichment scores were comparable to those determined from cfDNA samples (mean = 3.03), confirming successful 5mC enrichment (further quality metrics are listed in Table S2). To generate combined blood cell 5mC signals, sequencing depth normalized read density at 300-bp genomic windows was determined. Next, the read density of each window was multiplied by the cell type´s relative contribution to normal plasma, followed by summation. Then, I prepared an orthogonal 5mC dataset using 450k methylation array data from primary tumor tissue of LUAD patients (n = 469) (Weisenberger 2014) and cfDNA from healthy individuals (n = 4) (Moss et al. 2018). To increase the comparability between the two datasets, β-values falling into the same 300-bp genomic window were averaged. The mean β-value difference between LUAD tissue and healthy cfDNA was correlated to the $\log_2$-transformed fold-changes comparing cfMeDIP-seq data from NSCLC patients (n = 71) to controls (n = 13). A modest correlation was observed when the entire dataset was considered (Figure 15A; Spearman, $\rho = 0.11$). Spearman rank correlation coefficients increased when genomic regions hypermethylated in the combined blood cells were excluded. The highest correlation was observed when genomic loci with a 5mC blood cell signal <20% of the median 5mC signal in NSCLC plasma were retained (Figure 15B and Table 12; Spearman, $\rho = 0.26$). In the following, this cut-off was used to exclude genomic regions with high blood cell methylation. Notably, the correlation of the cfMeDIP-seq data to other cancer types (Illumina 450k methylation array data obtained from TCGA) resulted in similar high Spearman rank correlation coefficients, while non-malignant lung tissue showed lower correlation (Table 12).

A

Total cfMeDIP-seq dataset
n = 9.6 million genomic regions

B

Blood cell methylation <20%
n = 577,701 genomic regions
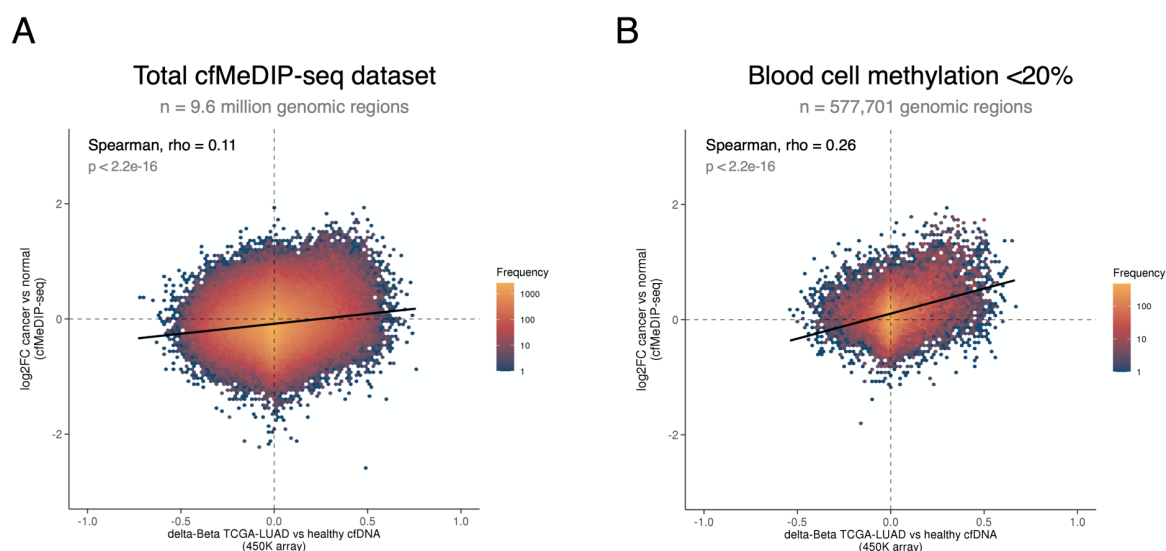
Figure 15: Correlation of differential methylation between cfDNA of NSCLC patients and primary LUAD tissue. Density plots illustrate the overlap between differentially methylated regions (DMRs) comparing plasma of NSCLC patient samples to healthy controls (cfMeDIP-seq; y-axis) and DMRs comparing lung adenocarcinoma tissue (LUAD) to healthy cell-free DNA (450k methylation array; x-axis), considering the entire cfMeDIP-seq dataset (A) and only genomic regions with blood cell methylation <20% (B). Hex bin colors indicate the number of overlaps at the given bin and linear regression lines are shown in black. cfDNA, cell-free DNA; cfMeDIP-seq, cell-free methylation DNA immunoprecipitation sequencing; log2FC, log$_2$-transformed fold-change; NSCLC, non-small cell lung cancer.

Table 12: Correlation of differential methylation between cfDNA of NSCLC patients and tissue from various primary tumors and normal lung. Filtering stringency refers to the comparison between the combined blood cell 5mC signal (i.e. monocytes, neutrophils and erythroid progenitors) and the median 5mC signal of NSCLC patient plasma samples at individual 300-bp genomic windows. For instance, at a cut-off of 50%, genomic windows with a blood cell 5mC level >50% of the median patient 5mC signals are excluded. Spearman rank correlation coefficients are listed for six cancer entities and non-malignant lung tissue.

Filtering stringency	Spearman correlation coefficient (ρ)							Remaining genomic regions
	LUAD	BRCA	COAD	PRAD	KIRC	STAD	Lung	
No filtering	0.11	0.11	0.09	0.12	0.11	0.08	0.10	9,603,454
90%	0.15	0.16	0.13	0.16	0.14	0.11	0.12	2,584,031
80%	0.16	0.16	0.13	0.17	0.15	0.11	0.13	2,204,550
70%	0.17	0.17	0.14	0.18	0.16	0.11	0.14	1,825,746
60%	0.18	0.19	0.15	0.20	0.17	0.13	0.15	1,484,681
50%	0.21	0.21	0.17	0.21	0.18	0.16	0.16	1,338,865
40%	0.23	0.23	0.20	0.23	0.20	0.18	0.17	937,531
30%	0.24	0.24	0.21	0.24	0.21	0.20	0.18	719,468
20%	0.26	0.26	0.23	0.25	0.22	0.22	0.18	577,701
10%	0.26	0.26	0.23	0.24	0.21	0.23	0.17	468,683

BRCA, breast carcinoma; cfDNA, cell-free DNA; COAD, colon adenocarcinoma; KIRC, kidney renal clear cell carcinoma; LUAD, lung adenocarcinoma; NSCLC, non-small cell lung cancer; PRAD, prostate adenocarcinoma; STAD, stomach adenocarcinoma.

In the next step, I investigated whether the blood cell 5mC filtering can facilitate the identification of cancer-specific 5mC signals from plasma. First, LUAD-specific 5mC alterations were identified from 450k methylation array data by differential analysis between primary tumor tissue of LUAD patients (n = 469) and various non-malignant tissues (i.e. lung, breast, colon, prostate, and kidney [n = 32 samples per tissue type]). I identified 3,276 hyper- and 5,637 hypomethylated regions specific to LUAD and intersected them with the cfMeDIP-seq dataset. Next, summed 5mC signals (cfMeDIP-seq) were computed per sample at the LUAD-specific hypermethylated sites and compared between NSCLC patient samples and healthy controls. No difference in the 5mC signals was observed between the two groups when all 3,276 hypermethylated sites were included (Figure 16A, left; p = 0.058). In contrast, when only the sites remaining after blood cell filtering (n = 207) were considered, significantly higher 5mC signals in NSCLC patients compared to the healthy controls were found (Figure 16A, right; p = 2.3e-05). LUAD-specific hypomethylated sites showed no difference in the 5mC signals between the two groups (data not shown). Furthermore, I assessed whether the blood cell-filtered cfMeDIP-seq data can discriminate NSCLC patient samples from other cancer entities. For this purpose, I selected the top 100 uniquely hyper- and hypomethylated genomic regions of the six most common cancer types (i.e. lung, breast, colon, prostate, kidney, and stomach cancer) from 450k array data (Figure 16B and Figure S6B). The summed cfMeDIP-seq signal was computed at each cancer entity-unique set of regions and compared between patient and control samples. The LUAD-unique hypermethylated regions showed a significantly higher 5mC signal in the patient samples compared to the controls (Figure 16C; p = 3.0e-04). For the other cancer entity-unique sets of regions, only PRAD demonstrated a significantly increased 5mC signal in NSCLC patient samples (p = 0.03). Without blood cell filtering, I also observed a significantly higher sum of 5mC signal in NSCLC patient samples at the LUAD-unique set of genomic regions. However, the difference between patients and controls was less pronounced (Figure S6A; p = 0.003). At hypomethylated sites, 5mC signals were lower in NSCLC patients compared to healthy individuals for all cancer entity-unique sets of regions when the entire cfMeDIP-seq data was considered. After blood cell filtering, there was no difference between patients and controls (Figure S6C), reflecting the lower accuracy of cfMeDIP-seq for the detection of hypomethylated regions.
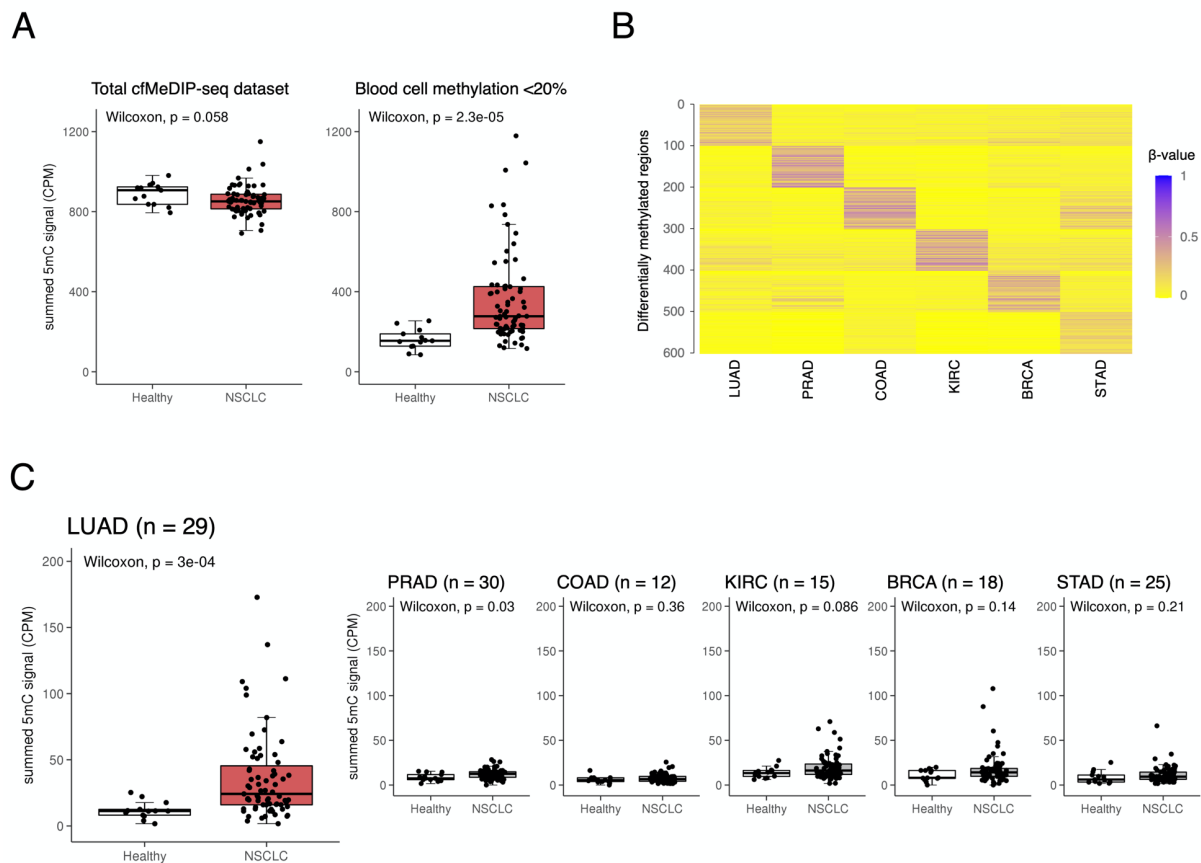
Figure 16: 5mC signal comparison between NSCLC patients and healthy controls at cancer-specific loci. Boxplots illustrate the summed cfMeDIP-seq signal per sample comparing healthy controls (n = 13) to NSCLC patient plasma (n = 71). (A) cfMeDIP-seq signal comparison at genomic windows specifically hypermethylated in LUAD tissue compared to normal tissue types before (left; n = 3,276 genomic regions) and after (right; n = 207 genomic regions) exclusion of genomic regions hypermethylated in blood cells (i.e. monocytes, neutrophils and erythroid progenitors). (B) Heatmap of the top 100 uniquely hypermethylated genomic regions of the six most common cancer types determined from TCGA methylation data. (C) cfMeDIP-seq signal comparison at the entity-unique hypermethylated regions illustrated in (B), following the exclusion of regions hypermethylated in blood cells. The number of genomic regions remaining after blood cell signal filtering is indicated above the graphs for each cancer entity. BRCA, breast carcinoma; cfMeDIP-seq, cell-free methylation DNA immunoprecipitation sequencing; COAD, colon adenocarcinoma; CPM, counts per million; KIRC, kidney renal clear cell carcinoma; LUAD, lung adenocarcinoma; NSCLC, non-small cell lung cancer; PRAD, prostate adenocarcinoma; STAD, stomach adenocarcinoma.

Finally, I assessed if the blood cell-filtered cfMeDIP-seq data was able to differentiate NSCLC patient samples from healthy individuals. To this end, patient samples were stratified based on chromosomal instability (t-MAD) scores determined from the same plasma samples by sWGS. The t-MAD score has been demonstrated to positively correlate with the tumor fraction in cfDNA (Dietz et al. 2020; Mouliere et al. 2018a). Therefore, high t-MAD score samples are potentially easier to differentiate from healthy controls. For this study, samples exceeding the third quartile of the cohort t-MAD score were defined as t-MAD$_{high}$ (0.0174), while samples below the first quartile were graded t-MAD$_{low}$ (0.0074). PCA using the 0.5% most variable genomic windows with low blood cell methylation (n = 2,889) could partition t-MAD$_{high}$ samples and most samples with an intermediate t-MAD score (0.0074 < t-MAD$_{mid}$ < 0.0174) from the healthy controls. t-MAD$_{low}$ samples clustered closest to the controls (Figure 17A). When the PCA clustering was repeated using the top 0.5% variable genomic regions without blood cell filtering (n = 35,016), only t-MAD$_{high}$ samples could be

discriminated from healthy individuals (Figure 17B). Combined, these results suggested that the exclusion of genomic regions hypermethylated in blood cells facilitates the identification of tumor-derived 5mC alterations.
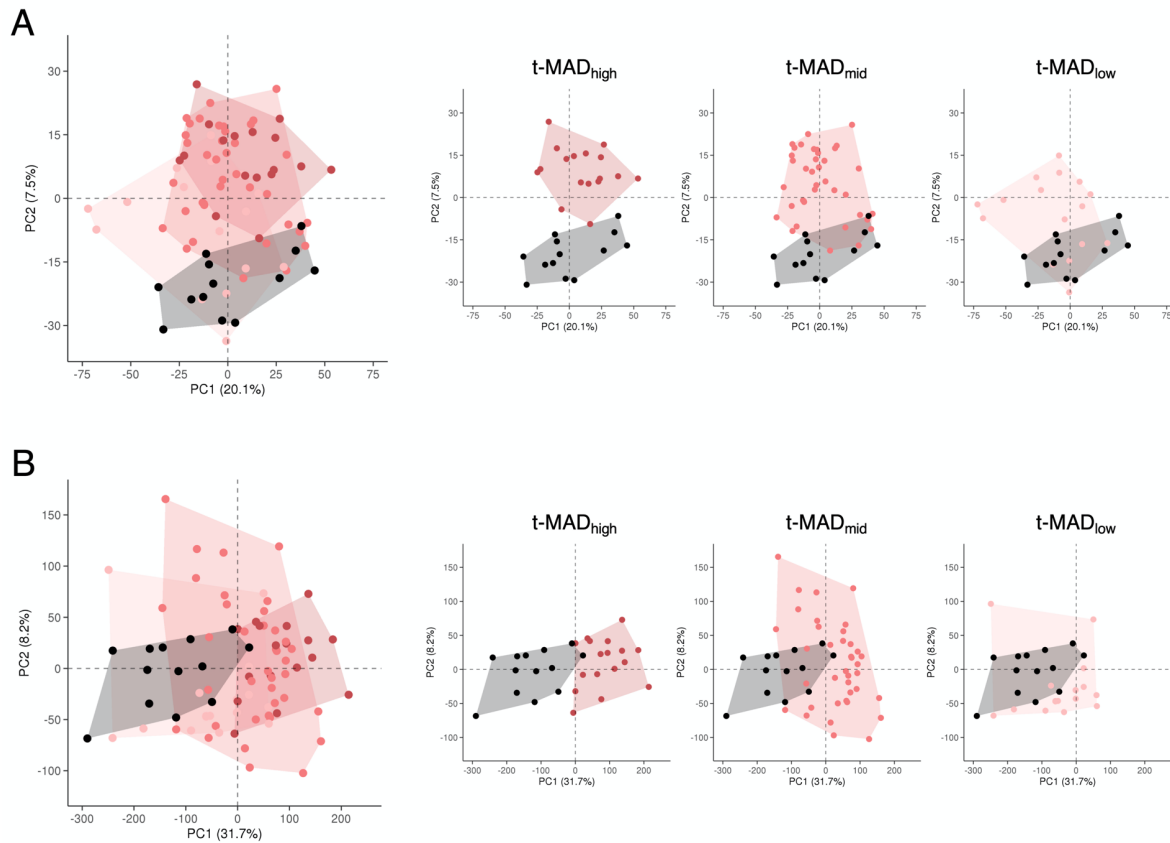
Figure 17: Principle component analysis of 5mC signals measured from NSCLC patient samples and healthy controls with and without blood cell filtering. Principle component analyses of the top 0.5% genomic regions with the most variable 5mC signals. (A) Plasma sample clustering after the exclusion of regions hypermethylated in blood cells (i.e. monocytes, neutrophils and erythroid progenitors) and (B) without blood cell filtering. Each dot represents one plasma sample from 13 healthy controls (black) and 71 NSCLC patients (red). The different shades of red indicate patient samples with high (>0.0174; dark red), intermediate (0.0074 < t-MAD < 0.0174; red) and low (<0.0074; light red) t-MAD scores. The comparison between control samples and individual patient groups stratified by their t-MAD score are illustrated separately for better visualization. t-MAD, trimmed median absolute deviation from copy number neutrality.

3.6.2 Identification of differentially methylated regions between NSCLC patients and healthy controls

Following the exclusion of genomic regions hypermethylated in non-tumor cfDNA, differential analysis was performed between NSCLC patient samples and healthy individuals to identify tumor-informative 5mC signals from the genome-wide cfMeDIP-seq data. For patients with longitudinal follow-up samples available, only the sample with the highest t-MAD score was included (n = 26). These samples contain the highest ctDNA fraction and are therefore easier to distinguish from healthy controls. The differential analysis was performed on 300-bp genomic windows remaining after blood cell 5mC signal filtering (n = 577,701), comparing 26 NSCLC patient to 13 control samples. In total, 5,297 hyper- and 202 hypomethylated DMRs were identified (Figure 18A). The majority of DMRs were associated with intergenic regions and introns (Figure 18B). However, the highest enrichment of DMRs was found within the first exon, at promoter regions, 5´-UTRs, exons and TESs. Introns, intergenic regions and 3´-UTRs were only slightly enriched or underrepresented (Figure 18C).
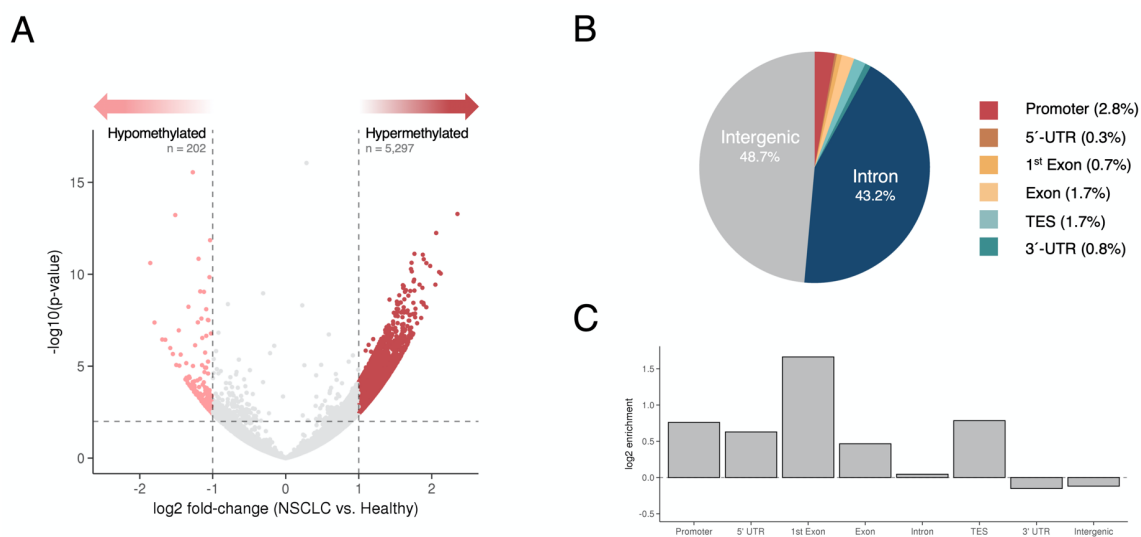
Figure 18: Differential analysis results comparing 5mC profiles of NSCLC patients to healthy individuals. (A) Volcano plot of differentially methylated regions (DMRs) determined from plasma samples of NSCLC patients (n = 26) *versus* healthy individuals (n = 13). Regions with p < 0.01 and |log$_2$(fold-change)| > 1 are indicated in red. (B) Relative genomic distribution of DMRs. (C) DMR abundance at genomic elements relative to that expected by random distribution throughout the genome. Expressed as log$_2$ enrichment with positive values indicating enriched more than expected. NSCLC, non-small cell lung cancer; TES, transcription end site; UTR, untranslated region.

Hierarchical clustering using the entire 5,499 DMRs clearly separated all NSCLC samples from those of healthy individuals (Figure 19). Furthermore, it could be observed that the longitudinally taken samples clustered according to their t-MAD scores rather than with samples from the same patient. For example, 14 serial plasma samples taken from patient P028 are separated into three clusters of high, intermediate and low t-MAD scores. Similar observations were made for patients P012, P025 or P044.
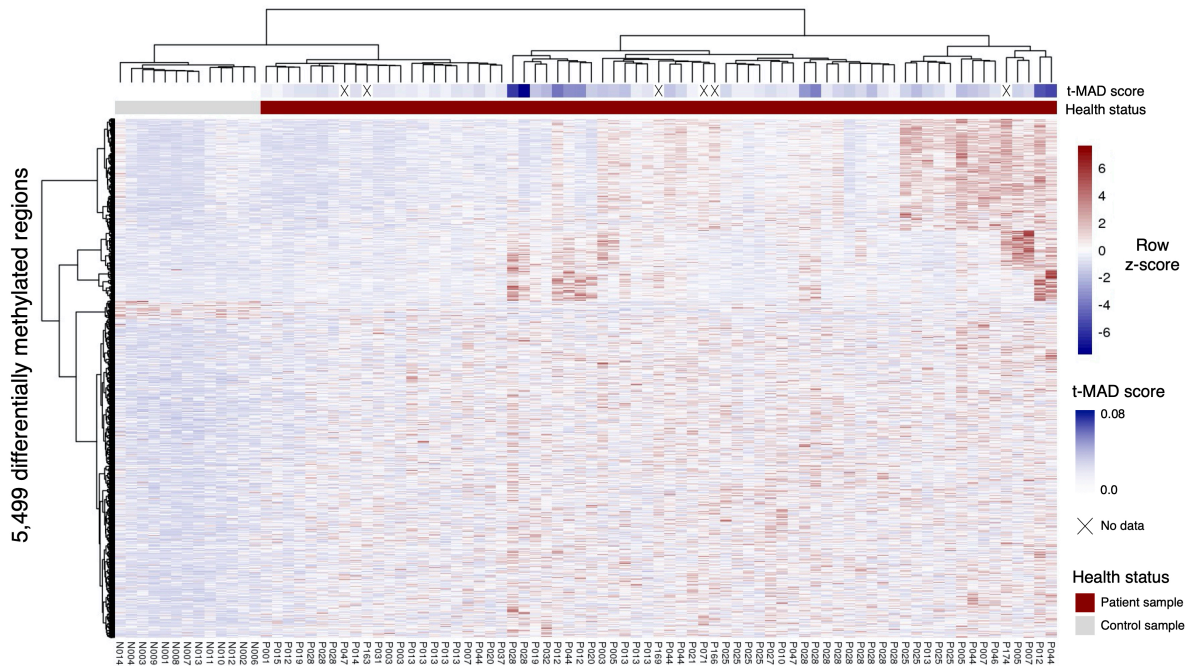
Figure 19: Heatmap of differentially methylated regions comparing NSCLC patients to healthy controls. The heatmap illustrates clustering of 71 NSCLC patient samples and 13 healthy controls (Euclidean distance and Ward clustering) based on 5,499 differentially methylated regions (DMRs) identified between cases and controls following exclusion of genomic regions hypermethylated in blood cells (i.e. monocytes, neutrophils and erythroid progenitors). Patient identifiers are provided for each column with "P" representing NSCLC patients and "N" controls. NSCLC, non-small cell lung cancer; t-MAD, trimmed median absolute deviation from copy number neutrality.

Earlier, I demonstrated that the exclusion of genomic regions hypermethylated in blood cells could identify cancer-associated 5mC alterations in plasma of NSCLC patients. Here, I investigated whether differential analysis could further enhance the detectability of tumor-derived 5mC signals. For this purpose, DMRs were intersected with the same genomic regions previously determined to be hypermethylated specifically in LUAD tissue (LUAD *vs.* non-cancerous tissue types; LUAD-specific) and uniquely hypermethylated in LUAD compared to other cancer entities (entity-unique). In total 25 LUAD-specific and three entity-unique genomic regions overlapped with the DMRs. Notably, none of the genomic regions uniquely hypermethylated in other cancer entities (i.e. breast, colon, prostate, kidney, and stomach cancer) could be intersected. Significantly higher summed 5mC signals were observed in NSCLC patient samples compared to healthy controls at both the LUAD-specific (Figure 20A; p = 1.3e-06) and entity-unique DMRs (Figure 20B; p = 9.1e-05). The difference in the summed 5mC signal between the two groups after selection of DMRs was higher when compared to the exclusion of blood cell hypermethylated regions alone (Figure 16A; p = 2.3e-05 and Figure 16C; p = 3.0e-04).

A

B
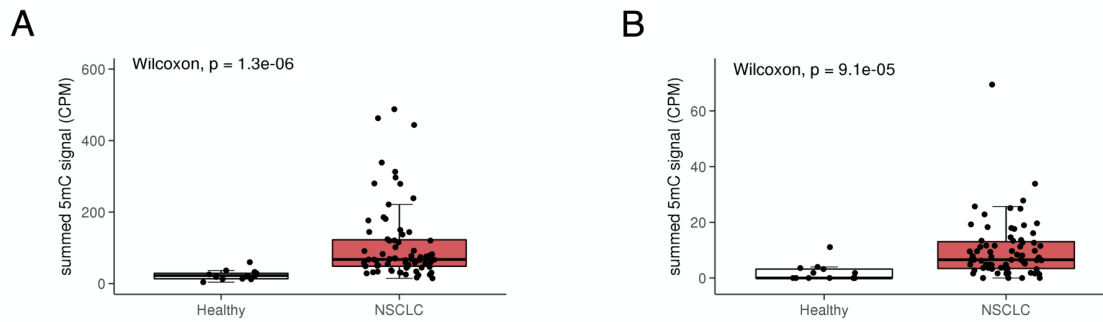
Figure 20: 5mC signals at LUAD-specific genomic regions. Boxplots show the summed 5mC signal – as normalized counts per million (CPM) – for 13 healthy controls and 71 NSCLC patients. (A) 5mC signal of DMRs at genomic regions hypermethylated specifically in LUAD tissue *versus* various non-malignant tissue types (n = 25) and (B) at regions uniquely hypermethylated in LUAD tissue compared to other cancer entities (n = 3). NSCLC, non-small cell lung cancer.

3.6.3 TSS-proximal hypermethylation is associated with low expression in lung adenocarcinoma tissue

Repression of tumor suppressor genes by promoter hypermethylation is a common event in cancer (Greenberg and Bourc'his 2019; Hansen et al. 2011). The previously identified DMRs were enriched at genomic elements proximal to TSSs (i.e. promoter regions, 5′-UTRs and first exons). Therefore, I investigated whether the 5mC signal at these elements would allow inferring the transcriptional status of associated genes from plasma of NSCLC patients. TCGA RNA sequencing data (Grossman et al. 2016) was used to determine LUAD-specific expression patterns comparing primary tumor tissue of LUAD patients (n = 533) to various non-cancerous tissues (i.e. lung, breast, colon, prostate, kidney, and stomach [n = 29 samples per tissue type]). This resulted in 7,133 differentially expressed genes (3,457 up- and 3,676 downregulated). To identify gene regulatory DMRs, LUAD-specific differentially expressed genes were overlapped with DMRs located within promoters, 5′-UTRs and first exons (Figure 21A). Fifty-four hypermethylated DMRs at 45 genes (Table S3) coincided with downregulated expression in LUAD tissue, while only 18 hypermethylated sites were located at genes with elevated expression. No hypomethylated DMRs overlapped with differentially expressed genes. Permutation testing was performed to assess the significance of overlap between hypermethylated DMRs and differentially regulated genes. This revealed significant enrichment of DMRs coinciding with downregulated (p = 1.0e-04) but not with upregulated expression (p = 0.284) at promoters, 5′-UTRs and first exons (Figure S7A). Other elements (i.e. exons, TESs, 3′-UTRs, introns, and intergenic regions) showed no enrichment of DMRs at genes down- or upregulated in LUAD tissue (Figure S7B).

The following investigations focused on the 54 DMRs associated with downregulated gene expression in LUAD tissue. These genes are likely to be epigenetically regulated by their TSS-proximal 5mC levels. The NSCLC-specific cfDNA hypermethylation of 35 DMRs was confirmed in LUAD tissue, using TCGA methylation data (LUAD *vs.* normal tissue). Six DMRs were hypo- or not differentially methylated in LUAD tissue and 13 DMRs were not covered by the 450k array (Table S3). The majority of the genes associated to downregulated

expression coded for transcription factors (n = 12) or long non-coding RNAs (lncRNA; n = 9). In total, 30 out of 45 genes were previously described in the context of cancer; with 24 genes associated with tumor suppressive functions (Figure 21B). Some of the most significantly hypermethylated DMRs included members of the *GATA* transcription factor family (i.e. *GATA3* and *GATA4*) as well as homeobox genes such as *HOXA9* (Figure 21C). This was in line with previous reports, showing promoter hypermethylation and consequential expression downregulation of these genes in NSCLC (Faryna et al. 2012; Gao et al. 2019; Zhao et al. 2019). *SOX9-AS1* was the most significantly hypermethylated lncRNA gene. It has been described as a regulator of *SOX9*, which promotes proliferation in NSCLC cell lines via activation of Wnt/β-catenin signaling. Epigenetic silencing of *SOX9-AS1* therefore favors NSCLC tumorigenesis (Barter et al. 2017; Guo et al. 2018). Interestingly, three of the identified genes (*HOXA9*, *GATA4* and *CDO1*) were recently included in an 8-gene-panel whose promoter hypermethylation measured from cfDNA was used for the diagnosis of stage I NSCLC (nodule size ≤ 3.0 cm) (Chen et al. 2020a). Table S3 summarizes literature search on the cancer association of all 45 genes.
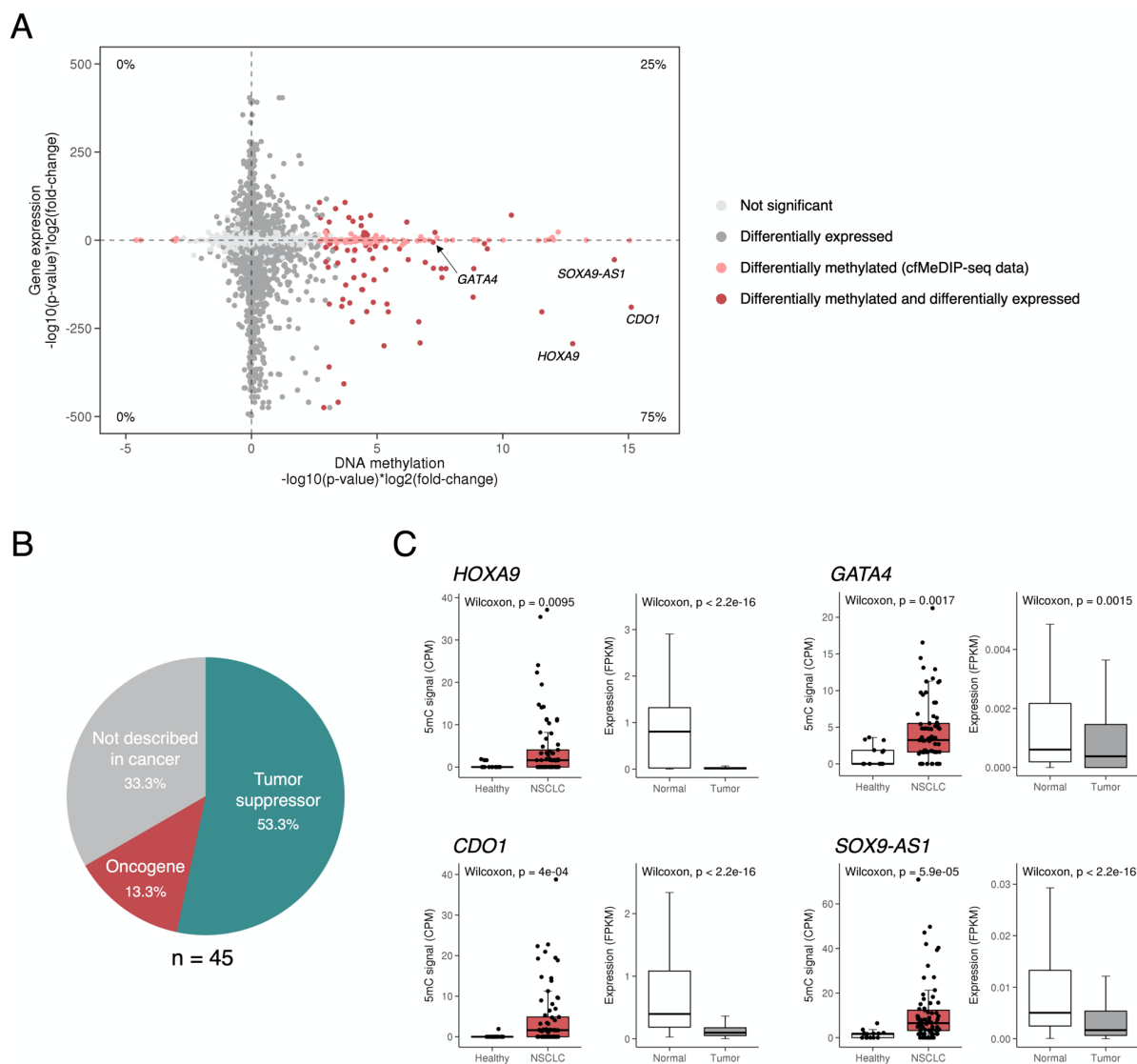
Figure 21: Relationship between TSS-proximal cfDNA methylation and gene expression in LUAD tissue. (A) Starburst plot comparing cfDNA methylation (NSCLC patients *vs.* healthy controls) and gene expression (LUAD tissue *vs.* non-cancerous lung, breast, colon, prostate, kidney, and stomach tissue). Each dot represents a TSS-associated genomic region and colors indicate significance of the differential expression/methylation analysis. Differentially methylated regions (DMRs) and genes with p < 0.01 and -1 < log₂(fold-change) < 1 were considered significant. Percentages refer to the relative number of both differentially expressed genes and differentially methylated regions within each quadrant of the plot. (B) Overview of the cancer association of the 45 genes with TSS-proximal hypermethylation in plasma of NSCLC patients and downregulated expression in LUAD tissue. (C) CfDNA 5mC signal and tissue expression of representative genes with TSS-associated DMRs. Plasma 5mC levels (left) are illustrated for healthy controls (n = 13) and NSCLC patient samples (n = 71) with each dot representing one sample. Gene expression boxplots (right) compare LUAD (n = 533 samples) to normal tissue (n = 174; same tissue types as in (A)). CfMeDIP-seq, cell-free methylation DNA immunoprecipitation sequencing; CPM, counts per million; FPKM, fragments per kilobase per million reads; LUAD, lung adenocarcinoma; NSCLC, non-small cell lung cancer; TSS, transcription start site.

3.6.4 Selection of diagnostic and prognostic 5mC biomarkers

To identify diagnostic and prognostic 5mC biomarkers from plasma, samples were divided into three groups: (1) NSCLC patient plasma taken at time points of radiologically progressive disease (PD; n = 28), (2) patient plasma sampled at non-PD time points (i.e. stable disease [SD], partial response [PR], and initial diagnosis; n = 37), and (3) plasma from healthy individuals (n = 13). A sensitive diagnostic 5mC marker should discriminate PD but also non-PD samples from healthy controls, while a prognostic marker should be able to differentiate PD from non-PD plasma samples. To reduce the number of potential biomarkers, I focused on hypermethylated DMRs associated with LUAD biology as determined from TCGA methylation and expression data. These included (i) TSS-proximal DMRs associated with downregulated gene expression in LUAD tissue, (ii) DMRs hypermethylated in LUAD but not in non-malignant tissue types, and (iii) DMRs uniquely hypermethylated in LUAD tissue compared to other cancer entities. LUAD-specific methylation and expression datasets were determined as described in sections 3.6.1 and 3.6.3 and overlapped with 71 DMRs at 57 genes (Figure 22A). The majority of genes were either associated with transcription factors (TFs; n = 16) or lncRNAs (n = 11). The 57 potential marker genes were not enriched at lung cancer associated gene sets obtained from MSigDB (data not shown). For biomarker identification, 5mC signals of DMRs mapping to the same gene were combined by summation.

The 5mC levels at 54 genes were significantly increased in PD samples, and 44 genes could additionally discriminate non-PD samples from healthy controls. Furthermore, 31 potential prognostic marker genes with significantly higher 5mC signals were found in PD compared to non-PD samples. A total of 21 5mC biomarkers differentiated between non-PD and healthy samples as well as PD and non-PD samples (Figure S8A). Among these, *HOXA10-AS*, *SOX9-AS1*, *PRAC1* and *PTGER4* presented the most significant differences between the sample groups (Figure 22B). TCGA LUAD data demonstrated transcriptional downregulation of *HOXA10-AS*, *SOX9-AS1* and *PRAC1* in tissue coinciding with TSS-proximal hypermethylation determined from plasma. LUAD-specific hypermethylation compared to non-malignant tissues was found at *HOXA10-AS*, *PRAC1* and *PTGER4*. Recent literature described *SOX9-AS1* to have tumor suppressive properties in NSCLC (Barter et al. 2017; Guo et al. 2018), while *PRAC1* was defined as tumor suppressor in renal, colon and prostate cancer (Hu et al. 2018). Notably, transcriptional repression of *PRAC1* by promoter hypermethylation in colon cancer was associated with co-regulation of *HOXB13* (Hu et al. 2018), a tumor suppressor that was also hypermethylated in the 5mC dataset of this study. While the role of *PTGER4* in NSCLC tumors remains elusive, its promoter hypermethylation – in combination with *SHOX2* hypermethylation – has been used as a plasma-based biomarker for the diagnosis of NSCLC (Weiss et al. 2017). *SHOX2* was also identified to be hypermethylated in plasma of NSCLC patients but was not associated with methylation or expression in LUAD tissue. Another study demonstrated that reduced *PTGER4* plasma methylation indicated therapy success in stage IV NSCLC patients (Schotten et al. 2021). The

lncRNA *HOXA10-AS* was found to be co-expressed with *HOXA10* and regulates its transcription initiation by recruitment of the SNF2L chromatin remodeling complex. Both *HOXA10* and *HOXA10-AS* were described to be highly expressed in various cancer entities – including NSCLC – and associated with oncogenic functions (i.e. increased cell proliferation, migration, invasiveness and inhibition of apoptosis) (Shao et al. 2018; Sheng et al. 2018). The consequences of *HOXA10-AS* hypermethylation in cancer have not yet been elucidated. Further potential diagnostic and prognostic 5mC biomarkers are illustrated in Figure S8, and literature search on their association to cancer is summarized in Table S3.
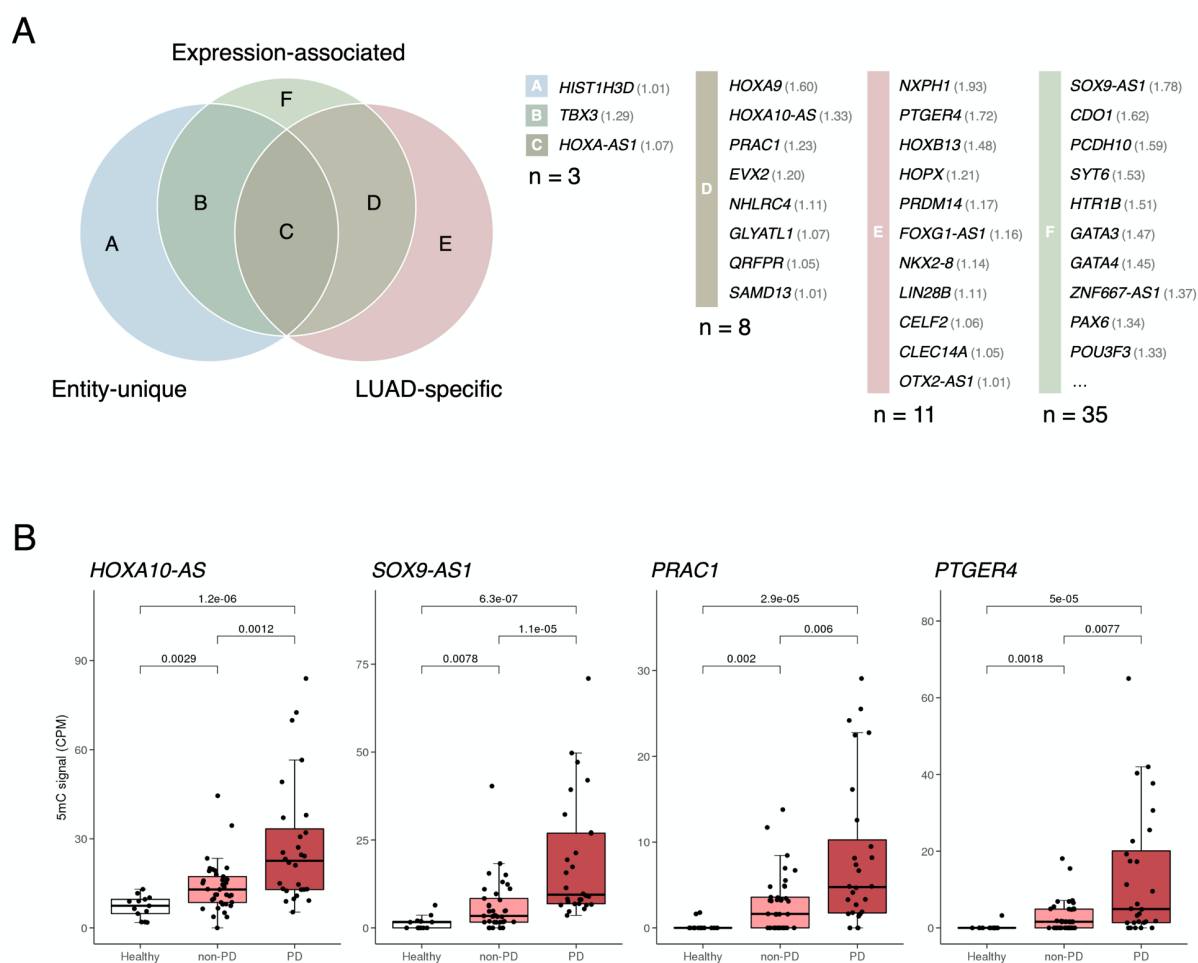
Figure 22: Representative diagnostic and prognostic 5mC biomarkers. (A) Venn diagram illustrating the overlap between genomic regions hypermethylated in NSCLC patient plasma compared to healthy controls at genes associated to downregulated expression in LUAD tissue (expression-associated), specifically hypermethylated in LUAD tissue compared to non-malignant tissue types (LUAD-specific) and uniquely hypermethylated in LUAD compared to other cancer entities (entity-unique). Genes are listed alongside their $\log_2$ fold-changes in brackets. (B) Representation of the top four most significant 5mC biomarker genes capable of differentiating healthy control samples (n = 13) from non-progressive (non-PD; n = 37) samples and non-PD from progressive (PD; n = 28) samples. LUAD, lung adenocarcinoma; NSCLC, non-small cell lung cancer.

