

RESEARCH

Open Access



Longitudinal monitoring of cell-free DNA methylation in *ALK*-positive non-small cell lung cancer patients

Florian Janke^{1,2*}, Arlou Kristina Angeles^{1,2}, Anja Lisa Riediger^{1,3,4,5}, Simone Bauer¹, Martin Reck⁶, Albrecht Stenzinger^{2,7,8}, Marc A. Schneider^{2,9}, Thomas Muley^{2,9}, Michael Thomas^{2,10}, Petros Christopoulos^{2,10} and Holger Sültmann^{1,2,8}

Abstract

Background: DNA methylation (5-mC) signals in cell-free DNA (cfDNA) of cancer patients represent promising biomarkers for minimally invasive tumor detection. The high abundance of cancer-associated 5-mC alterations permits parallel and highly sensitive assessment of multiple 5-mC biomarkers. Here, we performed genome-wide 5-mC profiling in the plasma of metastatic *ALK*-rearranged non-small cell lung cancer (NSCLC) patients receiving tyrosine kinase inhibitor therapy. We established a strategy to identify *ALK*-specific 5-mC changes from cfDNA and demonstrated the suitability of the identified markers for cancer detection, prognosis, and therapy monitoring.

Methods: Longitudinal plasma samples ($n = 79$) of 21 *ALK*-positive NSCLC patients and 13 healthy donors were collected alongside 15 *ALK*-positive tumor tissue and 10 healthy lung tissue specimens. All plasma and tissue samples were analyzed by cell-free DNA methylation immunoprecipitation sequencing to generate genome-wide 5-mC profiles. Information on genomic alterations (i.e., somatic mutations/fusions and copy number alterations) determined in matched plasma samples was available from previous studies.

Results: We devised a strategy that identified tumor-specific 5-mC biomarkers by reducing 5-mC background signals derived from hematopoietic cells. This was followed by differential methylation analysis (cases vs. controls) and biomarker validation using 5-mC profiles of *ALK*-positive tumor tissues. The resulting 245 differentially methylated regions were enriched for lung adenocarcinoma-specific 5-mC patterns in TCGA data and indicated transcriptional repression of several genes described to be silenced in NSCLC (e.g., *PCDH10*, *TBX2*, *CDO1*, and *HOXA9*). Additionally, 5-mC-based tumor DNA (5-mC score) was highly correlated with other genomic alterations in cell-free DNA (Spearman, $\rho > 0.6$), while samples with high 5-mC scores showed significantly shorter overall survival (log-rank $p = 0.025$). Longitudinal 5-mC scores reflected radiologic disease assessments and were significantly elevated at disease progression compared to the therapy start ($p = 0.0023$). In 7 out of 8 instances, rising 5-mC scores preceded imaging-based evaluation of disease progression.

Conclusion: We demonstrated a strategy to identify 5-mC biomarkers from the plasma of cancer patients and integrated them into a quantitative measure of cancer-associated 5-mC alterations. Using longitudinal plasma samples

*Correspondence: fjanke@dkfz-heidelberg.de

¹ Division of Cancer Genome Research, German Cancer Research Center, National Center for Tumor Diseases, Heidelberg, Germany
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

of *ALK*-positive NSCLC patients, we highlighted the suitability of cfDNA methylation for prognosis and therapy monitoring.

Keywords: *ALK*-positive NSCLC, Biomarkers, Cell-free DNA, cfMeDIP-seq, DNA methylation, Liquid biopsy, Epigenetics