3.6.5 5mC biomarkers correlate with tumor-specific genomic markers measured from the same plasma samples

Genomic alterations detected in cfDNA of cancer patients are often tumor-specific (van der Pol and Mouliere 2019). To confirm the tumor association of the identified 5mC biomarkers, 5mC signals were correlated to genomic alterations determined from the same plasma samples. As part of another project, CAPP-seq and sWGS was performed to capture the genomic landscape (i.e. SNVs and CNAs) in plasma of NSCLC patients (Dietz et al. 2020). The quantified genomic biomarkers included the *EML4-ALK* fusion gene as well as *ALK* resistance mutations, *TP53* co-mutations and chromosomal instability (t-MAD scores). Figure 23 depicts the correlation of four 5mC biomarkers – previously identified to be diagnostic and prognostic in the presented cohort – to the abundance of the *EML4-ALK* fusion gene, the maximal variant allele frequencies (VAF_{max}), and the t-MAD score. VAF_{max} represents the sum of all quantified mutations in the gene panel used. This metric provides an estimation of the total genomic tumor burden within a plasma sample and accounts for tumor heterogeneity and polyclonality (Velimirovic et al. 2020). Positive correlations were observed between all 5mC biomarkers and the genomic tumor-derived alterations. 5mC signals were most similar to the t-MAD score, with *PTGER4* demonstrating the overall highest Pearson correlation coefficient ($r = 0.86$). Other 5mC marker genes with high correlations to genomic alterations in plasma include *HTR1B*, *CELF2*, *CLEC14A* and *HOXA9* (Table S4). Table S4 additionally provides the Pearson correlation coefficients and corresponding p-values comparing 5mC biomarkers to the abundance of *TP53* and *ALK* resistance mutations (i.e. L1196M and G1202R) in cfDNA.
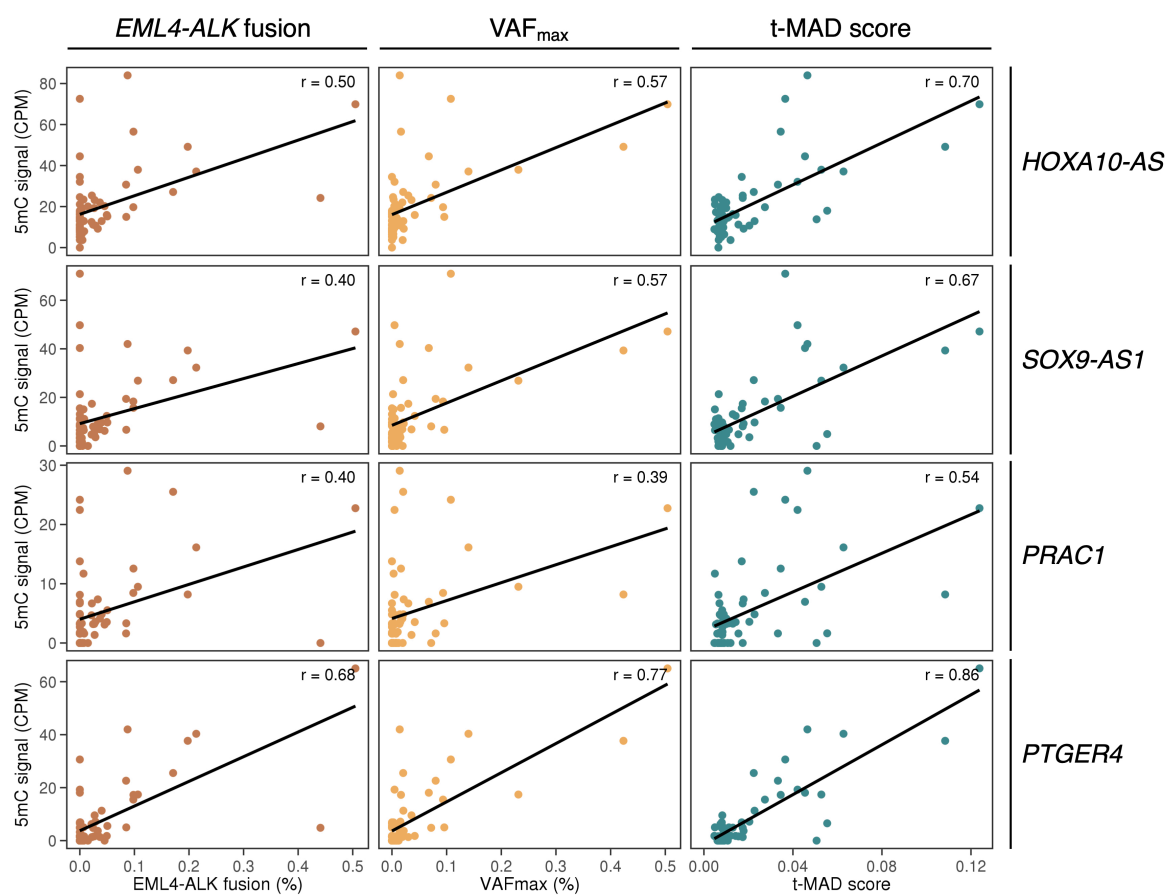
Figure 23: 5mC biomarker correlation to genomic markers measured from the same plasma samples. Correlation of the 5mC signal (cfMeDIP-seq) at four marker genes (*HOXA10-AS*, *SOX9-AS1*, *PRAC1*, and *PTGER4*) and genomic tumor biomarkers (*EML4-ALK* fusion gene, VAF_{max}, and t-MAD score) measured from the same plasma samples. VAF_{max} represents the summed up variant allele frequency of all mutations detected per plasma sample. Linear regression lines are displayed in black for each marker combination and Pearson correlation coefficients are illustrated. ALK, anaplastic lymphoma kinase; CPM, counts per million; EML4, echinoderm microtubule-associated protein-like 4; t-MAD, trimmed median absolute deviation from copy number neutrality; VAF_{max}, maximal variant allele frequency.

3.6.6 Plasma 5mC biomarkers reflect therapy response and disease progression in longitudinal samples

The presented cohort included eleven patients with longitudinally taken plasma samples available, ranging from two to 14 samples per patient (Figure 8). All plasma samples were analyzed retrospectively and did not influence therapy decision. The 5mC signals at selected marker genes were compared to the patients´ clinical courses and the co-measured genomic tumor biomarkers. In most cases, 5mC signals reflected the clinical status of the respective patient, rising at TKI failure and time points of radiologically progressive disease and decreasing at successful therapeutic interventions (Figure 24A-E and Figure S9-13). The cohort comprised nine instances of disease progression with available plasma samples from a previous SD time point. Rising 5mC levels from the SD to the PD time point indicated progression in 7/9 (*HOXA10-AS*), 5/9 (*PRAC1*), 8/9 (*SOX9-AS1*), and 9/9 (*PTGER4*) cases. Eight patients deceased shortly after the last sampling time point. Increased 5mC levels were

observed in the plasma samples taken before the patient´s death in the majority of cases (6/8, *HOXA10-AS*; 8/8, *SOX9-AS1*; 6/8 *PRAC1*; 7/8, *PTGER4*). Similar results were obtained when the kinetics of genomic tumor biomarkers were inspected. For patient P028 (Figure 24A), 14 serial plasma samples were available for 5mC analysis. This allowed narrow therapy surveillance over the course of multiple lines of TKI and chemotherapy. Steadily increasing *SOX9-AS1* 5mC levels correctly indicated both alectinib failure and ineffective ceritinib treatment, which was accompanied by the emergence of the *ALK* G1202R resistance mutation. 5mC levels dropped following lorlatinib administration (active against *ALK* G1202R), indicating initial response of the patient. The patient´s response was additionally reflected by the clearance of *ALK* G1202R and decreasing abundances of all genomic tumor markers. Lorlatinib failure coincided with an elevated 5mC signal and temporary disease control due to chemotherapy was reflected by low 5mC levels. At disease progression under chemotherapy, treatment was switched to brigatinib which stabilized the patient´s disease and resulted in a drop of both 5mC and genomic markers. Shortly after, steep increase of the 5mC signal together with rising genomic tumor biomarkers reflected rapid disease progression, followed by the patient´s death. The course of patient P028 suggested that 5mC signals could indicate molecular progression earlier than radiological disease evaluation. Clinical progression under alectinib treatment was first identified at day 161. However, *SOX9-AS1* 5mC levels already indicated progression by a 2.5-fold increase of the 5mC signal at day 49 (112 days earlier). Afterwards, the 5mC signal persistently rose until successful therapeutic intervention. 5mC levels of *HOXA10-AS* and *PTGER4* indicated progression at the same time point. Notably, genomic biomarkers also predicted progression prior to radiological disease assessment (emergence of the *EML4-ALK* fusion gene, *ALK* G1202R mutation and rising t-MAD score), albeit at a later time point (day 106). Other instances of early identification of disease progression were observed in patient P044 (*PTGER4*; day 481; Figure S11), patient P025 (*HOXA10-AS*; day 112) and patient P012 (*HOXA10-AS*; day 273; Figure S9), preceding radiology by 27, 47 and 119 days, respectively. Here, genomic biomarkers predicted progression at the same time points or earlier.
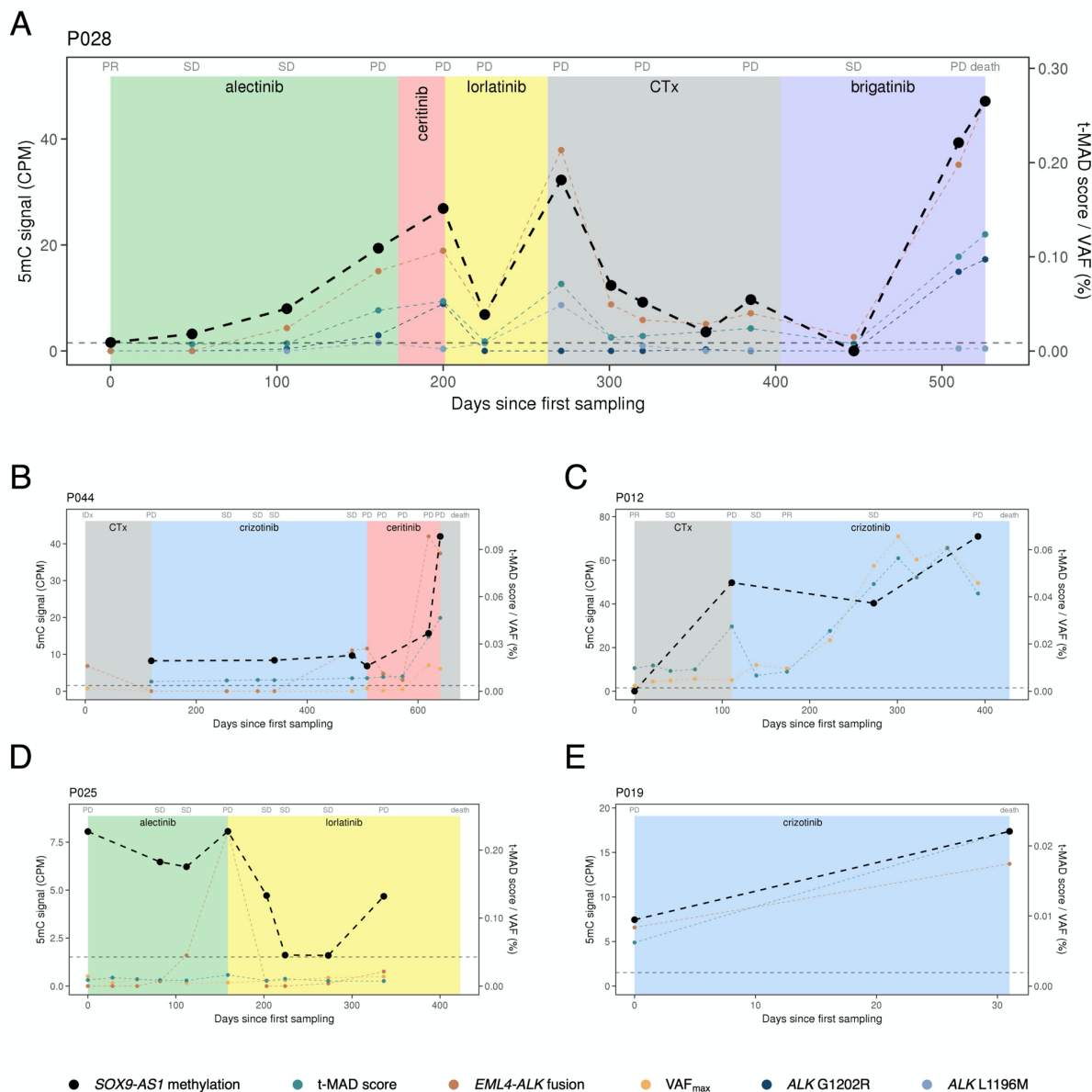
Figure 24: Representative profiles of patients illustrating the utility of *SOX9-AS1* hypermethylation in cfDNA as biomarker for therapy monitoring of ALK⁺ NSCLC patients. *SOX9-AS1* methylation (cfMeDIP-seq), molecular alterations (CAPP-seq) and t-MAD score (sWGS) dynamics in plasma cfDNA during sequential ALK-directed tyrosine kinase inhibitor treatment and chemotherapy of patients (A) P028, (B) P044, (C) P012, (D) P025, and (E) P019. The dotted horizontal line represents the median *SOX9-AS1* 5mC signal of 13 healthy individuals. Radiological disease assessments are indicated above each plot and administered therapy lines are shown by the background colors. Each dot represents one plasma sample. ALK, anaplastic lymphoma kinase; CAPP-seq, Cancer Personalized Profiling by deep sequencing; cfMeDIP-seq, cell-free methylation DNA immunoprecipitation; CPM, counts per million; CTx, chemotherapy; EML4, echinoderm microtubule-associated protein-like 4; IDx, initial diagnosis; PD, progressive disease; PR, partial response; SD, stable disease; t-MAD, trimmed median absolute deviation from copy number neutrality; VAF_max, maximal variant allele frequency.

3.7 Identification of NSCLC-specific 5hmC biomarkers by hMeSEAL-seq

To identify NSCLC-specific 5hmC signatures from cfDNA, I adapted the bioinformatic workflow previously described for the identification of tumor-informative 5mC alterations (Figure 14). This comprised the exclusion of genomic regions with high 5hmC levels in blood cells (i.e. monocytes, neutrophils and erythroid progenitors), followed by differential analysis between NSCLC patient and healthy control samples. As reference 5hmC profiles were not available for this study, blood cell filtering was performed based on the threshold established on the basis of the 5mC data. Moreover, D(h)MRs were only associated with LUAD tissue expression data.

Genome-wide 5hmC profiles were generated from primary monocytes, neutrophils and erythroid progenitor cells to identify genomic regions hyper-hydroxymethylated in non-tumor plasma cfDNA. On average, 55.7 million paired reads were obtained and CpG enrichment scores (mean = 1.73) confirmed successful 5hmC precipitation (further sequencing quality metrics are provided in Table S2). The blood cells′ 5hmC signals were integrated at 300-bp genomic windows as described in section 3.6.1 and regions with combined blood cell 5hmC signals <20% of the median signals in NSCLC plasma were retained for further analysis. This resulted in 499,681 out of 9,603,454 genomic regions remaining. PCA clustering analysis on the 0.5% genomic regions with the most variable 5hmC signals was performed with (Figure 25A; n = 2,498) and without (Figure 25B; n = 37,121) exclusion of blood cell signals. Hereby, it was evaluated whether the blood cell filtering facilitates the differentiation between patient and healthy control samples. Patients were stratified by their t-MAD scores to test if samples with high tumor-fractions (t-MAD$_{high}$) are easier to distinguish from controls compared to samples with lower ctDNA content (t-MAD$_{mid}$ and t-MAD$_{low}$). Patient samples without available t-MAD scores were excluded from this analysis, resulting in 68 patient and 14 control samples remaining. PCA clusters without blood cell filtering revealed no clear separation between NSCLC and control samples irrespective of the sample t-MAD score. However, an improved separation was observed when compared to the PCA results following the exclusion of hyper-hydroxymethylated regions in blood cells (Figure 25A and B). Notably, separation between patient and control samples did not improve in t-MAD$_{high}$ samples as observed in the 5mC-enriched sequencing data (Figure 17A and B). The improved PCA clustering after blood cell filtering suggested a higher tumor association of the remaining genomic regions compared to the consideration of all regions. Therefore, further analysis steps were carried out on blood cell-filtered 5hmC profiles.
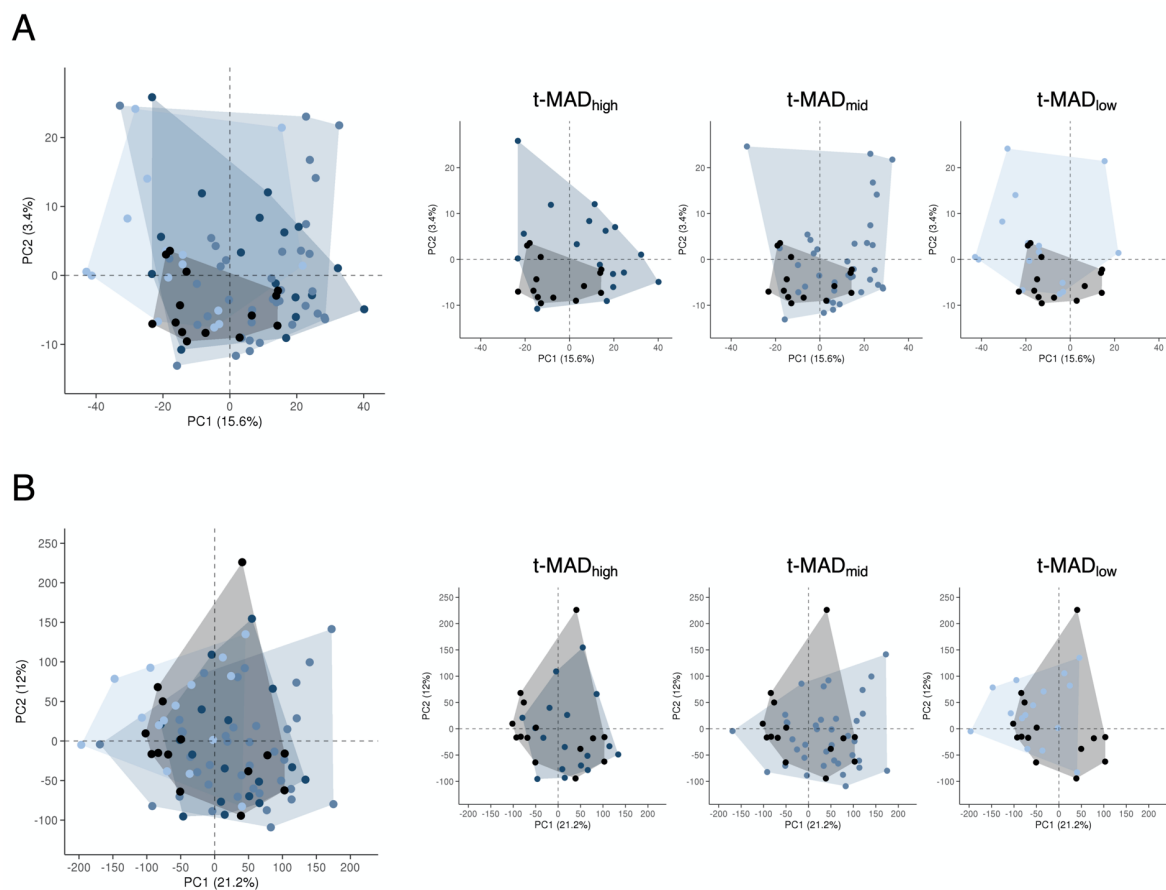
Figure 25: Principle component analysis of 5hmC signals measured from NSCLC patient samples and healthy controls with and without blood cell filtering. Principle component analyses of the top 0.5% genomic regions with the most variable 5hmC signals with (A) and without (B) the exclusion of genomic regions hyper-hydroxymethylated in blood cells (i.e. monocytes, neutrophils and erythroid progenitors). Each dot represents one plasma sample from 14 healthy controls (black) and 68 NSCLC patients (blue). The different shades of blue indicate patient samples with high (>0.0174; dark blue), intermediate (0.0074 < t-MAD < 0.0174; blue) and low (<0.0074; light blue) t-MAD scores. The comparison between control samples and individual patient groups stratified by their t-MAD score are illustrated separately for better visualization. t-MAD, trimmed median absolute deviation from copy number neutrality.

3.7.1 Identification of differentially hydroxymethylated regions between NSCLC patients and healthy controls

For the identification of tumor-informative 5hmC alterations, I performed differential analysis between NSCLC and healthy control samples. Analogous to the identification of DMRs (section 3.6.2), only the sample with the highest t-MAD score per patient was included, resulting in 28 patient and 14 control samples. Differential analysis was carried out on 499,681 300-bp genomic windows. A total of 424 hyper- and 71 hypo-hydroxymethylated DhMRs were identified (Figure 26A). Most DhMRs were found within or proximal to genes, with the majority mapping to introns. Sixteen TSS-proximal DhMRs (±1,500 bp) were identified and 21 DhMRs were associated with exons, TESs or 3′-UTRs (Figure 26B).
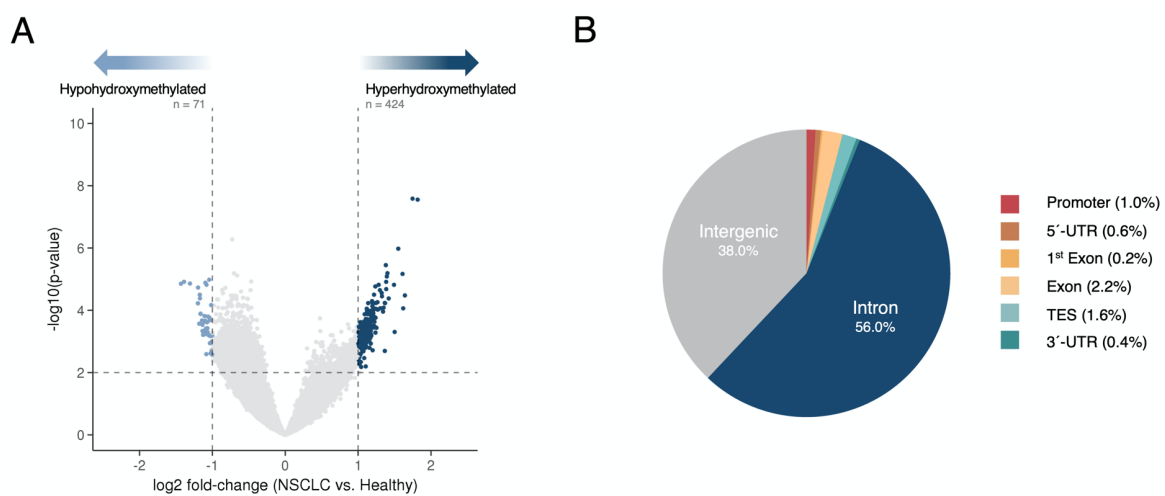
Figure 26: Differential analysis results comparing 5hmC profiles of NSCLC patients to healthy controls. (A) Volcano plot showing differentially hydroxymethylated regions (DhMRs) determined from plasma of NSCLC patients (n = 28) and healthy individuals (n = 14). Regions with p < 0.01 and |log$_2$(fold-change)| > 1 are indicated in blue. (B) Relative genomic distribution of DhMRs. NSCLC, non-small cell lung cancer; TES, transcription end site; UTR, untranslated region.

A separation between all NSCLC (n = 74) and healthy control samples (n = 14) was observed after hierarchical clustering using the 495 DhMRs (Figure 27). Similar to the clustering based on DMRs (Figure 19), samples with high t-MAD scores demonstrated close association with one another. This could be observed at several instances where longitudinally taken samples cluster closer to those with similar t-MAD scores than samples taken from the same patient. For example, the t-MAD$_{high}$ plasma samples of patients P012, P028 and P044 demonstrated similar 5hmC profiles at the 495 DhMRs. These samples were taken shortly before the respective patient′s death, highlighting their progressive disease status and the elevated ctDNA fraction. This suggests that the identified DhMRs capture tumor-derived 5hmC alterations.
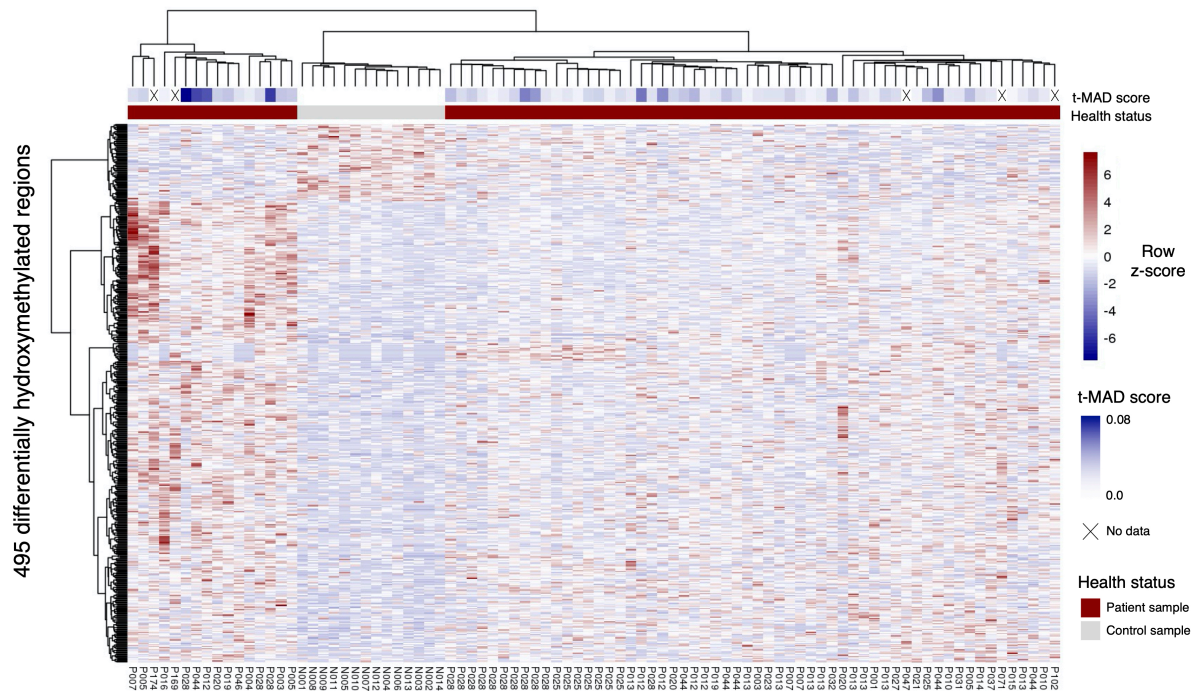
Figure 27: Heatmap of differentially hydroxymethylated regions comparing NSCLC patients to healthy controls. Clustering of 74 NSCLC patient and 14 healthy control samples (Euclidean distance and Ward clustering) based on 495 differentially hydroxymethylated regions (DhMRs). DhMRs were identified between patients and controls following exclusion of genomic regions hyper-hydroxymethylated in blood cells (i.e. monocytes, neutrophils and erythroid progenitors). Patient identifiers are provided for each column with "P" representing NSCLC patients and "N" controls. NSCLC, non-small cell lung cancer; t-MAD, trimmed median absolute deviation from copy number neutrality.

3.7.2 5hmC signals in cfDNA are associated with gene expression in lung adenocarcinoma tissue

The accumulation of 5hmC within gene bodies as well as immediate upstream regions was demonstrated in several studies and positively correlated with the expression of associated genes (Madzo et al. 2014; Neri et al. 2013; Song et al. 2011; Szulwach et al. 2011). Previously, I observed that genes with TSS-proximal (±1,500 bp) 5hmC peaks in cfDNA of healthy donors showed higher expression in whole blood compared to genes without 5hmC modifications near their TSS (see 3.5; Figure 13B). Here, I assessed whether the 5hmC signal of the identified DhMRs could inform about tumor-related transcriptional processes. For this purpose, 5hmC signals of TSS-proximal as well as gene body associated DhMRs were compared with LUAD-specific gene expression determined from TCGA RNA sequencing data (Grossman et al. 2016) as described in section 3.6.3 (Figure 28A). Two out of 16 TSS-proximal DhMRs were associated with differentially expressed genes. *KPNA7* and *KIF25* were both hydroxymethylated at a higher level in cfDNA of NSCLC patients and were upregulated in LUAD tissues. Both genes have not yet been described in the context of lung cancer, however, high expression of *KPNA7* was demonstrated to increase cell proliferation in pancreatic and breast cancer cell lines (Vuorinen et al. 2018), while *KIF25* – alongside other members of the kinesin gene family – was shown to be highly expressed in estrogen

receptor-positive breast cancer cells (Zou et al. 2014). The inclusion of all gene body associated DhMRs identified 14 additional hyper-hydroxymethylated DhMRs that coincide with elevated gene expression in LUAD tissues. Eleven out of 16 genes (including *KPNA7* and *KIF25*) have been described to be associated with oncogenic properties in different cancer entities. Two genes were related to tumor suppressive functions, and three were not yet described in a cancer context (Figure 28B and Table S5). The most significantly hyper-hydroxymethylated DhMRs were found within introns of *FOXD4L1*, *HECW2*, *ALCAM*, and *IL1RAP* and in the promoter region of *KPNA7*. Expression levels of the cell adhesion molecule *ALCAM* were previously described to be upregulated in NSCLC, breast, and prostate cancer patients and are associated with poor prognosis (Burkhardt et al. 2006; Münsterberg et al. 2020; Sanders et al. 2019). *In vitro* and *in vivo* models revealed reduced cell adhesion and tumor cell dissemination upon *ALCAM* knockout, demonstrating its role as a promoter of brain metastasis formation (Sanders et al. 2019). Interestingly, *ALCAM* expression was also detectable in circulating tumor cells (CTCs) of NSCLC patients and correlated with its expression in matched brain metastasis tissue (Sanders et al. 2019). *IL1RAP* codes for a co-receptor of the interleukin-1 receptor and is overexpressed in various solid tumors, including lung cancer (Lv et al. 2021; Millrud et al. 2020). Currently, two clinical trials investigate the efficacy of an IL1RAP-targeting antibody (CAN04) in combination with platinum-based chemotherapy (NCT03267316) and immune checkpoint inhibition by pembrolizumab (NCT04452214) in solid tumors (i.e. NSCLC, urothelial carcinoma, malignant melanoma, and head and neck squamous cell carcinoma). *FOXD4L1* and *HECW2* have not yet been described in a cancer context. Notably, the majority of hyper-hydroxymethylated DhMRs within gene bodies (48/75) coincided with downregulated gene expression, suggesting a weaker association to active gene expression compared to TSS-proximal DMRs.

Eight transcriptionally downregulated genes in LUAD contained hypo-hydroxymethylated DhMRs within their gene body, all of which were located within introns (Table S6). Three of these (*PDZRN4*, *ANGPTL1*, and *PRKG2*) have been reported to act as tumor suppressors in different cancer types, while only *FGF1* was associated with oncogenic properties (the others have not been described in cancer yet). The literature search on hyper- and hypo-hydroxymethylated DhMRs associated with upregulated and downregulated gene expression is summarized in Table S5 and Table S6, respectively.

A

B

C
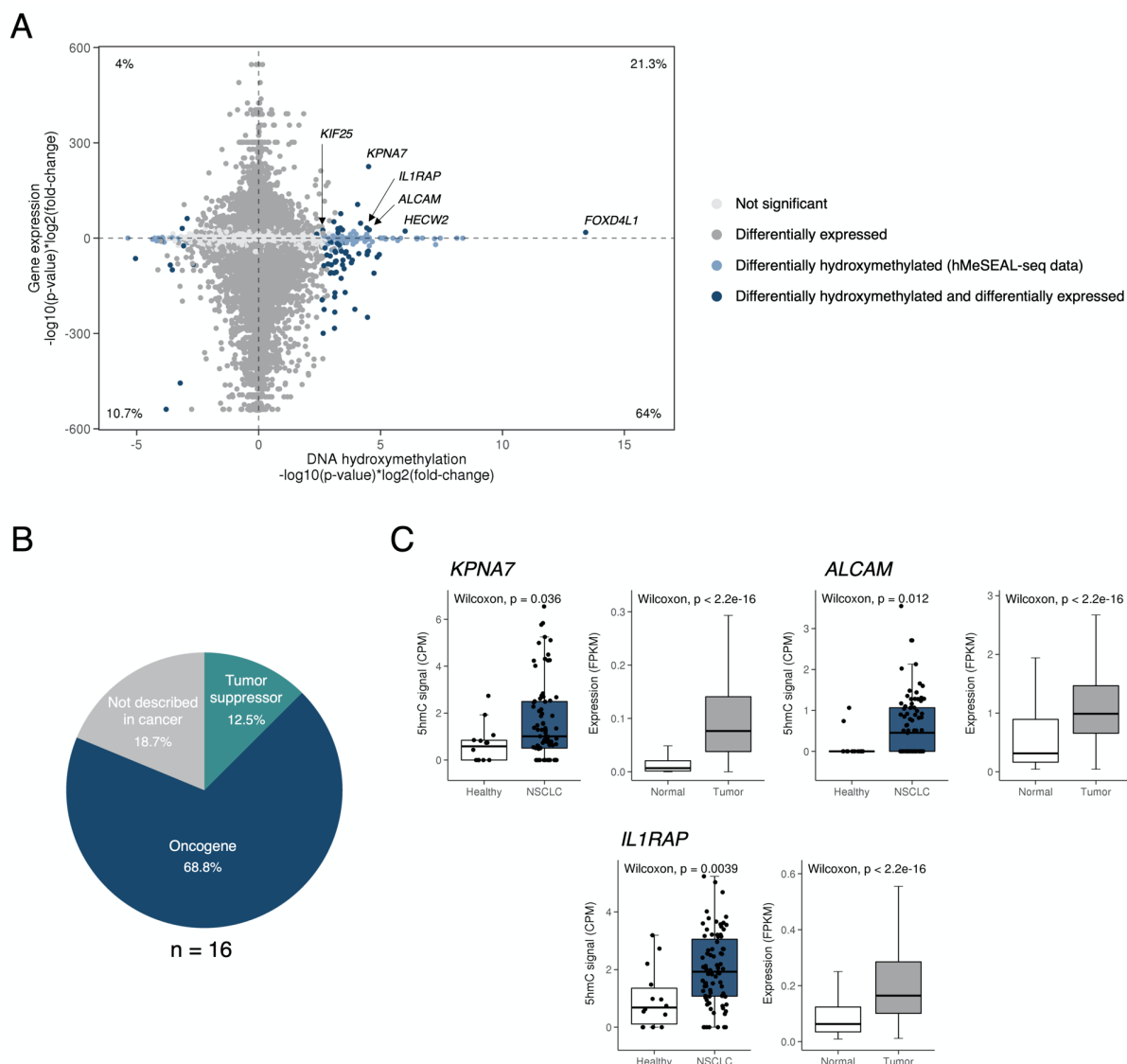
Figure 28: Relationship between TSS-proximal and gene body associated cfDNA hydroxymethylation and gene expression in LUAD tissue. (A) Starburst plot comparing cfDNA hydroxymethylation (NSCLC patients *vs.* healthy controls) and gene expression (LUAD tissue *vs.* non-cancerous lung, breast, colon, prostate, kidney, and stomach tissue). Each dot represents one TSS- or gene body-associated genomic region and colors indicate significance of the differential expression/hydroxymethylation analysis. Percentages refer to the relative number of both differentially expressed genes and differentially hydroxymethylated regions within each quadrant of the plot. (B) Overview of the cancer association of the 16 genes (Table S5) with upregulated gene expression in LUAD tissue and hyper-hydroxymethylated genomic regions in cfDNA of NSCLC patients. (C) Boxplots illustrate cfDNA 5hmC signals and LUAD tissue expression of representative genes. 5hmC levels (left) are illustrated for healthy controls (n = 14) and NSCLC patient samples (n = 74) with each dot representing one sample. Gene expression boxplots (right) compare LUAD (n = 533 samples) to normal tissue (n = 174; same tissue types as in (A)). CPM, counts per million; FPKM, fragments per kilobase per million reads; hMeSEAL, hydroxymethylation selective chemical labeling; LUAD, lung adenocarcinoma; NSCLC, non-small cell lung cancer; TSS, transcription start site.

3.7.3 Selection of 5hmC biomarkers

For the identification of potential 5hmC biomarkers, I focused on DhMRs mapping to genes previously associated with upregulated expression in LUAD tissue and validated in the recent literature to have oncogenic properties in NSCLC or other cancer entities. This resulted in eleven marker candidates listed in Table S5. The 5hmC signals of these DhMRs were correlated to the abundance of genomic alterations measured from the same plasma

samples (Dietz et al. 2020). The same genomic alterations as described in section 3.6.5 were used (i.e. SNVs and CNAs). This step aimed to validate the tumor association of the identified 5hmC markers. DhMRs at *KPNA7* and *KIF25* were the only 5hmC biomarkers with significant correlations to at least one of the assessed genomic alterations (Figure 29A and B; Table S5). The highest correlation was observed between *KPNA7* and the t-MAD score (Pearson, r = 0.40). Interestingly, DhMRs at *KPNA7* and *KIF25* were the only biomarker candidates located within promoter regions; all other candidate marker DhMRs mapped within introns. Compared to 5mC biomarkers, the 5hmC associated marker genes demonstrated lower correlations to the genomic alterations.

Analogous to the identification of 5mC biomarkers, I assessed whether the 5hmC signals at the selected marker candidates could distinguish between cfDNA samples from NSCLC patients and healthy controls as well as PD from non-PD patient samples. More than half of the marker genes (7/11) demonstrated significantly higher 5hmC signals in patient samples compared to the healthy controls. Differentiation between PD and non-PD samples was possible only based on the 5hmC levels at the *IL1RAP* gene locus (Figure 29C).
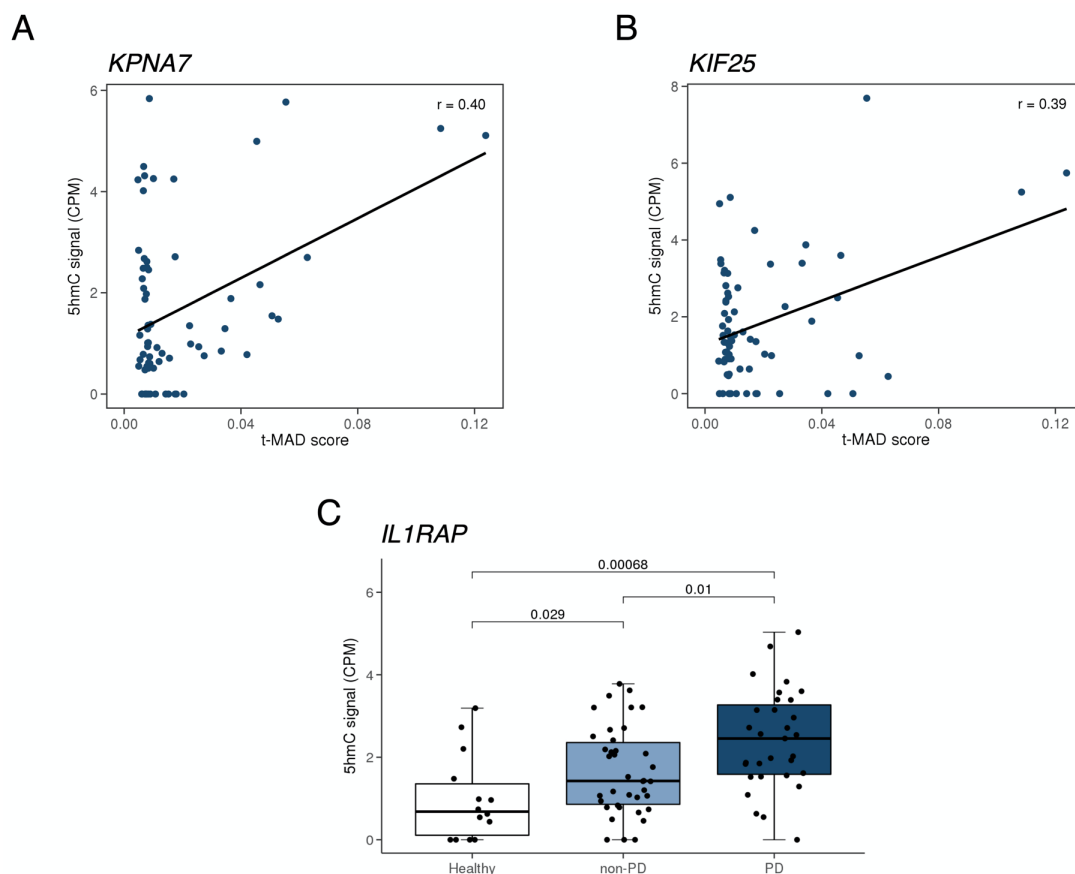
Figure 29: Representative 5hmC biomarker candidates. Correlation of (A) *KPNA7* and (B) *KIF25* 5hmC signals to t-MAD scores determined in the same plasma samples from shallow whole-genome sequencing data. Each dot represents one plasma sample and Pearson correlation coefficients are illustrated. (C) *IL1RAP* 5hmC signals compared between healthy controls (n = 14), patient samples taken at non-progressive time points (non-PD; n = 37) and samples taken at progressive time points (PD; n = 31). Each dot represents one plasma sample. CPM, copies per million; t-MAD, trimmed median absolute deviation from copy number neutrality.

3.7.4 5hmC biomarkers predict therapy failure in longitudinally taken plasma samples

To investigate whether 5hmC signals could predict therapy success in longitudinally collected LUAD samples, *KPNA7*, *KIF25* and *IL1RAP* were chosen as candidate biomarkers. The 5hmC signals at these genes were either significantly correlated with genomic alterations (i.e. t-MAD and VAF_{max}) or significantly elevated in PD compared to non-PD samples. In four patients, the radiologically assessed disease state was reflected by the abundance of 5hmC signals at the investigated marker regions (Figure 30A-D). The response of patient P044 (Figure 30A) to crizotinib treatment was accompanied by successively decreasing *IL1RAP* 5hmC signals. Disease progression at day 508 was indicated by increasing 5hmC levels and ineffective ceritinib treatment coincided with high 5hmC signals at day 640, shortly before the patient′s death. Genomic alterations (i.e. t-MAD score, VAF_{max} and the *EML4-ALK* fusion), the previously identified 5mC biomarkers and radiology captured ceritinib failure earlier compared to the 5hmC marker. Similarly, response to crizotinib of patient P005 (Figure 30B) coincided with decreasing *IL1RAP* 5hmC levels, while 5hmC signals increased in the plasma sample taken few days before the patient′s death. In total, eight patients deceased shortly after the last plasma sampling time point. The patients′ deaths were preceded by increasing *IL1RAP* 5hmC signals at 6/8 cases, while *KPNA7* and *KIF25* 5hmC levels increased in 3/8 and 4/8 patients, respectively. At ten instances, 5hmC profiles of PD samples and preceding SD samples were available. The 5hmC signal increase from the SD to the PD time point correctly indicated disease progression in 6/10 (*IL1RAP*), 5/10 (*KIF25*) and 2/10 (*KPNA7*) cases. Persistently rising *IL1RAP* 5hmC signals during alectinib treatment of patient P028 (Figure 30D) could predict therapy failure already at day 49, while radiological progression was first detected 112 days later. 5mC biomarkers predicted therapy failure at the same time point, and cfDNA genomic alterations (i.e. t-MAD score, *EML4-ALK* fusion and ALK G1202R) preceded radiological progression by 55 days. *KIF25* 5hmC levels indicated alectinib failure at the same time point as *IL1RAP* hydroxymethylation. 5hmC markers did not precede radiological disease assessment in other patients.
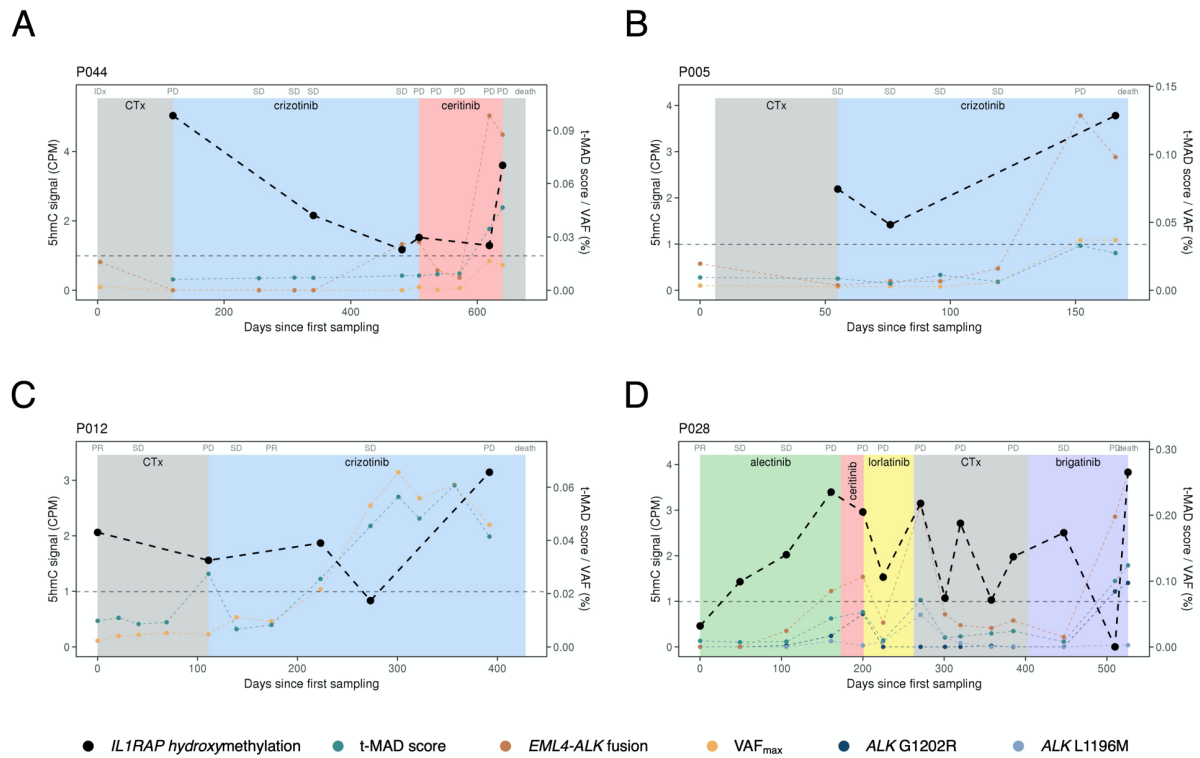
Figure 30: Representative patients illustrating the utility of *IL1RAP* 5hmC levels in cfDNA as biomarker for therapy monitoring of ALK⁺ NSCLC patients. *IL1RAP* 5hmC levels, molecular alterations (CAPP-seq) and t-MAD score (sWGS) dynamics in plasma cfDNA during sequential ALK-directed tyrosine kinase inhibitor treatment and chemotherapy of patients (A) P044, (B) P005, (C) P012, and (D) P028. The dotted horizontal line represents the median *IL1RAP* 5hmC signal of 14 healthy individuals. Radiological disease assessments are indicated above each plot and administered therapy lines are shown by the background colors. Each dot represents one plasma sample. ALK, anaplastic lymphoma kinase; CAPP-seq, Cancer Personalized Profiling by deep sequencing; cfMeDIP-seq, cell-free methylation DNA immunoprecipitation; CPM, counts per million; CTx, chemotherapy; EML4, echinoderm microtubule-associated protein-like 4; IDx, initial diagnosis; PD, progressive disease; PR, partial response; SD, stable disease; t-MAD, trimmed median absolute deviation from copy number neutrality; VAF$_{max}$, maximal variant allele frequency.

3.8 Inference of copy number alterations from 5mC data

Somatic CNAs are common in many metastatic cancers and can provide estimates for tumor burden based on cfDNA samples. Commonly, sWGS is used as a cost-effective method for the assessment of copy number changes from liquid biopsies (Dietz et al. 2020; Mouliere et al. 2018b; Smith et al. 2020). Here, I explored whether genome-wide 5mC and/or 5hmC profiles can also be used to detect CNAs from cfDNA. This would allow simultaneous genomic and epigenomic tumor assessment from the same dataset. As a reference, copy number profiles from sWGS data (Dietz et al. 2020) at 1 Mb bins were generated using the ichorCNA software (Adalsteinsson et al. 2017). Afterwards, CNAs in the corresponding 5mC- and 5hmC-enriched sequencing datasets were analyzed using the same algorithm. Matched CNA patterns were found in several patient samples when comparing copy number profiles from 5mC-enriched sequencing data to sWGS data. CNA profiles detected from 5hmC-enriched cfDNA were more diffuse and missed some of the copy number changes detected in the other two datasets (Figure 31A and Figure S13). Next, t-MAD scores were calculated from 5mC- and 5hmC-enriched as well as sWGS data for a quantitative comparison between the three data types. All three data types were downsampled to a common read coverage of 5 million paired reads per samples to mitigate the effect of varying sequencing depths on the t-MAD score calculation. Bin sizes were kept at 1 Mb as for CNA profiling by ichorCNA. A high positive correlation between t-MAD scores derived from sWGS and 5mC-enriched data was found (Pearson, r = 0.88; Figure 31B). The t-MAD scores derived from 5hmC-enriched data presented a lower correlation to sWGS data (Pearson, r = 0.39; Figure 31C). Especially the 5hmC-enriched dataset tended to underestimate chromosomal instability in samples with high t-MAD scores determined from sWGS data.
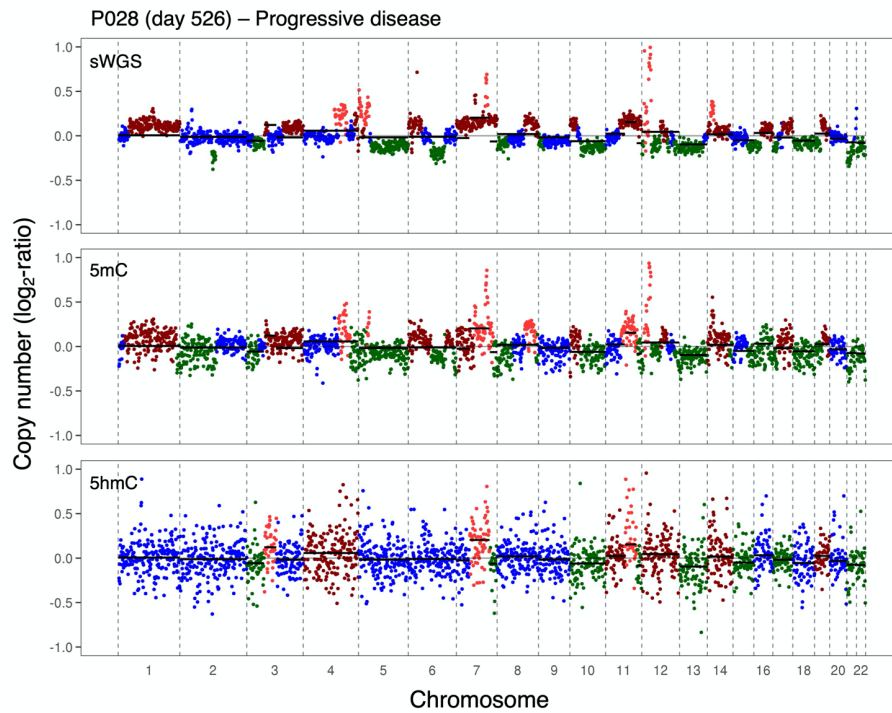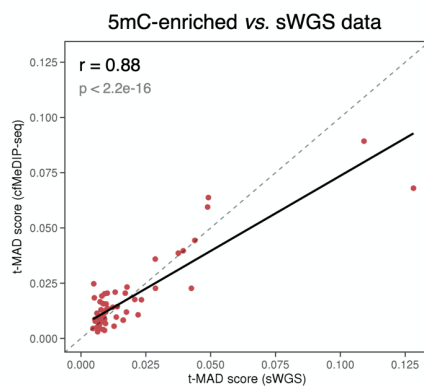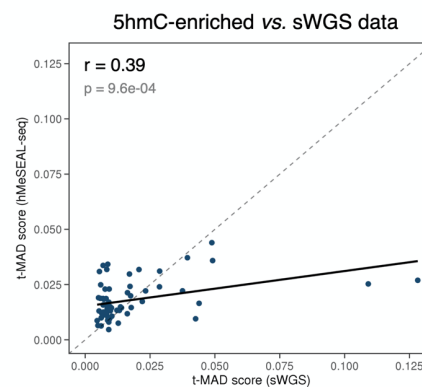
Figure 31: Inference of copy number alterations from sWGS, 5mC- and 5hmC-enriched sequencing data. (A) Exemplary copy number profiles inferred from shallow whole-genome (sWGS; top), 5mC-enriched (middle) and 5hmC-enriched sequencing data (bottom) of cfDNA from patient P028 at a time point of progressive disease. Colors indicate copy number neutrality (blue), deletions (green), copy number gain (brown; 3 copies), and amplifications (red; >3 copies). Scatter plots show the correlation of t-MAD scores derived from sWGS data to 5mC (B) and 5hmC data derived t-MAD scores (C). The dotted line indicates a perfect correlation between both datasets. Pearson correlation coefficients are displayed. cfMeDIP; cell-free methylation DNA immunoprecipitation; hMeSEAL, hydroxymethylation-selective chemical labeling; t-MAD, trimmed median absolute deviation from copy number neutrality.

4 Discussion

Targeted therapies are of high relevance for the treatment of advanced NSCLC with genomic alterations: Patients harboring tumors with actionable oncogenic drivers – such as *EGFR*, *ALK* or *ROS1* – demonstrate significant clinical benefits from directed TKI treatment (Maemondo et al. 2010; Shaw et al. 2014; Solomon et al. 2014; Soria et al. 2018). Multiple TKIs are approved for the management of advanced ALK⁺ tumors and immensely improved patient prognosis. Despite the benefits of ALK-directed therapies, clinical courses vary widely due to the development of drug resistances such as mutations in the *ALK* kinase domain, *ALK* amplifications or the activation of bypass signaling pathways (Shaw et al. 2013; Solomon et al. 2014). Some of these resistances can be overcome by second and third generation ALK-TKIs that allow sustained disease remission even after first TKI failure (Camidge et al. 2014). Therefore, it is crucial to detect therapy failure early to guide subsequent therapy lines and optimize the patient´s benefit from the treatment. Liquid biopsies can discover tumor-derived material from blood samples of cancer patients. The minimal-invasive sample collection procedure allows repeated sampling throughout patient treatment and provides the opportunity to capture early signs of disease progression (Wan et al. 2017). We and others measured cancer-specific genomic alterations in cfDNA (e.g. SNVs and CNAs) to monitor the therapy of ALK⁺ NSCLC patients (Dagogo-Jack et al. 2018; Dietz et al. 2020; Li et al. 2021; Zhang et al. 2020a). These alterations are commonly very specific to the patient´s tumor, yet suffer from limited sensitivity due to their low relative abundance in the cancer genome (van der Pol and Mouliere 2019). The profiling of epigenomic alterations in cfDNA presents a promising alternative. Cancer-specific changes in the methylome and hydroxymethylome are more prevalent and more pervasive compared to genomic alterations, providing a higher level of sensitivity (Kandoth et al. 2013; Li and Zhou 2020). Additionally, DNA modifications can inform about the tumor´s tissue-of-origin and their role in transcriptional regulation might identify active or repressed gene expression signatures (Moss et al. 2018; Song et al. 2017; Sun et al. 2015).

In this study, genome-wide 5mC and 5hmC profiles were analyzed from cfDNA of metastatic ALK⁺ NSCLC patients and healthy individuals. The first objective was the establishment of a workflow that enriches for tumor-derived 5(h)mC signatures by: (1) Exclusion of genomic regions with high 5(h)mC levels in cell types contributing to non-tumor cfDNA, followed by (2) differential analysis between patient and healthy donor samples. The second objective was to associate tumor-specific 5(h)mC patterns with activation and repression of genes in LUAD tissue. The third objective was to correlate 5(h)mC biomarkers with cancer-specific genomic alterations co-measured in the same plasma samples. Finally, dynamic changes in biomarker abundances were followed in longitudinal plasma samples to monitor the patients´ ALK-directed TKI therapy.

4.1 Technical performance of 5mC and 5hmC enrichment

Experimentally, cfMeDIP- and hMeSEAL-seq precipitate cfDNA fragments containing methylated or hydroxymethylated CpG dinucleotides (Shen et al. 2019b; Song et al. 2017). By sequencing only the enriched cfDNA fraction, (hydroxy-)methylome profiling is feasible at low sequencing costs. Additionally, bisulfite conversion is not needed, rendering the protocols applicable to samples with low cfDNA content (Shen et al. 2019b). These advantages outweigh the protocols′ lower resolution (down to ~100-bp genomic regions) compared to bisulfite conversion-based methods (single nucleotide level) and favor their application on cfDNA samples.

As part of this study, the technical performance of cfMeDIP- and hMeSEAL-seq was compared and revealed distinct differences. Although spike-in recovery was similar in both protocols, 5mC libraries demonstrated a comparatively higher enrichment for CpG sites than 5hmC libraries. While 5mC is commonly located within CGIs (Greenberg and Bourc'his 2019), 5hmC was reported to occur at less CpG dense regions (e.g. CGI borders) (Li et al. 2016; Skvortsova et al. 2019a). This likely results in high per fragment CpG densities in 5mC libraries, thereby explaining the different enrichments between the 5mC and 5hmC datasets. The determined CpG enrichment scores in 5mC libraries were coherent with those reported in the recent literature (Shen et al. 2018). For 5hmC-enriched data, this quality metric has not yet been applied in other studies. Saturation analysis was performed to address whether the given number of paired reads suffices to generate reproducible coverage profiles (Lienhard et al. 2014). Lower reproducibility at similar sequencing depths was observed in 5hmC compared to 5mC libraries, indicating that higher sequence coverage in the 5hmC dataset could improve the detection of tumor-derived 5hmC alterations. The comparatively lower reproducibility suggests a higher complexity of the generated 5hmC profiles. This is in contrast to the lower abundance of 5hmC compared to 5mC marks throughout the genome (Brazauskas and Kriaucionis 2014), which suggests that fewer sequencing reads are needed to sufficiently cover the hydroxymethylome. The reason for the reduced reproducibility of the 5hmC libraries remains unclear. However, a lower specificity of hMeSEAL-seq, compared to cfMeDIP-seq, is unlikely since similar quantities of unspecific binding events (number of reads covering no CpG site) and comparable enrichment efficiencies (spike-in recovery) were found. A more plausible explanation resides in the different binding mechanisms of the two protocols: While the enzymatic 5hmC labeling of hMeSEAL-seq captures hydroxymethylated cytosines irrespective of local 5hmC density or CpG-context (Thomson et al. 2013), the antibody-based cfMeDIP-seq has a higher specificity for regions of high or intermediate 5mC occupancy (Taiwo et al. 2012). This could result in cfMeDIP-seq missing single 5mC marks, which would reduce the number of genomic regions captured by this method. Consequently, fewer sequencing reads are needed to reproducibly cover these regions.

4.2 Identification of tumor-derived 5mC and 5hmC alterations

Tumor DNA can be detected in plasma of cancer patients and offers the possibility of minimal-invasive cancer assessment (Bettegowda et al. 2014; Newman et al. 2014). cfDNA is regarded as a mixture of DNA derived from different cell and tissue types, with ctDNA presenting only a minor fraction in most cancer patients (Moss et al. 2018; Sun et al. 2015). Therefore, a major challenge for cfDNA-based cancer detection is to identify the minute amounts of tumor-derived DNA fragments in the total cfDNA pool.

4.2.1 Generation of genome-wide 5mC and 5hmC reference profiles from primary blood cells

Recently it was demonstrated that the majority of non-tumor cfDNA is derived from hematopoietic cells (Lehmann-Werman et al. 2016; Sadeh et al. 2021; Sun et al. 2015). The most detailed study in this regard used methylation data to deconvolute the relative contributions of individual cell types to the cfDNA composition in healthy individuals (Moss et al. 2018).

Here, I used this information as a basis to identify and subtract non-tumor cfDNA from genome-wide 5mC and 5hmC profiles measured in cfDNA of ALK[+] NSCLC patients. By exclusion of hypermethylated or hyper-hydroxymethylated genomic regions in healthy cfDNA, I aimed to enrich for cancer-associated 5(h)mC alterations. For this purpose, 5mC and 5hmC profiles were generated from primary monocytes, neutrophils and erythroid progenitor cells. Combined, these cell types constitute 72.2% of the non-tumor cfDNA repertoire (Moss et al. 2018). Other contributors only account for minor fractions of healthy cfDNA and were not considered in this study. Creating in-house 5(h)mC reference profiles has conceptual advantages over the usage of publicly available datasets: (1) Publicly available blood cell reference data was generated using Illumina 450k methylation arrays which cover only a small fraction of the genome. In contrast, 5mC and 5hmC profiles determined by cfMeDIP- and hMeSEAL-seq, respectively, enable the identification and exclusion of blood cell-derived genomic regions on a genome-wide scale. (2) Using the same methodology for the analysis of cfDNA samples and blood cell gDNA improves data comparability. (3) Blood cell 5hmC profiles were not available from the literature and had to be generated to enable the filtering of hyper-hydroxymethylated genomic regions.

The analysis of individual blood cell types permitted to weight and combine 5(h)mC signals based on their relative abundance in healthy cfDNA. This presents a considerable advantage over the usage of bulk 5(h)mC profiles from peripheral mononuclear cell or whole blood. The contribution of different blood cells to cfDNA was shown to deviate from their relative abundance in the circulation (Moss et al. 2018). Therefore, weighted integration of individual blood cell 5(h)mC levels should provide a more accurate reflection of non-tumor cfDNA. However, there are limitations to the blood cell reference datasets used for this study: First,

the considered cell types only account for 72.2% of healthy cfDNA and cannot exclude the entire non-tumor cfDNA. Second, the list of cell types contributing to healthy cfDNA (Moss et al. 2018) might not be complete, and, third, reference datasets were generated from only one individual per cell type. Averaging the 5(h)mC profiles of multiple individuals would likely provide a more robust representation of the cell types' (hydroxy-)methylomes.

4.2.2 Enrichment of tumor-derived cfDNA by the exclusion of genomic regions associated with high 5(h)mC levels in blood cells

Methylation data from LUAD and non-malignant tissues (TCGA) were used as reference to (1) establish a cut-off at which genomic regions were excluded due to high blood cell 5(h)mC levels and (2) validate the tumor association of the remaining regions. Due to the unavailability of 5hmC tissue data, these steps were solely performed on the 5mC dataset. A moderate correlation between 5mC profiles of NSCLC cfDNA and LUAD tissue was evident which successively increased with rising blood cell filtering stringency. Similar observations were made in another study that correlated blood cell-filtered cfDNA and tissue 5mC profiles of glioma patients (Nassiri et al. 2020). This suggested increased tumor association of the 5mC signals remaining after the filtering step. A justified explanation for this result is that the partial removal of non-tumor cfDNA signals facilitates the identification of cancer-associated 5mC signatures from the huge amount of non-tumor derived cfDNA. Other studies demonstrated the utility of different approaches to unmask cancer-derived alterations from non-tumor cfDNA. Mouliere *et al.* enriched for ctDNA by focusing their analysis on short cfDNA fragments (90 – 150 bp). Thereby, they increased the sensitivity of their assay to identify otherwise undetectable CNAs (Mouliere et al. 2017). Larson and colleagues determined cancer-specific cell-free RNA (cfRNA) signatures by restricting their analysis to genes rarely detected in a non-cancer reference group. Using this approach, they could reduce the likelihood of false-positive cfRNA signals and increase sensitivity (Larson et al. 2021). I chose a filtering cut-off at blood cell 5mC levels <20% of the median NSCLC cfDNA signal in the same genomic region. This cut-off presented the closest tumor association (highest correlation to LUAD tissue methylation), while retaining the largest number of genomic regions for subsequent analyses. A limitation of deriving a blood cell filtering cut-off by correlation to LUAD tissue data is the employed reference. TCGA 5mC data was generated by Illumina 450k methylation arrays and contains only three patient samples with confirmed *EML4-ALK* fusions (Weisenberger 2014). The preparation of cfMeDIP-seq dataset from matched tumor tissue would increase dataset comparability and enable the identification of 5mC alterations specific to ALK[+] NSCLC patients.

Notably, removal of cfDNA from blood cells also increased the correlation between 5mC profiles in NSCLC cfDNA and tissues of various other tumors as well as non-malignant lung. The methylome of blood cells is highly distinctive from most somatic tissues (Lowe et al. 2013; Lowe et al. 2015; Varley et al. 2013). Therefore, it is reasonable that the partial

removal of blood cell 5mC profiles increases the association between NSCLC cfDNA and various tissues. Correlations between NSCLC cfDNA methylation and different tumor tissues were similar, while normal lung 5mC profiles demonstrated a comparatively lower association to NSCLC cfDNA. In conclusion, the majority of the tumor-derived 5mC signatures revealed after blood cell filtering reflected methylome perturbations common to several cancer entities and not exclusive to lung cancers.

To validate the increased tumor association of NSCLC cfDNA after blood cell filtering, 5mC signals in patient and control samples were compared. At hypermethylated genomic regions in LUAD tissue, 5mC signals were higher in patient compared to healthy control cfDNA. Without blood cell filtering, no difference between the two groups was noted. This again demonstrated that the blood cell filtering step facilitates the identification of tumor-derived 5mC alterations by removing uninformative background signals. Genomic regions hypermethylated in LUAD, but not in other cancer entities, also exhibited higher 5mC signals in NSCLC compared to healthy donor cfDNA. The difference between the groups was increased following blood cell filtering and absent at regions uniquely hypermethylated in other cancer entities. These observations underline the capability of cfMeDIP-seq to capture entity-specific 5mC alterations from patient cfDNA and suggest its suitability for tumor classification. This is in line with studies showing the application of 5mC signatures for tissue-of-origin inference from cfDNA of cancer patients (Guo et al. 2017; Lehmann-Werman et al. 2016; Moss et al. 2018; Sun et al. 2015). In addition, Shen and colleagues demonstrated that cfMeDIP-seq can distinguish tumors of patients harboring diverse cancer types (Shen et al. 2018). Moreover, blood cell filtering might be a reasonable approach to facilitate cancer classification by enhancing the tumor signal in cfDNA samples.

The analysis of hypomethylated genomic regions in LUAD tissue gave less conclusive results. This can be attributed to the limitation of cfMeDIP-seq to provide an "unmethylated signal". The absence of 5mC signals at genomic regions might either reflect a lack of methylation or a poor representation of the regions in the cfDNA sample (Shen et al. 2019b), impeding the interpretation of a missing signal.

Additionally, I observed that patient and control samples were easier to distinguish based on their 5mC profiles after blood cell filtering (PCA clustering), compared to the unfiltered data. The best separation was found between healthy controls and t-MAD$_{high}$ samples. Samples with high t-MAD scores commonly contain elevated tumor fractions in cfDNA (Dietz et al. 2020; Mouliere et al. 2018a). This suggests that the tumor-derived 5mC signal is the driving force differentiating patient samples from controls and shows that the ctDNA fraction is enriched after blood cell filtering. PCA clustering was also performed using the 5hmC-enriched dataset. Here, the same blood cell filtering cut-off as for the 5mC dataset was used. The separation between patients and controls improved after blood cell 5hmC signal filtering, however, differentiation between groups was less clear. In addition, t-MAD scores did not affect clustering, potentially reflecting a lower tumor association of the remaining 5hmC signals compared to the 5mC dataset.

4.2.3 Differential analysis identifies tumor-specific 5(h)mC alterations

The blood cell-filtered 5mC and 5hmC datasets were subjected to differential analysis between NSCLC patient and healthy control samples. This aimed to further enrich for tumor-derived epigenetic alterations. Differential analysis of 5mC data resulted in a considerably higher number of hyper- compared to hypomethylated DMRs. A similar imbalance was found for the 5hmC dataset. Other studies using cfMeDIP- or hMeSEAL-seq for cfDNA-based 5(h)mC profiling showed corresponding imbalances (Shen et al. 2018; Tian et al. 2018; Zhang et al. 2018b). A potential explanation for this observation is the high non-tumor background signal in cfDNA of cancer patients, masking the hypomethylation or hypo-hydroxymethylation status of the tumor fraction. Although hyper-D(h)MRs can also be masked by the non-tumor cfDNA signal, they can be more readily identified at regions of no or low 5(h)mC signals in healthy donors. Approximately 11-fold more DMRs, compared to DhMRs, were identified in this study. In part, this difference can be attributed to the higher sequence coverage of the 5mC dataset. High sequencing depth increases the statistical power of the negative-binomial test used by edgeR, resulting in a higher sensitivity of DMR detection (Gontarz et al. 2020; Liu et al. 2014). Additionally, the lower reproducibility of the 5hmC (compared to 5mC) dataset indicates that increased sequence coverage could improve the detection of cancer-associated 5hmC alterations (discussed in section 4.1). However, biological factors might also contribute to the detection of more DMRs. The abundance of 5hmC modifications throughout the human genome is much lower compared to 5mC (Brazauskas and Kriaucionis 2014). This potentially translates into fewer tumor-associated 5hmC alterations retrievable from cfDNA.

Hierarchical clustering based on all DMRs or DhMRs separated patient from healthy control samples. This represents an improvement compared to the sole blood cell filtering – which could only partially differentiate samples by disease status (PCA clustering) – and suggests that the differential analysis further enriches for tumor-informative 5(h)mC alterations. The preferential clustering of samples with similar t-MAD scores was already noted after blood cell filtering in the 5mC dataset. This was preserved after differential analysis and became additionally apparent in the 5hmC-enriched samples, further emphasizing that the ctDNA content in the cfDNA samples is a major factor contributing to the differentiation between patients and controls. 5mC signals of DMRs overlapping with LUAD tissue-specific sites were significantly higher in patients compared to controls and the difference between these two groups increased compared to performing blood cell filtering alone. By this means, it was further confirmed that the differential analysis after blood cell filtering is a reasonable approach to enrich for tumor-derived alterations in cfDNA.

In summary, non-tumor background signals can be efficiently reduced by the exclusion of genomic regions with high 5(h)mC signals in blood cells and subsequent differential analysis (patient *vs.* control samples). This might facilitate several downstream analysis steps, such as the identification of cancer- or tissue-specific biomarkers and inference of tumor expression

profiles. Especially, in clinical settings with low ctDNA content (e.g. MRD monitoring or early disease detection) filtering of blood cell-derived signals could enable a better tumor detection. For this study, a larger non-cancer cohort and the inclusion of patients with other lung-related diseases could further refine the set of identified D(h)MRs.

4.3 Inference of gene expression from 5mC and 5hmC signals in cfDNA

The role of DNA methylation and hydroxymethylation in the regulation of gene expression is well characterized. At TSS-proximal regions, DNA hypermethylation is commonly described to repress transcription, while high 5hmC levels are associated with active genes (Greenberg and Bourc'his 2019; Thomson et al. 2012). In this study, it was assessed whether 5mC and 5hmC profiles can infer gene expression from cfDNA samples.

4.3.1 Inference of whole blood expression

The inference of cancer-associated transcription from cfDNA is challenging due to the low tumor DNA content in most samples (Bettegowda et al. 2014; Diehl et al. 2008). Previous studies used plasma of healthy individuals to derive cfDNA characteristics informative of gene expression in hematopoietic cells (Snyder et al. 2016; Ulz et al. 2016). The high abundance of blood cell DNA in these samples (~85%) (Moss et al. 2018; Sadeh et al. 2021; Sun et al. 2015) allows a more straightforward inference of gene expression that can be transferred to patient cfDNA in the next step. I compared TSS-proximal (±1,500 bp) 5mC and 5hmC signals in cfDNA of healthy individuals to whole blood gene expression obtained from the GTEx project. DNA modifications around the TSS are frequently described to regulate transcription (Greenberg and Bourc'his 2019; Thomson et al. 2012). Additionally, this study showed an enrichment of 5hmC, but not 5mC, peaks proximal to the TSS of genes predicted to be active in the GM12878 lymphoblastoid cell line. Therefore, it is reasonable to assume that 5(h)mC levels at these regions might inform about gene expression.

Occupation of 5hmC sites near TSSs demonstrated higher expression levels compared to genes without DNA modification (i.e. 5mC or 5hmC) in the same region and 5hmC-associated TSSs were enriched in gene sets regulating various blood cell-related processes (e.g. leukocyte, neutrophil and macrophage migration). These findings match the activating function of 5hmC, which is described in the literature (Ehrlich and Ehrlich 2014; Mellén et al. 2012; Thomson et al. 2012) and propose that TSS-proximal 5hmC can inform about transcriptional processes occurring in blood cells. Against the expectations, genes with 5mC peaks near their TSS also demonstrated higher expression compared to genes without DNA modifications. However, expression levels of 5hmC-associated genes were 3.1-fold higher compared to 5mC genes. This suggests a closer connection of TSS-proximal 5hmC to active transcription and favors 5hmC as a proxy for gene expression inference in cfDNA.

A limitation of this approach to deduce gene expression is that the cell types giving rise to cfDNA in healthy individuals are heterogeneous and not accurately reflected by the relative abundance of cells in whole blood. For example, erythroid progenitor cells are believed to shed DNA into the circulation during the generation of enucleated red blood cells (i.e. during cell birth rather than cell death) and are likely to be underrepresented in whole blood data (Moss et al. 2018). This reduces the comparability of the cell-free 5(h)mC profiles to the utilized expression data, thereby impeding the inference of transcription activity. Additionally, it has to be considered that transcription regulation is not exclusively conferred by DNA modifications (Yilmaz and Grotewold 2010). Other processes, such as modifications on the histone level or expression regulation by small RNAs, cannot be inferred in this study.

4.3.2 Inference of lung adenocarcinoma tissue expression

To conclude about tumor-derived expression from cell-free 5(h)mC profiles in NSCLC patients, I restricted the analysis to genomic regions with low (or no) signals in cfDNA of healthy individuals (i.e. regions remaining after blood cell filtering and differential analysis). By focusing on these background-reduced regions, a high tumor-specificity is ensured and the challenge of the low ctDNA content in most patient samples might be overcome. The cancer-associated 5(h)mC profiles were compared to LUAD tissue expression data to evaluate their role in the regulation of gene activity.

DMRs were found to be enriched at TSS-proximal regions (i.e. promoters, 5′-UTRs and first exons). Most of them exhibited characteristics supporting their validity as tumor-specific and involvement in regulation of expression: (1) Genes associated to hypermethylated DMRs significantly overlapped with those repressed in LUAD tissue (45/63), (2) LUAD-specific hypermethylation was confirmed at 35/45 genes, and (3) previous reports associated 30/45 genes to cancer-related processes with 24/30 exhibiting tumor suppressive functions. *GATA4* is an example for a gene that matches all of these criteria. It belongs to the *GATA* zinc finger TF family. Loss of *GATA4* function has been linked to various malignancies (e.g. ovarian, gastric, colorectal, and lung cancer) (Akiyama et al. 2003; Bai et al. 2000; Gao et al. 2019; Lassus et al. 2001). In NSCLC, *GATA4* deficiency is accomplished by promoter hypermethylation (Gao et al. 2019; Guo et al. 2004). This is consistent with this study, which shows elevated *GATA4* 5mC levels in plasma of NSCLC patients, but not in healthy individuals. *GATA4* is considered a pioneer modifier, opening closed chromatin to facilitate TF binding (Cirillo et al. 2002). Its overexpression in NSCLC cell lines induces senescence *via* inhibition of TGF-β signaling, conferred by *GATA4*-mediated *TGFBR1* downregulation. This demonstrates the tumor suppressive role of *GATA4* in NSCLC. Interestingly, *GATA4* promoter hypermethylation is present in several molecular subtypes of LUAD (i.e. *EGFR*, *KRAS* and *EML4-ALK*-driven tumors) as well as in LUSC (Gao et al. 2019). This highlights the potential of *GATA4* methylation in cfDNA as universal lung cancer biomarker. Among the DMRs associated with downregulated expression in LUAD, many additional TFs (e.g.

GATA3, *HOXA9*, *NKX2-8*, *PAX6*, and *TBX3*; n = 12/45) were identified. These were predominantly depicted in a tumor suppressive context and TSS-proximal hypermethylation was confirmed in LUAD tissue for 10/12 genes. The overrepresentation of 5mC-mediated TF silencing might reflect the disruption of large-scale transcriptional processes in the NSCLC patient tumors. Commonly, TFs control the expression of several downstream targets. Dysregulation of established cell regulatory networks secondary to epigenetic TF silencing is frequently described during carcinogenesis (Akiyama et al. 2003; Suvà et al. 2014; Winslow et al. 2011).

The second most abundant group of genes within the identified DMRs were lncRNAs (e.g. *HOXA10-AS1*, *SOX9-AS1* and *ZNF667-AS1*; n = 9/45). Similar to TFs, lncRNAs participate in many regulatory pathways (e.g. chromatin organization and transcriptional regulation), frequently controlling multiple targets by acting as scaffolds, miRNA sponges or competitors to endogenous RNA (Yang et al. 2014). This further highlights the potential association of the identified DMRs with the disruption of global transcriptional processes. Dysregulation of lncRNA expression is a pervasive event during NSCLC pathogenesis and oftentimes conferred by epigenetic alterations (Ashouri et al. 2016; Yan et al. 2015). *SOX9-AS1* was the most significantly hypermethylated lncRNA in cfDNA of NSCLC patients within this study. It functions as a regulator of *SOX9*, a TF reported to induce proliferation by activating Wnt/β-catenin signaling in NSCLC and hepatocellular carcinoma cell lines (Guo et al. 2018; Zhang et al. 2019b). Although the role of *SOX9-AS1* has not yet been described in NSCLC, its expression downregulation in LUAD tissue proposes *SOX9-AS1*-mediated repression of *SOX9*. The TSS-proximal hypermethylation in NSCLC cfDNA found in this study indicates epigenetic silencing of *SOX9-AS1*, supporting this hypothesis.

After blood cell filtering and differential analysis of the 5hmC-enriched sequencing data, 16 genes with DhMRs near their TSS remained. Of these genes, two (*KPNA7* and *KIF25*) were also differentially expressed in LUAD tissue data from TCGA. In line with the frequently reported transcription activating role of 5hmC (Ehrlich and Ehrlich 2014; Mellén et al. 2012; Thomson and Meehan 2017) and its association with elevated expression in whole blood shown in this study, *KPNA7* and *KIF25* are overexpressed in LUAD compared to various normal tissues. Both genes were previously portrayed to confer oncogenic functions in cancer cell lines (Laurila et al. 2014; Vuorinen et al. 2018; Zou et al. 2014). While *KPNA7* is highly expressed during embryogenesis, transcription is drastically reduced or absent in most adult tissues. In pancreatic and breast cancer cell lines, *KPNA7* is re-expressed and its silencing reduces cell proliferation *via* p21 induction and G1 arrest (Laurila et al. 2014; Vuorinen et al. 2018). The kinesin family member, *KIF25*, is one of 19 kinesin genes upregulated in estrogen-receptor-positive breast cancer cells and associated with increased cell growth and survival (Zou et al. 2014). *KPNA7* and *KIF25* overexpression in LUAD tissue might indicate similar oncogenic properties in lung cancer, potentially regulated by TSS-proximal hyper-hydroxymethylation. Previous studies showed that 5hmC abundances throughout the gene body can also inform about the transcriptional status of genes (Guler et

al. 2020; Song et al. 2017; Tian et al. 2018; Zhang et al. 2018b). In contrast, the gene body-associated DhMRs of this study were not enriched at genes overexpressed in LUAD tissue, suggesting a more pronounced regulatory effect of 5hmC near TSSs.

The results of this study highlight that gene regulatory DNA modifications in tissues are detectable in cfDNA of cancer patients. By focusing on genomic regions with low 5(h)mC levels in blood cells, tumor-derived transcription can be inferred at individual genes. To verify the connection between the cell-free (hydroxy-)methylome and tumor transcription, expression profiling from matched tumor tissue could be employed. This would enable to depict transcription specific to the investigated ALK[+] molecular subtype.

4.4 Therapy monitoring using 5(h)mC biomarkers in cfDNA

Several studies have shown the utility of cfDNA-based SNV and CNA detection for therapy monitoring in cancer (Diehl et al. 2008; Abbosh et al. 2017; Dagogo-Jack et al. 2018; Dietz et al. 2020). In some patients, genomic alterations in cfDNA could anticipate disease relapse in advance of radiological assessments, highlighting the potential clinical benefit of treatment surveillance using liquid biopsies (Abbosh et al. 2017; Dietz et al. 2020). Epigenetic alterations occur at high frequencies in most cancers and their abundance translates into an increased sensitivity compared to genomic ctDNA profiling (Kandoth et al. 2013; Li and Zhou 2020). However, in many clinical settings (e.g. MRD monitoring after successful therapeutic intervention or early disease detection) the required sensitivity to detect tumor-derived alterations is not achievable by interrogating genomic changes.

The present study demonstrates that cancer-derived alterations in the cell-free (hydroxy-)methylome are suitable biomarkers for therapy monitoring of metastatic ALK[+] NSCLC patients receiving TKIs. Candidate markers for treatment surveillance were selected from the tumor-enriched D(h)MRs identified by blood cell filtering and differential analysis between patient and control samples. Further, public datasets with tumor tissue methylation and expression were leveraged to identify the most promising candidates associated with LUAD biology. Among these, I focused on D(h)MRs whose 5(h)mC signal could distinguish between cfDNA samples from healthy individuals and NSCLC as well as PD from non-PD patients. These D(h)MRs likely represent prognostic biomarkers suitable for therapy monitoring in longitudinal plasma samples. 5mC markers (n = 21) were highly correlated to tumor-specific genomic alterations (i.e. SNVs, *EML4-ALK* fusion and t-MAD scores) measured in matched cfDNA samples. This confirmed the cancer association of the selected DMRs. The highest correlations were noted when 5mC biomarkers were compared to the t-MAD score and VAF_{max}. Individual mutations (i.e. *ALK* resistance and *TP53* mutations) and the *EML4-ALK* fusion resulted in lower correlations. t-MAD and VAF_{max} are estimators of the total tumor burden within a cfDNA sample (Dietz et al. 2020; Mouliere et al. 2018a;

Velimirovic et al. 2020), supporting a view of cfDNA 5mC markers as a reflection of the entire tumor rather than informing about particular subclones. This is in line with the notion that 5mC alterations occur early during carcinogenesis (Greenberg and Bourc'his 2019). Consequently, methylome perturbations should be present in most (if not all) tumor subclones. The overall highest correlation was observed between the *PTGER4* 5mC levels and the t-MAD score. *PTGER4* hypermethylation in cfDNA was described to allow differentiation of lung cancer from non-malignant diseases (i.e. chronic obstructive pulmonary disease and benign lung lesions) and healthy controls (Schotten et al. 2021; Weiss et al. 2017). Additionally, *PTGER4* methylation was reported to be more sensitive in therapy response prediction compared to established protein markers (i.e. CA125, CEA, Cyfra211, and NSE) (Zhang et al. 2020c), highlighting the coherence of the identified 5mC biomarkers to the recent literature. In contrast to 5mC markers, 5hmC profiling revealed one potentially prognostic marker (*IL1RAP*; intron 3). Most DhMRs, including *IL1RAP*, were not correlated to the co-measured genomic alterations, suggesting a lower tumor association of 5hmC compared to 5mC markers. Only genes with TSS-proximal DhMRs (*KPNA7* and *KIF25*) were significantly associated with at least one of the genomic markers.

The utility for therapy monitoring was evaluated for four 5mC (i.e. *HOXA10-AS, SOX9-AS1, PTGER4,* and *PRAC1*) and three 5hmC (i.e. *IL1RAP, KPNA7* and *KIF25*) biomarkers. 5mC marker abundances in longitudinal plasma samples reflected therapy response in the majority of patients and corresponded with dynamic changes of the co-measured genomic alterations. Radiological disease progression was accompanied by rising 5mC levels and successful therapeutic interventions coincided with decreased marker abundances. Similarly, other studies demonstrated that 5mC markers can predict therapeutic success/failure from serial plasma samples in breast and renal cell carcinoma patients (Lasseter et al. 2020; Moss et al. 2020). Here, the evaluated 5hmC biomarkers presented a lower sensitivity for therapy surveillance, detecting disease progression in fewer cases compared to the 5mC markers. A combination of multiple 5hmC markers might increase the sensitivity and enhance the suitability of 5hmC profiling for treatment monitoring. The feasibility of this approach was shown by Song *et al.*, who used linear discriminant analysis to combine hepatocellular carcinoma-specific 5hmC alterations into a single score, which accurately tracked treatment and response in hepatocellular carcinoma patients following surgical tumor resection (Song et al. 2017). Moreover, increased sequencing depth will likely contribute to the detection of additional 5hmC markers and might enhance sensitivity. We previously demonstrated that genomic ctDNA profiling can reveal disease progression earlier than radiologic assessments (Dietz et al. 2020). This study showed similar results using 5(h)mC markers, with patient P028 representing an illustrative example. Alectinib failure in this patient was indicated by successively rising levels of genomic alterations (i.e. t-MAD, *EML4-ALK* fusion, *ALK* G1202R mutation) in cfDNA before imaging-based determination of PD. Elevated *SOX9-AS1, HOXA10-AS* and *PTGER4* methylation as well as *IL1RAP* hydroxymethylation levels suggested disease progression earlier than genomic ctDNA markers. Furthermore, crizotinib

failure in patients P012 and P044 as well as alectinib failure in P025 could be detected earlier than radiology from 5mC marker abundances. In these patients, epigenetic and genetic alterations detected relapse at the same time point.

Combined, these results illustrate that 5(h)mC alterations in cfDNA are valuable biomarkers for therapy monitoring and in some cases can anticipate tumor relapse earlier than radiological diagnosis. The case of patient P028 showed that 5(h)mC profiling might have sensitivity advantages over the assessment of genomic alterations. However, a systematic evaluation on more samples would be required to confirm this hypothesis. Targeted 5(h)mC sequencing panels could be designed for the identified marker regions to further increase the sensitivity for treatment monitoring. Here, (oxidative) bisulfite and bisulfite-free options have been proposed in literature (Chen et al. 2020c; Liu et al. 2019). Moreover, the integration of multiple 5mC and 5hmC markers together with genomic alterations is likely to further enhance cfDNA-based therapy monitoring.

4.5 Detection of copy number alterations from 5mC data

Somatic CNAs are common in metastatic tumors and their profiling from cfDNA has been applied in various cancer entities (Jensen et al. 2019; Mouliere et al. 2018a). The advancement of statistical and bioinformatic methods allows cost effective cfDNA-based analysis of CNAs from sWGS data (<0.5x genome coverage) (Adalsteinsson et al. 2017; Mouliere et al. 2018a). This study showed that algorithms trained to infer CNAs from sWGS data (i.e. ichorCNA and CNAclinic) are also capable to reliably detect copy number changes from 5mC-enriched sequencing reads (cfMeDIP-seq). Regions with abnormal copy numbers observed in cfMeDIP-seq data were largely coherent with those found by sWGS. Additionally, t-MAD scores of both analysis types were highly correlated. Hence, cfMeDIP-seq allows simultaneous assessment of epigenomic and genomic biomarkers without additional costs. The integration of tumor-derived 5mC alterations and CNAs might contribute to more sensitive cancer detection. In contrast, CNA inference from 5hmC-enriched sequencing data was less accurate: Only large chromosomal regions with copy number gains/losses were detectable and genome-wide chromosomal instability (t-MAD) was underestimated in t-MAD$_{high}$ samples. Song *et al.* similarly demonstrated that hMeSEAL-seq data can detect large CNAs in lung cancer patient cfDNA, albeit at a lower resolution compared to non-enriched WGS data (Song et al. 2017). The applied CNA-calling algorithms are based on depth of coverage and rely on an even distribution of sequencing reads throughout the genome (Adalsteinsson et al. 2017; Mouliere et al. 2018a; Raman et al. 2019). Genome-wide 5hmC abundances are low (0.1 – 1% of cytosines (Brazauskas and Kriaucionis 2014)) and predominantly associated with genes (Song et al. 2017; Zhang et al. 2018b). Additionally, 5hmC quantities were reported to be reduced in lung cancers (Song et al. 2017). Therefore, uniform genome coverage is unlikely to be achieved by hMeSEAL-seq. This might explain

the low resolution of CNA profiling from hMeSEAL-seq data. Although it is also depleted in most cancers (Jones and Baylin 2007; Skvortsova et al. 2019b), 5mC occurs at a higher abundance (3 – 4% of cytosines (Brazauskas and Kriaucionis 2014)) than 5hmC and provides a more accurate reflection of the entire genome. Consequently, the higher CNA-calling rate from cfMeDIP-seq data is expected.

4.6 Conclusion and outlook

The aim of this study was to identify tumor-derived 5(h)mC alterations from cfDNA of ALK[+] NSCLC patients, associate the cancer-specific 5(h)mC biomarkers to gene expression and assess their suitability for tracking cancer dynamics under TKI therapy in longitudinally taken plasma samples.

In summary, the findings of this study propose a ctDNA enrichment strategy that, first, determines and excludes genomic regions highly (hydroxy-)methylated in the non-tumor fraction of cfDNA and, second, performs differential analysis between patient and control samples to further enrich for regions harboring tumor-derived 5(h)mC signals. The comparison to tissue methylation data and hierarchical clustering confirmed the tumor association of the genomic regions remaining following ctDNA enrichment and identified potential 5mC biomarkers for cancer classification. This strategy was applied to facilitate the characterization of cancer-derived 5(h)mC alterations and improve the accuracy of cfDNA-based tumor detection and monitoring. Future studies might benefit from the genome-wide blood cell (hydroxy-)methylome data generated within this project and 5(h)mC profiling of additional cell types contributing to non-tumor cfDNA (e.g. lymphocytes or vascular endothelial cells) will likely improve the identification of cancer-associated 5(h)mC alterations. In particular, the tumor-enriched 5mC profiles were capable of inferring the transcriptional status of individual genes associated with LUAD biology, demonstrating that regulatory 5mC marks are preserved in cfDNA. However, cancer associated processes, such as regulation of signaling pathways, could not be derived in this study. Here, higher sequencing depth – especially for the 5hmC dataset – would increase the sensitivity and likely result in the identification of additional gene regulatory DNA modifications. Moreover, expression data from matched tumor tissue could allow for the determination of transcriptional processes specific to the investigated ALK[+] molecular tumor subtype. This was not feasible with the given reference data derived from LUAD tumors of various molecular drivers. Cancer-specific 5(h)mC biomarkers correlated with genomic alterations measured from the same plasma samples and their quantification in serial plasma samples enabled continuous monitoring of ALK-directed TKI therapy. In some cases, 5(h)mC alterations could be detected prior to clinical progression and, in one patient, 5(h)mC biomarkers were superior to both imaging and genomic alterations in detecting therapy relapse.

In conclusion, this study demonstrates that tumor-derived alterations in cfDNA can be identified from genome-wide 5(h)mC profiles and their assessment in longitudinal plasma samples tracks cancer dynamics during therapy. The high prevalence of 5(h)mC alterations could translate into sensitive tumor monitoring and allow therapy surveillance in individuals without detectable genomic alterations. Further, the tissue specificity and gene regulatory roles of 5(h)mC provide information beyond genomic markers. In the future, the integrated analysis of epigenetic and genomic features retrievable from cfDNA will be crucial to improve cancer detection, especially in challenging clinical scenarios where tissue samples are not available. Hereby, the high specificity of genomic alterations could complement the sensitivity and tissue specificity of epigenomic aberrations.

5 Summary

Targeted therapies improve the prognosis of advanced anaplastic lymphoma kinase-driven non-small cell lung cancer (ALK+ NSCLC) patients. However, clinical courses vary considerably due to acquired drug resistance. Thus, timely detection of treatment failure is crucial to guide subsequent therapies and optimize patient outcome. The analysis of tumor alterations in cell-free DNA (cfDNA) represents a novel approach to monitor cancer dynamics during therapy in longitudinal plasma samples. Besides mutations and copy number alterations, cancer-specific epigenomic changes have emerged as promising biomarkers for cfDNA-based tumor assessment.

This thesis aimed to identify tumor-derived methylation (5mC) and hydroxymethylation (5hmC) alterations in cfDNA of metastatic ALK+ NSCLC patients, associate these epigenetic biomarkers to gene expression in lung cancer and assess their suitability for monitoring of tyrosine kinase inhibitor therapy in serial plasma samples.

To this end, 79 longitudinal plasma samples from 31 patients were collected alongside plasma of 14 healthy individuals. Genome-wide 5mC and 5hmC profiles were generated by cell-free methylation immunoprecipitation and 5hmC selective chemical labeling, followed by sequencing. Additionally, 5(h)mC profiles of primary monocytes, neutrophils and erythroid progenitor cells were prepared using the same methods. These hematopoietic cells constitute the major non-tumor contributors (72.2%) to cfDNA of cancer patients.