Background

Liquid biopsies from circulating cell-free DNA (cfDNA) have demonstrated their utility for minimally invasive cancer detection, tumor genotyping, resistance as well as residual disease monitoring during therapy [1–7]. The analysis of cfDNA molecules carrying genomic aberrations (i.e., somatic mutations or copy number alterations [CNAs]) is highly tumor-specific and allows accurate detection and longitudinal assessment of cancers. However, this approach poses several challenges, including the low abundance of mutated cfDNA fragments, a lack of common mutations across patient groups and the necessity to have a priori knowledge of a tumor's molecular profile [3, 8]. DNA methylation, occurring at the 5'-carbon of cytosines (5-mC), was demonstrated to be preserved in cfDNA [9–11] and represents a biomarker with the potential to overcome some of these limitations. Aberrant methylation at cytosine–guanine dinucleotides (CpGs) is central to carcinogenesis and usually occurs genome-wide [12–14]. This allows parallel assessment of multiple 5-mC sites, thereby increasing the probability to capture cancer-derived signals in the circulation. In addition, tumors without known genomic alterations (i.e., mutations or CNAs) may be detected utilizing cancer-specific 5-mC signatures. With its gene regulatory function, 5-mC contains additional information about the tumor that cannot be derived from genomic cfDNA alterations. Its presence at regulatory regions, such as promoters or enhancers, represses transcription of associated genes. Previous reports were able to deduce silenced tumor suppressors [15] and cell type-specific gene regulation (i.e., tumor localization) from cell-free 5-mC profiles [9, 11, 16–19]. Cell-free methylation DNA immunoprecipitation followed by high-throughput sequencing (cfMeDIP-seq) is a sensitive approach to detect 5-mC signals from low amounts of DNA (> 1 ng). The enrichment of methylated cfDNA molecules allows genome-scale 5-mC profiling without error-prone bisulfite conversion [10, 20]. In principle, cfMeDIP-seq enables concurrent assessment of numerous 5-mC tumor biomarkers [10, 21–23]. Yet, their identification in cfDNA presents a challenge, because cfDNA is regarded as a mixture of DNA fragments released from various cell and tissue types. Most cfDNA is derived from hematopoietic cells, while tumor-derived DNA molecules commonly compose a minor fraction (< 1%) [11, 18, 19]. This poses the difficulty in identifying tumor-informative

5-mC signals within the vast amount of non-cancer background DNA.

So far, few studies addressed the capability of cfMeDIP-seq for the assessment of longitudinal therapy kinetics [21, 24, 25]. Here, we applied cfMeDIP-seq to longitudinally sampled plasma of non-small cell lung cancer (NSCLC) patients with oncogenic rearrangements of the anaplastic lymphoma kinase (*ALK*) gene. These patients are susceptible to *ALK* tyrosine kinase inhibitor (TKI) therapy and can experience long survival under serial treatment with multiple targeted drugs [26–28]. However, therapy failure due to acquired drug resistance is common [27]. Therefore, timely recognition of disease progression and consequential adaptation of therapy strategy is desirable for better disease management. We implemented a strategy to identify tumor-specific 5-mC biomarkers from cfMeDIP-seq data of cancer patient cfDNA. Our approach uses public whole-genome bisulfite sequencing (WGBS) datasets of cell types composing the non-tumor fraction of cfDNA [18] to identify and reduce confounding 5-mC background signals. We validated the tumor specificity of the resulting 5-mC biomarkers using lung cancer tissue methylation and gene expression data, systematically compared them to genomic alterations previously determined in matched plasma [2, 4], and followed their abundances in serial plasma samples taken during TKI therapy. The results of this study highlight the complementarity of epigenomic and genomic cfDNA analysis and, for the first time, demonstrate the applicability of cfMeDIP-seq for longitudinal treatment monitoring in *ALK*-positive NSCLC. Additionally, we provide a strategy for 5-mC biomarker identification that can be applied in future studies.

Results

Patient characteristics

A total of 66 plasma specimens from 21 metastatic *ALK*-positive NSCLC lung adenocarcinoma (LUAD) patients were included in this study (Table 1). Longitudinal plasma was available for eleven patients ranging from 2 to 14 consecutive samples. Baseline tissue biopsies identified *EML4-ALK* fusion variant 1 (V1; E13:A20) in 43% (9/21), V2 (E20:A20) in 10% (2/21), V3 (E6:A20) in 33% (7/21), and other variants in 10% (2/21) of patients. *TP53* mutations were detected in baseline tissue biopsies of 29% (6/21) of cases. All patients received one or multiple lines of *ALK*-directed TKI therapy, and