A technical novelty of this study was the enrichment for tumor-derived 5(h)mC signals in cfDNA by excluding genomic regions highly (hydroxy-)methylated in the reference blood cell types. Of 9,603,454 300-bp genomic loci, 577,701 (5mC; 6.0%) and 499,681 (5hmC; 5.2%) exhibited low (or no) signal in the profiled blood cells. The blood cell signal-reduced 5mC regions demonstrated an increased correlation to lung cancer tissue methylation (Spearman, $\rho = 0.26$), compared to the entire dataset ($\rho = 0.11$), and revealed cancer- as well as tissue-specific 5mC signals. Cancer *versus* control analysis at the remaining genomic regions identified 5,499 differentially methylated (DMRs) and 495 differentially hydroxymethylated regions (DhMRs). Hierarchical clustering analysis based on the D(h)MRs cleanly separated patient from control samples and clustered patient cfDNA according to the inferred tumor burden within the samples. This suggests that sample separation is primarily driven by tumor-derived signals and confirms that the identified D(h)MRs are enriched for cancer 5(h)mC alterations. DMRs proximal to transcription start sites were enriched at genes downregulated in lung cancer tissue, demonstrating that cancer-specific gene regulatory 5mC marks can be retrieved from cfDNA. Many of these genes (e.g. *GATA4* and *HOXA9*) were previously described to confer tumor suppressive functions in NSCLC. 5(h)mC levels in cfDNA correlated with tumor-derived genomic alterations (e.g. *EML4-ALK* fusion and global chromosomal instability [t-MAD score]) determined in matched plasma samples. The highest correlation was observed between *PTGER4* methylation and t-MAD scores (Pearson, $r = 0.86$). Four 5mC (*SOX9-AS1*, *HOXA10-AS*, *PRAC1*, and *PTGER4*) and three 5hmC

biomarkers (*IL1RAP*, *KPNA7* and *KIF25*) were employed for therapy monitoring in ten patients with available longitudinal samples (≥2). In particular 5mC biomarkers mirrored cancer dynamics found by radiologic imaging and genomic tumor alterations in cfDNA. At four instances, cfDNA 5mC levels anticipated therapy relapse in advance of imaging with a maximum lead time of 481 days. In one patient, both 5mC and 5hmC biomarkers detected disease progression ahead of imaging and genomic alterations in cfDNA, highlighting the sensitivity of 5(h)mC-based tumor assessment.

In conclusion, 5mC and 5hmC profiling from cfDNA provides an opportunity for sensitive cancer detection and therapy monitoring. The tissue-specificity and the regulatory functions of these DNA modifications provide data about the tumors that currently cannot be obtained by copy number or single nucleotide variation profiling.

6 Zusammenfassung

Zielgerichtete Therapien verbessern die Prognose von fortgeschrittenen anaplastische Lymphomkinase-getriebenen nicht-kleinzelligen Bronchialkarzinom (ALK⁺ NSCLC)-Patienten. Dennoch variiert das Ansprechen, aufgrund des Auftretens von Therapieresistenzen während der Behandlung, erheblich. Die zeitnahe Detektion von Therapieversagen ist essentiell, um Folgebehandlungen zu leiten und dadurch den klinischen Ausgang zu optimieren. Die Analyse von tumorassoziierten Veränderungen in zellfreier DNA (cfDNA) bietet die Möglichkeit Tumordynamiken und Therapieansprechen in longitudinal gesammelten Plasmaproben zu überwachen. Neben Mutationen und Kopienzahlveränderungen, stellen epigenetische Veränderungen im Tumor vielversprechende Biomarker für cfDNA-basierte Analysen dar.

Das Ziel dieser Arbeit war die Identifizierung von tumorassoziierten Methylierungs- (5mC) und Hydroxymethylierungsveränderungen (5hmC) in cfDNA von metastasierten ALK⁺ NSCLC Patienten. Zudem wurden diese epigenetischen Marker mit Genexpressionsdaten aus Lungenkrebsgewebe assoziiert und deren Eignung für die Überwachung von Tyrosinkinaseinhibitor-Therapien in seriellen Plasmaproben geprüft.

Zu diesem Zweck wurden 79 longitudinale Plasmaproben von 31 Patienten und 14 Proben von gesunden Spendern gesammelt. Das Erstellen von genomweiten 5mC und 5hmC Profilen erfolgte durch Immunpräzipitation von methylierter cfDNA, sowie chemischer Markierung und anschließender Präzipitation von hydroxymethylierter cfDNA. Die angereicherten cfDNA-Fraktionen wurden anschließend sequenziert. Zudem wurden 5(h)mC-Profile von primären Monozyten, neutrophilen Granulozyten und erythroiden Vorläuferzellen generiert. Diese hämatopoetischen Zelltypen generieren den Großteil (72.2%) der nicht-tumorassoziierten cfDNA in Krebspatienten.

Eine technische Neuheit dieser Studie war die Anreicherung von tumorassoziierten 5(h)mC Veränderungen in cfDNA durch Ausschluss von genomischen Regionen mit hohen (Hydroxy-)methylierungssignalen in den analysierten Blutzelltypen. Von 9.603.454 300-bp Regionen wiesen 577.701 (5mC; 6.0%) und 499.681 (5hmC; 5.2%) geringe (oder keine) Signale in den Blutzellen auf. Im Vergleich zu dem gesamten Datensatz zeigten die blutzellgefilterten Regionen eine erhöhte Korrelation zu Methylierungsdaten aus Lungenkrebsgewebe (Spearman, $\rho = 0.11$ *vs.* $\rho = 0.26$). Zudem wurden krebs- und gewebespezifische 5mC-Veränderungen identifiziert. Die Gegenüberstellung der blutzellgefilterten genomischen Regionen von Patienten- und Kontrollproben resultierte in 5.499 differenziell methylierten (DMRs) und 495 differenziell hydroxymethylierten Regionen (DhMRs). Anhand der D(h)MRs konnte eine eindeutige Separation der Proben von Patienten und gesunden Spendern erzielt werden (hierarchische Clusteranalyse). Patientenproben mit vergleichbarem Tumorgehalt in der cfDNA wiesen die höchste Ähnlichkeit auf. Dies deutet darauf hin, dass vom Tumor stammende 5(h)mC Signale die Probenseparation maßgeblich beeinflussen und die identifizierten D(h)MRs größtenteils tumorassoziiert sind. Die

Anreicherung von DMRs nahe Transkriptionsstartpunkten von Genen mit herunterregulierter Expression in Lungenkrebsgewebe zeigte, dass regulatorische 5mC-Markierungen aus cfDNA abgeleitet werden können. Viele dieser 5mC-regulierten Gene (z.B. *GATA4* oder *HOXA9*) wurden zuvor als Tumor-Suppressoren in NSCLC beschrieben. 5(h)mC-Signale korrelierten mit genomischen Alterationen (z.B. *EML4-ALK* Fusion und globale chromosomale Instabiliät [t-MAD score]), gemessen aus den gleichen Plasmaproben. *PTGER4*-Methylierung und der t-MAD score zeigten die höchste Korrelation (Pearson, $r = 0.86$). Die Eignung von vier 5mC (*SOX9-AS1*, *HOXA10-AS*, *PRAC1*, und *PTGER4*) und drei 5hmC (*IL1RAP*, *KPNA7* und *KIF25*) Biomarkern zur Therapieüberwachung von zehn Patienten mit longitudinalen Plasmaproben (≥2) wurde getestet. Besonders 5mC-Marker spiegelten dynamische Veränderungen des Tumors wider, die bereits durch bildgebende Verfahren und in cfDNA-basierter genomischer Analyse gefunden wurden. In vier Fällen ließen 5mC Signale das Rezidiv des Tumors vor radiologischem Progress erkennen (max. 481 Tage früher). Der Progress eines Patienten wurde sowohl durch 5mC- als auch 5hmC-Biomarker vor der Bildgebung und der genomischen cfDNA Analyse erkannt.

Zusammenfassend wurde gezeigt, dass die Analyse von tumorassoziierten 5mC- und 5hmC-Veränderungen in cfDNA sensitive Möglichkeiten zur Detektion von Tumoren und Therapieüberwachung darstellen. Die Gewebespezifität dieser DNA-Modifikationen sowie deren Rolle in der Regulierung von Genexpression ermöglicht Einblicke in den Tumor, die gegenwärtig durch die Detektion von Kopienzahlveränderungen oder Mutation nicht gewährleistet werden kann.

7 Attributions

High-throughput DNA sequencing was performed by the High-Throughput Sequencing Unit of the German Cancer Research Center core facility.

8 References

Abbosh, C., Birkbak, N. J., Wilson, G. A., Jamal-Hanjani, M., Constantin, T., Salari, R., Le Quesne, J., Moore, D. A., Veeriah, S., Rosenthal, R., Marafioti, T., Kirkizlar, E., Watkins, T. B. K., McGranahan, N., Ward, S., Martinson, L., Riley, J., Fraioli, F., Al Bakir, M., Grönroos, E., Zambrana, F., Endozo, R., Bi, W. L., Fennessy, F. M., Sponer, N., Johnson, D., Laycock, J., Shafi, S., Czyzewska-Khan, J., Rowan, A., Chambers, T., Matthews, N., Turajlic, S., Hiley, C., Lee, S. M., Forster, M. D., Ahmad, T., Falzon, M., Borg, E., Lawrence, D., Hayward, M., Kolvekar, S., Panagiotopoulos, N., Janes, S. M., Thakrar, R., Ahmed, A., Blackhall, F., Summers, Y., Hafez, D., Naik, A., Ganguly, A., Kareht, S., Shah, R., Joseph, L., Marie Quinn, A., Crosbie, P. A., Naidu, B., Middleton, G., Langman, G., Trotter, S., Nicolson, M., Remmen, H., Kerr, K., Chetty, M., Gomersall, L., Fennell, D. A., Nakas, A., Rathinam, S., Anand, G., Khan, S., Russell, P., Ezhil, V., Ismail, B., Irvin-Sellers, M., Prakash, V., Lester, J. F., Kornaszewska, M., Attanoos, R., Adams, H., Davies, H., Oukrif, D., Akarca, A. U., Hartley, J. A., Lowe, H. L., Lock, S., Iles, N., Bell, H., Ngai, Y., Elgar, G., Szallasi, Z., Schwarz, R. F., Herrero, J., Stewart, A., Quezada, S. A., Peggs, K. S., Van Loo, P., Dive, C., Lin, C. J., Rabinowitz, M., Aerts, H., Hackshaw, A., Shaw, J. A., Zimmermann, B. G. and Swanton, C. (2017). **Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution**. Nature *545 (7655)*, 446-451, doi: 10.1038/nature22364.

Ackermann, C. J., Stock, G., Tay, R., Dawod, M., Gomes, F. and Califano, R. (2019). **Targeted Therapy For RET-Rearranged Non-Small Cell Lung Cancer: Clinical Development And Future Directions**. Onco Targets Ther *12*, 7857-7864, doi: 10.2147/ott.S171665.

Adalsteinsson, V. A., Ha, G., Freeman, S. S., Choudhury, A. D., Stover, D. G., Parsons, H. A., Gydush, G., Reed, S. C., Rotem, D., Rhoades, J., Loginov, D., Livitz, D., Rosebrock, D., Leshchiner, I., Kim, J., Stewart, C., Rosenberg, M., Francis, J. M., Zhang, C. Z., Cohen, O., Oh, C., Ding, H., Polak, P., Lloyd, M., Mahmud, S., Helvie, K., Merrill, M. S., Santiago, R. A., O'Connor, E. P., Jeong, S. H., Leeson, R., Barry, R. M., Kramkowski, J. F., Zhang, Z., Polacek, L., Lohr, J. G., Schleicher, M., Lipscomb, E., Saltzman, A., Oliver, N. M., Marini, L., Waks, A. G., Harshman, L. C., Tolaney, S. M., Van Allen, E. M., Winer, E. P., Lin, N. U., Nakabayashi, M., Taplin, M. E., Johannessen, C. M., Garraway, L. A., Golub, T. R., Boehm, J. S., Wagle, N., Getz, G., Love, J. C. and Meyerson, M. (2017). **Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors**. Nat Commun *8 (1)*, 1324, doi: 10.1038/s41467-017-00965-y.

Aguet, F., Brown, A. A., Castel, S. E., Davis, J. R., He, Y., Jo, B., Mohammadi, P., Park, Y., Parsana, P., Segrè, A. V., Strober, B. J., Zappala, Z., Cummings, B. B., Gelfand, E. T., Hadley, K., Huang, K. H., Lek, M., Li, X., Nedzel, J. L., Nguyen, D. Y., Noble, M. S., Sullivan, T. J., Tukiainen, T., MacArthur, D. G., Getz, G., Addington, A., Guan, P., Koester, S., Little, A. R., Lockhart, N. C., Moore, H. M., Rao, A., Struewing, J. P., Volpi, S., Brigham, L. E., Hasz, R., Hunter, M., Johns, C., Johnson, M., Kopen, G., Leinweber, W. F., Lonsdale, J. T., McDonald, A., Mestichelli, B., Myer, K., Roe, B., Salvatore, M., Shad, S., Thomas, J. A., Walters, G., Washington, M., Wheeler, J., Bridge, J., Foster, B. A., Gillard, B. M., Karasik, E., Kumar, R., Miklos, M., Moser, M. T., Jewell, S. D., Montroy, R. G., Rohrer, D. C., Valley, D., Mash, D. C., Davis, D. A., Sobin, L., Barcus, M. E., Branton, P. A., Abell, N. S., Balliu, B., Delaneau, O., Frésard, L., Gamazon, E. R., Garrido-Martín, D., Gewirtz, A. D. H., Gliner, G., Gloudemans, M. J., Han, B., He, A. Z., Hormozdiari, F., Li, X., Liu, B., Kang, E. Y., McDowell, I. C., Ongen, H., Palowitch, J. J., Peterson, C. B., Quon, G., Ripke, S., Saha, A., Shabalin, A. A., Shimko, T. C., Sul, J. H., Teran, N. A., Tsang, E. K., Zhang, H., Zhou, Y.-H., Bustamante, C. D., Cox, N. J., Guigó, R., Kellis, M., McCarthy, M. I., Conrad, D. F., Eskin, E., Li, G., Nobel, A. B., Sabatti, C., Stranger, B. E., Wen, X., Wright, F. A., Ardlie, K. G., Dermitzakis, E. T., Lappalainen, T., Aguet, F., Ardlie, K. G., Cummings, B. B., Gelfand, E. T., Getz, G., Hadley, K., Handsaker, R. E., Huang, K. H., Kashin, S., Karczewski, K. J., Lek, M., Li, X., MacArthur, D. G., Nedzel, J. L., Nguyen, D. T., Noble, M. S., Segrè, A. V., Trowbridge, C. A., Tukiainen, T., Abell, N. S., Balliu, B., Barshir, R., Basha, O., Battle, A., Bogu, G. K., Brown, A., Brown, C. D., Castel, S. E., Chen, L. S., Chiang, C., Conrad, D. F., Cox, N. J., Damani, F. N., Davis, J. R., Delaneau, O., Dermitzakis, E. T., Engelhardt, B. E., Eskin, E., Ferreira, P. G., Frésard, L., Gamazon, E. R., Garrido-Martín, D., Gewirtz, A. D. H., Gliner, G., Gloudemans, M. J., Guigo, R., Hall, I. M., Han, B., He, Y., Hormozdiari, F., Howald, C., Kyung Im, H., Jo, B., Yong Kang, E., Kim, Y., Kim-Hellmuth, S., Lappalainen, T., Li, G., Li, X., Liu, B., Mangul, S., McCarthy, M. I., McDowell, I. C., Mohammadi, P., Monlong, J., Montgomery, S. B., Muñoz-Aguirre, M., Ndungu, A. W., Nicolae, D. L., Nobel, A. B., Oliva, M., Ongen, H., Palowitch, J. J., Panousis, N., Papasaikas, P., Park, Y., Parsana, P., Payne, A. J., Peterson, C. B., Quan, J., Reverter, F., Sabatti, C., Saha, A., Sammeth, M., Scott, A. J., Shabalin, A. A., Sodaei, R., Stephens, M., Stranger, B. E., Strober, B. J., Sul, J. H., Tsang, E. K., Urbut, S., van de Bunt, M., Wang, G., Wen, X., Wright, F. A., Xi, H. S., Yeger-Lotem, E., Zappala, Z., Zaugg, J. B., Zhou, Y.-H., Akey, J. M., Bates, D., Chan, J., Chen, L. S., Claussnitzer, M., Demanelis, K., Diegel, M., Doherty, J. A., Feinberg, A. P., Fernando, M. S., Halow, J., Hansen, K. D., Haugen, E., Hickey, P. F., Hou, L., Jasmine, F., Jian, R., Jiang, L., Johnson, A., Kaul, R., Kellis, M., Kibriya,

M. G., Lee, K., Billy Li, J., Li, Q., Li, X., Lin, J., Lin, S., Linder, S., Linke, C., Liu, Y., Maurano, M. T., Molinie, B., Montgomery, S. B., Nelson, J., Neri, F. J., Oliva, M., Park, Y., Pierce, B. L., Rinaldi, N. J., Rizzardi, L. F., Sandstrom, R., Skol, A., Smith, K. S., Snyder, M. P., Stamatoyannopoulos, J., Stranger, B. E., Tang, H., Tsang, E. K., Wang, L., Wang, M., Van Wittenberghe, N., Wu, F., Zhang, R., Nierras, C. R., Branton, P. A., Carithers, L. J., Guan, P., Moore, H. M., Rao, A., Vaught, J. B., Gould, S. E., Lockart, N. C., Martin, C., Struewing, J. P., Volpi, S., Addington, A. M., Koester, S. E., Little, A. R., Consortium, G. T., Lead, a., Laboratory, D. A., Coordinating, C., management, N. I. H. p., Biospecimen, c., Pathology, e, Q. T. L. m. w. g., Laboratory, D. A., Coordinating Center —Analysis Working, G., Statistical Methods groups— Analysis Working, G., Enhancing, G. g., Fund, N. I. H. C., Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida and Biospecimen Collection Source Site—, N. (2017). **Genetic effects on gene expression across human tissues**. Nature *550 (7675)*, 204-213, doi: 10.1038/nature24277.

Ahrendt, S. A., Chow, J. T., Xu, L. H., Yang, S. C., Eisenberger, C. F., Esteller, M., Herman, J. G., Wu, L., Decker, P. A., Jen, J. and Sidransky, D. (1999). **Molecular detection of tumor cells in bronchoalveolar lavage fluid from patients with early stage lung cancer**. J Natl Cancer Inst *91 (4)*, 332-339, doi: 10.1093/jnci/91.4.332.

Akiyama, Y., Watkins, N., Suzuki, H., Jair, K. W., van Engeland, M., Esteller, M., Sakai, H., Ren, C. Y., Yuasa, Y., Herman, J. G. and Baylin, S. B. (2003). **GATA-4 and GATA-5 transcription factor genes and potential downstream antitumor target genes are epigenetically silenced in colorectal and gastric cancer**. Mol Cell Biol *23 (23)*, 8429-8439, doi: 10.1128/mcb.23.23.8429-8439.2003.

Alberg, A. J., Brock, M. V., Ford, J. G., Samet, J. M. and Spivack, S. D. (2013). **Epidemiology of Lung Cancer: Diagnosis and Management of Lung Cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines**. CHEST *143 (5)*, e1S-e29S, doi: 10.1378/chest.12-2345.

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A. J. R., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Børresen-Dale, A.-L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjörd, J. E., Foekens, J. A., Greaves, M., Hosoda, F., Hutter, B., Ilicic, T., Imbeaud, S., Imielinski, M., Jäger, N., Jones, D. T. W., Jones, D., Knappskog, S., Kool, M., Lakhani, S. R., López-Otín, C., Martin, S., Munshi, N. C., Nakamura, H., Northcott, P. A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J. V., Puente, X. S., Raine, K., Ramakrishna, M., Richardson, A. L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T. N., Span, P. N., Teague, J. W., Totoki, Y., Tutt, A. N. J., Valdés-Mas, R., van Buuren, M. M., van 't Veer, L., Vincent-Salomon, A., Waddell, N., Yates, L. R., Australian Pancreatic Cancer Genome, I., Consortium, I. B. C., Consortium, I. M.-S., PedBrain, I., Zucman-Rossi, J., Futreal, P. A., McDermott, U., Lichter, P., Meyerson, M., Grimmond, S. M., Siebert, R., Campo, E., Shibata, T., Pfister, S. M., Campbell, P. J. and Stratton, M. R. (2013). **Signatures of mutational processes in human cancer**. Nature *500 (7463)*, 415-421, doi: 10.1038/nature12477.

Ali, S. M., Hensing, T., Schrock, A. B., Allen, J., Sanford, E., Gowen, K., Kulkarni, A., He, J., Suh, J. H., Lipson, D., Elvin, J. A., Yelensky, R., Chalmers, Z., Chmielecki, J., Peled, N., Klempner, S. J., Firozvi, K., Frampton, G. M., Molina, J. R., Menon, S., Brahmer, J. R., MacMahon, H., Nowak, J., Ou, S. H., Zauderer, M., Ladanyi, M., Zakowski, M., Fischbach, N., Ross, J. S., Stephens, P. J., Miller, V. A., Wakelee, H., Ganesan, S. and Salgia, R. (2016). **Comprehensive Genomic Profiling Identifies a Subset of Crizotinib-Responsive ALK-Rearranged Non-Small Cell Lung Cancer Not Detected by Fluorescence In Situ Hybridization**. Oncologist *21 (6)*, 762-770, doi: 10.1634/theoncologist.2015-0497.

Alix-Panabières, C. and Pantel, K. (2021). **Liquid Biopsy: From Discovery to Clinical Application**. Cancer Discov *11 (4)*, 858-873, doi: 10.1158/2159-8290.Cd-20-1311.

Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C. and Wingett, S. (2012). **FastQC: a quality control tool for high throughput sequence data** (

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D. and Irizarry, R. A. (2014). **Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays**. Bioinformatics *30 (10)*, 1363-1369, doi: 10.1093/bioinformatics/btu049.

Asada, K., Kobayashi, K., Joutard, S., Tubaki, M., Takahashi, S., Takasawa, K., Komatsu, M., Kaneko, S., Sese, J. and Hamamoto, R. (2020). **Uncovering Prognosis-Related Genes and Pathways by Multi-Omics Analysis in Lung Cancer**. Biomolecules *10 (4)*, doi: 10.3390/biom10040524.

Ashouri, A., Sayin, V. I., Van den Eynden, J., Singh, S. X., Papagiannakopoulos, T. and Larsson, E. (2016). **Pan-cancer transcriptomic analysis associates long non-coding RNAs with key mutational driver events**. Nat Commun *7*, 13197, doi: 10.1038/ncomms13197.

References

Awad, M. M., Oxnard, G. R., Jackman, D. M., Savukoski, D. O., Hall, D., Shivdasani, P., Heng, J. C., Dahlberg, S. E., Jänne, P. A., Verma, S., Christensen, J., Hammerman, P. S. and Sholl, L. M. (2016). **MET Exon 14 Mutations in Non-Small-Cell Lung Cancer Are Associated With Advanced Age and Stage-Dependent MET Genomic Amplification and c-Met Overexpression**. J Clin Oncol 34 (7), 721-730, doi: 10.1200/jco.2015.63.4600.

Azzi, A., Dallmann, R., Casserly, A., Rehrauer, H., Patrignani, A., Maier, B., Kramer, A. and Brown, S. A. (2014). **Circadian behavior is light-reprogrammed by plastic DNA methylation**. Nat Neurosci 17 (3), 377-382, doi: 10.1038/nn.3651.

Bachman, M., Uribe-Lewis, S., Yang, X., Williams, M., Murrell, A. and Balasubramanian, S. (2014). **5-Hydroxymethylcytosine is a predominantly stable DNA modification**. Nat Chem 6 (12), 1049-1055, doi: 10.1038/nchem.2064.

Bai, Y., Akiyama, Y., Nagasaki, H., Yagi, O. K., Kikuchi, Y., Saito, N., Takeshita, K., Iwai, T. and Yuasa, Y. (2000). **Distinct expression of CDX2 and GATA4/5, development-related genes, in human gastric cancer cell lines**. Mol Carcinog 28 (3), 184-188, doi: 10.1002/1098-2744(200007)28:3<184::aid-mc7>3.0.co;2-6.

Bai, Y., Lang, L., Zhao, W. and Niu, R. (2019). **Long Non-Coding RNA HOXA11-AS Promotes Non-Small Cell Lung Cancer Tumorigenesis Through microRNA-148a-3p/DNMT1 Regulatory Axis**. Onco Targets Ther 12, 11195-11206, doi: 10.2147/ott.S198367.

Barter, M. J., Gomez, R., Hyatt, S., Cheung, K., Skelton, A. J., Xu, Y., Clark, I. M. and Young, D. A. (2017). **The long non-coding RNA ROCR contributes to SOX9 expression and chondrogenic differentiation of human mesenchymal stem cells**. Development 144 (24), 4510-4521, doi: 10.1242/dev.152504.

Baylin, S. B. and Jones, P. A. (2011). **A decade of exploring the cancer epigenome — biological and translational implications**. Nature Reviews Cancer 11 (10), 726-734, doi: 10.1038/nrc3130.

Baylin, S. B. and Ohm, J. E. (2006). **Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction?** Nat Rev Cancer 6 (2), 107-116, doi: 10.1038/nrc1799.

Beaver, J. A., Jelovac, D., Balukrishna, S., Cochran, R., Croessmann, S., Zabransky, D. J., Wong, H. Y., Toro, P. V., Cidado, J., Blair, B. G., Chu, D., Burns, T., Higgins, M. J., Stearns, V., Jacobs, L., Habibi, M., Lange, J., Hurley, P. J., Lauring, J., VanDenBerg, D., Kessler, J., Jeter, S., Samuels, M. L., Maar, D., Cope, L., Cimino-Mathews, A., Argani, P., Wolff, A. C. and Park, B. H. (2014). **Detection of cancer DNA in plasma of patients with early-stage breast cancer**. Clin Cancer Res 20 (10), 2643-2650, doi: 10.1158/1078-0432.Ccr-13-2933.

Belinsky, S. A., Nikula, K. J., Palmisano, W. A., Michels, R., Saccomanno, G., Gabrielson, E., Baylin, S. B. and Herman, J. G. (1998). **Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis**. Proc Natl Acad Sci U S A 95 (20), 11891-11896, doi: 10.1073/pnas.95.20.11891.

Bergamaschi, A., Ning, Y., Ku, C.-J., Ellison, C., Collin, F., Guler, G., Phillips, T., McCarthy, E., Wang, W., Antoine, M., Scott, A., Lloyd, P., Ashworth, A., Quake, S. and Levy, S. (2020). **Pilot study demonstrating changes in DNA hydroxymethylation enable detection of multiple cancers in plasma cell-free DNA**. medRxiv, 2020.2001.2022.20018382, doi: 10.1101/2020.01.22.20018382.

Bergethon, K., Shaw, A. T., Ou, S. H., Katayama, R., Lovly, C. M., McDonald, N. T., Massion, P. P., Siwak-Tapp, C., Gonzalez, A., Fang, R., Mark, E. J., Batten, J. M., Chen, H., Wilner, K. D., Kwak, E. L., Clark, J. W., Carbone, D. P., Ji, H., Engelman, J. A., Mino-Kenudson, M., Pao, W. and Iafrate, A. J. (2012). **ROS1 rearrangements define a unique molecular class of lung cancers**. J Clin Oncol 30 (8), 863-870, doi: 10.1200/jco.2011.35.6345.

Berrondo, C., Flax, J., Kucherov, V., Siebert, A., Osinski, T., Rosenberg, A., Fucile, C., Richheimer, S. and Beckham, C. J. (2016). **Expression of the Long Non-Coding RNA HOTAIR Correlates with Disease Progression in Bladder Cancer and Is Contained in Bladder Cancer Patient Urinary Exosomes**. PLoS One 11 (1), e0147236, doi: 10.1371/journal.pone.0147236.

Best, M. G., Sol, N., Kooi, I., Tannous, J., Westerman, B. A., Rustenburg, F., Schellen, P., Verschueren, H., Post, E., Koster, J., Ylstra, B., Ameziane, N., Dorsman, J., Smit, E. F., Verheul, H. M., Noske, D. P.,

Reijneveld, J. C., Nilsson, R. J. A., Tannous, B. A., Wesseling, P. and Wurdinger, T. (2015). **RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics**. Cancer Cell *28 (5)*, 666-676, doi: 10.1016/j.ccell.2015.09.018.

Best, M. G., Wesseling, P. and Wurdinger, T. (2018). **Tumor-Educated Platelets as a Noninvasive Biomarker Source for Cancer Detection and Progression Monitoring**. Cancer Res *78 (13)*, 3407-3412, doi: 10.1158/0008-5472.Can-18-0887.

Bettegowda, C., Sausen, M., Leary, R. J., Kinde, I., Wang, Y., Agrawal, N., Bartlett, B. R., Wang, H., Luber, B., Alani, R. M., Antonarakis, E. S., Azad, N. S., Bardelli, A., Brem, H., Cameron, J. L., Lee, C. C., Fecher, L. A., Gallia, G. L., Gibbs, P., Le, D., Giuntoli, R. L., Goggins, M., Hogarty, M. D., Holdhoff, M., Hong, S. M., Jiao, Y., Juhl, H. H., Kim, J. J., Siravegna, G., Laheru, D. A., Lauricella, C., Lim, M., Lipson, E. J., Marie, S. K., Netto, G. J., Oliner, K. S., Olivi, A., Olsson, L., Riggins, G. J., Sartore-Bianchi, A., Schmidt, K., Shih l, M., Oba-Shinjo, S. M., Siena, S., Theodorescu, D., Tie, J., Harkins, T. T., Veronese, S., Wang, T. L., Weingart, J. D., Wolfgang, C. L., Wood, L. D., Xing, D., Hruban, R. H., Wu, J., Allen, P. J., Schmidt, C. M., Choti, M. A., Velculescu, V. E., Kinzler, K. W., Vogelstein, B., Papadopoulos, N. and Diaz, L. A., Jr. (2014). **Detection of circulating tumor DNA in early- and late-stage human malignancies**. Sci Transl Med *6 (224)*, 224ra224, doi: 10.1126/scitranslmed.3007094.

Bostick, M., Kim, J. K., Estève, P. O., Clark, A., Pradhan, S. and Jacobsen, S. E. (2007). **UHRF1 plays a role in maintaining DNA methylation in mammalian cells**. Science *317 (5845)*, 1760-1764, doi: 10.1126/science.1147939.

Botezatu, I., Serdyuk, O., Potapova, G., Shelepov, V., Alechina, R., Molyaka, Y., Ananév, V., Bazin, I., Garin, A., Narimanov, M., Knysh, V., Melkonyan, H., Umansky, S. and Lichtenstein, A. (2000). **Genetic analysis of DNA excreted in urine: a new approach for detecting specific genomic DNA sequences from cells dying in an organism**. Clin Chem *46 (8 Pt 1)*, 1078-1084.

Brahmer, J., Reckamp, K. L., Baas, P., Crinò, L., Eberhardt, W. E., Poddubskaya, E., Antonia, S., Pluzanski, A., Vokes, E. E., Holgado, E., Waterhouse, D., Ready, N., Gainor, J., Arén Frontera, O., Havel, L., Steins, M., Garassino, M. C., Aerts, J. G., Domine, M., Paz-Ares, L., Reck, M., Baudelet, C., Harbison, C. T., Lestini, B. and Spigel, D. R. (2015). **Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer**. N Engl J Med *373 (2)*, 123-135, doi: 10.1056/NEJMoa1504627.

Brandeis, M., Ariel, M. and Cedar, H. (1993). **Dynamics of DNA methylation during development**. Bioessays *15 (11)*, 709-713, doi: 10.1002/bies.950151103.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. and Jemal, A. (2018). **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries**. CA: A Cancer Journal for Clinicians *68 (6)*, 394-424, doi: 10.3322/caac.21492.

Brazauskas, P. and Kriaucionis, S. (2014). **DNA modifications: Another stable base in DNA**. Nat Chem *6 (12)*, 1031-1033, doi: 10.1038/nchem.2115.

Burkhardt, M., Mayordomo, E., Winzer, K. J., Fritzsche, F., Gansukh, T., Pahl, S., Weichert, W., Denkert, C., Guski, H., Dietel, M. and Kristiansen, G. (2006). **Cytoplasmic overexpression of ALCAM is prognostic of disease progression in breast cancer**. J Clin Pathol *59 (4)*, 403-409, doi: 10.1136/jcp.2005.028209.

Camidge, D. R., Kim, H. R., Ahn, M. J., Yang, J. C. H., Han, J. Y., Hochmair, M. J., Lee, K. H., Delmonte, A., García Campelo, M. R., Kim, D. W., Griesinger, F., Felip, E., Califano, R., Spira, A., Gettinger, S. N., Tiseo, M., Lin, H. M., Gupta, N., Hanley, M. J., Ni, Q., Zhang, P. and Popat, S. (2020). **Brigatinib Versus Crizotinib in Advanced ALK Inhibitor-Naive ALK-Positive Non-Small Cell Lung Cancer: Second Interim Analysis of the Phase III ALTA-1L Trial**. J Clin Oncol *38 (31)*, 3592-3603, doi: 10.1200/jco.20.00505.

Camidge, D. R., Pao, W. and Sequist, L. V. (2014). **Acquired resistance to TKIs in solid tumours: learning from lung cancer**. Nat Rev Clin Oncol *11 (8)*, 473-481, doi: 10.1038/nrclinonc.2014.104.

Campbell, J. D., Alexandrov, A., Kim, J., Wala, J., Berger, A. H., Pedamallu, C. S., Shukla, S. A., Guo, G., Brooks, A. N., Murray, B. A., Imielinski, M., Hu, X., Ling, S., Akbani, R., Rosenberg, M., Cibulskis, C., Ramachandran, A., Collisson, E. A., Kwiatkowski, D. J., Lawrence, M. S., Weinstein, J. N., Verhaak, R. G. W., Wu, C. J., Hammerman, P. S., Cherniack, A. D., Getz, G., Cancer Genome Atlas Research, N., Artyomov, M. N., Schreiber, R., Govindan, R. and Meyerson, M. (2016). **Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas**. Nature genetics *48 (6)*, 607-616, doi: 10.1038/ng.3564.

References

Cao, L., Wang, S., Zhang, Y., Wong, K. C., Nakatsu, G., Wang, X., Wong, S., Ji, J. and Yu, J. (2018). **Zinc-finger protein 471 suppresses gastric cancer through transcriptionally repressing downstream oncogenic PLS3 and TFAP2A**. Oncogene *37 (26)*, 3601-3616, doi: 10.1038/s41388-018-0220-5.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C. and Schultz, N. (2012). **The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data**. Cancer Discov *2 (5)*, 401-404, doi: 10.1158/2159-8290.Cd-12-0095.

Chan, K. C. A., Jiang, P., Sun, K., Cheng, Y. K. Y., Tong, Y. K., Cheng, S. H., Wong, A. I. C., Hudecova, I., Leung, T. Y., Chiu, R. W. K. and Lo, Y. M. D. (2016). **Second generation noninvasive fetal genome analysis reveals de novo mutations, single-base parental inheritance, and preferred DNA ends**. Proceedings of the National Academy of Sciences *113 (50)*, E8159-E8168, doi: 10.1073/pnas.1615800113.

Chaudhuri, A. A., Chabon, J. J., Lovejoy, A. F., Newman, A. M., Stehr, H., Azad, T. D., Khodadoust, M. S., Esfahani, M. S., Liu, C. L., Zhou, L., Scherer, F., Kurtz, D. M., Say, C., Carter, J. N., Merriott, D. J., Dudley, J. C., Binkley, M. S., Modlin, L., Padda, S. K., Gensheimer, M. F., West, R. B., Shrager, J. B., Neal, J. W., Wakelee, H. A., Loo, B. W., Jr., Alizadeh, A. A. and Diehn, M. (2017). **Early Detection of Molecular Residual Disease in Localized Lung Cancer by Circulating Tumor DNA Profiling**. Cancer Discov *7 (12)*, 1394-1403, doi: 10.1158/2159-8290.Cd-17-0716.

Chen, C., Huang, X., Yin, W., Peng, M., Wu, F., Wu, X., Tang, J., Chen, M., Wang, X., Hulbert, A., Brock, M. V., Liu, W., Herman, J. G. and Yu, F. (2020a). **Ultrasensitive DNA hypermethylation detection using plasma for early detection of NSCLC: a study in Chinese patients with very small nodules**. Clinical Epigenetics *12 (1)*, 39, doi: 10.1186/s13148-020-00828-2.

Chen, H. and Boutros, P. C. (2011). **VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R**. BMC Bioinformatics *12 (1)*, 35, doi: 10.1186/1471-2105-12-35.

Chen, H. A., Kuo, T. C., Tseng, C. F., Ma, J. T., Yang, S. T., Yen, C. J., Yang, C. Y., Sung, S. Y. and Su, J. L. (2016). **Angiopoietin-like protein 1 antagonizes MET receptor activity to repress sorafenib resistance and cancer stemness in hepatocellular carcinoma**. Hepatology *64 (5)*, 1637-1651, doi: 10.1002/hep.28773.

Chen, T. J., Gao, F., Yang, T., Li, H., Li, Y., Ren, H. and Chen, M. W. (2020b). **LncRNA HOTAIRM1 Inhibits the Proliferation and Invasion of Lung Adenocarcinoma Cells via the miR-498/WWOX Axis**. Cancer Manag Res *12*, 4379-4390, doi: 10.2147/cmar.S244573.

Chen, X., Gole, J., Gore, A., He, Q., Lu, M., Min, J., Yuan, Z., Yang, X., Jiang, Y., Zhang, T., Suo, C., Li, X., Cheng, L., Zhang, Z., Niu, H., Li, Z., Xie, Z., Shi, H., Zhang, X., Fan, M., Wang, X., Yang, Y., Dang, J., McConnell, C., Zhang, J., Wang, J., Yu, S., Ye, W., Gao, Y., Zhang, K., Liu, R. and Jin, L. (2020c). **Non-invasive early detection of cancer four years before conventional diagnosis using a blood test**. Nature Communications *11 (1)*, 3475, doi: 10.1038/s41467-020-17316-z.

Chen, Y., Löhr, M. and Jesnowski, R. (2010). **Inhibition of ankyrin-B expression reduces growth and invasion of human pancreatic ductal adenocarcinoma**. Pancreatology *10 (5)*, 586-596, doi: 10.1159/000308821.

Chen, Y., Yang, L., Cui, T., Pacyna-Gengelbach, M. and Petersen, I. (2015). **HOPX is methylated and exerts tumour-suppressive function through Ras-induced senescence in human lung cancer**. J Pathol *235 (3)*, 397-407, doi: 10.1002/path.4469.

Chen, Y. H., Hancock, B. A., Solzak, J. P., Brinza, D., Scafe, C., Miller, K. D. and Radovich, M. (2017). **Next-generation sequencing of circulating tumor DNA to predict recurrence in triple-negative breast cancer patients with residual disease after neoadjuvant chemotherapy**. NPJ Breast Cancer *3*, 24, doi: 10.1038/s41523-017-0028-4.

Chiarle, R., Voena, C., Ambrogio, C., Piva, R. and Inghirami, G. (2008). **The anaplastic lymphoma kinase in the pathogenesis of cancer**. Nat Rev Cancer *8 (1)*, 11-23, doi: 10.1038/nrc2291.

Childress, M. A., Himmelberg, S. M., Chen, H., Deng, W., Davies, M. A. and Lovly, C. M. (2018). **ALK Fusion Partners Impact Response to ALK Inhibition: Differential Effects on Sensitivity, Cellular Phenotypes, and Biochemical Properties**. Mol Cancer Res *16 (11)*, 1724-1736, doi: 10.1158/1541-7786.Mcr-18-0171.

Choi, Y. L., Lira, M. E., Hong, M., Kim, R. N., Choi, S. J., Song, J. Y., Pandy, K., Mann, D. L., Stahl, J. A., Peckham, H. E., Zheng, Z., Han, J., Mao, M. and Kim, J. (2014). **A novel fusion of TPR and ALK in lung adenocarcinoma**. J Thorac Oncol 9 *(4)*, 563-566, doi: 10.1097/jto.0000000000000093.

Christopoulos, P., Endris, V., Bozorgmehr, F., Elsayed, M., Kirchner, M., Ristau, J., Buchhalter, I., Penzel, R., Herth, F. J., Heussel, C. P., Eichhorn, M., Muley, T., Meister, M., Fischer, J. R., Rieken, S., Warth, A., Bischoff, H., Schirmacher, P., Stenzinger, A. and Thomas, M. (2018). **EML4-ALK fusion variant V3 is a high-risk feature conferring accelerated metastatic spread, early treatment failure and worse overall survival in ALK(+) non-small cell lung cancer**. Int J Cancer *142 (12)*, 2589-2598, doi: 10.1002/ijc.31275.

Christopoulos, P., Kirchner, M., Bozorgmehr, F., Endris, V., Elsayed, M., Budczies, J., Ristau, J., Penzel, R., Herth, F. J., Heussel, C. P., Eichhorn, M., Muley, T., Meister, M., Fischer, J. R., Rieken, S., Lasitschka, F., Bischoff, H., Sotillo, R., Schirmacher, P., Thomas, M. and Stenzinger, A. (2019). **Identification of a highly lethal V3(+) TP53(+) subset in ALK(+) lung adenocarcinoma**. Int J Cancer *144 (1)*, 190-199, doi: 10.1002/ijc.31893.

Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M. and Zaret, K. S. (2002). **Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4**. Mol Cell 9 *(2)*, 279-289, doi: 10.1016/s1097-2765(02)00459-8.

Collisson, E. A., Campbell, J. D., Brooks, A. N., Berger, A. H., Lee, W., Chmielecki, J., Beer, D. G., Cope, L., Creighton, C. J., Danilova, L., Ding, L., Getz, G., Hammerman, P. S., Neil Hayes, D., Hernandez, B., Herman, J. G., Heymach, J. V., Jurisica, I., Kucherlapati, R., Kwiatkowski, D., Ladanyi, M., Robertson, G., Schultz, N., Shen, R., Sinha, R., Sougnez, C., Tsao, M.-S., Travis, W. D., Weinstein, J. N., Wigle, D. A., Wilkerson, M. D., Chu, A., Cherniack, A. D., Hadjipanayis, A., Rosenberg, M., Weisenberger, D. J., Laird, P. W., Radenbaugh, A., Ma, S., Stuart, J. M., Averett Byers, L., Baylin, S. B., Govindan, R., Meyerson, M., Rosenberg, M., Gabriel, S. B., Cibulskis, K., Sougnez, C., Kim, J., Stewart, C., Lichtenstein, L., Lander, E. S., Lawrence, M. S., Getz, G., Kandoth, C., Fulton, R., Fulton, L. L., McLellan, M. D., Wilson, R. K., Ye, K., Fronick, C. C., Maher, C. A., Miller, C. A., Wendl, M. C., Cabanski, C., Ding, L., Mardis, E., Govindan, R., Creighton, C. J., Wheeler, D., Balasundaram, M., Butterfield, Y. S. N., Carlsen, R., Chu, A., Chuah, E., Dhalla, N., Guin, R., Hirst, C., Lee, D., Li, H. I., Mayo, M., Moore, R. A., Mungall, A. J., Schein, J. E., Sipahimalani, P., Tam, A., Varhol, R., Gordon Robertson, A., Wye, N., Thiessen, N., Holt, R. A., Jones, S. J. M., Marra, M. A., Campbell, J. D., Brooks, A. N., Chmielecki, J., Imielinski, M., Onofrio, R. C., Hodis, E., Zack, T., Sougnez, C., Helman, E., Sekhar Pedamallu, C., Mesirov, J., Cherniack, A. D., Saksena, G., Schumacher, S. E., Carter, S. L., Hernandez, B., Garraway, L., Beroukhim, R., Gabriel, S. B., Getz, G., Meyerson, M., Hadjipanayis, A., Lee, S., Mahadeshwar, H. S., Pantazi, A., Protopopov, A., Ren, X., Seth, S., Song, X., Tang, J., Yang, L., Zhang, J., Chen, P.-C., Parfenov, M., Wei Xu, A., Santoso, N., Chin, L., Park, P. J., Kucherlapati, R., Hoadley, K. A., Todd Auman, J., Meng, S., Shi, Y., Buda, E., Waring, S., Veluvolu, U., Tan, D., Mieczkowski, P. A., Jones, C. D., Simons, J. V., Soloway, M. G., Bodenheimer, T., Jefferys, S. R., Roach, J., Hoyle, A. P., Wu, J., Balu, S., Singh, D., Prins, J. F., Marron, J. S., Parker, J. S., Neil Hayes, D., Perou, C. M., Liu, J., Cope, L., Danilova, L., Weisenberger, D. J., Maglinte, D. T., Lai, P. H., Bootwalla, M. S., Van Den Berg, D. J., Triche Jr, T., Baylin, S. B., Laird, P. W., Rosenberg, M., Chin, L., Zhang, J., Cho, J., DiCara, D., Heiman, D., Lin, P., Mallard, W., Voet, D., Zhang, H., Zou, L., Noble, M. S., Lawrence, M. S., Saksena, G., Gehlenborg, N., Thorvaldsdottir, H., Mesirov, J., Nazaire, M.-D., Robinson, J., Getz, G., Lee, W., Arman Aksoy, B., Ciriello, G., Taylor, B. S., Dresdner, G., Gao, J., Gross, B., Seshan, V. E., Ladanyi, M., Reva, B., Sinha, R., Onur Sumer, S., Weinhold, N., Schultz, N., Shen, R., Sander, C., Ng, S., Ma, S., Zhu, J., Radenbaugh, A., Stuart, J. M., Benz, C. C., Yau, C., Haussler, D., Spellman, P. T., Wilkerson, M. D., Parker, J. S., Hoadley, K. A., Kimes, P. K., Neil Hayes, D., Perou, C. M., Broom, B. M., Wang, J., Lu, Y., Kwok Shing Ng, P., Diao, L., Averett Byers, L., Liu, W., Heymach, J. V., Amos, C. I., Weinstein, J. N., Akbani, R., Mills, G. B., Curley, E., Paulauskis, J., Lau, K., Morris, S., Shelton, T., Mallery, D., Gardner, J., Penny, R., Saller, C., Tarvin, K., Richards, W. G., Cerfolio, R., Bryant, A., Raymond, D. P., Pennell, N. A., Farver, C., Czerwinski, C., Huelsenbeck-Dill, L., Iacocca, M., Petrelli, N., Rabeno, B., Brown, J., Bauer, T., Dolzhanskiy, O., Potapova, O., Rotin, D., Voronina, O., Nemirovich-Danchenko, E., Fedosenko, K. V., Gal, A., Behera, M., Ramalingam, S. S., Sica, G., Flieder, D., Boyd, J., Weaver, J., Kohl, B., Huy Quoc Thinh, D., Sandusky, G., Juhl, H., The Cancer Genome Atlas Research, N., Disease analysis working, g., Genome sequencing centres: The, E., Edythe, L. B. I., Washington University in St, L., Baylor College of, M., Genome characterization centres: Canada's Michael Smith Genome Sciences Centre, B. C. C. A., The, E., Edythe, L. B. I., Harvard Medical, S. B., Women's Hospital, M. D. A. C. C., University of North Carolina, C. H., University of, K., The, U. S. C. J. H. U. E. C. C., Genome data analysis centres: The, E.,

References

Edythe, L. B. I., Memorial Sloan-Kettering Cancer, C., University of California, S. C. B. I., Oregon, H., Sciences, U., The University of Texas, M. D. A. C. C., Biospecimen core resource: International Genomics, C., Tissue source sites: Analytical Biological Service, I., Brigham, Women's, H., University of Alabama at, B., Cleveland, C., Christiana, C., Cureline, Emory, U., Fox Chase Cancer, C., Ilsbio, Indiana, U., Indivumed and John Flynn, H. (2014). **Comprehensive molecular profiling of lung adenocarcinoma**. Nature *511 (7511)*, 543-550, doi: 10.1038/nature13385.