Table 1 Patient characteristics

ALK-positive NSCLC patients (n = 21; n = 66 plasma specimens)	
Age, median (range)	56 (42–80)
Sex, male	11/21
Stage IV	21/21
<i>ALK fusion variant, patient number¹</i>	
EML4-ALKV1/V2	11
EML4-ALKV3	7
Other ²	2
No data	1
<i>TP53 status, patient number¹</i>	
Positive	6
Negative	14
No data	1
<i>Treatment, sample number</i>	
Crizotinib	19
Ceritinib, Alectinib, Brigatinib	27
Lorlatinib	5
Chemotherapy	10
Naïve	4
No data	1
Number of samples per patient, mean (range)	3.1 (1–14)
<i>Radiological evaluation at sampling, number of samples</i>	
Extracranial PD	27
Intracranial PD	4
SD	30
PR	2
No data	3

ALK anaplastic lymphoma kinase, EML4 echinoderm microtubule-associated protein-like 4, KCL1 kinesin light chain 1, NGS next-generation sequencing, PD progressive disease, PR partial response, SD stable disease, TP53 tumor protein 53

¹ Data available for 20/21 patients from NGS of tissue biopsies at diagnosis of stage IV disease

² One patient with a K9A20 (KCL1) and one with an E9A10 fusion

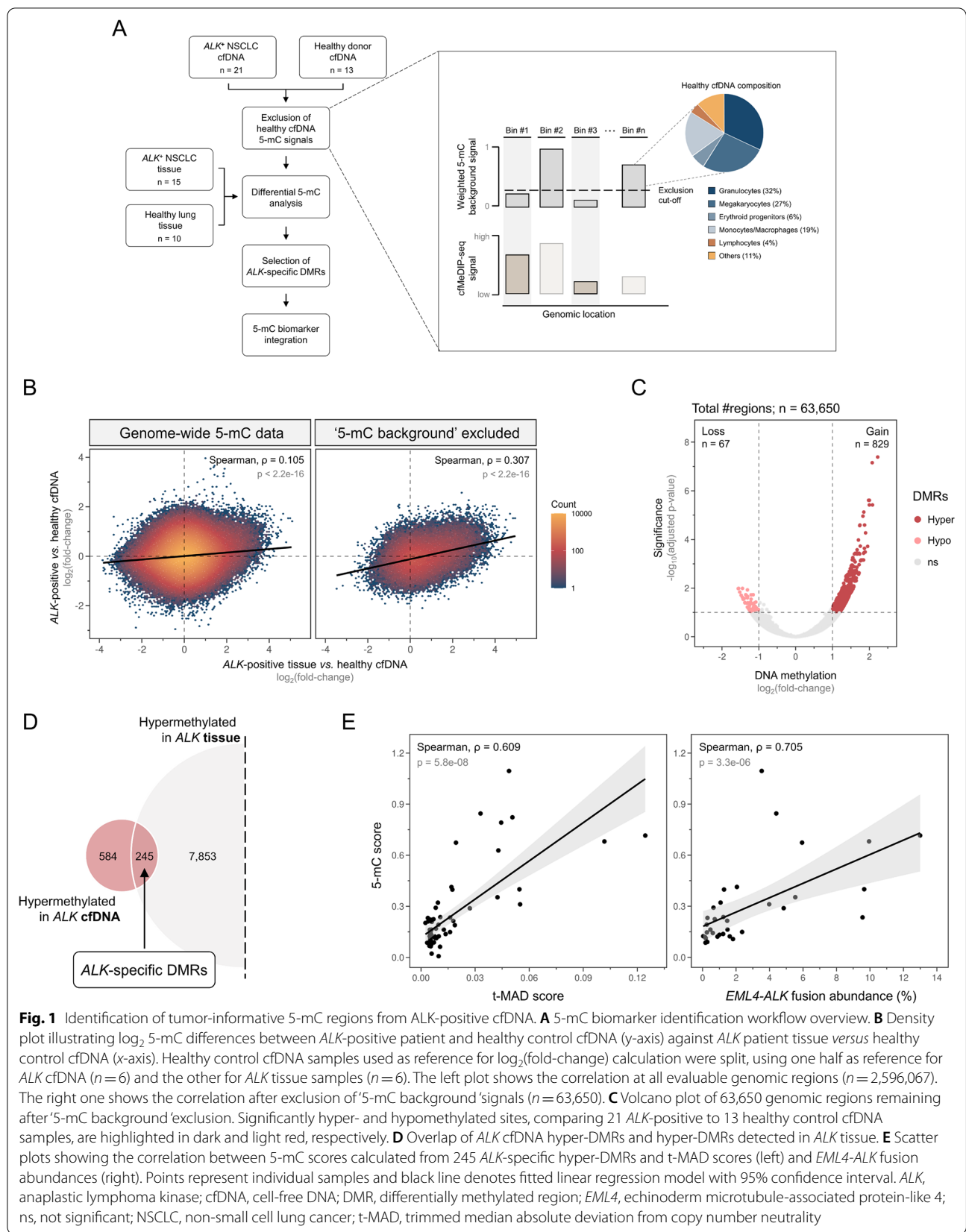
30 plasma samples were taken at time points of disease recurrence. At the last follow-up date, 10/21 patients had deceased (Additional file 1: Figure S1). Information on genomic alterations of all plasma samples was available from previously published work [2, 4]. This included cell-free abundances of the *EML4-ALK* fusion gene and somatic mutations, determined by hybrid-capture-based sequencing as well as genome-scale chromosomal instability assessment by shallow whole-genome sequencing (sWGS), summarized as trimmed median absolute deviation from copy number neutrality (t-MAD) scores [29].