Dagogo-Jack, I., Brannon, A. R., Ferris, L. A., Campbell, C. D., Lin, J. J., Schultz, K. R., Ackil, J., Stevens, S., Dardaei, L., Yoda, S., Hubbeling, H., Digumarthy, S. R., Riester, M., Hata, A. N., Sequist, L. V., Lennes, I. T., Iafrate, A. J., Heist, R. S., Azzoli, C. G., Farago, A. F., Engelman, J. A., Lennerz, J. K., Benes, C. H., Leary, R. J., Shaw, A. T. and Gainor, J. F. (2018). **Tracking the Evolution of Resistance to ALK Tyrosine Kinase Inhibitors through Longitudinal Analysis of Circulating Tumor DNA**. JCO Precis Oncol *2018*, doi: 10.1200/po.17.00160.

Dawson, S.-J., Tsui, D. W. Y., Murtaza, M., Biggs, H., Rueda, O. M., Chin, S.-F., Dunning, M. J., Gale, D., Forshew, T., Mahler-Araujo, B., Rajan, S., Humphray, S., Becq, J., Halsall, D., Wallis, M., Bentley, D., Caldas, C. and Rosenfeld, N. (2013). **Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer**. New England Journal of Medicine *368 (13)*, 1199-1209, doi: 10.1056/NEJMoa1213261.

de Koning, H. J., van der Aalst, C. M., de Jong, P. A., Scholten, E. T., Nackaerts, K., Heuvelmans, M. A., Lammers, J.-W. J., Weenink, C., Yousaf-Khan, U., Horeweg, N., van 't Westeinde, S., Prokop, M., Mali, W. P., Mohamed Hoesein, F. A. A., van Ooijen, P. M. A., Aerts, J. G. J. V., den Bakker, M. A., Thunnissen, E., Verschakelen, J., Vliegenthart, R., Walter, J. E., ten Haaf, K., Groen, H. J. M. and Oudkerk, M. (2020). **Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial**. New England Journal of Medicine *382 (6)*, 503-513, doi: 10.1056/NEJMoa1911793.

Diehl, F., Schmidt, K., Choti, M. A., Romans, K., Goodman, S., Li, M., Thornton, K., Agrawal, N., Sokoll, L., Szabo, S. A., Kinzler, K. W., Vogelstein, B. and Diaz Jr, L. A. (2008). **Circulating mutant DNA to assess tumor dynamics**. Nature Medicine *14 (9)*, 985-990, doi: 10.1038/nm.1789.

Dietz, S., Christopoulos, P., Yuan, Z., Angeles, A. K., Gu, L., Volckmar, A.-L., Ogrodnik, S. J., Janke, F., Dalle Fratte, C. and Zemojtel, T. (2020). **Longitudinal therapy monitoring of ALK-positive lung cancer by combined copy number and targeted mutation profiling of cell-free DNA**. EBioMedicine 62, 103103.

Dor, Y. and Cedar, H. (2018). **Principles of DNA methylation and their implications for biology and medicine**. Lancet *392 (10149)*, 777-786, doi: 10.1016/s0140-6736(18)31268-6.

Duruisseaux, M., Besse, B., Cadranel, J., Pérol, M., Mennecier, B., Bigay-Game, L., Descourt, R., Dansin, E., Audigier-Valette, C., Moreau, L., Hureaux, J., Veillon, R., Otto, J., Madroszyk-Flandin, A., Cortot, A., Guichard, F., Boudou-Rouquette, P., Langlais, A., Missy, P., Morin, F. and Moro-Sibilot, D. (2017). **Overall survival with crizotinib and next-generation ALK inhibitors in ALK-positive non-small-cell lung cancer (IFCT-1302 CLINALK): a French nationwide cohort retrospective study**. Oncotarget *8 (13)*, 21903-21917, doi: 10.18632/oncotarget.15746.

Duruisseaux, M. and Esteller, M. (2018). **Lung cancer epigenetics: From knowledge to applications**. Semin Cancer Biol *51*, 116-128, doi: 10.1016/j.semcancer.2017.09.005.

Eden, S. and Cedar, H. (1994). **Role of DNA methylation in the regulation of transcription**. Curr Opin Genet Dev *4 (2)*, 255-259, doi: 10.1016/s0959-437x(05)80052-8.

Ehrlich, M. and Ehrlich, K. C. (2014). **DNA cytosine methylation and hydroxymethylation at the borders**. Epigenomics *6 (6)*, 563-566, doi: 10.2217/epi.14.48.

Ernst, J. and Kellis, M. (2012). **ChromHMM: automating chromatin-state discovery and characterization**. Nat Methods *9 (3)*, 215-216, doi: 10.1038/nmeth.1906.

Excellence;, N. I. f. H. a. C. (2019). **Lung cancer: diagnosis and management (NICE Guideline 122)**.

Faryna, M., Konermann, C., Aulmann, S., Bermejo, J. L., Brugger, M., Diederichs, S., Rom, J., Weichenhan, D., Claus, R., Rehli, M., Schirmacher, P., Sinn, H. P., Plass, C. and Gerhauser, C. (2012). **Genome-wide methylation screen in low-grade breast cancer identifies novel epigenetically altered genes as potential biomarkers for tumor diagnosis**. Faseb j *26 (12)*, 4937-4950, doi: 10.1096/fj.12-209502.

Feinberg, A. P. and Vogelstein, B. (1983). **Hypomethylation distinguishes genes of some human cancers from their normal counterparts**. Nature *301 (5895)*, 89-92, doi: 10.1038/301089a0.

Field, J. K., Duffy, S. W., Baldwin, D. R., Whynes, D. K., Devaraj, A., Brain, K. E., Eisen, T., Gosney, J., Green, B. A., Holemans, J. A., Kavanagh, T., Kerr, K. M., Ledson, M., Lifford, K. J., McRonald, F. E., Nair, A., Page, R. D., Parmar, M. K. B., Rassl, D. M., Rintoul, R. C., Screaton, N. J., Wald, N. J., Weller, D., Williamson, P. R., Yadegarfar, G. and Hansell, D. M. (2016). **UK Lung Cancer RCT Pilot Screening Trial: baseline findings from the screening arm provide evidence for the potential implementation of lung cancer screening**. Thorax *71 (2)*, 161-170, doi: 10.1136/thoraxjnl-2015-207140.

Forloni, M., Gupta, R., Nagarajan, A., Sun, L. S., Dong, Y., Pirazzoli, V., Toki, M., Wurtz, A., Melnick, M. A., Kobayashi, S., Homer, R. J., Rimm, D. L., Gettinger, S. J., Politi, K., Dogra, S. K. and Wajapeyee, N. (2016). **Oncogenic EGFR Represses the TET1 DNA Demethylase to Induce Silencing of Tumor Suppressors in Cancer Cells**. Cell Rep *16 (2)*, 457-471, doi: 10.1016/j.celrep.2016.05.087.

Frampton, G. M., Ali, S. M., Rosenzweig, M., Chmielecki, J., Lu, X., Bauer, T. M., Akimov, M., Bufill, J. A., Lee, C., Jentz, D., Hoover, R., Ou, S. H., Salgia, R., Brennan, T., Chalmers, Z. R., Jaeger, S., Huang, A., Elvin, J. A., Erlich, R., Fichtenholtz, A., Gowen, K. A., Greenbowe, J., Johnson, A., Khaira, D., McMahon, C., Sanford, E. M., Roels, S., White, J., Greshock, J., Schlegel, R., Lipson, D., Yelensky, R., Morosini, D., Ross, J. S., Collisson, E., Peters, M., Stephens, P. J. and Miller, V. A. (2015). **Activation of MET via diverse exon 14 splicing alterations occurs in multiple tumor types and confers clinical sensitivity to MET inhibitors**. Cancer Discov *5 (8)*, 850-859, doi: 10.1158/2159-8290.Cd-15-0285.

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I. T., García Girón, C., Gonzalez, J. M., Grego, T., Hardy, M., Hourlier, T., Hunt, T., Izuogu, O. G., Lagarde, J., Martin, F. J., Martínez, L., Mohanan, S., Muir, P., Navarro, F. C. P., Parker, A., Pei, B., Pozo, F., Ruffier, M., Schmitt, B. M., Stapleton, E., Suner, M.-M., Sycheva, I., Uszczynska-Ratajczak, B., Xu, J., Yates, A., Zerbino, D., Zhang, Y., Aken, B., Choudhary, J. S., Gerstein, M., Guigó, R., Hubbard, T. J. P., Kellis, M., Paten, B., Reymond, A., Tress, M. L. and Flicek, P. (2019). **GENCODE reference annotation for the human and mouse genomes**. Nucleic acids research *47 (D1)*, D766-D773, doi: 10.1093/nar/gky955.

Gainor, J. F., Dardaei, L., Yoda, S., Friboulet, L., Leshchiner, I., Katayama, R., Dagogo-Jack, I., Gadgeel, S., Schultz, K., Singh, M., Chin, E., Parks, M., Lee, D., DiCecca, R. H., Lockerman, E., Huynh, T., Logan, J., Ritterhouse, L. L., Le, L. P., Muniappan, A., Digumarthy, S., Channick, C., Keyes, C., Getz, G., Dias-Santagata, D., Heist, R. S., Lennerz, J., Sequist, L. V., Benes, C. H., Iafrate, A. J., Mino-Kenudson, M., Engelman, J. A. and Shaw, A. T. (2016). **Molecular Mechanisms of Resistance to First- and Second-Generation ALK Inhibitors in ALK-Rearranged Lung Cancer**. Cancer Discov 6 *(10)*, 1118-1133, doi: 10.1158/2159-8290.Cd-16-0596.

Galluzzi, L., Vitale, I., Aaronson, S. A., Abrams, J. M., Adam, D., Agostinis, P., Alnemri, E. S., Altucci, L., Amelio, I., Andrews, D. W., Annicchiarico-Petruzzelli, M., Antonov, A. V., Arama, E., Baehrecke, E. H., Barlev, N. A., Bazan, N. G., Bernassola, F., Bertrand, M. J. M., Bianchi, K., Blagosklonny, M. V., Blomgren, K., Borner, C., Boya, P., Brenner, C., Campanella, M., Candi, E., Carmona-Gutierrez, D., Cecconi, F., Chan, F. K. M., Chandel, N. S., Cheng, E. H., Chipuk, J. E., Cidlowski, J. A., Ciechanover, A., Cohen, G. M., Conrad, M., Cubillos-Ruiz, J. R., Czabotar, P. E., D'Angiolella, V., Dawson, T. M., Dawson, V. L., De Laurenzi, V., De Maria, R., Debatin, K.-M., DeBerardinis, R. J., Deshmukh, M., Di Daniele, N., Di Virgilio, F., Dixit, V. M., Dixon, S. J., Duckett, C. S., Dynlacht, B. D., El-Deiry, W. S., Elrod, J. W., Fimia, G. M., Fulda, S., García-Sáez, A. J., Garg, A. D., Garrido, C., Gavathiotis, E., Golstein, P., Gottlieb, E., Green, D. R., Greene, L. A., Gronemeyer, H., Gross, A., Hajnoczky, G., Hardwick, J. M., Harris, I. S., Hengartner, M. O., Hetz, C., Ichijo, H., Jäättelä, M., Joseph, B., Jost, P. J., Juin, P. P., Kaiser, W. J., Karin, M., Kaufmann, T., Kepp, O., Kimchi, A., Kitsis, R. N., Klionsky, D. J., Knight, R. A., Kumar, S., Lee, S. W., Lemasters, J. J., Levine, B., Linkermann, A., Lipton, S. A., Lockshin, R. A., López-Otín, C., Lowe, S. W., Luedde, T., Lugli, E., MacFarlane, M., Madeo, F., Malewicz, M., Malorni, W., Manic, G., Marine, J.-C., Martin, S. J., Martinou, J.-C., Medema, J. P., Mehlen, P., Meier, P., Melino, S., Miao, E. A., Molkentin, J. D., Moll, U. M., Muñoz-Pinedo, C., Nagata, S., Nuñez, G., Oberst, A., Oren, M., Overholtzer, M., Pagano, M., Panaretakis, T., Pasparakis, M., Penninger, J. M., Pereira, D. M., Pervaiz, S., Peter, M. E., Piacentini, M., Pinton, P., Prehn, J. H. M., Puthalakath, H., Rabinovich, G. A., Rehm, M., Rizzuto, R., Rodrigues, C. M. P., Rubinsztein, D. C., Rudel, T., Ryan, K. M., Sayan, E., Scorrano, L., Shao, F., Shi, Y., Silke, J., Simon, H.-U., Sistigu, A., Stockwell, B. R., Strasser, A., Szabadkai, G., Tait, S. W. G., Tang, D., Tavernarakis, N., Thorburn, A., Tsujimoto, Y., Turk, B., Vanden Berghe, T., Vandenabeele, P., Vander Heiden, M. G., Villunger, A., Virgin, H. W., Vousden, K. H., Vucic, D., Wagner, E. F., Walczak, H., Wallach, D., Wang, Y., Wells, J. A., Wood, W., Yuan, J., Zakeri, Z., Zhivotovsky, B., Zitvogel, L.,

References

Melino, G. and Kroemer, G. (2018). **Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018**. Cell Death & Differentiation *25 (3)*, 486-541, doi: 10.1038/s41418-017-0012-4.

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C. and Schultz, N. (2013). **Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal**. Sci Signal *6 (269)*, pl1, doi: 10.1126/scisignal.2004088.

Gao, L., Hu, Y., Tian, Y., Fan, Z., Wang, K., Li, H., Zhou, Q., Zeng, G., Hu, X., Yu, L., Zhou, S., Tong, X., Huang, H., Chen, H., Liu, Q., Liu, W., Zhang, G., Zeng, M., Zhou, G., He, Q., Ji, H. and Chen, L. (2019). **Lung cancer deficient in the tumor suppressor GATA4 is sensitive to TGFBR1 inhibition**. Nat Commun *10 (1)*, 1665, doi: 10.1038/s41467-019-09295-7.

Garcia-Romero, N., Esteban-Rubio, S., Rackov, G., Carrión-Navarro, J., Belda-Iniesta, C. and Ayuso-Sacido, A. (2018). **Extracellular vesicles compartment in liquid biopsies: Clinical application**. Molecular Aspects of Medicine *60*, 27-37, doi: https://doi.org/10.1016/j.mam.2017.11.009.

Gel, B., Díez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M. A. and Malinverni, R. (2015). **regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests**. Bioinformatics *32 (2)*, 289-291, doi: 10.1093/bioinformatics/btv562.

Gerlinger, M., Rowan, A. J., Horswell, S., Math, M., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., Varela, I., Phillimore, B., Begum, S., McDonald, N. Q., Butler, A., Jones, D., Raine, K., Latimer, C., Santos, C. R., Nohadani, M., Eklund, A. C., Spencer-Dene, B., Clark, G., Pickering, L., Stamp, G., Gore, M., Szallasi, Z., Downward, J., Futreal, P. A. and Swanton, C. (2012). **Intratumor heterogeneity and branched evolution revealed by multiregion sequencing**. N Engl J Med *366 (10)*, 883-892, doi: 10.1056/NEJMoa1113205.

Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W. J. and Nekrutenko, A. (2005). **Galaxy: a platform for interactive large-scale genome analysis**. Genome Res *15 (10)*, 1451-1455, doi: 10.1101/gr.4086505.

Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M. and Carell, T. (2010). **Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates**. PLoS One *5 (12)*, e15367, doi: 10.1371/journal.pone.0015367.

Goldstraw, P., Chansky, K., Crowley, J., Rami-Porta, R., Asamura, H., Eberhardt, W. E., Nicholson, A. G., Groome, P., Mitchell, A. and Bolejack, V. (2016). **The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer**. J Thorac Oncol *11 (1)*, 39-51, doi: 10.1016/j.jtho.2015.09.009.

Gontarz, P., Fu, S., Xing, X., Liu, S., Miao, B., Bazylianska, V., Sharma, A., Madden, P., Cates, K., Yoo, A., Moszczynska, A., Wang, T. and Zhang, B. (2020). **Comparison of differential accessibility analysis strategies for ATAC-seq data**. Scientific Reports *10 (1)*, 10150, doi: 10.1038/s41598-020-66998-4.

González-Masiá, J. A., García-Olmo, D. and García-Olmo, D. C. (2013). **Circulating nucleic acids in plasma and serum (CNAPS): applications in oncology**. Onco Targets Ther *6*, 819-832, doi: 10.2147/ott.S44668.

Greenberg, M. V. C. and Bourc'his, D. (2019). **The diverse roles of DNA methylation in mammalian development and disease**. Nat Rev Mol Cell Biol *20 (10)*, 590-607, doi: 10.1038/s41580-019-0159-6.

Grossman, R. L., Heath, A. P., Ferretti, V., Varmus, H. E., Lowy, D. R., Kibbe, W. A. and Staudt, L. M. (2016). **Toward a Shared Vision for Cancer Genomic Data**. New England Journal of Medicine *375 (12)*, 1109-1112, doi: 10.1056/NEJMp1607591.

Guan, R., Hong, W., Huang, J., Peng, T., Zhao, Z., Lin, Y., Yu, M. and Jian, Z. (2020). **The expression and prognostic value of GLYATL1 and its potential role in hepatocellular carcinoma**. J Gastrointest Oncol *11 (6)*, 1305-1321, doi: 10.21037/jgo-20-186.

Guler, G. D., Ning, Y., Ku, C. J., Phillips, T., McCarthy, E., Ellison, C. K., Bergamaschi, A., Collin, F., Lloyd, P., Scott, A., Antoine, M., Wang, W., Chau, K., Ashworth, A., Quake, S. R. and Levy, S. (2020). **Detection**

of early stage pancreatic cancer using 5-hydroxymethylcytosine signatures in circulating cell free DNA. Nat Commun 11 (1), 5270, doi: 10.1038/s41467-020-18965-w.

Guo, M., Akiyama, Y., House, M. G., Hooker, C. M., Heath, E., Gabrielson, E., Yang, S. C., Han, Y., Baylin, S. B., Herman, J. G. and Brock, M. V. (2004). **Hypermethylation of the GATA genes in lung cancer**. Clin Cancer Res 10 (23), 7917-7924, doi: 10.1158/1078-0432.Ccr-04-1140.

Guo, M., Zhou, X., Han, X., Zhang, Y. and Jiang, L. (2019). **SPINK1 is a prognosis predicting factor of non-small cell lung cancer and regulates redox homeostasis**. Oncol Lett 18 (6), 6899-6908, doi: 10.3892/ol.2019.11005.

Guo, S., Diep, D., Plongthongkum, N., Fung, H. L., Zhang, K. and Zhang, K. (2017). **Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA**. Nat Genet 49 (4), 635-642, doi: 10.1038/ng.3805.

Guo, Y. Z., Xie, X. L., Fu, J. and Xing, G. L. (2018). **SOX9 regulated proliferation and apoptosis of human lung carcinoma cells by the Wnt/β-catenin signaling pathway**. Eur Rev Med Pharmacol Sci 22 (15), 4898-4907, doi: 10.26355/eurrev_201808_15626.

Haffner, M. C., Chaux, A., Meeker, A. K., Esopi, D. M., Gerber, J., Pellakuru, L. G., Toubaji, A., Argani, P., Iacobuzio-Donahue, C., Nelson, W. G., Netto, G. J., De Marzo, A. M. and Yegnasubramanian, S. (2011). **Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers**. Oncotarget 2 (8), 627-637, doi: 10.18632/oncotarget.316.

Hansen, K. D., Timp, W., Bravo, H. C., Sabunciyan, S., Langmead, B., McDonald, O. G., Wen, B., Wu, H., Liu, Y., Diep, D., Briem, E., Zhang, K., Irizarry, R. A. and Feinberg, A. P. (2011). **Increased methylation variation in epigenetic domains across cancer types**. Nat Genet 43 (8), 768-775, doi: 10.1038/ng.865.

Harada, H., Miyamoto, K., Yamashita, Y., Taniyama, K., Mihara, K., Nishimura, M. and Okada, M. (2015). **Prognostic signature of protocadherin 10 methylation in curatively resected pathological stage I non-small-cell lung cancer**. Cancer Med 4 (10), 1536-1546, doi: 10.1002/cam4.507.

Hays, J. T. and Ebbert, J. O. (2008). **Varenicline for Tobacco Dependence**. New England Journal of Medicine 359 (19), 2018-2024, doi: 10.1056/NEJMct0800146.

He, S., Ma, X., Zheng, N., Wang, G., Wang, M., Xia, W. and Yu, D. (2021). **PRDM14 mediates chemosensitivity and glycolysis in drug-resistant A549/cisplatin cells and their progenitor A549 human lung adenocarcinoma cells**. Mol Med Rep 23 (2), doi: 10.3892/mmr.2020.11788.

He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C. X., Zhang, K., He, C. and Xu, G. L. (2011). **Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA**. Science 333 (6047), 1303-1307, doi: 10.1126/science.1210944.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. and Glass, C. K. (2010). **Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities**. Mol Cell 38 (4), 576-589, doi: 10.1016/j.molcel.2010.05.004.

Hirsch, F. R., Suda, K., Wiens, J. and Bunn, P. A., Jr. (2016). **New and emerging targeted treatments in advanced non-small-cell lung cancer**. Lancet 388 (10048), 1012-1024, doi: 10.1016/s0140-6736(16)31473-8.

Hong Yan, C. Z., Tingting Liu, Jianzhong Liu (2020). **Mifepristone regulates the multidrug resistance via miR-758/MEPE of Hela cell lines**. European Journal of Gynaecological Oncology 41 (2), 214-220, doi: 10.31083/j.ejgo.2020.02.5052.

Howlader, N., Noone, A., Krapcho, M., Miller, D., Brest, A., Yu, M., Ruhl, J., Tatalovich, Z., Mariotto, A., Lewis, D., Chen, H., Feuer, E. and Cronin, K. (2017). **SEER Cancer Statistics Review 1975-2017**. URL: https://seer.cancer.gov/csr/1975_2017/ [as of17.04.2020].

Hu, T., Yang, H. and Han, Z. G. (2015). **PDZRN4 acts as a suppressor of cell proliferation in human liver cancer cell lines**. Cell Biochem Funct 33 (7), 443-449, doi: 10.1002/cbf.3130.

References

Hu, W., Yang, Y., Li, X., Huang, M., Xu, F., Ge, W., Zhang, S. and Zheng, S. (2018). **Multi-omics Approach Reveals Distinct Differences in Left- and Right-Sided Colon Cancer**. Mol Cancer Res *16 (3)*, 476-485, doi: 10.1158/1541-7786.Mcr-17-0483.

Huisgen, R., Grashey, R. and Steingruber, E. (1963). **Azomethin-ylide und ihre 1.3-dipolaren cycloadditionen**. Tetrahedron Letters *4 (22)*, 1441-1445, doi: https://doi.org/10.1016/S0040-4039(01)90848-2.

Hwang, J. A., Lee, B. B., Kim, Y., Hong, S. H., Kim, Y. H., Han, J., Shim, Y. M., Yoon, C. Y., Lee, Y. S. and Kim, D. H. (2015). **HOXA9 inhibits migration of lung cancer cells and its hypermethylation is associated with recurrence in non-small cell lung cancer**. Mol Carcinog *54 Suppl 1*, E72-80, doi: 10.1002/mc.22180.

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C. and Zhang, Y. (2011). **Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine**. Science *333 (6047)*, 1300-1303, doi: 10.1126/science.1210597.

Jahr, S., Hentze, H., Englisch, S., Hardt, D., Fackelmayer, F. O., Hesch, R. D. and Knippers, R. (2001). **DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells**. Cancer Res *61 (4)*, 1659-1665.

Jensen, T. J., Goodman, A. M., Kato, S., Ellison, C. K., Daniels, G. A., Kim, L., Nakashe, P., McCarthy, E., Mazloom, A. R., McLennan, G., Grosu, D. S., Ehrich, M. and Kurzrock, R. (2019). **Genome-Wide Sequencing of Cell-Free DNA Identifies Copy-Number Alterations That Can Be Used for Monitoring Response to Immunotherapy in Cancer Patients**. Mol Cancer Ther *18 (2)*, 448-458, doi: 10.1158/1535-7163.Mct-18-0535.

Jeppesen, D. K., Fenix, A. M., Franklin, J. L., Higginbotham, J. N., Zhang, Q., Zimmerman, L. J., Liebler, D. C., Ping, J., Liu, Q., Evans, R., Fissell, W. H., Patton, J. G., Rome, L. H., Burnette, D. T. and Coffey, R. J. (2019). **Reassessment of Exosome Composition**. Cell *177 (2)*, 428-445.e418, doi: https://doi.org/10.1016/j.cell.2019.02.029.

Jiang, P., Chan, C. W. M., Chan, K. C. A., Cheng, S. H., Wong, J., Wong, V. W.-S., Wong, G. L. H., Chan, S. L., Mok, T. S. K., Chan, H. L. Y., Lai, P. B. S., Chiu, R. W. K. and Lo, Y. M. D. (2015). **Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients**. Proceedings of the National Academy of Sciences *112 (11)*, E1317-E1325, doi: 10.1073/pnas.1500076112.

Jiang, P., Sun, K., Peng, W., Cheng, S. H., Ni, M., Yeung, P. C., Heung, M. M. S., Xie, T., Shang, H., Zhou, Z., Chan, R. W. Y., Wong, J., Wong, V. W. S., Poon, L. C., Leung, T. Y., Lam, W. K. J., Chan, J. Y. K., Chan, H. L. Y., Chan, K. C. A., Chiu, R. W. K. and Lo, Y. M. D. (2020). **Plasma DNA End-Motif Profiling as a Fragmentomic Marker in Cancer, Pregnancy, and Transplantation**. Cancer Discov *10 (5)*, 664-673, doi: 10.1158/2159-8290.Cd-19-0622.

Joehanes, R., Just, A. C., Marioni, R. E., Pilling, L. C., Reynolds, L. M., Mandaviya, P. R., Guan, W., Xu, T., Elks, C. E., Aslibekyan, S., Moreno-Macias, H., Smith, J. A., Brody, J. A., Dhingra, R., Yousefi, P., Pankow, J. S., Kunze, S., Shah, S. H., McRae, A. F., Lohman, K., Sha, J., Absher, D. M., Ferrucci, L., Zhao, W., Demerath, E. W., Bressler, J., Grove, M. L., Huan, T., Liu, C., Mendelson, M. M., Yao, C., Kiel, D. P., Peters, A., Wang-Sattler, R., Visscher, P. M., Wray, N. R., Starr, J. M., Ding, J., Rodriguez, C. J., Wareham, N. J., Irvin, M. R., Zhi, D., Barrdahl, M., Vineis, P., Ambatipudi, S., Uitterlinden, A. G., Hofman, A., Schwartz, J., Colicino, E., Hou, L., Vokonas, P. S., Hernandez, D. G., Singleton, A. B., Bandinelli, S., Turner, S. T., Ware, E. B., Smith, A. K., Klengel, T., Binder, E. B., Psaty, B. M., Taylor, K. D., Gharib, S. A., Swenson, B. R., Liang, L., DeMeo, D. L., O'Connor, G. T., Herceg, Z., Ressler, K. J., Conneely, K. N., Sotoodehnia, N., Kardia, S. L., Melzer, D., Baccarelli, A. A., van Meurs, J. B., Romieu, I., Arnett, D. K., Ong, K. K., Liu, Y., Waldenberger, M., Deary, I. J., Fornage, M., Levy, D. and London, S. J. (2016). **Epigenetic Signatures of Cigarette Smoking**. Circ Cardiovasc Genet *9 (5)*, 436-447, doi: 10.1161/circgenetics.116.001506.

Jones, P. A. and Baylin, S. B. (2007). **The epigenomics of cancer**. Cell *128 (4)*, 683-692, doi: 10.1016/j.cell.2007.01.029.

Jordan, E. J., Kim, H. R., Arcila, M. E., Barron, D., Chakravarty, D., Gao, J., Chang, M. T., Ni, A., Kundra, R., Jonsson, P., Jayakumaran, G., Gao, S. P., Johnsen, H. C., Hanrahan, A. J., Zehir, A., Rekhtman, N., Ginsberg, M. S., Li, B. T., Yu, H. A., Paik, P. K., Drilon, A., Hellmann, M. D., Reales, D. N., Benayed, R., Rusch, V. W., Kris, M. G., Chaft, J. E., Baselga, J., Taylor, B. S., Schultz, N., Rudin, C. M., Hyman, D. M.,

Berger, M. F., Solit, D. B., Ladanyi, M. and Riely, G. J. (2017). **Prospective Comprehensive Molecular Characterization of Lung Adenocarcinomas for Efficient Patient Matching to Approved and Emerging Therapies**. Cancer Discov 7 (6), 596-609, doi: 10.1158/2159-8290.Cd-16-1337.

Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J. F., Wyczalkowski, M. A., Leiserson, M. D. M., Miller, C. A., Welch, J. S., Walter, M. J., Wendl, M. C., Ley, T. J., Wilson, R. K., Raphael, B. J. and Ding, L. (2013). **Mutational landscape and significance across 12 major cancer types**. Nature 502 (7471), 333-339, doi: 10.1038/nature12634.

Khalil, A., Dekmak, B., Boulos, F., Kantrowitz, J., Spira, A., Fujimoto, J., Kadara, H., El-Hachem, N. and Nemer, G. (2018). **Transcriptomic Alterations in Lung Adenocarcinoma Unveil New Mechanisms Targeted by the TBX2 Subfamily of Tumor Suppressor Genes**. Front Oncol 8, 482, doi: 10.3389/fonc.2018.00482.

Khier, S. and Lohan, L. (2018). **Kinetics of circulating cell-free DNA for biomedical applications: critical appraisal of the literature**. Future science 4 (4), FSO295.

Kim, H., Kwon, Y. M., Kim, J. S., Lee, H., Park, J. H., Shim, Y. M., Han, J., Park, J. and Kim, D. H. (2004). **Tumor-specific methylation in bronchial lavage for the early detection of non-small-cell lung cancer**. J Clin Oncol 22 (12), 2363-2370, doi: 10.1200/jco.2004.10.077.

Kiselev, Y., Andersen, S., Johannessen, C., Fjukstad, B., Standahl Olsen, K., Stenvold, H., Al-Saad, S., Donnem, T., Richardsen, E., Bremnes, R. M. and Rasmussen Busund, L. T. (2018). **Transcription factor PAX6 as a novel prognostic factor and putative tumour suppressor in non-small cell lung cancer**. Sci Rep 8 (1), 5059, doi: 10.1038/s41598-018-23417-z.

Knight, L. J., Burrage, J., Bujac, S. R., Haggerty, C., Graham, A., Gibson, N. J., Ellison, G., Growcott, J. W., Brooks, A. N., Hughes, A. M., Xinarianos, G., Nikolaidis, G., Field, J. K. and Liloglou, T. (2009). **Epigenetic silencing of the endothelin-B receptor gene in non-small cell lung cancer**. Int J Oncol 34 (2), 465-471.

Koc, A., Goksel, T., Pelit, L., Korba, K., Dizdas, T. N., Baysal, E., Uzun, U. C., Kaya, O. O., Ozyilmaz, B., Kutbay, Y. B., Ozdemir, T. R., Kirbiyik, O., Erdogan, K. M., Guvenc, M. S., Kocal, G. C. and Basbinar, Y. (2019). **cfDNA in exhaled breath condensate (EBC) and contamination by ambient air: toward volatile biopsies**. J Breath Res 13 (3), 036006, doi: 10.1088/1752-7163/ab17ff.

Kolde, R. (2013). **pheatmap: Pretty Heatmaps. R package version 1.0.12**.

Kron, A., Alidousty, C., Scheffler, M., Merkelbach-Bruse, S., Seidel, D., Riedel, R., Ihle, M. A., Michels, S., Nogova, L., Fassunke, J., Heydt, C., Kron, F., Ueckeroth, F., Serke, M., Krüger, S., Grohe, C., Koschel, D., Benedikter, J., Kaminsky, B., Schaaf, B., Braess, J., Sebastian, M., Kambartel, K. O., Thomas, R., Zander, T., Schultheis, A. M., Büttner, R. and Wolf, J. (2018). **Impact of TP53 mutation status on systemic treatment outcome in ALK-rearranged non-small-cell lung cancer**. Ann Oncol 29 (10), 2068-2075, doi: 10.1093/annonc/mdy333.

Kumegawa, K., Maruyama, R., Yamamoto, E., Ashida, M., Kitajima, H., Tsuyada, A., Niinuma, T., Kai, M., Yamano, H.-o., Sugai, T., Tokino, T., Shinomura, Y., Imai, K. and Suzuki, H. (2016). **A genomic screen for long noncoding RNA genes epigenetically silenced by aberrant DNA methylation in colorectal cancer**. Scientific Reports 6 (1), 26699, doi: 10.1038/srep26699.

Lai, D., Ha, G. and Shah, S. (2021). **HMMcopy: Copy number prediction with correction for GC and mappability bias for HTS data. R package version 1.34.0.**

Langer, C. J., Gadgeel, S. M., Borghaei, H., Papadimitrakopoulou, V. A., Patnaik, A., Powell, S. F., Gentzler, R. D., Martins, R. G., Stevenson, J. P., Jalal, S. I., Panwalkar, A., Yang, J. C., Gubens, M., Sequist, L. V., Awad, M. M., Fiore, J., Ge, Y., Raftopoulos, H. and Gandhi, L. (2016). **Carboplatin and pemetrexed with or without pembrolizumab for advanced, non-squamous non-small-cell lung cancer: a randomised, phase 2 cohort of the open-label KEYNOTE-021 study**. Lancet Oncol 17 (11), 1497-1508, doi: 10.1016/s1470-2045(16)30498-3.

Langmead, B. and Salzberg, S. L. (2012). **Fast gapped-read alignment with Bowtie 2**. Nature Methods 9 (4), 357-359, doi: 10.1038/nmeth.1923.

Larson, M. H., Pan, W., Kim, H. J., Mauntz, R. E., Stuart, S. M., Pimentel, M., Zhou, Y., Knudsgaard, P., Demas, V., Aravanis, A. M. and Jamshidi, A. (2021). **A comprehensive characterization of the cell-

free transcriptome reveals tissue- and subtype-specific biomarkers for cancer detection. Nat Commun 12 (1), 2357, doi: 10.1038/s41467-021-22444-1.

Lasseter, K., Nassar, A. H., Hamieh, L., Berchuck, J. E., Nuzzo, P. V., Korthauer, K., Shinagare, A. B., Ogorek, B., McKay, R., Thorner, A. R., Lee, G. M., Braun, D. A., Bhatt, R. S., Freedman, M., Choueiri, T. K. and Kwiatkowski, D. J. (2020). Plasma cell-free DNA variant analysis compared with methylated DNA analysis in renal cell carcinoma. Genet Med 22 (8), 1366-1373, doi: 10.1038/s41436-020-0801-x.

Lassus, H., Laitinen, M. P., Anttonen, M., Heikinheimo, M., Aaltonen, L. A., Ritvos, O. and Butzow, R. (2001). Comparison of serous and mucinous ovarian carcinomas: distinct pattern of allelic loss at distal 8p and expression of transcription factor GATA-4. Lab Invest 81 (4), 517-526, doi: 10.1038/labinvest.3780260.

Laurila, E., Vuorinen, E., Savinainen, K., Rauhala, H. and Kallioniemi, A. (2014). KPNA7, a nuclear transport receptor, promotes malignant properties of pancreatic cancer cells in vitro. Exp Cell Res 322 (1), 159-167, doi: 10.1016/j.yexcr.2013.11.014.

Law, C. W., Chen, Y., Shi, W. and Smyth, G. K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biology 15 (2), R29, doi: 10.1186/gb-2014-15-2-r29.

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T. and Carey, V. J. (2013a). Software for computing and annotating genomic ranges. PLoS Comput Biol 9 (8), e1003118, doi: 10.1371/journal.pcbi.1003118.

Lawrence, M. S., Stojanov, P., Mermel, C. H., Robinson, J. T., Garraway, L. A., Golub, T. R., Meyerson, M., Gabriel, S. B., Lander, E. S. and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 505 (7484), 495-501, doi: 10.1038/nature12912.

Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., Carter, S. L., Stewart, C., Mermel, C. H., Roberts, S. A., Kiezun, A., Hammerman, P. S., McKenna, A., Drier, Y., Zou, L., Ramos, A. H., Pugh, T. J., Stransky, N., Helman, E., Kim, J., Sougnez, C., Ambrogio, L., Nickerson, E., Shefler, E., Cortés, M. L., Auclair, D., Saksena, G., Voet, D., Noble, M., DiCara, D., Lin, P., Lichtenstein, L., Heiman, D. I., Fennell, T., Imielinski, M., Hernandez, B., Hodis, E., Baca, S., Dulak, A. M., Lohr, J., Landau, D.-A., Wu, C. J., Melendez-Zajgla, J., Hidalgo-Miranda, A., Koren, A., McCarroll, S. A., Mora, J., Lee, R. S., Crompton, B., Onofrio, R., Parkin, M., Winckler, W., Ardlie, K., Gabriel, S. B., Roberts, C. W. M., Biegel, J. A., Stegmaier, K., Bass, A. J., Garraway, L. A., Meyerson, M., Golub, T. R., Gordenin, D. A., Sunyaev, S., Lander, E. S. and Getz, G. (2013b). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499 (7457), 214-218, doi: 10.1038/nature12213.

Lehmann-Werman, R., Neiman, D., Zemmour, H., Moss, J., Magenheim, J., Vaknin-Dembinsky, A., Rubertsson, S., Nellgård, B., Blennow, K., Zetterberg, H., Spalding, K., Haller, M. J., Wasserfall, C. H., Schatz, D. A., Greenbaum, C. J., Dorrell, C., Grompe, M., Zick, A., Hubert, A., Maoz, M., Fendrich, V., Bartsch, D. K., Golan, T., Ben Sasson, S. A., Zamir, G., Razin, A., Cedar, H., Shapiro, A. M., Glaser, B., Shemer, R. and Dor, Y. (2016). Identification of tissue-specific cell death using methylation patterns of circulating DNA. Proc Natl Acad Sci U S A 113 (13), E1826-1834, doi: 10.1073/pnas.1519286113.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Subgroup, G. P. D. P. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25 (16), 2078-2079, doi: 10.1093/bioinformatics/btp352.

Li, J., Dong, W., Liu, L. N., Huang, Y. J. and Xiao, M. F. (2021). Liquid biopsy for ALK-positive early non-small-cell lung cancer predicts disease relapse. Future Oncol 17 (1), 81-90, doi: 10.2217/fon-2020-0554.

Li, J., Wei, Z., Li, H., Dang, Q., Zhang, Z., Wang, L., Gao, W., Zhang, P., Yang, D., Liu, J., Sun, Y. and Gao, W. (2015). Clinicopathological significance of fibroblast growth factor 1 in non-small cell lung cancer. Hum Pathol 46 (12), 1821-1828, doi: 10.1016/j.humpath.2015.07.022.

Li, W. and Zhou, X. J. (2020). Methylation extends the reach of liquid biopsy in cancer detection. Nat Rev Clin Oncol 17 (11), 655-656, doi: 10.1038/s41571-020-0420-0.

Li, X., Liu, Y., Salz, T., Hansen, K. D. and Feinberg, A. (2016). Whole-genome analysis of the methylome and hydroxymethylome in normal and malignant lung and liver. Genome Res 26 (12), 1730-1741, doi: 10.1101/gr.211854.116.

Lian, C. G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., Xu, W., Tan, L., Hu, Y., Zhan, Q., Lee, C. W., Hu, D., Lian, B. Q., Kleffel, S., Yang, Y., Neiswender, J., Khorasani, A. J., Fang, R., Lezcano, C., Duncan, L. M., Scolyer, R. A., Thompson, J. F., Kakavand, H., Houvras, Y., Zon, L. I., Mihm, M. C., Jr., Kaiser, U. B., Schatton, T., Woda, B. A., Murphy, G. F. and Shi, Y. G. (2012a). **Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma**. Cell *150 (6)*, 1135-1146, doi: 10.1016/j.cell.2012.07.033.

Lian, Y. X., Chen, R., Xu, Y. H., Peng, C. L. and Hu, H. C. (2012b). **Effect of protein-tyrosine phosphatase 4A3 by small interfering RNA on the proliferation of lung cancer**. Gene *511 (2)*, 169-176, doi: 10.1016/j.gene.2012.09.079.

Liao, Y., Smyth, G. K. and Shi, W. (2014). **featureCounts: an efficient general purpose program for assigning sequence reads to genomic features**. Bioinformatics *30 (7)*, 923-930, doi: 10.1093/bioinformatics/btt656.

Licchesi, J. D., Westra, W. H., Hooker, C. M. and Herman, J. G. (2008a). **Promoter hypermethylation of hallmark cancer genes in atypical adenomatous hyperplasia of the lung**. Clin Cancer Res *14 (9)*, 2570-2578, doi: 10.1158/1078-0432.Ccr-07-2033.

Licchesi, J. D., Westra, W. H., Hooker, C. M., Machida, E. O., Baylin, S. B. and Herman, J. G. (2008b). **Epigenetic alteration of Wnt pathway antagonists in progressive glandular neoplasia of the lung**. Carcinogenesis *29 (5)*, 895-904, doi: 10.1093/carcin/bgn017.

Lienhard, M. and Chavez, L. (2016). **Quantitative Comparison of Large-Scale DNA Enrichment Sequencing Data**. In: Statistical Genomics: Methods and Protocols, eds. Mathé, E. and Davis, S., Springer New York, New York, NY, pp. 191-208.

Lienhard, M., Grimm, C., Morkel, M., Herwig, R. and Chavez, L. (2014). **MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments**. Bioinformatics *30 (2)*, 284-286, doi: 10.1093/bioinformatics/btt650.

Lin, J. J., Riely, G. J. and Shaw, A. T. (2017). **Targeting ALK: Precision Medicine Takes on Drug Resistance**. Cancer Discov *7 (2)*, 137-155, doi: 10.1158/2159-8290.Cd-16-1123.

Lin, X., Shen, J., Dan, P., He, X., Xu, C., Chen, X., Tanyi, J. L., Montone, K., Fan, Y., Huang, Q., Zhang, L. and Zhong, X. (2018). **RNA-binding protein LIN28B inhibits apoptosis through regulation of the AKT2/FOXO3A/BIM axis in ovarian cancer cells**. Signal Transduction and Targeted Therapy *3 (1)*, 23, doi: 10.1038/s41392-018-0026-5.

Liu, M. C., Oxnard, G. R., Klein, E. A., Swanton, C. and Seiden, M. V. (2020). **Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA**. Ann Oncol *31 (6)*, 745-759, doi: 10.1016/j.annonc.2020.02.011.

Liu, Y., Siejka-Zielińska, P., Velikova, G., Bi, Y., Yuan, F., Tomkova, M., Bai, C., Chen, L., Schuster-Böckler, B. and Song, C. X. (2019). **Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution**. Nat Biotechnol *37 (4)*, 424-429, doi: 10.1038/s41587-019-0041-2.

Liu, Y., Zhou, J. and White, K. P. (2014). **RNA-seq differential expression studies: more sequence or more replication?** Bioinformatics *30 (3)*, 301-304, doi: 10.1093/bioinformatics/btt688.

Lowe, R., Gemma, C., Beyan, H., Hawa, M. I., Bazeos, A., Leslie, R. D., Montpetit, A., Rakyan, V. K. and Ramagopalan, S. V. (2013). **Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies**. Epigenetics *8 (4)*, 445-454, doi: 10.4161/epi.24362.

Lowe, R., Slodkowicz, G., Goldman, N. and Rakyan, V. K. (2015). **The human blood DNA methylome displays a highly distinctive profile compared with other somatic tissues**. Epigenetics *10 (4)*, 274-281, doi: 10.1080/15592294.2014.1003744.

Lu, T., Liu, H. and You, G. (2018). **Long non-coding RNA C5orf66-AS1 prevents oral squamous cell carcinoma through inhibiting cell growth and metastasis**. Int J Mol Med *42 (6)*, 3291-3299, doi: 10.3892/ijmm.2018.3913.