Identification of tumor-informative cell-free 5-mC biomarkers

To identify tumor-associated 5-mC biomarkers in cfDNA, we first generated genome-wide DNA

methylation profiles from 66 *ALK*-positive and 13 healthy control plasma samples by cfMeDIP-seq. We chose 21 patient samples for the biomarker identification process by selecting one sample per individual in cases with available serial plasma. Hereby, longitudinal samples with detectable genomic alterations (previously measured in the same plasma [2, 4]) were favored, reasoning that these samples were expected to contain elevated amounts of tumor-derived cfDNA and were therefore well suited for biomarker identification (selected samples are specified in Additional file 1: Table S1). In addition to plasma, *ALK*-positive tumor ($n = 15$) and normal lung tissue ($n = 10$) samples were subjected to cfMeDIP-seq. The marker identification strategy established in this study interrogated 5-mC coverage profiles at 7,264,350 non-overlapping genomic windows [18] and selected tumor-informative regions by excluding non-tumor 5-mC signals. This was followed by differential methylation analysis between cases and controls and biomarker validation in *ALK*-positive tumor tissues (Fig. 1A and Additional file 1: Fig. S2).

As most cfDNA molecules originate from various hematopoietic cells [11, 18, 19], we reasoned that frequently methylated genomic regions in these cells contain confounding background 5-mC signals rather than carrying information about a patient's tumor. To test this, we used publicly available WGBS data of cell types described to compose cfDNA in healthy individuals (i.e., granulocytes, megakaryocytes, erythroid progenitors, monocytes, macrophages, lymphocytes, and other non-hematopoietic cells) and combined them according to their relative contribution to cfDNA (Fig. 1A; Methods) [18]. We found that focusing on genomic regions commonly unmethylated in the combined '5-mC background' (i.e., methylated in the combination of hematopoietic cells) increased the correlation between *ALK*-positive cfDNA and *ALK* tissue 5-mC signals compared to all evaluable regions. The highest correlation was observed at an exclusion threshold of $\beta < 0.15$ (Fig. 1B; Spearman, $\rho = 0.307$), indicating an enhanced tumor association of the 5-mC signals at the remaining 63,650 sites. Next, we identified cancer-derived differentially methylated regions (DMRs) from cfDNA at '5-mC background'-depleted sites by comparing *ALK*-positive patients to healthy controls. 829 hyper- and 67 hypomethylated DMRs were detected in *ALK*-positive cfDNA (Fig. 1C). To validate their tumor association, we overlapped the cfDNA-derived DMRs to differentially methylated sites found by comparing *ALK*-positive tumor tissue to normal lung tissue (Additional file 1: Figure S3A). Two hundred and forty-five of 829 (29.6%) DMRs were concordantly hypermethylated in *ALK*-positive cfDNA and tumor tissue samples (Fig. 1D), and the significance of overlaps