References

Lv, Q., Xia, Q., Li, A. and Wang, Z. (2021). **The Potential Role of IL1RAP on Tumor Microenvironment-Related Inflammatory Factors in Stomach Adenocarcinoma**. Technol Cancer Res Treat *20*, 1533033821995282, doi: 10.1177/1533033821995282.

Lynch, T. J., Bell, D. W., Sordella, R., Gurubhagavatula, S., Okimoto, R. A., Brannigan, B. W., Harris, P. L., Haserlat, S. M., Supko, J. G., Haluska, F. G., Louis, D. N., Christiani, D. C., Settleman, J. and Haber, D. A. (2004). **Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib**. N Engl J Med *350 (21)*, 2129-2139, doi: 10.1056/NEJMoa040938.

Ma, Q., Niu, R., Huang, W., Da, L., Tang, Y., Jiang, D., Xi, Y. and Zhang, C. (2020). **Long Noncoding RNA PTPRG Antisense RNA 1 Reduces Radiosensitivity of Nonsmall Cell Lung Cancer Cells Via Regulating MiR-200c-3p/TCF4**. Technol Cancer Res Treat *19*, 1533033820942615, doi: 10.1177/1533033820942615.

Madzo, J., Liu, H., Rodriguez, A., Vasanthakumar, A., Sundaravel, S., Caces, D. B. D., Looney, T. J., Zhang, L., Lepore, J. B., Macrae, T., Duszynski, R., Shih, A. H., Song, C. X., Yu, M., Yu, Y., Grossman, R., Raumann, B., Verma, A., He, C., Levine, R. L., Lavelle, D., Lahn, B. T., Wickrema, A. and Godley, L. A. (2014). **Hydroxymethylation at gene regulatory regions directs stem/early progenitor cell commitment during erythropoiesis**. Cell Rep 6 *(1)*, 231-244, doi: 10.1016/j.celrep.2013.11.044.

Maemondo, M., Inoue, A., Kobayashi, K., Sugawara, S., Oizumi, S., Isobe, H., Gemma, A., Harada, M., Yoshizawa, H., Kinoshita, I., Fujita, Y., Okinaga, S., Hirano, H., Yoshimori, K., Harada, T., Ogura, T., Ando, M., Miyazawa, H., Tanaka, T., Saijo, Y., Hagiwara, K., Morita, S. and Nukiwa, T. (2010). **Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR**. N Engl J Med *362 (25)*, 2380-2388, doi: 10.1056/NEJMoa0909530.

Manicone, M., Scaini, M. C., Rodriquenz, M. G., Facchinetti, A., Tartarone, A., Aieta, M., Zamarchi, R. and Rossi, E. (2017). **Liquid biopsy for monitoring anaplastic lymphoma kinase inhibitors in non-small cell lung cancer: two cases compared**. J Thorac Dis 9 *(Suppl 13)*, S1391-s1396, doi: 10.21037/jtd.2017.08.151.

Martin, M. (2011). **Cutadapt removes adapter sequences from high-throughput sequencing reads**. 2011 *17 (1)*, 3, doi: 10.14806/ej.17.1.200.

McGranahan, N. and Swanton, C. (2015). **Biological and therapeutic impact of intratumor heterogeneity in cancer evolution**. Cancer Cell *27 (1)*, 15-26, doi: 10.1016/j.ccell.2014.12.001.

Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. and Heintz, N. (2012). **MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system**. Cell *151 (7)*, 1417-1430, doi: 10.1016/j.cell.2012.11.022.

Millrud, C. R., von Wachenfeldt, K., Falk, H. H., Forsberg, G. and Liberg, D. (2020). **Abstract 2269: The anti-IL1RAP antibody CAN04 increases tumor sensitivity to platinum-based chemotherapy**. Cancer Research *80 (16 Supplement)*, 2269-2269, doi: 10.1158/1538-7445.Am2020-2269.

Mitchell, K., Barreyro, L., Todorova, T. I., Taylor, S. J., Antony-Debré, I., Narayanagari, S. R., Carvajal, L. A., Leite, J., Piperdi, Z., Pendurti, G., Mantzaris, I., Paietta, E., Verma, A., Gritsman, K. and Steidl, U. (2018). **IL1RAP potentiates multiple oncogenic signaling pathways in AML**. J Exp Med *215 (6)*, 1709-1727, doi: 10.1084/jem.20180147.

Mithani, S. K., Smith, I. M., Zhou, S., Gray, A., Koch, W. M., Maitra, A. and Califano, J. A. (2007). **Mitochondrial resequencing arrays detect tumor-specific mutations in salivary rinses of patients with head and neck cancer**. Clin Cancer Res *13 (24)*, 7335-7340, doi: 10.1158/1078-0432.Ccr-07-0220.

Mok, T., Wu, Y. L., Lee, J. S., Yu, C. J., Sriuranpong, V., Sandoval-Tan, J., Ladrera, G., Thongprasert, S., Srimuninnimit, V., Liao, M., Zhu, Y., Zhou, C., Fuerte, F., Margono, B., Wen, W., Tsai, J., Truman, M., Klughammer, B., Shames, D. S. and Wu, L. (2015). **Detection and Dynamic Changes of EGFR Mutations from Circulating Tumor DNA as a Predictor of Survival Outcomes in NSCLC Patients Treated with First-line Intercalated Erlotinib and Chemotherapy**. Clin Cancer Res *21 (14)*, 3196-3203, doi: 10.1158/1078-0432.Ccr-14-2594.

Mok, T. S., Wu, Y. L., Ahn, M. J., Garassino, M. C., Kim, H. R., Ramalingam, S. S., Shepherd, F. A., He, Y., Akamatsu, H., Theelen, W. S., Lee, C. K., Sebastian, M., Templeton, A., Mann, H., Marotti, M.,

Ghiorghiu, S. and Papadimitrakopoulou, V. A. (2017). **Osimertinib or Platinum-Pemetrexed in EGFR T790M-Positive Lung Cancer**. N Engl J Med 376 (7), 629-640, doi: 10.1056/NEJMoa1612674.

Mok, T. S., Wu, Y. L., Thongprasert, S., Yang, C. H., Chu, D. T., Saijo, N., Sunpaweravong, P., Han, B., Margono, B., Ichinose, Y., Nishiwaki, Y., Ohe, Y., Yang, J. J., Chewaskulyong, B., Jiang, H., Duffield, E. L., Watkins, C. L., Armour, A. A. and Fukuoka, M. (2009). **Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma**. N Engl J Med 361 (10), 947-957, doi: 10.1056/NEJMoa0810699.

Morgensztern, D., Ng, S. H., Gao, F. and Govindan, R. (2010). **Trends in Stage Distribution for Patients with Non-small Cell Lung Cancer: A National Cancer Database Survey**. Journal of Thoracic Oncology 5 (1), 29-33, doi: https://doi.org/10.1097/JTO.0b013e3181c5920c.

Morris, S. W., Kirstein, M. N., Valentine, M. B., Dittmer, K. G., Shapiro, D. N., Saltman, D. L. and Look, A. T. (1994). **Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma**. Science 263 (5151), 1281-1284, doi: 10.1126/science.8122112.

Moss, J., Magenheim, J., Neiman, D., Zemmour, H., Loyfer, N., Korach, A., Samet, Y., Maoz, M., Druid, H., Arner, P., Fu, K. Y., Kiss, E., Spalding, K. L., Landesberg, G., Zick, A., Grinshpun, A., Shapiro, A. M. J., Grompe, M., Wittenberg, A. D., Glaser, B., Shemer, R., Kaplan, T. and Dor, Y. (2018). **Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease**. Nat Commun 9 (1), 5068, doi: 10.1038/s41467-018-07466-6.

Moss, J., Zick, A., Grinshpun, A., Carmon, E., Maoz, M., Ochana, B. L., Abraham, O., Arieli, O., Germansky, L., Meir, K., Glaser, B., Shemer, R., Uziely, B. and Dor, Y. (2020). **Circulating breast-derived DNA allows universal detection and monitoring of localized breast cancer**. Ann Oncol 31 (3), 395-403, doi: 10.1016/j.annonc.2019.11.014.

Mouliere, F., Chandrananda, D., Piskorz, A. M., Moore, E. K., Morris, J., Ahlborn, L. B., Mair, R., Goranova, T., Marass, F., Heider, K., Wan, J. C. M., Supernat, A., Hudecova, I., Gounaris, I., Ros, S., Jimenez-Linan, M., Garcia-Corbacho, J., Patel, K., Østrup, O., Murphy, S., Eldridge, M. D., Gale, D., Stewart, G. D., Burge, J., Cooper, W. N., van der Heijden, M. S., Massie, C. E., Watts, C., Corrie, P., Pacey, S., Brindle, K. M., Baird, R. D., Mau-Sørensen, M., Parkinson, C. A., Smith, C. G., Brenton, J. D. and Rosenfeld, N. (2018a). **Enhanced detection of circulating tumor DNA by fragment size analysis**. Sci Transl Med 10 (466), doi: 10.1126/scitranslmed.aat4921.

Mouliere, F., Mair, R., Chandrananda, D., Marass, F., Smith, C. G., Su, J., Morris, J., Watts, C., Brindle, K. M. and Rosenfeld, N. (2018b). **Detection of cell-free DNA fragmentation and copy number alterations in cerebrospinal fluid from glioma patients**. EMBO Mol Med 10 (12), doi: 10.15252/emmm.201809323.

Mouliere, F., Piskorz, A. M., Chandrananda, D., Moore, E., Morris, J., Smith, C. G., Goranova, T., Heider, K., Mair, R., Supernat, A., Gounaris, I., Ros, S., Wan, J. C. M., Jimenez-Linan, M., Gale, D., Brindle, K., Massie, C. E., Parkinson, C. A., Brenton, J. D. and Rosenfeld, N. (2017). **Selecting short DNA fragments in plasma improves detection of circulating tumour DNA**. bioRxiv, 134437, doi: 10.1101/134437.

Münsterberg, J., Loreth, D., Brylka, L., Werner, S., Karbanova, J., Gandrass, M., Schneegans, S., Besler, K., Hamester, F., Robador, J. R., Bauer, A. T., Schneider, S. W., Wrage, M., Lamszus, K., Matschke, J., Vashist, Y., Uzunoglu, G., Steurer, S., Horst, A. K., Oliveira-Ferrer, L., Glatzel, M., Schinke, T., Corbeil, D., Pantel, K., Maire, C. and Wikman, H. (2020). **ALCAM contributes to brain metastasis formation in non-small-cell lung cancer through interaction with the vascular endothelium**. Neuro Oncol 22 (7), 955-966, doi: 10.1093/neuonc/noaa028.

Murtaza, M., Dawson, S. J., Pogrebniak, K., Rueda, O. M., Provenzano, E., Grant, J., Chin, S. F., Tsui, D. W. Y., Marass, F., Gale, D., Ali, H. R., Shah, P., Contente-Cuomo, T., Farahani, H., Shumansky, K., Kingsbury, Z., Humphray, S., Bentley, D., Shah, S. P., Wallis, M., Rosenfeld, N. and Caldas, C. (2015). **Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer**. Nat Commun 6, 8760, doi: 10.1038/ncomms9760.

Murtaza, M., Dawson, S. J., Tsui, D. W., Gale, D., Forshew, T., Piskorz, A. M., Parkinson, C., Chin, S. F., Kingsbury, Z., Wong, A. S., Marass, F., Humphray, S., Hadfield, J., Bentley, D., Chin, T. M., Brenton, J. D., Caldas, C. and Rosenfeld, N. (2013). **Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA**. Nature 497 (7447), 108-112, doi: 10.1038/nature12065.

Nassiri, F., Chakravarthy, A., Feng, S., Shen, S. Y., Nejad, R., Zuccato, J. A., Voisin, M. R., Patil, V., Horbinski, C., Aldape, K., Zadeh, G. and De Carvalho, D. D. (2020). **Detection and discrimination of intracranial**

tumors using plasma cell-free DNA methylomes. Nat Med 26 *(7)*, 1044-1047, doi: 10.1038/s41591-020-0932-2.

National Institute for Health and Care Excellence, T. (2019). **Lung cancer: diagnosis and management (NICE Guideline 122)**. URL: https://www.nice.org.uk/guidance/ng122 [as of17.04.2020].

National Lung Screening Trial Research, T., Aberle, D. R., Adams, A. M., Berg, C. D., Black, W. C., Clapp, J. D., Fagerstrom, R. M., Gareen, I. F., Gatsonis, C., Marcus, P. M. and Sicks, J. D. (2011). **Reduced lung-cancer mortality with low-dose computed tomographic screening**. The New England journal of medicine 365 *(5)*, 395-409, doi: 10.1056/NEJMoa1102873.

Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Pagnani, A., Zecchina, R., Parlato, C. and Oliviero, S. (2013). **Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells**. Genome Biol 14 *(8)*, R91, doi: 10.1186/gb-2013-14-8-r91.

Newman, A. M., Bratman, S. V., To, J., Wynne, J. F., Eclov, N. C., Modlin, L. A., Liu, C. L., Neal, J. W., Wakelee, H. A., Merritt, R. E., Shrager, J. B., Loo, B. W., Jr., Alizadeh, A. A. and Diehn, M. (2014). **An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage**. Nat Med 20 *(5)*, 548-554, doi: 10.1038/nm.3519.

Newman, A. M., Lovejoy, A. F., Klass, D. M., Kurtz, D. M., Chabon, J. J., Scherer, F., Stehr, H., Liu, C. L., Bratman, S. V., Say, C., Zhou, L., Carter, J. N., West, R. B., Sledge, G. W., Shrager, J. B., Loo, B. W., Jr., Neal, J. W., Wakelee, H. A., Diehn, M. and Alizadeh, A. A. (2016). **Integrated digital error suppression for improved detection of circulating tumor DNA**. Nat Biotechnol 34 *(5)*, 547-555, doi: 10.1038/nbt.3520.

Ng, S. B., Chua, C., Ng, M., Gan, A., Poon, P. S., Teo, M., Fu, C., Leow, W. Q., Lim, K. H., Chung, A., Koo, S. L., Choo, S. P., Ho, D., Rozen, S., Tan, P., Wong, M., Burkholder, W. F. and Tan, I. B. (2017). **Individualised multiplexed circulating tumour DNA assays for monitoring of tumour presence in patients after colorectal cancer surgery**. Sci Rep 7, 40737, doi: 10.1038/srep40737.

Noh, K. W., Lee, M. S., Lee, S. E., Song, J. Y., Shin, H. T., Kim, Y. J., Oh, D. Y., Jung, K., Sung, M., Kim, M., An, S., Han, J., Shim, Y. M., Zo, J. I., Kim, J., Park, W. Y., Lee, S. H. and Choi, Y. L. (2017). **Molecular breakdown: a comprehensive view of anaplastic lymphoma kinase (ALK)-rearranged non-small cell lung cancer**. J Pathol 243 *(3)*, 307-319, doi: 10.1002/path.4950.

Nuzzo, P. V., Berchuck, J. E., Korthauer, K., Spisak, S., Nassar, A. H., Abou Alaiwi, S., Chakravarthy, A., Shen, S. Y., Bakouny, Z., Boccardo, F., Steinharter, J., Bouchard, G., Curran, C. R., Pan, W., Baca, S. C., Seo, J. H., Lee, G. M., Michaelson, M. D., Chang, S. L., Waikar, S. S., Sonpavde, G., Irizarry, R. A., Pomerantz, M., De Carvalho, D. D., Choueiri, T. K. and Freedman, M. L. (2020). **Detection of renal cell carcinoma using plasma and urine cell-free DNA methylomes**. Nat Med 26 *(7)*, 1041-1043, doi: 10.1038/s41591-020-0933-1.

Okamoto, J., Hirata, T., Chen, Z., Zhou, H. M., Mikami, I., Li, H., Yagui-Beltran, A., Johansson, M., Coussens, L. M., Clement, G., Shi, Y., Zhang, F., Koizumi, K., Shimizu, K., Jablons, D. and He, B. (2010). **EMX2 is epigenetically silenced and suppresses growth in human lung cancer**. Oncogene 29 *(44)*, 5969-5975, doi: 10.1038/onc.2010.330.

Okano, M., Bell, D. W., Haber, D. A. and Li, E. (1999). **DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development**. Cell 99 *(3)*, 247-257, doi: 10.1016/s0092-8674(00)81656-6.

Okano, M., Xie, S. and Li, E. (1998). **Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases**. Nat Genet 19 *(3)*, 219-220, doi: 10.1038/890.

Overman, M. J., Modak, J., Kopetz, S., Murthy, R., Yao, J. C., Hicks, M. E., Abbruzzese, J. L. and Tam, A. L. (2013). **Use of research biopsies in clinical trials: are risks and benefits adequately discussed?** J Clin Oncol 31 *(1)*, 17-22, doi: 10.1200/jco.2012.43.1718.

Paez, J. G., Jänne, P. A., Lee, J. C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F. J., Lindeman, N., Boggon, T. J., Naoki, K., Sasaki, H., Fujii, Y., Eck, M. J., Sellers, W. R., Johnson, B. E. and Meyerson, M. (2004). **EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy**. Science 304 *(5676)*, 1497-1500, doi: 10.1126/science.1099314.

Pan, W., Gu, W., Nagpal, S., Gephart, M. H. and Quake, S. R. (2015). **Brain tumor mutations detected in cerebral spinal fluid**. Clin Chem *61 (3)*, 514-522, doi: 10.1373/clinchem.2014.235457.

Pantel, K. and Alix-Panabières, C. (2013). **Real-time liquid biopsy in cancer patients: fact or fiction?** Cancer Res *73 (21)*, 6384-6388, doi: 10.1158/0008-5472.Can-13-2030.

Pilkington, G., Boland, A., Brown, T., Oyee, J., Bagust, A. and Dickson, R. (2015). **A systematic review of the clinical effectiveness of first-line chemotherapy for adult patients with locally advanced or metastatic non-small cell lung cancer**. Thorax *70 (4)*, 359-367, doi: 10.1136/thoraxjnl-2014-205914.

Pillai, R. N., Behera, M., Berry, L. D., Rossi, M. R., Kris, M. G., Johnson, B. E., Bunn, P. A., Ramalingam, S. S. and Khuri, F. R. (2017). **HER2 mutations in lung adenocarcinomas: A report from the Lung Cancer Mutation Consortium**. Cancer *123 (21)*, 4099-4105, doi: 10.1002/cncr.30869.

Planchard, D., Smit, E. F., Groen, H. J. M., Mazieres, J., Besse, B., Helland, Å., Giannone, V., D'Amelio, A. M., Jr., Zhang, P., Mookerjee, B. and Johnson, B. E. (2017). **Dabrafenib plus trametinib in patients with previously untreated BRAF(V600E)-mutant metastatic non-small-cell lung cancer: an open-label, phase 2 trial**. Lancet Oncol *18 (10)*, 1307-1316, doi: 10.1016/s1470-2045(17)30679-4.

Quinlan, A. R. and Hall, I. M. (2010). **BEDTools: a flexible suite of utilities for comparing genomic features**. Bioinformatics *26 (6)*, 841-842, doi: 10.1093/bioinformatics/btq033.

Ramalingam, S. S., Vansteenkiste, J., Planchard, D., Cho, B. C., Gray, J. E., Ohe, Y., Zhou, C., Reungwetwattana, T., Cheng, Y., Chewaskulyong, B., Shah, R., Cobo, M., Lee, K. H., Cheema, P., Tiseo, M., John, T., Lin, M.-C., Imamura, F., Kurata, T., Todd, A., Hodge, R., Saggese, M., Rukazenkov, Y. and Soria, J.-C. (2019). **Overall Survival with Osimertinib in Untreated, EGFR-Mutated Advanced NSCLC**. New England Journal of Medicine *382 (1)*, 41-50, doi: 10.1056/NEJMoa1913662.

Raman, L., Dheedene, A., De Smet, M., Van Dorpe, J. and Menten, B. (2019). **WisecondorX: improved copy number detection for routine shallow whole-genome sequencing**. Nucleic Acids Res *47 (4)*, 1605-1614, doi: 10.1093/nar/gky1263.

Reck, M., Rodríguez-Abreu, D., Robinson, A. G., Hui, R., Csőszi, T., Fülöp, A., Gottfried, M., Peled, N., Tafreshi, A., Cuffe, S., O'Brien, M., Rao, S., Hotta, K., Leiby, M. A., Lubiniecki, G. M., Shentu, Y., Rangwala, R. and Brahmer, J. R. (2016). **Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer**. N Engl J Med *375 (19)*, 1823-1833, doi: 10.1056/NEJMoa1606774.

Riediger, A. L., Dietz, S., Schirmer, U., Meister, M., Heinzmann-Groth, I., Schneider, M., Muley, T., Thomas, M. and Sültmann, H. (2016). **Mutation analysis of circulating plasma DNA to determine response to EGFR tyrosine kinase inhibitor therapy of lung adenocarcinoma patients**. Scientific Reports *6 (1)*, 33505, doi: 10.1038/srep33505.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. and Smyth, G. K. (2015). **limma powers differential expression analyses for RNA-sequencing and microarray studies**. Nucleic Acids Research *43 (7)*, e47-e47, doi: 10.1093/nar/gkv007.

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2009). **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data**. Bioinformatics *26 (1)*, 139-140, doi: 10.1093/bioinformatics/btp616.

Rui, Y., Peng, W. J., Wang, M., Wang, Q., Liu, Z. L., Chen, Y. Q. and Huang, L. N. (2017). **HIST1H3D: A promising therapeutic target for lung cancer**. Int J Oncol *50 (3)*, 815-822, doi: 10.3892/ijo.2017.3856.

Sadeh, R., Sharkia, I., Fialkoff, G., Rahat, A., Gutin, J., Chappleboim, A., Nitzan, M., Fox-Fisher, I., Neiman, D., Meler, G., Kamari, Z., Yaish, D., Peretz, T., Hubert, A., Cohen, J. E., Salah, A., Temper, M., Grinshpun, A., Maoz, M., Abu-Gazala, S., Ben Ya'acov, A., Shteyer, E., Safadi, R., Kaplan, T., Shemer, R., Planer, D., Galun, E., Glaser, B., Zick, A., Dor, Y. and Friedman, N. (2021). **ChIP-seq of plasma cell-free nucleosomes identifies gene expression programs of the cells of origin**. Nature Biotechnology *39 (5)*, 586-598, doi: 10.1038/s41587-020-00775-6.

Sanders, A. J., Owen, S., Morgan, L. D., Ruge, F., Collins, R. J., Ye, L., Mason, M. D. and Jiang, W. G. (2019). **Importance of activated leukocyte cell adhesion molecule (ALCAM) in prostate cancer progression and metastatic dissemination**. Oncotarget *10 (59)*, 6362-6377, doi: 10.18632/oncotarget.27279.

Sanders, H. R., Li, H. R., Bruey, J. M., Scheerle, J. A., Meloni-Ehrig, A. M., Kelly, J. C., Novick, C. and Albitar, M. (2011). **Exon scanning by reverse transcriptase-polymerase chain reaction for detection of known and novel EML4-ALK fusion variants in non-small cell lung cancer**. Cancer Genet 204 (1), 45-52, doi: 10.1016/j.cancergencyto.2010.08.024.

Schotten, L. M., Darwiche, K., Seweryn, M., Yildiz, V., Kneuertz, P. J., Eberhardt, W. E. E., Eisenmann, S., Welter, S., Sisson, B. E., Pietrzak, M., Wiesweg, M., Ploenes, T., Hager, T., He, K., Freitag, L., Aigner, C., Taube, C. and Oezkan, F. (2021). **DNA methylation of PTGER4 in peripheral blood plasma helps to distinguish between lung cancer, benign pulmonary nodules and chronic obstructive pulmonary disease patients**. European Journal of Cancer 147, 142-150, doi: https://doi.org/10.1016/j.ejca.2021.01.032.

Selamat, S. A., Chung, B. S., Girard, L., Zhang, W., Zhang, Y., Campan, M., Siegmund, K. D., Koss, M. N., Hagen, J. A., Lam, W. L., Lam, S., Gazdar, A. F. and Laird-Offringa, I. A. (2012). **Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression**. Genome research 22 (7), 1197-1211, doi: 10.1101/gr.132662.111.

Selamat, S. A., Galler, J. S., Joshi, A. D., Fyfe, M. N., Campan, M., Siegmund, K. D., Kerr, K. M. and Laird-Offringa, I. A. (2011). **DNA methylation changes in atypical adenomatous hyperplasia, adenocarcinoma in situ, and lung adenocarcinoma**. PLoS One 6 (6), e21443, doi: 10.1371/journal.pone.0021443.

Serpas, L., Chan, R. W. Y., Jiang, P., Ni, M., Sun, K., Rashidfarrokhi, A., Soni, C., Sisirak, V., Lee, W.-S., Cheng, S. H., Peng, W., Chan, K. C. A., Chiu, R. W. K., Reizis, B. and Lo, Y. M. D. (2019). **Dnase1l3 deletion causes aberrations in length and end-motif frequencies in plasma DNA**. Proceedings of the National Academy of Sciences 116 (2), 641-649, doi: 10.1073/pnas.1815031116.

Shao, M., Yang, Q., Zhu, W., Jin, H., Wang, J., Song, J., Kong, Y. and Lv, X. (2018). **LncHOXA10 drives liver TICs self-renewal and tumorigenesis via HOXA10 transcription activation**. Molecular Cancer 17 (1), 173, doi: 10.1186/s12943-018-0921-y.

Shaw, A. T., Felip, E., Bauer, T. M., Besse, B., Navarro, A., Postel-Vinay, S., Gainor, J. F., Johnson, M., Dietrich, J., James, L. P., Clancy, J. S., Chen, J., Martini, J. F., Abbattista, A. and Solomon, B. J. (2017). **Lorlatinib in non-small-cell lung cancer with ALK or ROS1 rearrangement: an international, multicentre, open-label, single-arm first-in-man phase 1 trial**. Lancet Oncol 18 (12), 1590-1599, doi: 10.1016/s1470-2045(17)30680-0.

Shaw, A. T., Kim, D. W., Nakagawa, K., Seto, T., Crinó, L., Ahn, M. J., De Pas, T., Besse, B., Solomon, B. J., Blackhall, F., Wu, Y. L., Thomas, M., O'Byrne, K. J., Moro-Sibilot, D., Camidge, D. R., Mok, T., Hirsh, V., Riely, G. J., Iyer, S., Tassell, V., Polli, A., Wilner, K. D. and Jänne, P. A. (2013). **Crizotinib versus chemotherapy in advanced ALK-positive lung cancer**. N Engl J Med 368 (25), 2385-2394, doi: 10.1056/NEJMoa1214886.

Shaw, A. T., Ou, S. H., Bang, Y. J., Camidge, D. R., Solomon, B. J., Salgia, R., Riely, G. J., Varella-Garcia, M., Shapiro, G. I., Costa, D. B., Doebele, R. C., Le, L. P., Zheng, Z., Tan, W., Stephenson, P., Shreeve, S. M., Tye, L. M., Christensen, J. G., Wilner, K. D., Clark, J. W. and Iafrate, A. J. (2014). **Crizotinib in ROS1-rearranged non-small-cell lung cancer**. N Engl J Med 371 (21), 1963-1971, doi: 10.1056/NEJMoa1406766.

Shaw, A. T. and Solomon, B. J. (2015). **Crizotinib in ROS1-rearranged non-small-cell lung cancer**. N Engl J Med 372 (7), 683-684, doi: 10.1056/NEJMc1415359.

Shen, N., Zhang, D., Yin, L., Qiu, Y., Liu, J., Yu, W., Fu, X., Zhu, B., Xu, X., Duan, A., Chen, Z., Wang, X., Cao, X., Zhao, T., Zhou, Z., Yu, L., Qin, H., Fang, Z., Li, J. Y., Liu, Y., Xiong, L., Yuan, B., Li, F. and Zhang, Y. (2019a). **Bile cell-free DNA as a novel and powerful liquid biopsy for detecting somatic variants in biliary tract cancer**. Oncol Rep 42 (2), 549-560, doi: 10.3892/or.2019.7177.

Shen, S. Y., Burgener, J. M., Bratman, S. V. and De Carvalho, D. D. (2019b). **Preparation of cfMeDIP-seq libraries for methylome profiling of plasma cell-free DNA**. Nat Protoc 14 (10), 2749-2780, doi: 10.1038/s41596-019-0202-2.

Shen, S. Y., Singhania, R., Fehringer, G., Chakravarthy, A., Roehrl, M. H. A., Chadwick, D., Zuzarte, P. C., Borgida, A., Wang, T. T., Li, T., Kis, O., Zhao, Z., Spreafico, A., Medina, T. D. S., Wang, Y., Roulois, D.,

Ettayebi, I., Chen, Z., Chow, S., Murphy, T., Arruda, A., O'Kane, G. M., Liu, J., Mansour, M., McPherson, J. D., O'Brien, C., Leighl, N., Bedard, P. L., Fleshner, N., Liu, G., Minden, M. D., Gallinger, S., Goldenberg, A., Pugh, T. J., Hoffman, M. M., Bratman, S. V., Hung, R. J. and De Carvalho, D. D. (2018). **Sensitive tumour detection and classification using plasma cell-free DNA methylomes**. Nature *563 (7732)*, 579-583, doi: 10.1038/s41586-018-0703-0.

Shen, Z., Chen, B., Gan, X., Hu, W., Zhong, G., Li, H., Xie, X., Liu, Y., Li, H., Xu, X., Huang, Z. and Chen, J. (2016). **Methylation of neurofilament light polypeptide promoter is associated with cell invasion and metastasis in NSCLC**. Biochem Biophys Res Commun *470 (3)*, 627-634, doi: 10.1016/j.bbrc.2016.01.094.

Sheng, K., Lu, J. and Zhao, H. (2018). **ELK1-induced upregulation of lncRNA HOXA10-AS promotes lung adenocarcinoma progression by increasing Wnt/β-catenin signaling**. Biochemical and Biophysical Research Communications *501 (3)*, 612-618, doi: https://doi.org/10.1016/j.bbrc.2018.04.224.

Siravegna, G., Mussolin, B., Venesio, T., Marsoni, S., Seoane, J., Dive, C., Papadopoulos, N., Kopetz, S., Corcoran, R. B., Siu, L. L. and Bardelli, A. (2019). **How liquid biopsies can change clinical practice in oncology**. Ann Oncol *30 (10)*, 1580-1590, doi: 10.1093/annonc/mdz227.

Skoulidis, F. and Heymach, J. V. (2019). **Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy**. Nat Rev Cancer *19 (9)*, 495-509, doi: 10.1038/s41568-019-0179-8.

Skvortsova, K., Masle-Farquhar, E., Luu, P. L., Song, J. Z., Qu, W., Zotenko, E., Gould, C. M., Du, Q., Peters, T. J., Colino-Sanguino, Y., Pidsley, R., Nair, S. S., Khoury, A., Smith, G. C., Miosge, L. A., Reed, J. H., Kench, J. G., Rubin, M. A., Horvath, L., Bogdanovic, O., Lim, S. M., Polo, J. M., Goodnow, C. C., Stirzaker, C. and Clark, S. J. (2019a). **DNA Hypermethylation Encroachment at CpG Island Borders in Cancer Is Predisposed by H3K4 Monomethylation Patterns**. Cancer Cell *35 (2)*, 297-314.e298, doi: 10.1016/j.ccell.2019.01.004.

Skvortsova, K., Stirzaker, C. and Taberlay, P. (2019b). **The DNA methylation landscape in cancer**. Essays Biochem *63 (6)*, 797-811, doi: 10.1042/ebc20190037.

Smith, C. G., Moser, T., Mouliere, F., Field-Rayner, J., Eldridge, M., Riediger, A. L., Chandrananda, D., Heider, K., Wan, J. C. M., Warren, A. Y., Morris, J., Hudecova, I., Cooper, W. N., Mitchell, T. J., Gale, D., Ruiz-Valdepenas, A., Klatte, T., Ursprung, S., Sala, E., Riddick, A. C. P., Aho, T. F., Armitage, J. N., Perakis, S., Pichler, M., Seles, M., Wcislo, G., Welsh, S. J., Matakidou, A., Eisen, T., Massie, C. E., Rosenfeld, N., Heitzer, E. and Stewart, G. D. (2020). **Comprehensive characterization of cell-free tumor DNA in plasma and urine of patients with renal tumors**. Genome Medicine *12 (1)*, 23, doi: 10.1186/s13073-020-00723-8.

Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M. and Shendure, J. (2016). **Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin**. Cell *164 (1-2)*, 57-68.

Soda, M., Choi, Y. L., Enomoto, M., Takada, S., Yamashita, Y., Ishikawa, S., Fujiwara, S., Watanabe, H., Kurashina, K., Hatanaka, H., Bando, M., Ohno, S., Ishikawa, Y., Aburatani, H., Niki, T., Sohara, Y., Sugiyama, Y. and Mano, H. (2007). **Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer**. Nature *448 (7153)*, 561-566, doi: 10.1038/nature05945.

Solomon, B. J., Mok, T., Kim, D. W., Wu, Y. L., Nakagawa, K., Mekhail, T., Felip, E., Cappuzzo, F., Paolini, J., Usari, T., Iyer, S., Reisman, A., Wilner, K. D., Tursi, J. and Blackhall, F. (2014). **First-line crizotinib versus chemotherapy in ALK-positive lung cancer**. N Engl J Med *371 (23)*, 2167-2177, doi: 10.1056/NEJMoa1408440.

Song, C. X., Szulwach, K. E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C. H., Zhang, W., Jian, X., Wang, J., Zhang, L., Looney, T. J., Zhang, B., Godley, L. A., Hicks, L. M., Lahn, B. T., Jin, P. and He, C. (2011). **Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine**. Nat Biotechnol *29 (1)*, 68-72, doi: 10.1038/nbt.1732.

Song, C. X., Yin, S., Ma, L., Wheeler, A., Chen, Y., Zhang, Y., Liu, B., Xiong, J., Zhang, W., Hu, J., Zhou, Z., Dong, B., Tian, Z., Jeffrey, S. S., Chua, M. S., So, S., Li, W., Wei, Y., Diao, J., Xie, D. and Quake, S. R. (2017). **5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages**. Cell Res *27 (10)*, 1231-1242, doi: 10.1038/cr.2017.106.

Soria, J. C., Ohe, Y., Vansteenkiste, J., Reungwetwattana, T., Chewaskulyong, B., Lee, K. H., Dechaphunkul, A., Imamura, F., Nogami, N., Kurata, T., Okamoto, I., Zhou, C., Cho, B. C., Cheng, Y.,

References

Cho, E. K., Voon, P. J., Planchard, D., Su, W. C., Gray, J. E., Lee, S. M., Hodge, R., Marotti, M., Rukazenkov, Y. and Ramalingam, S. S. (2018). **Osimertinib in Untreated EGFR-Mutated Advanced Non-Small-Cell Lung Cancer**. N Engl J Med *378 (2)*, 113-125, doi: 10.1056/NEJMoa1713137.

Sriram, K. B., Relan, V., Clarke, B. E., Duhig, E. E., Windsor, M. N., Matar, K. S., Naidoo, R., Passmore, L., McCaul, E., Courtney, D., Yang, I. A., Bowman, R. V. and Fong, K. M. (2012). **Pleural fluid cell-free DNA integrity index to identify cytologically negative malignant pleural effusions including mesotheliomas**. BMC Cancer *12*, 428, doi: 10.1186/1471-2407-12-428.

Starr, A., Greif, J., Vexler, A., Ashkenazy-Voghera, M., Gladesh, V., Rubin, C., Kerber, G., Marmor, S., Lev-Ari, S., Inbar, M., Yarden, Y. and Ben-Yosef, R. (2006). **ErbB4 increases the proliferation potential of human lung cancer cells and its blockage can be used as a target for anti-cancer therapy**. Int J Cancer *119 (2)*, 269-274, doi: 10.1002/ijc.21818.

Su, C., Shi, K., Cheng, X., Han, Y., Li, Y., Yu, D. and Liu, Z. (2019). **Methylation of CLEC14A is associated with its expression and lung adenocarcinoma progression**. J Cell Physiol *234 (3)*, 2954-2962, doi: 10.1002/jcp.27112.

Sui, B. Q., Zhang, C. D., Liu, J. C., Wang, L. and Dai, D. Q. (2018). **HOXB13 expression and promoter methylation as a candidate biomarker in gastric cancer**. Oncol Lett *15 (6)*, 8833-8840, doi: 10.3892/ol.2018.8371.

Sun, K., Jiang, P., Chan, K. C., Wong, J., Cheng, Y. K., Liang, R. H., Chan, W. K., Ma, E. S., Chan, S. L., Cheng, S. H., Chan, R. W., Tong, Y. K., Ng, S. S., Wong, R. S., Hui, D. S., Leung, T. N., Leung, T. Y., Lai, P. B., Chiu, R. W. and Lo, Y. M. (2015). **Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments**. Proc Natl Acad Sci U S A *112 (40)*, E5503-5512, doi: 10.1073/pnas.1508736112.

Sun, R., Xiang, T., Tang, J., Peng, W., Luo, J., Li, L., Qiu, Z., Tan, Y., Ye, L., Zhang, M., Ren, G. and Tao, Q. (2020). **19q13 KRAB zinc-finger protein ZNF471 activates MAPK10/JNK3 signaling but is frequently silenced by promoter CpG methylation in esophageal cancer**. Theranostics *10 (5)*, 2243-2259, doi: 10.7150/thno.35861.

Suvà, M. L., Rheinbay, E., Gillespie, S. M., Patel, A. P., Wakimoto, H., Rabkin, S. D., Riggi, N., Chi, A. S., Cahill, D. P., Nahed, B. V., Curry, W. T., Martuza, R. L., Rivera, M. N., Rossetti, N., Kasif, S., Beik, S., Kadri, S., Tirosh, I., Wortman, I., Shalek, A. K., Rozenblatt-Rosen, O., Regev, A., Louis, D. N. and Bernstein, B. E. (2014). **Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells**. Cell *157 (3)*, 580-594, doi: 10.1016/j.cell.2014.02.030.

Szulwach, K. E., Li, X., Li, Y., Song, C. X., Han, J. W., Kim, S., Namburi, S., Hermetz, K., Kim, J. J., Rudd, M. K., Yoon, Y. S., Ren, B., He, C. and Jin, P. (2011). **Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells**. PLoS Genet *7 (6)*, e1002154, doi: 10.1371/journal.pgen.1002154.

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L. and Rao, A. (2009). **Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1**. Science *324 (5929)*, 930-935, doi: 10.1126/science.1170116.

Taiwo, O., Wilson, G. A., Morris, T., Seisenberger, S., Reik, W., Pearce, D., Beck, S. and Butcher, L. M. (2012). **Methylome analysis using MeDIP-seq with low DNA concentrations**. Nat Protoc *7 (4)*, 617-636, doi: 10.1038/nprot.2012.012.

Takai, D., Yagi, Y., Wakazono, K., Ohishi, N., Morita, Y., Sugimura, T. and Ushijima, T. (2001). **Silencing of HTR1B and reduced expression of EDN1 in human lung cancers, revealed by methylation-sensitive representational difference analysis**. Oncogene *20 (51)*, 7505-7513, doi: 10.1038/sj.onc.1204940.

Takeshita, H., Sato, M., Shiwaku, H. O., Semba, S., Sakurada, A., Hoshi, M., Hayashi, Y., Tagawa, Y., Ayabe, H. and Horii, A. (1999). **Expression of the DMBT1 gene is frequently suppressed in human lung cancer**. Jpn J Cancer Res *90 (9)*, 903-908, doi: 10.1111/j.1349-7006.1999.tb00833.x.

Takeuchi, K., Choi, Y. L., Togashi, Y., Soda, M., Hatano, S., Inamura, K., Takada, S., Ueno, T., Yamashita, Y., Satoh, Y., Okumura, S., Nakagawa, K., Ishikawa, Y. and Mano, H. (2009). **KIF5B-ALK, a novel fusion oncokinase identified by an immunohistochemistry-based diagnostic system for ALK-positive lung cancer**. Clin Cancer Res *15 (9)*, 3143-3149, doi: 10.1158/1078-0432.Ccr-08-3248.

Tang, X., Yin, X., Xiang, T., Li, H., Li, F., Chen, L. and Ren, G. (2012). **Protocadherin 10 is frequently downregulated by promoter methylation and functions as a tumor suppressor gene in non-small cell lung cancer**. Cancer Biomark *12 (1)*, 11-19, doi: 10.3233/cbm-2012-00280.

Tang, Z., Wang, L., Dai, P., Ruan, P., Liu, D. and Tan, Y. (2019). **TAL1-mediated epigenetic modifications down-regulate the expression of GIMAP family in non-small cell lung cancer**. bioRxiv, 739870, doi: 10.1101/739870.

Thierry, A. R., Mouliere, F., El Messaoudi, S., Mollevi, C., Lopez-Crapez, E., Rolet, F., Gillet, B., Gongora, C., Dechelotte, P., Robert, B., Del Rio, M., Lamy, P. J., Bibeau, F., Nouaille, M., Loriot, V., Jarrousse, A. S., Molina, F., Mathonnet, M., Pezet, D. and Ychou, M. (2014). **Clinical validation of the detection of KRAS and BRAF mutations from circulating tumor DNA**. Nat Med *20 (4)*, 430-435, doi: 10.1038/nm.3511.

Thomson, J. P., Hunter, J. M., Nestor, C. E., Dunican, D. S., Terranova, R., Moggs, J. G. and Meehan, R. R. (2013). **Comparative analysis of affinity-based 5-hydroxymethylation enrichment techniques**. Nucleic Acids Res *41 (22)*, e206, doi: 10.1093/nar/gkt1080.

Thomson, J. P., Lempiäinen, H., Hackett, J. A., Nestor, C. E., Müller, A., Bolognani, F., Oakeley, E. J., Schübeler, D., Terranova, R., Reinhardt, D., Moggs, J. G. and Meehan, R. R. (2012). **Non-genotoxic carcinogen exposure induces defined changes in the 5-hydroxymethylome**. Genome Biol *13 (10)*, R93, doi: 10.1186/gb-2012-13-10-r93.

Thomson, J. P. and Meehan, R. R. (2017). **The application of genome-wide 5-hydroxymethylcytosine studies in cancer research**. Epigenomics *9 (1)*, 77-91, doi: 10.2217/epi-2016-0122.

Tian, X., Sun, B., Chen, C., Gao, C., Zhang, J., Lu, X., Wang, L., Li, X., Xing, Y., Liu, R., Han, X., Qi, Z., Zhang, X., He, C., Han, D., Yang, Y. G. and Kan, Q. (2018). **Circulating tumor DNA 5-hydroxymethylcytosine as a novel diagnostic biomarker for esophageal cancer**. Cell Res *28 (5)*, 597-600, doi: 10.1038/s41422-018-0014-x.

Tie, J., Kinde, I., Wang, Y., Wong, H. L., Roebert, J., Christie, M., Tacey, M., Wong, R., Singh, M., Karapetis, C. S., Desai, J., Tran, B., Strausberg, R. L., Diaz, L. A., Papadopoulos, N., Kinzler, K. W., Vogelstein, B. and Gibbs, P. (2015). **Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer**. Annals of Oncology *26 (8)*, 1715-1722, doi: https://doi.org/10.1093/annonc/mdv177.

Tsou, J. A., Galler, J. S., Siegmund, K. D., Laird, P. W., Turla, S., Cozen, W., Hagen, J. A., Koss, M. N. and Laird-Offringa, I. A. (2007). **Identification of a panel of sensitive and specific DNA methylation markers for lung adenocarcinoma**. Mol Cancer 6, 70, doi: 10.1186/1476-4598-6-70.

Ulz, P., Belic, J., Graf, R., Auer, M., Lafer, I., Fischereder, K., Webersinke, G., Pummer, K., Augustin, H., Pichler, M., Hoefler, G., Bauernhofer, T., Geigl, J. B., Heitzer, E. and Speicher, M. R. (2016). **Whole-genome plasma sequencing reveals focal amplifications as a driving force in metastatic prostate cancer**. Nature Communications *7 (1)*, 12008, doi: 10.1038/ncomms12008.

Vaishnavi, A., Capelletti, M., Le, A. T., Kako, S., Butaney, M., Ercan, D., Mahale, S., Davies, K. D., Aisner, D. L., Pilling, A. B., Berge, E. M., Kim, J., Sasaki, H., Park, S., Kryukov, G., Garraway, L. A., Hammerman, P. S., Haas, J., Andrews, S. W., Lipson, D., Stephens, P. J., Miller, V. A., Varella-Garcia, M., Jänne, P. A. and Doebele, R. C. (2013). **Oncogenic and drug-sensitive NTRK1 rearrangements in lung cancer**. Nat Med *19 (11)*, 1469-1472, doi: 10.1038/nm.3352.

van der Pol, Y. and Mouliere, F. (2019). **Toward the Early Detection of Cancer by Decoding the Epigenetic and Environmental Fingerprints of Cell-Free DNA**. Cancer Cell *36 (4)*, 350-368, doi: 10.1016/j.ccell.2019.09.003.

Vansteenkiste, J., Crinò, L., Dooms, C., Douillard, J. Y., Faivre-Finn, C., Lim, E., Rocco, G., Senan, S., Van Schil, P., Veronesi, G., Stahel, R., Peters, S., Felip, E., Stahel, R., Felip, E., Peters, S., Kerr, K., Besse, B., Vansteenkiste, J., Eberhardt, W., Edelman, M., Mok, T., O'Byrne, K., Novello, S., Bubendorf, L., Marchetti, A., Baas, P., Reck, M., Syrigos, K., Paz-Ares, L., Smit, E. F., Meldgaard, P., Adjei, A., Nicolson, M., Crinò, L., Van Schil, P., Senan, S., Faivre-Finn, C., Rocco, G., Veronesi, G., Douillard, J.-Y., Lim, E., Dooms, C., Weder, W., De Ruysscher, D., Le Pechoux, C., De Leyn, P. and Westeel, V. (2014). **2nd ESMO Consensus Conference on Lung Cancer: early-stage non-small-cell lung cancer consensus on**

diagnosis, treatment and follow-up. Annals of Oncology *25 (8)*, 1462-1474, doi: https://doi.org/10.1093/annonc/mdu089.

Varley, K. E., Gertz, J., Bowling, K. M., Parker, S. L., Reddy, T. E., Pauli-Behn, F., Cross, M. K., Williams, B. A., Stamatoyannopoulos, J. A., Crawford, G. E., Absher, D. M., Wold, B. J. and Myers, R. M. (2013). **Dynamic DNA methylation across diverse human cell lines and tissues**. Genome Res *23 (3)*, 555-567, doi: 10.1101/gr.147942.112.

Velimirovic, M., Juric, D., Niemierko, A., Spring, L., Vidula, N., Wander, S. A., Medford, A., Parikh, A., Malvarosa, G., Yuen, M., Corcoran, R., Moy, B., Isakoff, S. J., Ellisen, L. W., Iafrate, A., Chabner, B. and Bardia, A. (2020). **Rising Circulating Tumor DNA As a Molecular Biomarker of Early Disease Progression in Metastatic Breast Cancer**. JCO Precision Oncology *(4)*, 1246-1262, doi: 10.1200/po.20.00117.

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr. and Kinzler, K. W. (2013). **Cancer genome landscapes**. Science *339 (6127)*, 1546-1558, doi: 10.1126/science.1235122.

Vuorinen, E. M., Rajala, N. K., Ihalainen, T. O. and Kallioniemi, A. (2018). **Depletion of nuclear import protein karyopherin alpha 7 (KPNA7) induces mitotic defects and deformation of nuclei in cancer cells**. BMC Cancer *18 (1)*, 325, doi: 10.1186/s12885-018-4261-5.

Wan, J. C. M., Massie, C., Garcia-Corbacho, J., Mouliere, F., Brenton, J. D., Caldas, C., Pacey, S., Baird, R. and Rosenfeld, N. (2017). **Liquid biopsies come of age: towards implementation of circulating tumour DNA**. Nat Rev Cancer *17 (4)*, 223-238, doi: 10.1038/nrc.2017.7.

Wan, S., Pan, Q., Yang, G., Kuang, J. and Luo, S. (2020). **Role of CYP4F2 as a novel biomarker regulating malignant phenotypes of liver cancer cells via the Nrf2 signaling axis**. Oncol Lett *20 (4)*, 13, doi: 10.3892/ol.2020.11874.

Wang, Q., Wu, J., Huang, H., Jiang, Y., Huang, Y., Fang, H., Zheng, G., Zhou, X., Wu, Y., Lei, C. and Hu, D. (2020). **lncRNA LIFR-AS1 suppresses invasion and metastasis of non-small cell lung cancer via the miR-942-5p/ZNF471 axis**. Cancer Cell Int *20*, 180, doi: 10.1186/s12935-020-01228-5.

Wang, R., Hu, H., Pan, Y., Li, Y., Ye, T., Li, C., Luo, X., Wang, L., Li, H., Zhang, Y., Li, F., Lu, Y., Lu, Q., Xu, J., Garfield, D., Shen, L., Ji, H., Pao, W., Sun, Y. and Chen, H. (2012a). **RET fusions define a unique molecular and clinicopathologic subtype of non-small-cell lung cancer**. J Clin Oncol *30 (35)*, 4352-4359, doi: 10.1200/jco.2012.44.1477.

Wang, T., Pei, X., Zhan, J., Hu, J., Yu, Y. and Zhang, H. (2012b). **FERM-containing protein FRMD5 is a p120-catenin interacting protein that regulates tumor progression**. FEBS Lett *586 (19)*, 3044-3050, doi: 10.1016/j.febslet.2012.07.019.

Wang, Y., Springer, S., Zhang, M., McMahon, K. W., Kinde, I., Dobbyn, L., Ptak, J., Brem, H., Chaichana, K., Gallia, G. L., Gokaslan, Z. L., Groves, M. L., Jallo, G. I., Lim, M., Olivi, A., Quinones-Hinojosa, A., Rigamonti, D., Riggins, G. J., Sciubba, D. M., Weingart, J. D., Wolinsky, J. P., Ye, X., Oba-Shinjo, S. M., Marie, S. K., Holdhoff, M., Agrawal, N., Diaz, L. A., Jr., Papadopoulos, N., Kinzler, K. W., Vogelstein, B. and Bettegowda, C. (2015). **Detection of tumor-derived DNA in cerebrospinal fluid of patients with primary tumors of the brain and spinal cord**. Proc Natl Acad Sci U S A *112 (31)*, 9704-9709, doi: 10.1073/pnas.1511694112.

Watanabe, T., Miura, T., Degawa, Y., Fujita, Y., Inoue, M., Kawaguchi, M. and Furihata, C. (2010). **Comparison of lung cancer cell lines representing four histopathological subtypes with gene expression profiling using quantitative real-time PCR**. Cancer Cell Int *10*, 2, doi: 10.1186/1475-2867-10-2.

Wei, F., Ge, Y., Li, W., Wang, X. and Chen, B. (2020). **Role of endothelin receptor type B (EDNRB) in lung adenocarcinoma**. Thorac Cancer *11 (7)*, 1885-1890, doi: 10.1111/1759-7714.13474.

Weisenberger, D. J. (2014). **Characterizing DNA methylation alterations from The Cancer Genome Atlas**. J Clin Invest *124 (1)*, 17-23, doi: 10.1172/jci69740.

Weiss, G., Schlegel, A., Kottwitz, D., König, T. and Tetzner, R. (2017). **Validation of the SHOX2/PTGER4 DNA Methylation Marker Panel for Plasma-Based Discrimination between Patients with Malignant and Nonmalignant Lung Disease**. J Thorac Oncol *12 (1)*, 77-84, doi: 10.1016/j.jtho.2016.08.123.

Wickham, H. (2016). **ggplot2: Elegant Graphics for Data Analysis.** Springer-Verlag New York.

Winslow, M. M., Dayton, T. L., Verhaak, R. G., Kim-Kiselak, C., Snyder, E. L., Feldser, D. M., Hubbard, D. D., DuPage, M. J., Whittaker, C. A., Hoersch, S., Yoon, S., Crowley, D., Bronson, R. T., Chiang, D. Y., Meyerson, M. and Jacks, T. (2011). **Suppression of lung adenocarcinoma progression by Nkx2-1**. Nature *473 (7345)*, 101-104, doi: 10.1038/nature09881.

Wolf, J., Seto, T., Han, J.-Y., Reguart, N., Garon, E. B., Groen, H. J. M., Tan, D. S. W., Hida, T., de Jonge, M., Orlov, S. V., Smit, E. F., Souquet, P.-J., Vansteenkiste, J., Hochmair, M., Felip, E., Nishio, M., Thomas, M., Ohashi, K., Toyozawa, R., Overbeck, T. R., de Marinis, F., Kim, T.-M., Laack, E., Robeva, A., Le Mouhaer, S., Waldron-Lynch, M., Sankaran, B., Balbin, O. A., Cui, X., Giovannini, M., Akimov, M. and Heist, R. S. (2020). **Capmatinib in MET Exon 14–Mutated or MET-Amplified Non–Small-Cell Lung Cancer**. New England Journal of Medicine *383 (10)*, 944-957, doi: 10.1056/NEJMoa2002787.

Woo, C. G., Seo, S., Kim, S. W., Jang, S. J., Park, K. S., Song, J. Y., Lee, B., Richards, M. W., Bayliss, R., Lee, D. H. and Choi, J. (2017). **Differential protein stability and clinical responses of EML4-ALK fusion variants to various ALK inhibitors in advanced ALK-rearranged non-small cell lung cancer**. Ann Oncol *28 (4)*, 791-797, doi: 10.1093/annonc/mdw693.

Worm Ørntoft, M. B., Jensen, S., Hansen, T. B., Bramsen, J. B. and Andersen, C. L. (2017). **Comparative analysis of 12 different kits for bisulfite conversion of circulating cell-free DNA**. Epigenetics *12 (8)*, 626-636, doi: 10.1080/15592294.2017.1334024.

Wu, X. and Zhang, Y. (2017). **TET-mediated active DNA demethylation: mechanism, function and beyond**. Nat Rev Genet *18 (9)*, 517-534, doi: 10.1038/nrg.2017.33.

Wu, Y., Davison, J., Qu, X., Morrissey, C., Storer, B., Brown, L., Vessella, R., Nelson, P. and Fang, M. (2016). **Methylation profiling identified novel differentially methylated markers including OPCML and FLRT2 in prostate cancer**. Epigenetics *11 (4)*, 247-258, doi: 10.1080/15592294.2016.1148867.

Xu, L., Du, H., Zhang, Q., Wang, C., Yan, L., Tian, G. and Fu, X. (2019). **BAI1-associated protein 2-like 2 is a potential biomarker in lung cancer**. Oncol Rep *41 (2)*, 1304-1312, doi: 10.3892/or.2018.6883.

Yan, X., Hu, Z., Feng, Y., Hu, X., Yuan, J., Zhao, S. D., Zhang, Y., Yang, L., Shan, W., He, Q., Fan, L., Kandalaft, L. E., Tanyi, J. L., Li, C., Yuan, C. X., Zhang, D., Yuan, H., Hua, K., Lu, Y., Katsaros, D., Huang, Q., Montone, K., Fan, Y., Coukos, G., Boyd, J., Sood, A. K., Rebbeck, T., Mills, G. B., Dang, C. V. and Zhang, L. (2015). **Comprehensive Genomic Characterization of Long Non-coding RNAs across Human Cancers**. Cancer Cell *28 (4)*, 529-540, doi: 10.1016/j.ccell.2015.09.006.

Yang, G., Lu, X. and Yuan, L. (2014). **LncRNA: a link between RNA and cancer**. Biochim Biophys Acta *1839 (11)*, 1097-1109, doi: 10.1016/j.bbagrm.2014.08.012.

Yang, H., Liu, Y., Bai, F., Zhang, J. Y., Ma, S. H., Liu, J., Xu, Z. D., Zhu, H. G., Ling, Z. Q., Ye, D., Guan, K. L. and Xiong, Y. (2013). **Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation**. Oncogene *32 (5)*, 663-669, doi: 10.1038/onc.2012.67.

Yang, H., Ye, D., Guan, K. L. and Xiong, Y. (2012). **IDH1 and IDH2 mutations in tumorigenesis: mechanistic insights and clinical perspectives**. Clin Cancer Res *18 (20)*, 5562-5571, doi: 10.1158/1078-0432.Ccr-12-1773.

Yang, J. C., Wu, Y. L., Schuler, M., Sebastian, M., Popat, S., Yamamoto, N., Zhou, C., Hu, C. P., O'Byrne, K., Feng, J., Lu, S., Huang, Y., Geater, S. L., Lee, K. Y., Tsai, C. M., Gorbunova, V., Hirsh, V., Bennouna, J., Orlov, S., Mok, T., Boyer, M., Su, W. C., Lee, K. H., Kato, T., Massey, D., Shahidi, M., Zazulina, V. and Sequist, L. V. (2015). **Afatinib versus cisplatin-based chemotherapy for EGFR mutation-positive lung adenocarcinoma (LUX-Lung 3 and LUX-Lung 6): analysis of overall survival data from two randomised, phase 3 trials**. Lancet Oncol *16 (2)*, 141-151, doi: 10.1016/s1470-2045(14)71173-8.

Yeung, Y. T., Fan, S., Lu, B., Yin, S., Yang, S., Nie, W., Wang, M., Zhou, L., Li, T., Li, X., Bode, A. M. and Dong, Z. (2020). **CELF2 suppresses non-small cell lung carcinoma growth by inhibiting the PREX2-PTEN interaction**. Carcinogenesis *41 (3)*, 377-389, doi: 10.1093/carcin/bgz113.

Yilmaz, A. and Grotewold, E. (2010). **Components and mechanisms of regulation of gene expression**. Methods Mol Biol *674*, 23-32, doi: 10.1007/978-1-60761-854-6_2.

References

Yin, W., Wang, X., Li, Y., Wang, B., Song, M., Hulbert, A., Chen, C. and Yu, F. (2020). **Promoter hypermethylation of cysteine dioxygenase type 1 in patients with non-small cell lung cancer**. Oncol Lett *20 (1)*, 967-973, doi: 10.3892/ol.2020.11592.

Yokoyama, S., Higashi, M., Tsutsumida, H., Wakimoto, J., Hamada, T., Wiest, E., Matsuo, K., Kitazono, I., Goto, Y., Guo, X., Hamada, T., Yamada, S., Hiraki, T., Yonezawa, S., Batra, S. K., Hollingsworth, M. A. and Tanimoto, A. (2017). **TET1-mediated DNA hypomethylation regulates the expression of MUC4 in lung cancer**. Genes Cancer *8 (3-4)*, 517-527, doi: 10.18632/genesandcancer.139.

Yu, C., Liu, Z., Chen, Q., Li, Y., Jiang, L., Zhang, Z. and Zhou, F. (2018). **Nkx2.8 Inhibits Epithelial-Mesenchymal Transition in Bladder Urothelial Carcinoma via Transcriptional Repression of Twist1**. Cancer Res *78 (5)*, 1241-1252, doi: 10.1158/0008-5472.Can-17-1545.

Yu, G., Wang, L. G., Han, Y. and He, Q. Y. (2012). **clusterProfiler: an R package for comparing biological themes among gene clusters**. Omics *16 (5)*, 284-287, doi: 10.1089/omi.2011.0118.

Zeng, Q., Dai, Y., Duan, C., Zeng, R., Zeng, Q. and Wei, C. (2020). **Long noncoding RNA POU3F3 enhances cancer cell proliferation, migration and invasion in non-small cell lung cancer (adenocarcinoma) by downregulating microRNA-30d-5p**. BMC Pulm Med *20 (1)*, 185, doi: 10.1186/s12890-020-01218-3.

Zhang, D.-L., Qu, L.-W., Ma, L., Zhou, Y.-C., Wang, G.-Z., Zhao, X.-C., Zhang, C., Zhang, Y.-F., Wang, M., Zhang, M.-Y., Yu, H., Sun, B.-B., Gao, S.-H., Cheng, X., Guo, M.-Z., Huang, Y.-C. and Zhou, G.-B. (2018a). **Genome-wide identification of transcription factors that are critical to non-small cell lung cancer**. Cancer Letters *434*, 132-143, doi: https://doi.org/10.1016/j.canlet.2018.07.020.

Zhang, E. W., Dagogo-Jack, I., Kuo, A., Rooney, M. M., Shaw, A. T. and Digumarthy, S. R. (2020a). **Association between circulating tumor DNA burden and disease burden in patients with ALK-positive lung cancer**. Cancer *126 (20)*, 4473-4484, doi: 10.1002/cncr.33118.

Zhang, G., Scarborough, H., Kim, J., Rozhok, A. I., Chen, Y. A., Zhang, X., Song, L., Bai, Y., Fang, B., Liu, R. Z., Koomen, J., Tan, A. C., Degregori, J. and Haura, E. B. (2016). **Coupling an EML4-ALK-centric interactome with RNA interference identifies sensitizers to ALK inhibitors**. Sci Signal *9 (450)*, rs12, doi: 10.1126/scisignal.aaf5011.

Zhang, J., Han, X., Gao, C., Xing, Y., Qi, Z., Liu, R., Wang, Y., Zhang, X., Yang, Y. G., Li, X., Sun, B. and Tian, X. (2018b). **5-Hydroxymethylome in Circulating Cell-free DNA as A Potential Biomarker for Non-small-cell Lung Cancer**. Genomics Proteomics Bioinformatics *16 (3)*, 187-199, doi: 10.1016/j.gpb.2018.06.002.

Zhang, J., Lee, D., Dhiman, V., Jiang, P., Xu, J., McGillivray, P., Yang, H., Liu, J., Meyerson, W., Clarke, D., Gu, M., Li, S., Lou, S., Xu, J., Lochovsky, L., Ung, M., Ma, L., Yu, S., Cao, Q., Harmanci, A., Yan, K. K., Sethi, A., Gürsoy, G., Schoenberg, M. R., Rozowsky, J., Warrell, J., Emani, P., Yang, Y. T., Galeev, T., Kong, X., Liu, S., Li, X., Krishnan, J., Feng, Y., Rivera-Mulia, J. C., Adrian, J., Broach, J. R., Bolt, M., Moran, J., Fitzgerald, D., Dileep, V., Liu, T., Mei, S., Sasaki, T., Trevilla-Garcia, C., Wang, S., Wang, Y., Zang, C., Wang, D., Klein, R. J., Snyder, M., Gilbert, D. M., Yip, K., Cheng, C., Yue, F., Liu, X. S., White, K. P. and Gerstein, M. (2020b). **An integrative ENCODE resource for cancer genomics**. Nat Commun *11 (1)*, 3696, doi: 10.1038/s41467-020-14743-w.

Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G. L., Luo, C., Jiang, H. and He, C. (2012). **Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA**. Nat Chem Biol *8 (4)*, 328-330, doi: 10.1038/nchembio.914.

Zhang, W., Wang, S., Zhang, X., Liu, K., Song, J., Leng, X., Luo, R. and Ran, L. (2019a). **Transmembrane Channel-Like 5 (TMC5) promotes prostate cancer cell proliferation through cell cycle regulation**. Biochimie *165*, 115-122, doi: 10.1016/j.biochi.2019.07.017.

Zhang, W., Wu, Y., Hou, B., Wang, Y., Deng, D., Fu, Z. and Xu, Z. (2019b). **A SOX9-AS1/miR-5590-3p/SOX9 positive feedback loop drives tumor growth and metastasis in hepatocellular carcinoma through the Wnt/β-catenin pathway**. Mol Oncol *13 (10)*, 2194-2210, doi: 10.1002/1878-0261.12560.

Zhang, Y., Huang, J., Zou, Q., Che, J., Yang, K., Fan, Q., Qian, D., Wu, J., Bao, E., Song, L. and Zhang, F. (2020c). **Methylated PTGER4 is better than CA125, CEA, Cyfra211 and NSE as a therapeutic response assessment marker in stage IV lung cancer**. Oncol Lett *19 (4)*, 3229-3238, doi: 10.3892/ol.2020.11434.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W. and Liu, X. S. (2008). **Model-based analysis of ChIP-Seq (MACS)**. Genome Biol 9 *(9)*, R137, doi: 10.1186/gb-2008-9-9-r137.

Zhao, L., Mao, C., Xiao, H., Fu, P., Xiao, H., Chen, C. and Wang, G. (2019). **Simultaneous expression of TTF1 and GATA3 in a lung biopsy sample: confusion in diagnostic pathology**. Int J Clin Exp Pathol *12 (9)*, 3613-3619.

Zhuang, L., Ding, W., Ding, W., Zhang, Q., Xu, X. and Xi, D. (2021). **lncRNA ZNF667-AS1 (NR_036521.1) inhibits the progression of colorectal cancer via regulating ANK2/JAK2 expression**. J Cell Physiol 236 *(3)*, 2178-2193, doi: 10.1002/jcp.30004.

Zou, J. X., Duan, Z., Wang, J., Sokolov, A., Xu, J., Chen, C. Z., Li, J. J. and Chen, H. W. (2014). **Kinesin family deregulation coordinated by bromodomain protein ANCCA and histone methyltransferase MLL for breast cancer cell growth, survival, and tamoxifen resistance**. Mol Cancer Res *12 (4)*, 539-549, doi: 10.1158/1541-7786.Mcr-13-0459.

9 Publications

Müller, S.*, **Janke, F.***, Dietz S, Sültmann, H. (2020). **Circulating MicroRNAs as Potential Biomarkers for Lung Cancer.** *Recent Results Cancer Res. 215:299-318*. https://doi.org/10.1007/978-3-030-26439-0_16

Janke, F.*, Bozorgmehr, F.*, Wrenger, S., Dietz, S., Heussel, C. P., Heussel, G., Silva, C. F., Rheinheimer, S., Feisst, M., Thomas, M., Golpon, H., Günther, A., Sültmann, H., Muley, T., Janciauskiene, S., Meister, M., & Schneider, M. A. (2020). **Novel Liquid Biomarker Panels for A Very Early Response Capturing of NSCLC Therapies in Advanced Stages.** *Cancers*, *12*(4), 954. https://doi.org/10.3390/cancers12040954

Dietz, S., Christopoulos, P., Yuan, Z., Angeles, A. K., Gu, L., Volckmar, A. L., Ogrodnik, S. J., **Janke, F.**, Fratte, C. D., Zemojtel, T., Schneider, M. A., Kazdal, D., Endris, V., Meister, M., Muley, T., Cecchin, E., Reck, M., Schlesner, M., Thomas, M., Stenzinger, A., … Sültmann, H. (2020). **Longitudinal therapy monitoring of ALK-positive lung cancer by combined copy number and targeted mutation profiling of cell-free DNA.** *EbioMedicine*, *62*, 103103. https://doi.org/10.1016/j.ebiom.2020.103103

Christopoulos, P., Dietz, S., Angeles, A. K., Rheinheimer, S., Kazdal, D., Volckmar, A. L., **Janke, F.**, Endris, V., Meister, M., Kriegsmann, M., Zemojtel, T., Reck, M., Stenzinger, A., Thomas, M., & Sültmann, H. (2021). **Earlier extracranial progression and shorter survival in ALK-rearranged lung cancer with positive liquid rebiopsies.** *Translational lung cancer research*, *10*(5), 2118–2131. https://doi.org/10.21037/tlcr-21-32

Publications under revision

Angeles, A. K., Christopoulos, P., Yuan, Z., Bauer, S., **Janke, F.**, Ogrodnik S. J., Schlesner, M., Meister, M., Schneider, M. A., Dietz, S., Stenzinger, A., Thomas, M., Sültmann, H. (2021). **Identification of early disease progression in ALK-rearranged lung cancer using circulating tumor DNA analysis.** Under review.

Publications in preparation

Angeles, A. K.*, **Janke, F.***, Bauer, S.*, Riediger, A. L., Christopoulos, P., Sültmann, H. (2021) **Liquid biopsies beyond mutations: Genomic and epigenomic features of cell-free DNA in cancer.** Publication in preparation.

Feuerstein-Akgoz, C., **Janke, F.**, Kessler, L., Miehle, F., Aviles-Huerta, D., Hey, J., Toth, R., Breuer, K., Assenov, Y., Adamski, J., Sültmann, H., Plass, C., Gerhauser, C. (2021). **Whole-genome DNA hydroxymethylation dynamics of human adipocyte differentiation.** Publication in preparation.

* shared first authorship

10 Appendix
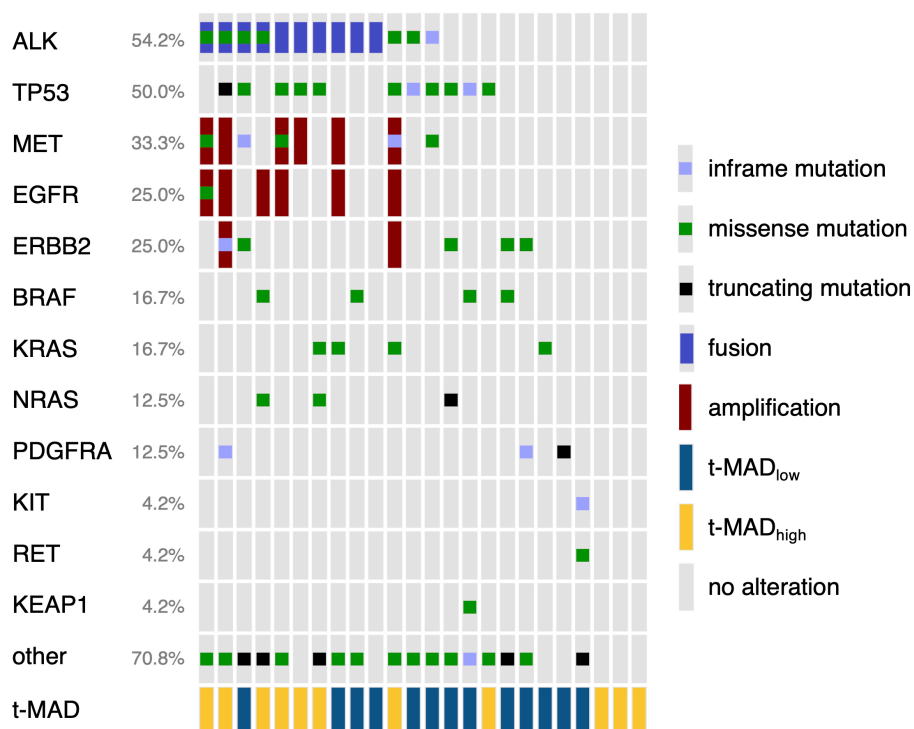
Figure S1: Oncoprint of molecular alterations determined by targeted panel and shallow whole-genome sequencing (Dietz et al. 2020). Molecular alterations and t-MAD scores are represented for 24 NSCLC patients for whom both genomic and 5(h)mC data are available. Alterations are noted if they were present in at least one plasma sample. Patients exceeding the third quartile of the cohort t-MAD (0.0174) in at least one sample were graded t-MAD$_{high}$ and otherwise t-MAD$_{low}$. Percentages indicate the frequency of molecular alterations in the cohort. t-MAD, trimmed median absolute deviation from copy number neutrality.
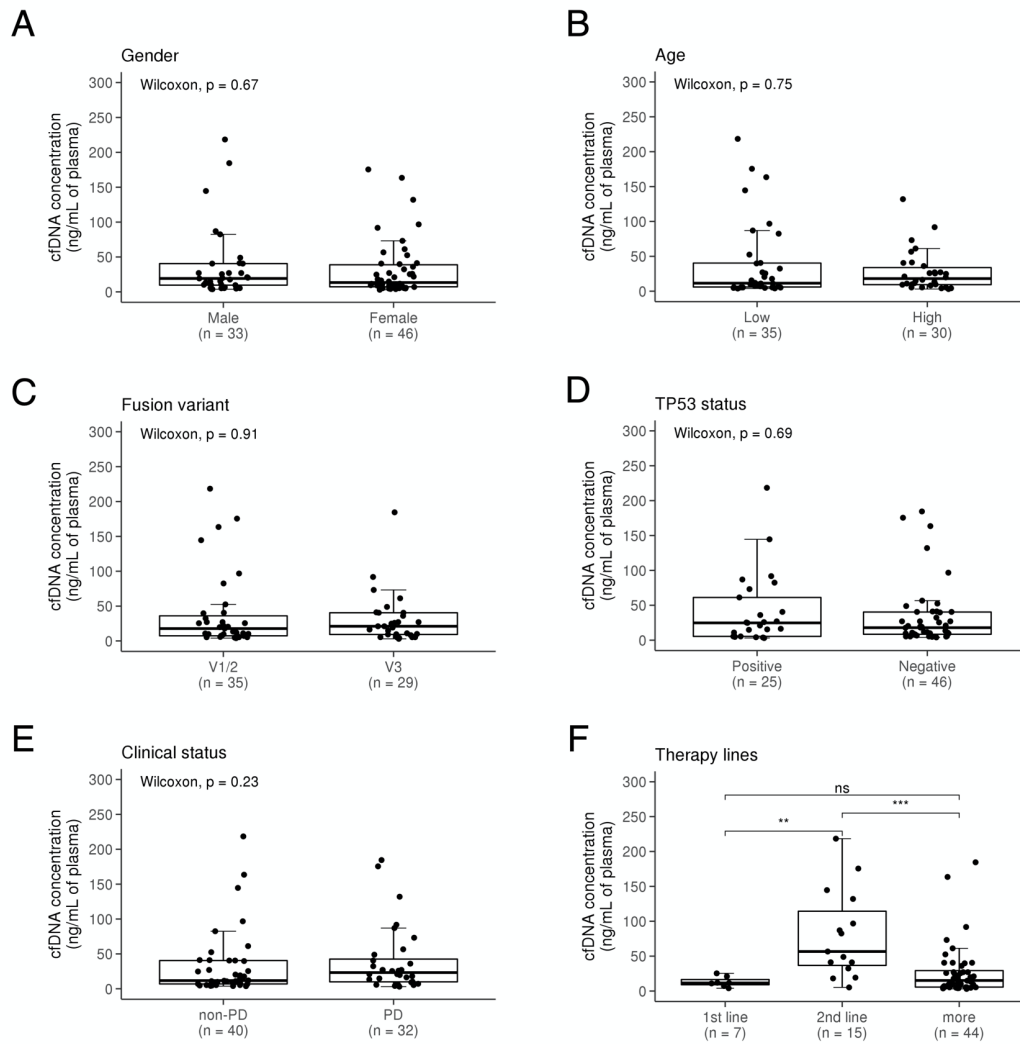
Figure S2: cfDNA concentration comparison. cfDNA concentration (ng/mL of plasma) by (A) gender, (B) age, (C) ALK fusion variant, (D) TP53 status, (E) clinical status and (F) number of previous therapy lines. Age groups refer to the mean cohort age of 57.5 years (Low: ≤57.5 / High: >57.5). Each dot represents one cfDNA sample. cfDNA, cell-free DNA; PD, progressive disease; V1/2, ALK fusion variant 1 or 2; V3, ALK fusion variant 3.

A

B
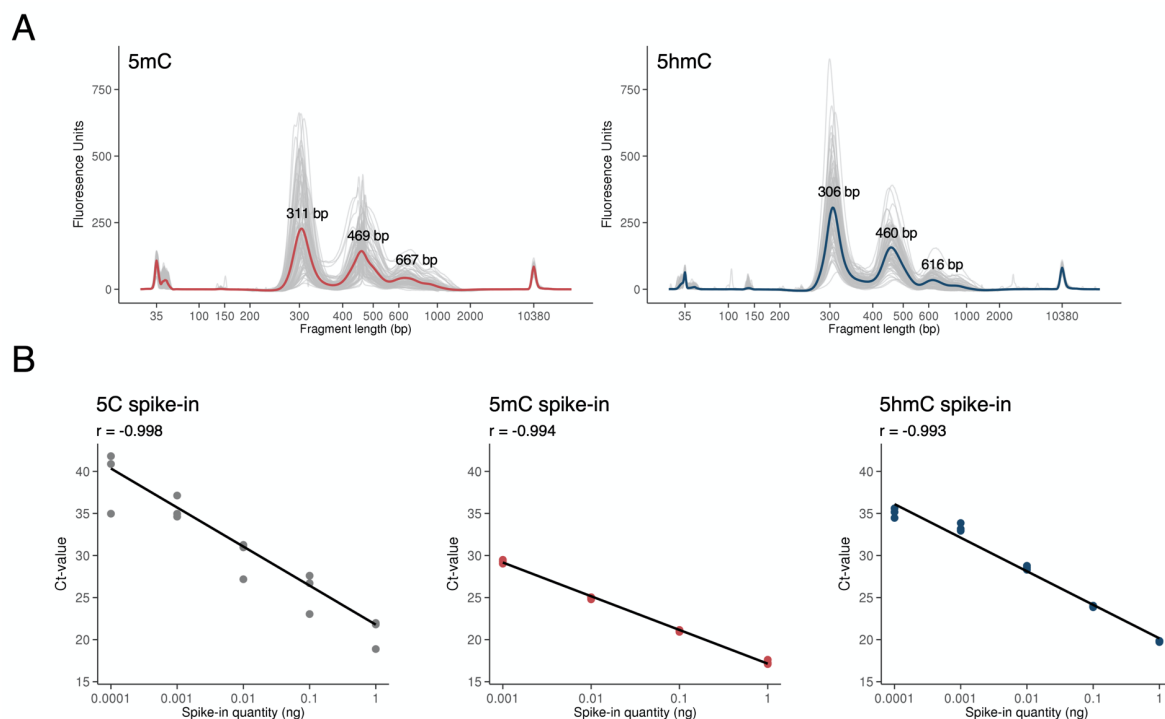
Figure S3: Integrity of 5(h)mC sequencing libraries and spike-in amplicon quantification standard curves. (A) Integrity and fragment length distribution of 5mC (left) and 5hmC (right) sequencing libraries. The median fragment length of all cfDNA samples is indicated in red (n = 87 samples) and blue (n = 93). Gray lines show size profiles of individual samples. Illustrated peak sizes refer to the median of all samples. (B) Standard curves for the quantification of unmodified (5C), methylated (5mC) and hydroxymethylated (5hmC) spike-ins. Each dot represents a technical replicate, linear repression lines refer to their median, and Pearson correlation coefficients are illustrated. cfDNA, cell-free DNA; Ct, cycle threshold.
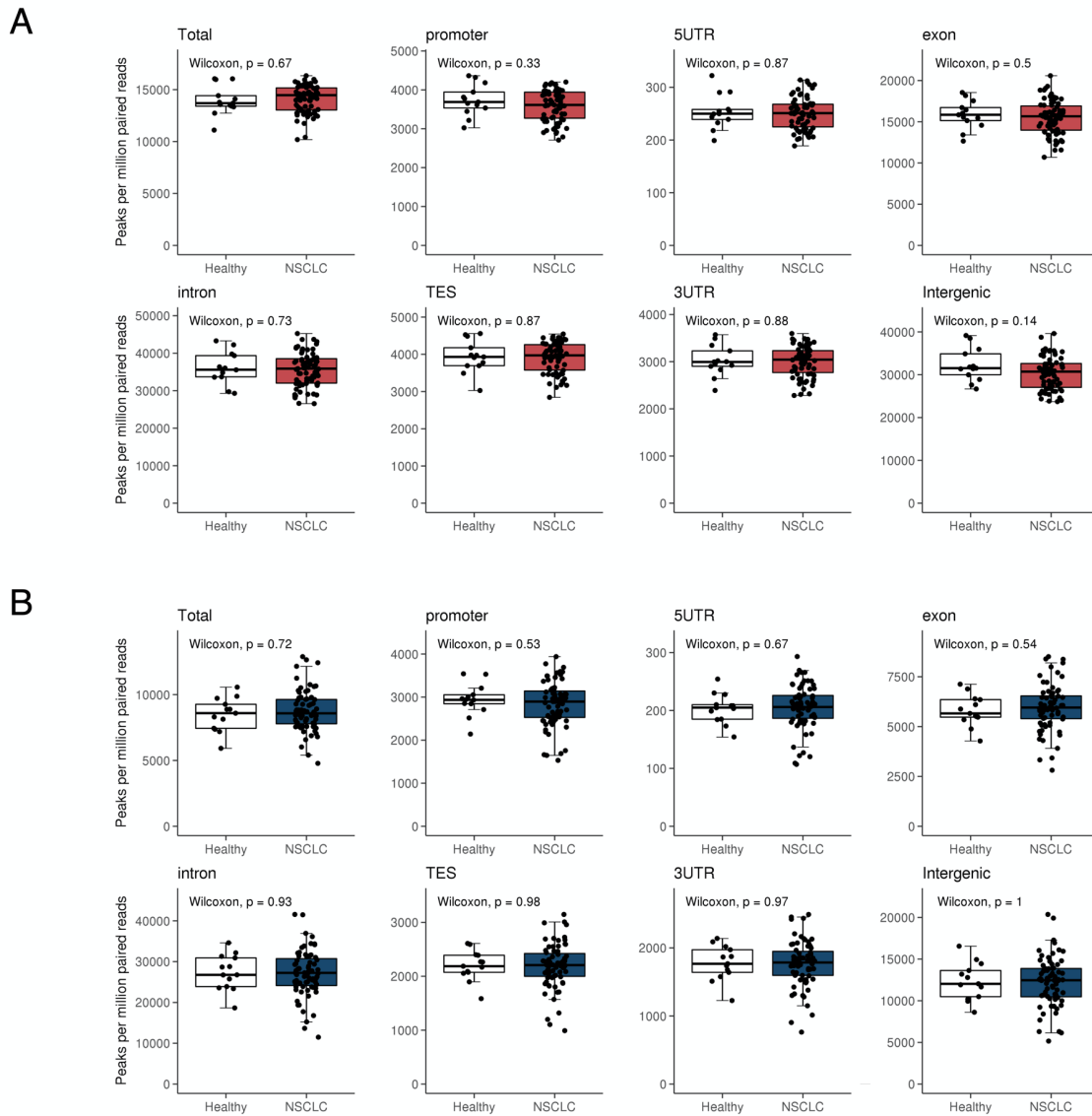
Figure S4: Peak number comparison between healthy control and NSCLC patient samples at genomic features. Individual boxplots illustrate the number of 5mC (A) and 5hmC (B) peaks per million paired reads throughout the entire genome (total) and at different genomic features. Plasma from healthy controls (n = 13) is compared to NSCLC patient samples (n = 71). TES, transcription end site; UTR, untranslated region.
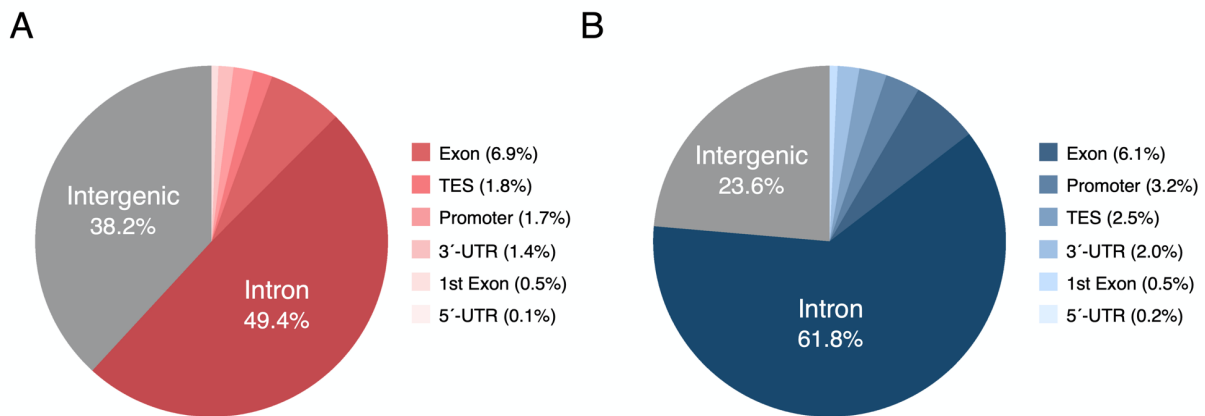
Figure S5: Relative abundance of 5(h)mC peaks at genomic features. Relative abundance of 5mC (A) and 5hmC (B) peaks at gene associated and intergenic regions. TES, transcription end site; UTR, untranslated region.
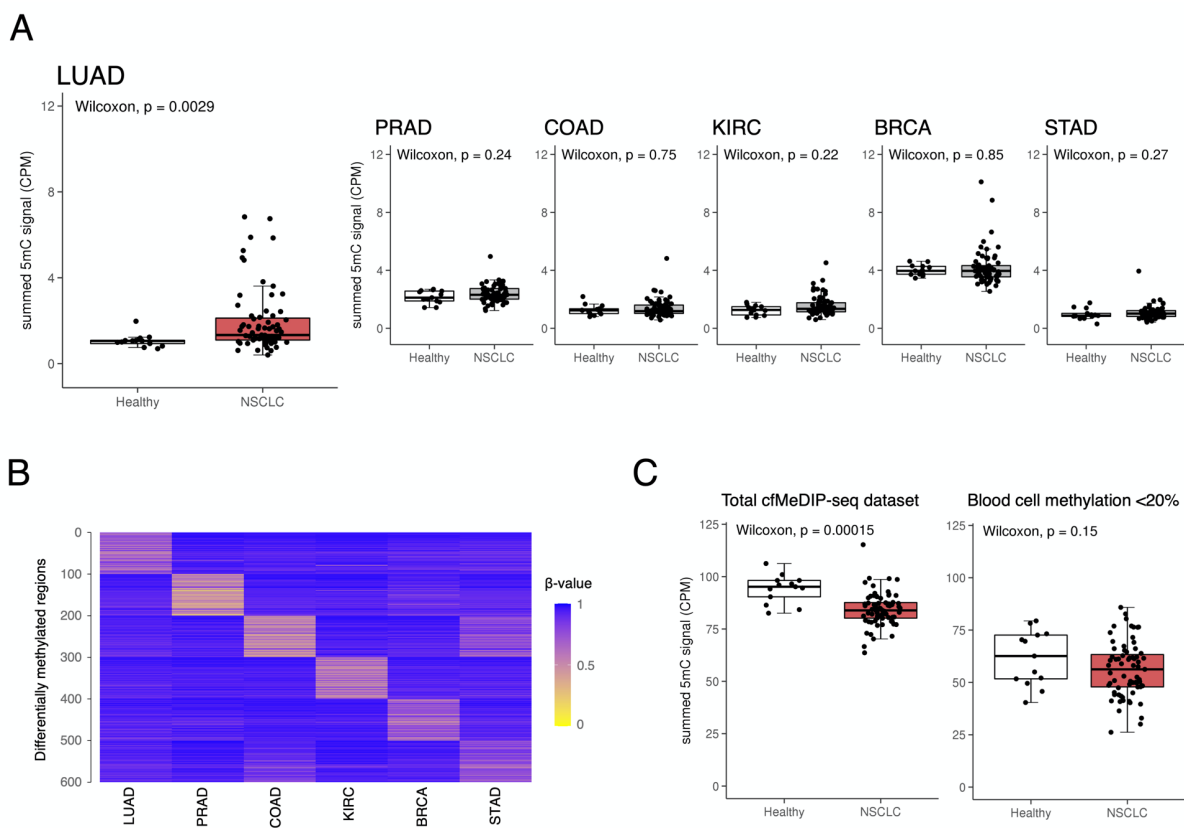
Figure S6: Cancer association of 5mC signals without the exclusion of genomic regions hypermethylated in blood cells. (A) Summed 5mC signal at the top 100 entity-unique hypermethylated genomic regions (illustrated in Figure 16B) for the six most common cancer types determined from TCGA 450k methylation array data. All 100 sites per entity were considered without exclusion of genomic regions hypermethylated in blood cells (i.e. monocytes, neutrophils and erythroid progenitors). Boxplots compare the 5mC signal between healthy controls (n = 13) and NSCLC patient plasma (n = 71) with each dot representing one sample. (B) Heatmap of the top 100 uniquely hypomethylated genomic regions in the six most common cancer entities determined from TCGA methylation data. (C) The summed 5mC signal at genomic regions uniquely hypomethylated in LUAD without (left) and with (right) exclusion of genomic windows hypermethylated in blood cells. BRCA, breast carcinoma; cfMeDIP-seq, cell-free methylation DNA immunoprecipitation sequencing; COAD, colon adenocarcinoma; CPM, counts per million; KIRC, kidney renal clear cell carcinoma; LUAD, lung adenocarcinoma; NSCLC, non-small cell lung cancer; PRAD, prostate adenocarcinoma; STAD, stomach adenocarcinoma; TCGA, The Cancer Genome Atlas.
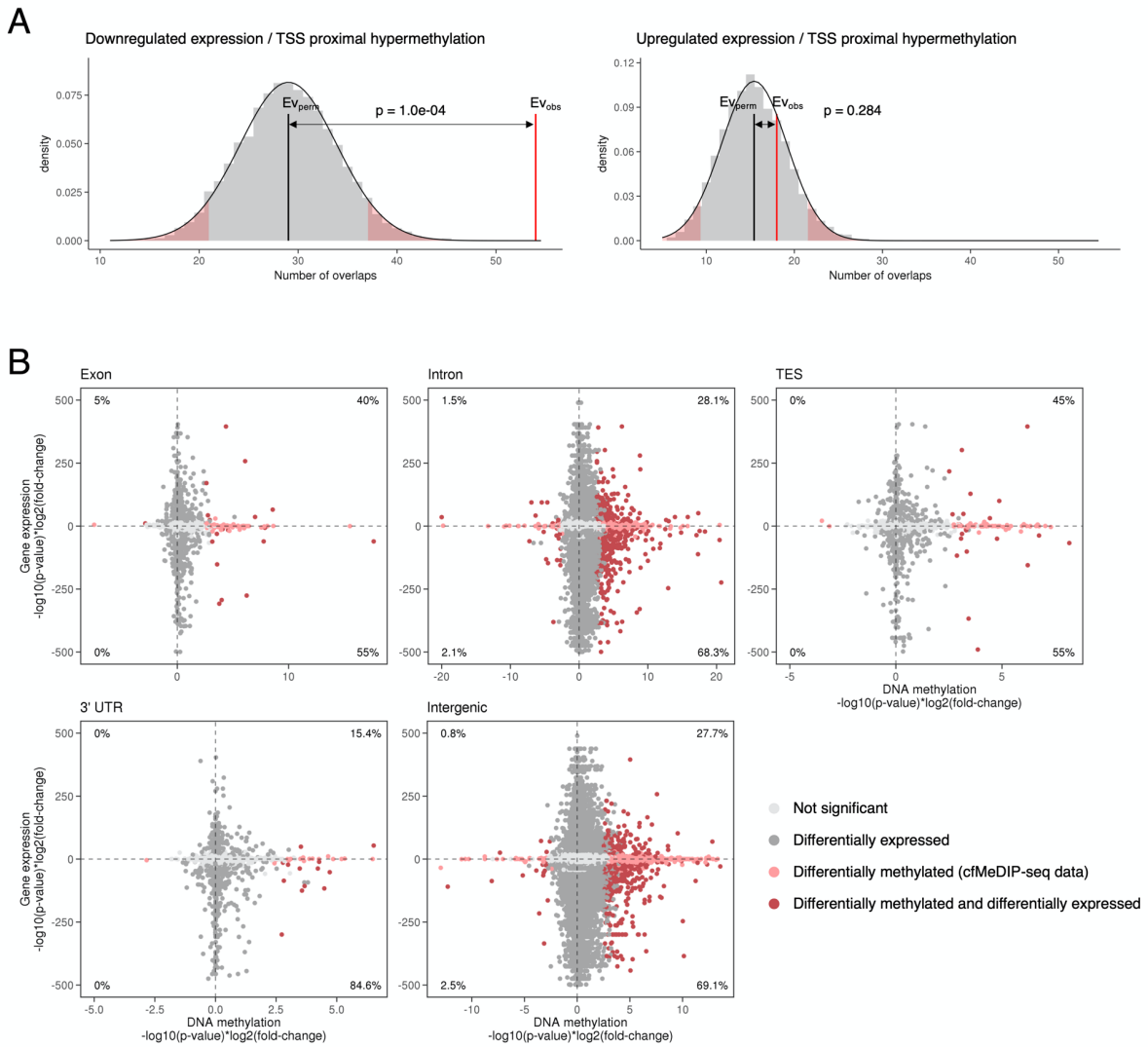
Figure S7: Permutation test and relationship between 5mC signals and gene expression in LUAD tissue. (A) Permutation analysis of the overlap between differentially hypermethylated regions in NSCLC patient samples (mapping to promoter regions, 5′-UTRs and first exons) and genes down- (left) or upregulated in LUAD compared to normal tissue types (right). Histograms show the expected null distribution of overlaps from 10,000 permutations. The mean number of overlaps from all permutation events (Ev_{perm}) is indicated as black line and the observed number of overlaps (Ev_{obs}) is represented as red line. (B) Starburst plots showing the number of differentially methylated regions (DMRs) with respect to the expression status of associated genes in LUAD compared to normal tissue types. Each plot illustrates DMRs mapping to the indicated genomic feature. cfMeDIP, cell-free methylation DNA immunoprecipitation sequencing; TSS, transcription start site; UTR, untranslated region.