was confirmed by permutation testing ($p < 0.0001$). Hypomethylated DMRs in cfDNA did not overlap with tumor tissue DNA hypomethylation, most likely due to the limitation of cfMeDIP-seq to detect hypomethylated regions in cfDNA samples. Importantly, we found that differential methylation analysis without prior '5-mC background' depletion identified fewer DMRs whose methylation status could be confirmed in tumor tissue (65/3,948 [1.6%] hyper-DMRs). This suggested that the '5-mC background' depletion facilitates the detection of cancer-derived DMRs. We then developed a metric, termed '5-mC score,' to quantitatively assess the extent of cancer-derived 5-mC changes in each cfDNA sample. The 5-mC score combined the 5-mC signal at the 245 hyper-DMRs with confirmed tumor tissue association by calculating their absolute median coverage deviation from a healthy cfDNA control panel (Methods). A high concordance between 5-mC scores and chromosomal instability (t-MAD score; Spearman, $\rho = 0.609$) as well as 5-mC scores and *EML4-ALK* fusion abundances (Spearman, $\rho = 0.705$) was observed, suggesting that cfMeDIP-seq profiles can inform about the abundance of tumor-derived cfDNA molecules in plasma samples (Fig. 1E). Similar correlations with the t-MAD score and the *EML4-ALK* fusion abundance were found when all 829 hyper-DMRs were considered for the calculation of the 5-mC score (Additional file 1: Figure S3B; Spearman, $\rho = 0.537$ [t-MAD] and $\rho = 0.657$ [*EML4-ALK* fusion]). This suggested that the presented strategy allows the identification of 5-mC biomarkers from cfDNA alone, without additional information from tumor tissue.

Cell-free 5-mC markers are enriched for lung adenocarcinoma-specific methylation and inform about tissue-specific gene expression

DNA methylation occurs in a tissue-specific manner and plays a part in transcriptional regulation [12]. We used publicly available reference datasets (i.e., Illumina 450 k methylation array and RNA-seq data) from The Cancer Genome Atlas (TCGA) [30, 31] to investigate whether the identified 245 *ALK*-specific hyper-DMRs were informative on LUAD biology. In total, 189 of the 245 hyper-DMRs were covered by at least one cytosine probed by the Illumina 450 k methylation array and 78/189 (41.3%) were concordantly hypermethylated in LUAD ($n = 455$) versus adjacent normal lung tissue ($n = 75$). Permutation testing confirmed significant enrichment of LUAD-specific hypermethylation within the hyper-DMRs identified from *ALK*-positive cfDNA ($p < 0.0001$). Interestingly, we observed a similar number of overlapping hyper-DMRs when TCGA-LUAD samples were stratified by pathologic stage or molecular driver (i.e., *EGFR*, *KRAS*, and *EML4-ALK*; Additional file 1: Table S2). This suggested

that some of the hyper-DMRs found in cfDNA might be informative of localized cancer in LUAD patients independent of the *ALK*-positive subtype addressed in this study. We next examined whether the identified hyper-DMRs were indicative of the transcriptional status of proximal genes. Thirty-five out of 189 (18.5%) hyper-DMRs, corresponding to 31/135 (23.0%) genes, demonstrated a significant inverse correlation between DNA methylation and gene expression in non-cancer TCGA tissue samples ($n = 150$), suggesting their 5-mC-dependent transcriptional repression. The majority of genomic regions with gene regulatory 5-mC signals resided in CpG islands (31/35) and many were located proximal to promoters (24/35; i.e., ≤ 5 kb upstream of the transcription start site and 5'-untranslated regions). Among the associated genes, 23/31 (74.2%) were transcriptionally downregulated in TCGA-LUAD ($n = 507$) versus normal lung tissues ($n = 288$) obtained from the Genotype-Tissue Expression project (GTEx; Additional file 1: Table S3). Promoter hypermethylation and/or transcriptional silencing in NSCLC was previously reported for some of these genes (e.g., *PCDH10*, *TBX2*, *CDO1*, and *HOXA9*). Interestingly, *PCDH10*, *TBX2*, and *CDO1* were described as potential biomarkers for early-stage lung cancers [32–34]. *PCDH10* hypermethylation was associated with adverse disease outcomes after surgery of stage I NSCLC [33], while *TBX2* expression was demonstrated to progressively decrease across premalignant lesions with respect to normal lungs [34]. Of note, promoter 5-mC levels of *CDO1* and *HOXA9* were previously utilized for plasma-based disease assessment in both early and advanced lung cancers [32, 35].