A

Diagnostic markers

23   21   10

Prognostic markers

Diagnostic and prognostic 5mC biomarkers

1. *HOXA10-AS*   9. *SYT6*   17. *EVX2*
2. *SOX9-AS1*   10. *HOXA9*   18. *HOXA-AS1*
3. *PRAC1*   11. *CDO1*   19. *FOXG1-AS1*
4. *PTGER4*   12. *POU3F3*   20. *HOXA11-AS*
5. *PCDH10*   13. *PRDM14*   21. *ST8SIA3*
6. *NXPH1*   14. *QRFPR*
7. *CELF2*   15. *PAX6*
8. *HTR1B*   16. *CLEC14A*

B
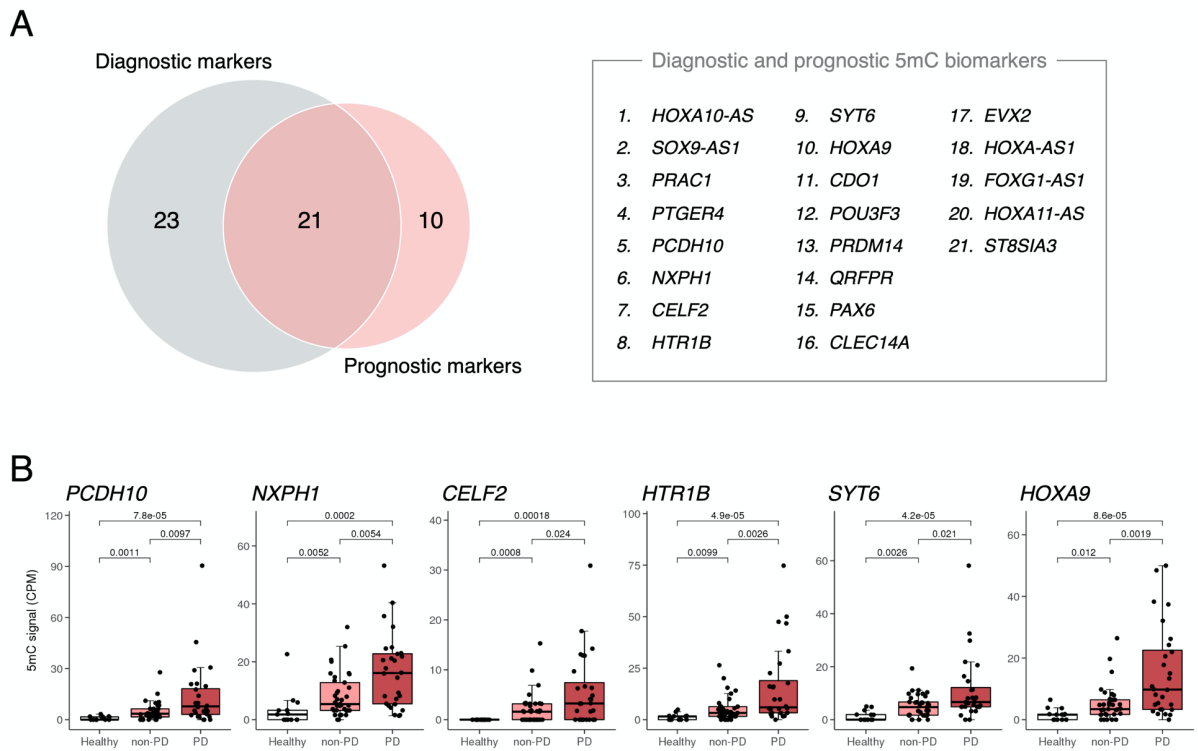
Figure S8: Diagnostic and prognostic 5mC biomarker candidates. (A) Overlap between diagnostic and prognostic 5mC biomarker candidates. Diagnostic markers were defined to present a significantly increased 5mC signal at PD but also non-PD plasma samples compared to samples from healthy controls. Prognostic markers show a higher 5mC signal in PD compared to non-PD samples. Genes associated with 5mC biomarkers with both diagnostic and prognostic value (n = 21) are listed alongside the Venn diagram (ordered by ascending p-values). (B) Illustration of further 5mC biomarker candidates capable of differentiating healthy samples (n = 13) from non-PD (n = 37) and non-PD from PD (n = 28) samples, extending the markers shown in Figure 22B. PD, progressive disease.
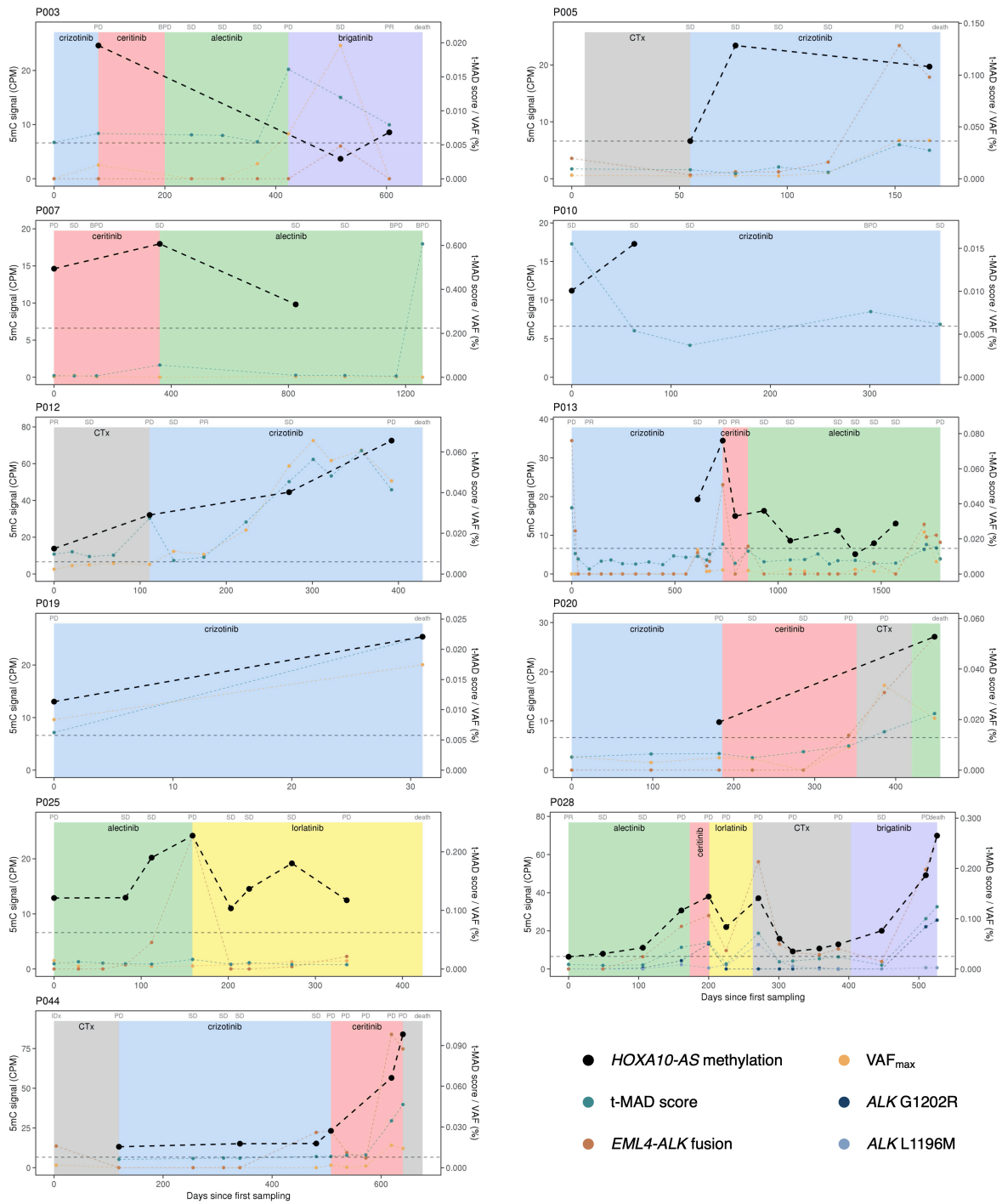
Figure S9: *HOXA10-AS* 5mC signal kinetics followed throughout therapy of ALK⁺ NSCLC patients. Dynamic changes of the plasma 5mC signal at *HOXA10-AS* throughout ALK-directed tyrosine kinase inhibitor treatment and chemotherapy are illustrated for eleven patients with at least two longitudinal plasma samples available. Kinetics of molecular alterations and the t-MAD score – measured from the same plasma samples – are shown alongside the 5mC signal. Each dot represents one plasma sample. Radiological disease assessment is indicated above each plot and therapy lines are shown by the background colors. ALK, anaplastic lymphoma kinase; BPD, brain progressive disease; CPM, counts per million; CTx, chemotherapy; EML4, echinoderm microtubule-associated protein-like 4; Idx, initial diagnosis; PD, progressive disease; PR, partial response; SD, stable disease; t-MAD, trimmed median absolute deviation from copy number neutrality; VAF_max, maximal variant allele frequency.

Figure S10: *PRAC1* 5mC signal kinetics followed throughout therapy of ALK+ NSCLC patients. Dynamic changes of the plasma 5mC signal at *PRAC1* throughout ALK-directed tyrosine kinase inhibitor treatment and chemotherapy are illustrated for eleven patients with at least two longitudinal plasma samples available. Kinetics of molecular alterations and the t-MAD score – measured from the same plasma samples – are shown alongside the 5mC signal. Each dot represents one plasma sample. Radiological disease assessment is indicated above each plot and therapy lines are shown by the background colors. ALK, anaplastic lymphoma kinase; BPD, brain progressive disease; CPM, counts per million; CTx, chemotherapy; EML4, echinoderm microtubule-associated protein-like 4; Idx, initial diagnosis; PD, progressive disease; PR, partial response; SD, stable disease; t-MAD, trimmed median absolute deviation from copy number neutrality; VAF$_{max}$, maximal variant allele frequency.

Figure S11: *PTGER4* 5mC signal kinetics followed throughout therapy of ALK⁺ NSCLC patients. Dynamic changes of the plasma 5mC signal at *PTGER4* throughout ALK-directed tyrosine kinase inhibitor treatment and chemotherapy are illustrated for eleven patients with at least two longitudinal plasma samples available. Kinetics of molecular alterations and the t-MAD score – measured from the same plasma samples – are shown alongside the 5mC signal. Each dot represents one plasma sample. Radiological disease assessment is indicated above each plot and therapy lines are shown by the background colors. ALK, anaplastic lymphoma kinase; BPD, brain progressive disease; CPM, counts per million; CTx, chemotherapy; EML4, echinoderm microtubule-associated protein-like 4; IDx, initial diagnosis; PD, progressive disease; PR, partial response; SD, stable disease; t-MAD, trimmed median absolute deviation from copy number neutrality; VAFmax, maximal variant allele frequency.
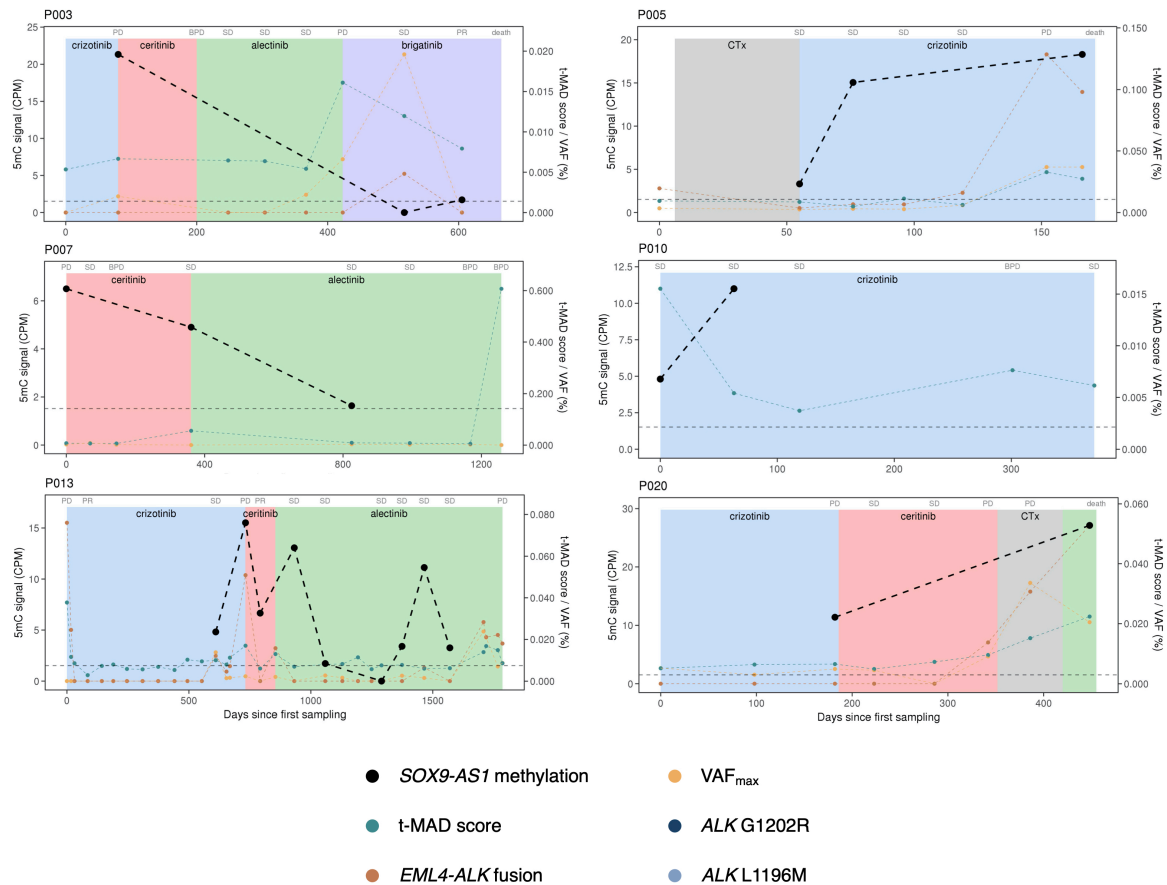
Figure S12: *SOX9-AS1* 5mC signal kinetics followed throughout therapy of ALK⁺ NSCLC patients. Dynamic changes of the plasma 5mC signal at *SOX9-AS1* throughout ALK-directed tyrosine kinase inhibitor treatment and chemotherapy are illustrated for six out of eleven patients with at least two longitudinal plasma samples available. The remaining five patient time courses are shown in Figure 24. Kinetics of molecular alterations and the t-MAD score – measured from the same plasma samples – are shown alongside the 5mC signal. Each dot represents one plasma sample. Radiological disease assessment is indicated above each plot and therapy lines are shown by the background colors. ALK, anaplastic lymphoma kinase; BPD, brain progressive disease; CPM, counts per million; CTx, chemotherapy; EML4, echinoderm microtubule-associated protein-like 4; IDx, initial diagnosis; PD, progressive disease; PR, partial response; SD, stable disease; t-MAD, trimmed median absolute deviation from copy number neutrality; VAF_max, maximal variant allele frequency.
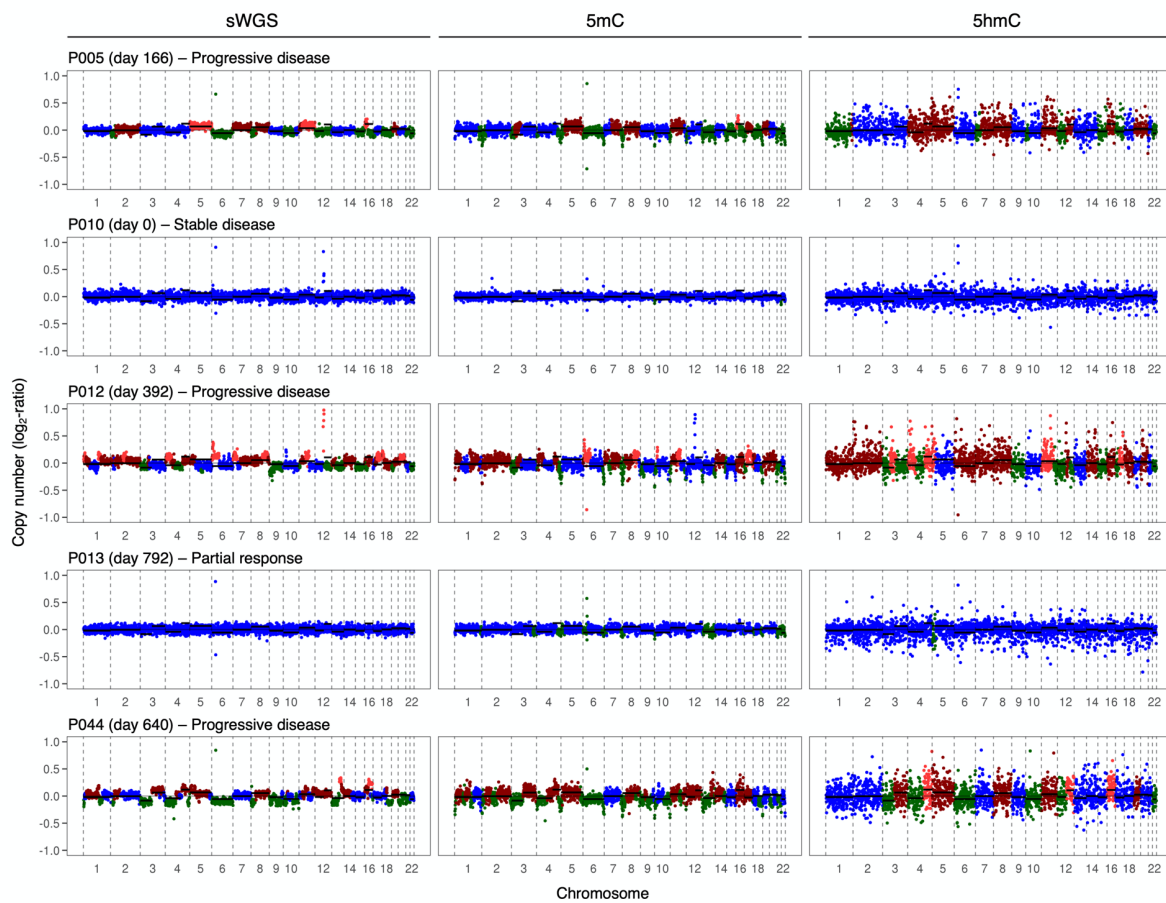
Figure S13: Copy number profiles inferred from sWGS, 5mC- and 5hmC-enriched sequencing data. Copy number profiles generated from shallow whole-genome (sWGS; left), 5mC-enriched (middle) and 5hmC-enriched sequencing datasets (right) of patients P005, P010, P012, P013, and P044. Sampling time points were chosen to exemplify progressive, stable and responsive disease states. Colors indicate copy number neutrality (blue), deletions (green), copy number gain (brown; 3 copies), and amplifications (red; >3 copies).

Table S1: Individual patient characteristics

Patient ID	Gender (m/f)	Age	Stage	*EML4-ALK* fusion variant	TP53 status	Number of samples	5hmC	5mC	SNV	CNA
P001	m	54	IV	V3	positive	1	X	X	X	X
P002	f	51	IV	V1	negative	1	X	X	X	X
P003	m	46	IV	E9A10	positive	4	X	X	X	X
P004	f	74	IV	V3	negative	1	X	-	X	X
P005	m	56	IV	V1	positive	3	X	X	X	X
P007	f	56	IV	V2	negative	3	X	X	X	X
P010	f	56	IV	V3	negative	2	X	X	X	X
P012	f	67	IV	K9A20	negative	5	X	X	X	X
P013	f	53	IV	V1	negative	9	X	X	X	X
P014	m	63	IV	NA	NA	1	X	X	X	X
P015	f	48	IV	V1	negative	1	X	X	X	X
P016	m	71	IV	V1	negative	1	X	-	X	X
P019	f	69	IV	V3	positive	2	X	X	X	X
P020	m	64	IV	V1	negative	3	X	X	X	X
P021	f	80	IV	V2	positive	1	X	X	X	X
P023	m	47	IV	V2	negative	1	X	X	X	X
P025	m	52	IV	V1	negative	8	X	X	X	X
P027	m	51	IV	V1	negative	1	X	X	X	X
P028	f	57	IV	V3	positive	14	X	X	X	X
P031	f	44	IV	V1	negative	1	X	X	X	X
P032	m	65	IV	V1	negative	1	X	X	X	X
P037	m	42	IV	V3	negative	1	X	X	X	X
P044	m	NA	NA	V3	negative	6	X	X	X	X
P046	m	NA	NA	V3	negative	1	X	X	X	X
P047	f	NA	IV	V1	NA	1	X	X	-	-
P071	f	NA	IV	V3	NA	1	X	X	-	-
P102	f	NA	IV	NA	NA	1	X	X	-	-
P163	f	NA	IV	NA	NA	1	X	X	-	-
P166	m	NA	IV	NA	NA	1	X	X	-	-
P169	f	NA	IV	NA	NA	1	X	X	-	-
P174	f	NA	IV	NA	NA	1	X	X	-	-

ALK, anaplastic lymphoma kinase; EML4, echinoderm microtubule-associated protein-like 4, NA, data not available.

Table S2: Quality metrics of 5(h)mC-enriched sequencing data from primary blood cells

		5mC	5hmC
Number of raw paired reads, x1e-06	Monocytes	76.4	83.4
	Neutrophils	71.2	98.8
	Erythroid progenitors	84.4	100.4
Number of paired reads after QC, x1e-06	Monocytes	43.2	58.6
	Neutrophils	37.5	61.5
	Erythroid progenitors	52.9	47.1
Non-duplicate mapping rate, %	Monocytes	56.5	70.3
	Neutrophils	52.7	62.2
	Erythroid progenitors	62.7	46.9
Saturation analysis, r	Monocytes	0.99	0.98
	Neutrophils	0.99	0.99
	Erythroid progenitors	0.99	0.98
CpGs covered, %	Monocytes	80.0	74.7
	Neutrophils	72.0	83.6
	Erythroid progenitors	77.1	85.9
CpGs covered >5x, %	Monocytes	46.5	32.9
	Neutrophils	45.3	35.8
	Erythroid progenitors	49.9	24.9
CpG enrichment score	Monocytes	2.9	1.9
	Neutrophils	3.1	1.8
	Erythroid progenitors	3.1	1.5
Reads covering no CpG site, %	Monocytes	9.9	8.2
	Neutrophils	2.8	13.6
	Erythroid progenitors	4.5	34.2

CpG, cytosine-guanine dinucleotide; QC, quality control.

Table S3: Literature search summary of genes associated with differentially methylated regions in plasma of NSCLC patients. This table includes genes with associated differentially hypermethylated regions in NSCLC patient plasma that were either associated with downregulated gene expression in LUAD tissue (inclusion criterion 1; Incl.), specifically hypermethylated LUAD compared to normal tissue types (2), and/or uniquely hypermethylated in LUAD compared to other cancer entities (3). Genes not previously described in the context of cancer were excluded from the table (*C8orf88, FAM181B, FAM181A, FOXG1-AS1, LINC01443, LMO2, NHLRC4, PDE1C, PRAC1, QRFPR, SERPINA6, SKIDA1, SMAD13, SYT6,* and *ZNF366*). The Δβ-values refer to the mean difference between β-values in LUAD and non-malignant lung, breast, prostate, colon, and kidney tissue at the differentially methylated region in NSCLC plasma. Genes without Δβ-value (NA) are not covered by the 450k methylation array.

Gene	Class	Role in cancer	Description	Δβ	Incl.
AFF3	Transcription factor	Tumor suppressor	Expression downregulation enhances proliferatory capacity in NSCLC cells (Zhang et al. 2018a).	-0.23	1
ANK2	Transporter	Oncogene	Low expression reduces proliferation of pancreatic cancer cells *in vitro* and *in vivo* via negative regulation of FAK, ERK1/2 and p38 (Chen et al. 2010).	NA	1
C5orf66-AS1	lncRNA	Tumor suppressor	Expression downregulation increases proliferation, invasion and migration in oral squamous cell carcinoma cells (Lu et al. 2018).	0.08	1
CDO1	Enzyme	Tumor suppressor	Promoter hypermethylation is found in tissue of patients (Yin et al. 2020). 5mC levels in cfDNA of stage I NSCLC patients are significantly higher compared to healthy controls (Chen et al. 2020a).	NA	1
CELF2	RNA-binding protein	Tumor suppressor	Negative regulator of PI3K/AKT signaling counteracting the inhibition of PTEN. Expression downregulation is associated with poor survival in NSCLC patients (Yeung et al. 2020).	0.24	2
CLEC14A	Plasma protein	Tumor suppressor	Promoter hypermethylation and reduced expression in NSCLC tissue. High expression correlates with improved clinical outcome (Su et al. 2019).	0.33	2
CYP4F2	Enzyme	Tumor suppressor	Low expression levels are associated with poor prognosis in liver cancer patients. Overexpression inhibits cell proliferation and migration in liver cancer cell lines (Wan et al. 2020).	NA	1
EDNRB	Receptor	Tumor suppressor	Downregulated by promoter hypermethylation in NSCLC (Knight et al. 2009). Inhibits proliferation and migration of NSCLC cell lines by regulating the ERK pathway (Wei et al. 2020).	-0.14	1
EMX2	Transcription factor	Tumor suppressor	Reduced expression and promoter hypermethylation in NSCLC tissue. Knockdown promotes proliferation, invasiveness and Wnt signaling in NSCLC cell lines (Okamoto et al. 2010).	0.22	1
ERBB4	Receptor	Oncogene	Overexpression increases proliferation rates in NSCLC cells (Starr et al. 2006).	0.02	1
EVX2	Transcription factor	Tumor suppressor	Reduced expression and promoter hypermethylation in NSCLC tissue. Knockdown promotes proliferation, invasiveness and Wnt signaling (Okamoto et al. 2010).	0.22	1/2
GATA3	Transcription factor	Unknown	Low expression in NSCLC tissue (Zhao et al. 2019).	0.22	1
GATA4	Transcription factor	Tumor suppressor	Promoter hypermethylation and low expression in NSCLC tissue. Overexpression in NSCLC cells downregulates TGFβ-signaling and thereby enhances senescence (Gao et al. 2019). Plasma 5mC levels are significantly increased in stage I NSCLC patients compared to healthy control (Chen et al. 2020a).	0.17	1
GLYATL1	Enzyme	Tumor suppressor	Downregulated expression by promoter hypermethylation in liver cancer patients (Guan et al. 2020).	0.28	1/2
HIST1H3D	Histone protein	Oncogene	Upregulated expression in NSCLC tissue and cell lines. Drives proliferation, cell cycle progression and inhibits apoptosis in NSCLC cells (Rui et al. 2017).	0.24	3

HOPX	Transcription factor	Tumor suppressor	Downregulated expression in NSCLC cell lines mediated by promoter hypermethylation. Overexpression activates MAPK signaling via oncogenic Ras and, thereby increases proliferation, migration, and invasion (Chen et al. 2015).	0.48	2
HOXA10-AS	lncRNA	Oncogene	Upregulated expression in NSCLC tissue. Silencing in NSCLC cells inhibits tumor progression, while high expression increases Wnt/β-catenin signaling (Sheng et al. 2018).	0.37	1/2
HOXA11-AS	lncRNA	Oncogene	Increased expression in NSCLC tissue and cells. Knockdown inhibits proliferation and promotes apoptosis in NSCLC cell lines (Bai et al. 2019).	0.24	1
HOXA9	Transcription factor	Tumor suppressor	Promoter methylation is a negative predictor for recurrence-free survival in NSCLC patients (Hwang et al. 2015). Plasma 5mC levels are significantly increased in stage I NSCLC patients compared to healthy control (Chen et al. 2020a).	0.38	1/2
HOXA-AS1	lncRNA	Tumor suppressor	Decreased expression in NSCLC tissue. *In vitro* knockdown promotes cell cycle progression, growth and invasiveness in NSCLC cells (Chen et al. 2020b).	0.32	1/2/3
HOXB13	Transcription factor	Tumor suppressor	Promoter hypermethylation is associated with high proliferation as well as high expression of estrogen and proesterone receptors in breast cancer patients (Sui et al. 2018).	0.13	2
HTR1B	Receptor	Tumor suppressor	Promoter methylation and decreased expression is observed in NSCLC but not normal lung tissue (Takai et al. 2001).	0.13	1
KCTD8	Not specified	Unknown	Increased methylation levels in low-grade breast cancer patients (Faryna et al. 2012).	0.17	1
LIN28B	RNA-binding protein	Oncogene	Inhibits apoptosis in ovarian cancer cells (Lin et al. 2018).	0.34	2
MAB21L1	Not specified	Tumor suppressor	Promoter hypermethylation in prostate cancer (Wu et al. 2016).	0.16	1
NEFL	Not specified	Tumor suppressor	Promoter hypermethylation in NSCLC cells. Overexpression inhibits NF-κB signaling and thereby decreases invasiveness and migration of NSCLC cells (Shen et al. 2016).	0.10	1
NKX2-8	Transcription factor	Tumor suppressor	Overexpression in bladder cancer cells inhibits EMT, thereby reducing motility and invasiveness (Yu et al. 2018).	0.21	1
NXPH1	Not specified	Unknown	Increased methylation levels in low-grade breast cancer patients (Faryna et al. 2012).	0.31	2
OTX2-AS1	lncRNA	Tumor suppressor	Increased abundance in exosomes of bladder cancer patients compared to healthy controls (Berrondo et al. 2016).	0.30	2
PAX6	Transcription factor	Tumor suppressor	Promoter hypermethylation is associated with poor overall survival and correlates with the presence of distance metastasis and TNM stage in NSCLC patients (Kiselev et al. 2018).	0.21	1
PCDH10	Receptor	Tumor suppressor	Promoter methylation predicts poor prognosis following curative resection of stage I NSCLC (Harada et al. 2015). Overexpression reduces proliferation and migration in NSCLC cells (Tang et al. 2012).	0.16	1
POU3F3	lncRNA	Oncogene	Overexpressed in NSCLC compared to adjacent normal tissue. Positively regulates proliferation, migration and invasion in NSCLC cells (Zeng et al. 2020).	0.16	1
PRAC1	Not specified	Tumor suppressor	Promoter hypermethylation and consequential expression downregulation in prostate, renal and colon cancer (Hu et al. 2018).	0.31	1/2
PRDM14	Transcription factor	Oncogene	Confers chemotherapy resistance in NSCLC cell lines (He et al. 2021).	0.38	2
PTGER4	Receptor	Unknown	Plasma hypermethylation can differentiate between NSCLC and non-malignant diseases (as part of a 2-marker-panel together with *SHOX2*) (Schotten et al. 2021; Weiss et al. 2017).	0.20	2

SLC7A8	Transporter	Tumor suppressor	Low expression levels in NSCLC patients are associated with shorter overall survival (Asada et al. 2020).	NA	1
SOX9-AS1	lncRNA	Unknown	Regulator of SOX9 (Barter et al. 2017), which promotes proliferation in NSCLC cell lines *via* activation of the Wnt/β-catenin signaling pathway (Guo et al. 2018).	0.1	1
ST8SIA3	Enzyme	Oncogene	Elevated expression in breast cancer *versus* adjacent normal tissue [276].	0.19	1
TAL1	Transcription factor	Tumor suppressor	Silenced by promoter hypermethylation in NSCLC patients. High expression is associated with good prognosis (Tang et al. 2019)	0.04	1
TBX3	Transcription factor	Tumor suppressor	Downregulated in NSCLC. *TBX* family members suppress cell cycle progression and positively regulate genes implicated in cancer development (Khalil et al. 2018).	0.18	1/3
ZNF471	Transcription factor	Tumor suppressor	Downregulated expression by promoter hypermethylation in various cancer types (Cao et al. 2018; Sun et al. 2020). Overexpression reduces invasiveness in NSCLC cells (Wang et al. 2020).	0.2	1
ZNF582-AS1	lncRNA	Tumor suppressor	Promoter hypermethylation and consequential expression downregulation in colon cancer tissue correlates with poor survival (Kumegawa et al. 2016).	0.13	1
ZNF667-AS1	lncRNA	Tumor suppressor	Reduced expression in colon cancer tumor tissue is associated to poor prognosis. Overexpression in colon cancer cells inhibits proliferation, migration and invasion by downregulation of JAK2 (Zhuang et al. 2021).	0.17	1

LINC, long intergenic non-coding; lncRNA, long non-coding RNA; NSCLC, non-small cell lung cancer; TNM, tumor-nodes-metastases.

Table S4: Correlation coefficients between diagnostic/prognostic 5mC biomarkers and molecular alterations measured in the same plasma samples. Pearson correlation coefficients and corresponding p-values (gray) are listed for the comparison of each 5mC biomarker to co-measured molecular alterations.

5mC biomarker	Molecular alteration											
	EML4-ALK fusion		G1202R		L1196M		TP53 mutation		VAF$_{max}$		t-MAD score	
HOXA10-AS	0.50	<0.001	0.60	0.005	0.21	0.472	0.48	<0.001	0.57	<0.001	0.70	<0.001
SOX9-AS1	0.40	0.001	0.76	<0.001	0.34	0.238	0.51	<0.001	0.57	<0.001	0.67	<0.001
PRAC1	0.40	0.001	0.43	0.056	0.41	0.148	0.38	0.002	0.39	0.001	0.54	<0.001
PTGER4	0.68	<0.001	0.74	<0.001	0.33	0.244	0.75	<0.001	0.77	<0.001	0.86	<0.001
PCDH10	0.35	0.004	0.17	0.475	0.11	0.709	0.11	0.365	0.23	0.069	0.41	0.001
NXPH1	0.35	0.004	0.62	0.004	0.10	0.745	0.30	0.014	0.46	<0.001	0.53	<0.001
CELF2	0.64	<0.001	0.81	<0.001	0.33	0.253	0.78	<0.001	0.79	<0.001	0.83	<0.001
HTR1B	0.61	<0.001	0.82	<0.001	0.36	0.200	0.80	<0.001	0.83	<0.001	0.84	<0.001
SYT6	0.43	<0.001	0.37	0.112	0.19	0.517	0.33	0.006	0.39	0.001	0.48	<0.001
HOXA9	0.45	<0.001	0.60	0.005	0.19	0.518	0.45	<0.001	0.51	<0.001	0.64	<0.001
CDO1	0.48	<0.001	0.41	0.076	0.06	0.832	0.33	0.007	0.42	<0.001	0.59	<0.001
POU3F3	0.49	<0.001	0.46	0.042	0.41	0.144	0.47	<0.001	0.49	<0.001	0.67	<0.001
PRDM14	0.38	0.002	0.47	0.035	0.46	0.099	0.34	0.006	0.40	0.001	0.63	<0.001
QRFPR	0.20	0.108	0.13	0.582	0.35	0.222	0.08	0.552	0.21	0.100	0.32	0.010
PAX6	0.44	<0.001	0.62	0.003	-0.12	0.67	0.46	<0.001	0.50	<0.001	0.52	<0.001
CLEC14A	0.51	<0.001	0.76	<0.001	-0.14	0.636	0.52	<0.001	0.67	<0.001	0.70	<0.001
EVX2	0.47	<0.001	0.54	0.013	0.16	0.573	0.39	0.001	0.47	<0.001	0.60	<0.001
HOXA-AS1	0.22	0.074	0.03	0.894	0.20	0.501	0.06	0.655	0.12	0.354	0.27	0.028
FOXG1-AS1	0.37	0.002	0.64	0.002	0.04	0.879	0.42	0.001	0.53	<0.001	0.63	<0.001
HOXA11-AS	0.38	0.002	0.52	0.020	0.11	0.698	0.32	0.009	0.44	<0.001	0.56	<0.001
ST8SIA3	0.20	0.105	0.36	0.118	-0.23	0.434	0.12	0.342	0.21	0.092	0.33	0.008

ALK, anaplastic lymphoma kinase; EML4, echinoderm microtubule-associated protein-like 4; t-MAD, trimmed median absolute deviation from copy number neutrality; VAF$_{max}$, maximal variant allele frequency.

Table S5: Literature search summary of genes with TSS-proximal or gene body associated hyper-hydroxymethylated DhMRs. Cancer association of genes with TSS-proximal (±1,500 bp) or gene body associated differentially hydroxymethylated regions (DhMRs) that coincide with upregulated gene expression in lung adenocarcinoma tissue. Genes not previously described in the context of cancer were excluded from the table (*FOXD4L1, HECW2* and *UGT2B4*).

Gene	Class	Role in cancer	Description
ALCAM	Plasma protein	Oncogene	*ALCAM* expression promotes brain metastasis formation in NSCLC by increased cell dissemination and interaction with brain endothelial cells (Münsterberg et al. 2020).
BAIAP2L1	Not specified	Oncogene	Upregulated expression in NSCLC tissue and cell lines compared to non-cancerous tissue. In vitro silencing decreases cell viability and colony formation capacity (Xu et al. 2019).
DMBT1	Not specified	Tumor suppressor	Low expression in NSCLC tissue and cell lines (Takeshita et al. 1999).
FRMD5	Not specified	Tumor suppressor	*In vitro FRMD5* knockdown promotes cell migration and invasion in NSCLC cell lines (Wang et al. 2012b).
IL1RAP	Plasma protein	Oncogene	Elevated expression in stomach cancer tissue and acute myeloid leukemia. Knockdown reduces proliferation, migration and invasion *in vivo* and *in vitro* (Lv et al. 2021; Mitchell et al. 2018).
KIF25	Kinesin-like protein	Oncogene	Highly expressed in estrogen receptor-positive breast cancer cells (Zou et al. 2014).
KPNA7	Transporter	Oncogene	*KPNA7* silencing reduces cell proliferation in breast and pancreatic cancer cell lines (Vuorinen et al. 2018).
MALL	Not specified	Oncogene	Upregulated expression in lung adenocarcinoma cell lines (A549, ABC-1) (Watanabe et al. 2010).
MEPE	Not specified	Oncogene	*MEPE* expression downregulation decreases proliferation, colony formation and invasion in Hela cells (Hong Yan 2020).
PTP4A3	Enzyme	Oncogene	*PTP4A3* knockdown in the ALK+ NSCLC cell line H1299 inhibits cell growth and colony formation (Lian et al. 2012b).
PTPRG-AS1	lncRNA	Oncogene	*PTPRG-AS1* is highly expressed in NSCLC tissues. Its overexpression promotes viability and radioresistance under X-ray irradiation (Ma et al. 2020).
SPINK1	Plasma protein	Oncogene	Elevated expression in NSCLC compared to adjacent normal tissue. *SPINK1* promotes tumor cell growth and inhibits apoptosis *in vitro* (Guo et al. 2019).
TMC5	Not specified	Oncogene	High expression in prostate cancer tissue compared to normal controls. *TMC5* knockdown inhibits cell proliferation by inducing cell cycle arrest (Zhang et al. 2019a).

ALK, anaplastic lymphoma kinase; TSS, transcription start site; NSCLC, non-small cell lung cancer.

Table S6: Literature summary of genes with TSS-proximal or gene body associated hypo-hydroxymethylated DhMRs. Cancer association of genes with TSS-proximal (±1,500 bp) or gene body associated differentially hypo-hydroxymethylated regions that coincide with downregulated gene expression in lung adenocarcinoma tissue. Genes not previously described in the context of cancer were excluded from the table (*ANO2, APCDD1L, GPR146, PRKG2,* and *SV2C*).

Gene	Class	Role in cancer	Description
ANGPTL1	Not specified	Tumor suppressor	Anti-angiogenic factor that inhibits endothelial cell proliferation, migration and adhesion. High expression in liver cancer patients is associated with improved clinical outcome and reduced metastasis formation (Chen et al. 2016).
FGF1	Growth factor	Oncogene	High expression in NSCLC tissue is associated with larger tumor size and poor prognosis (Li et al. 2015).
PDZRN4	Not specified	Tumor suppressor	Downregulated expression in liver and colon cancer tissue. Ectopic expression inhibits cell proliferation and colony formation in liver cancer cell lines (Hu et al. 2015).

TSS, transcription start site; NSCLC, non-small cell lung cancer.

Table S7: Correlation coefficients between 5hmC biomarkers and molecular alterations measured in the same plasma samples. Pearson correlation coefficients and corresponding p-values (gray) are listed for the comparison of each 5mC biomarker to co-measured molecular alterations.

5hmC biomarker	EML4-ALK fusion		G1202R		L1196M		TP53 mutation		VAF$_{max}$		t-MAD score	
ALCAM	0.11	0.361	0.14	0.558	0.19	0.519	0.15	0.220	0.08	0.532	0.12	0.349
BAIAP2L1	0.10	0.400	0.51	0.023	0.04	0.905	0.07	0.554	0.14	0.265	0.20	0.098
IL1RAP	0.07	0.573	0.07	0.767	0.29	0.311	0.17	0.169	0.10	0.434	0.14	0.257
KIF25	0.20	0.103	0.69	0.001	-0.32	0.260	0.26	0.033	0.32	0.007	0.39	0.001
KPNA7	0.12	0.337	0.70	0.001	0.09	0.769	0.26	0.028	0.32	0.007	0.40	0.001
MALL	0.17	0.170	0.26	0.259	0.41	0.146	0.21	0.080	0.23	0.059	0.15	0.221
MEPE	-0.02	0.860	0.04	0.852	0.09	0.747	0.14	0.258	0.08	0.496	0.11	0.381
PTP4A3	0.21	0.080	0.41	0.071	0.17	0.558	0.10	0.397	0.16	0.200	0.22	0.070
PTPRG-AS1	0.11	0.353	0.37	0.113	-0.19	0.526	0.06	0.645	0.15	0.234	0.07	0.563
SPINK1	0.31	0.010	0.55	0.013	0.01	0.969	0.26	0.034	0.30	0.013	0.27	0.023
TMC5	0.01	0.956	0.35	0.134	0.09	0.756	0.12	0.318	0.12	0.333	0.07	0.551

ALK, anaplastic lymphoma kinase; EML4, echinoderm microtubule-associated protein-like 4; t-MAD, trimmed median absolute deviation from copy number neutrality; VAF$_{max}$, maximal variant allele frequency.

11 Curriculum vitae

Name:	Florian Janke
Date of birth:	September 28, 1991
Place of birth:	Mannheim
Nationality:	German

Education

12/2017 – present

Doctor of Philosophy (Ph.D.), Helmholtz International Graduate School for Cancer Research, German Cancer Research Center (DKFZ), Heidelberg

Medical faculty, Ruperto Carola University Heidelberg, course of studies: Doctor scientiarum humanarum (Dr. sc. hum.)

09/2015 – 11/2017

Master of Science (M.Sc.), Mannheim University of Applied Sciences, Mannheim

09/2011 – 06/2015

Bachelor of Science (B.Sc.), Mannheim University of Applied Sciences, Mannheim

09/2002 – 05/2011

Abitur, Peter-Petersen-Gymnasium, Mannheim

Professional experience

12/2017 – present

Doctoral Researcher, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg

09/2013 – 06/2015

Student assistant, Roche Diagnostics Deutschland GmbH, Mannheim

07/2012 – 09/2012

Intern, Roche Diagnostics Deutschland GmbH, Mannheim

12 Acknowledgments

The conduction of this thesis would not have been possible without the support and guidance I received from many people.

First and foremost, I would like to thank Prof. Dr. Holger Sültmann for giving me the opportunity to write my thesis in his group. I am very grateful for all the support and encouragement throughout my studies. Without his guidance and continuous feedback this PhD would not have been achievable.

Furthermore, I want to thank PD Dr. Petros Christopoulos, Dr. Marc A. Schneider, Dr. Thomas Muley, and Dr. Michael Meister for providing all the plasma samples used for this study as well as comprehensive clinical information. I also would like to thank Dr. Martina Seiffert and Prof. Dr. Albrecht Stenzinger for their scientific support and guidance as members of my thesis advisory committee.

Moreover, I am very grateful for the advices and support from Dr. Steffen Dietz, Dr. Arlou Angeles and Dr. Simone Bauer during our weekly liquid biopsy meetings. Also, I would like to thank Louise Appenheimer for her great work while performing the experimental work for her MD thesis which supported my project. Thanks for the nice time, I really enjoyed working with you.

Many thanks also to PD Dr. Sabine Klauck for her very helpful comments while proofreading this manuscript.

Special thanks also to Sabrina Gerhardt and Simon Ogrodnik for their great experimental support throughout the entire project. Especially, I would like to thank all current and former members of the group for their input and all the scientific and non-scientific discussions: Dr. Ann-Kathrin Daum, Dr. Anja Riediger, Prof. Dr. Peter Altevogt, Lisa Gu, Sabrina Müller, and Sebastian Weingart. Many thanks to all of you!

Finally, I want to thank my family and girlfriend, Carolin, for the constant support and patience especially during the challenging periods of this thesis!

13 Declaration

EIDESSTATTLICHE VERSICHERUNG

1. Bei der eingereichten Dissertation zu dem Thema „**Genome-wide epigenomic analyses of cell-free DNA from anaplastic lymphoma kinase-rearranged non-small cell lung cancer patients**" handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

_____

Ort und Datum   Unterschrift