5-mC profiling of cfDNA complements CNA and mutation analysis

Cancer-associated alterations of the methylome are prevalent and pervasive across patients [36]. This allows simultaneous profiling of multiple 5-mC biomarkers, potentially overcoming sensitivity limitations posed by the analysis of less abundant genomic alterations. Here, we assessed whether the combined evaluation of the previously identified 245 hyper-DMRs (5-mC score) could identify tumor-derived signals in cfDNA samples without detectable genomic alterations (i.e., t-MAD score, focal amplifications, mutations, or fusions). To define a detection threshold, we calculated 5-mC scores from cfMeDIP-seq data of 13 healthy individuals and used the maximum value (median 0.0224; range 0–0.6870) to determine tumor DNA positivity. The t-MAD score detection threshold was established likewise using sWGS data of 16 healthy individuals (median 0.0051; range 0.0028–0.0081). We identified tumor-derived 5-mC signals in 92.4% (61/66) of cfDNA samples and 90.5%

(19/21) of *ALK*-positive patients (Fig. 2 and Additional file 1: Table S4). Hybrid-capture sequencing and sWGS found cancer-associated genomic alterations in 86.4% (57/66) of the samples, constituting 66.7% (14/21) of patients [2, 4]. The most recurrently altered genes were *ALK* (51.5%; 34/66 samples) and *TP53* (39.4%; 26/66 samples). Focal amplifications were detected in 16.7% (11/66) and t-MAD scores exceeding the detection threshold in 69.7% (46/66) of the samples. 5-mC analysis identified tumor-derived methylation changes in 6 samples from 6 patients without reported genomic alterations, whereas tumor DNA in 2 samples (2 patients) was detectable by hybrid-capture sequencing and/or sWGS only. Comparing the 5-mC analysis results to hybrid-capture sequencing or sWGS alone resulted in the identification of 14 and 17 additional samples (14 and 11 patients), respectively, with detectable tumor-derived alterations. Combining all three analysis types, we found tumor DNA in 95.5% (63/66) of all cfDNA samples and in at least one sample in 90.5% (19/21) of patients. This highlighted the added value of a multimodal approach for the detection of cancer signals in cfDNA samples.

5-mC score predicts poor overall survival and indicates molecular risk of *ALK*-positive lung cancer

To investigate whether the 5-mC score holds prognostic value, we compared it to the clinical outcomes in our *ALK*-positive patient cohort. Overall survival (OS) from the time of plasma sampling was significantly shorter in cases exceeding the cohort's median 5-mC score (median

7.7 vs. 14.0 months; $p=0.0253$; Fig. 3A). *EML4-ALK* fusion abundances and t-MAD scores were also predictive of OS (Additional file 1: Figure S4A/B and previously shown [4]). Recent studies identified the presence of the *EML4-ALK* fusion V3 and/or *TP53* mutations as molecular risk factors associated with shorter progression-free survival and OS in *ALK*-positive patients [37–40]. We found significantly higher 5-mC scores in samples of *EML4-ALK* V3 compared to V1/2 patients ($p=0.0034$; Fig. 3B), while no association between the 5-mC score and *TP53* mutation status was observed ($p=0.5543$; Additional file 1: Figure S4C). This was in contrast to the elevated *EML4-ALK* fusion levels and higher t-MAD scores detected in both V3 versus V1/2 and *TP53*-positive versus *TP53*-negative patients within a previous study [4].

5-mC scores reflect disease kinetics under *ALK* TKI therapy in longitudinal cfDNA samples

Targeted treatment of *ALK*-positive patients is characterized by high incidences of acquired drug resistance and consequent patient relapse [41, 42]. Regular disease surveillance is therefore instrumental for the early detection of tumor progression and guidance of subsequent therapy decisions. We and others previously demonstrated the feasibility of cfDNA mutation and CNA profiling for the monitoring of *ALK*-positive NSCLC [2, 4, 43–45]. In this study, we assessed whether the 5-mC scores reflected therapy-associated tumor DNA dynamics in the plasma of *ALK*-positive patients. Representative

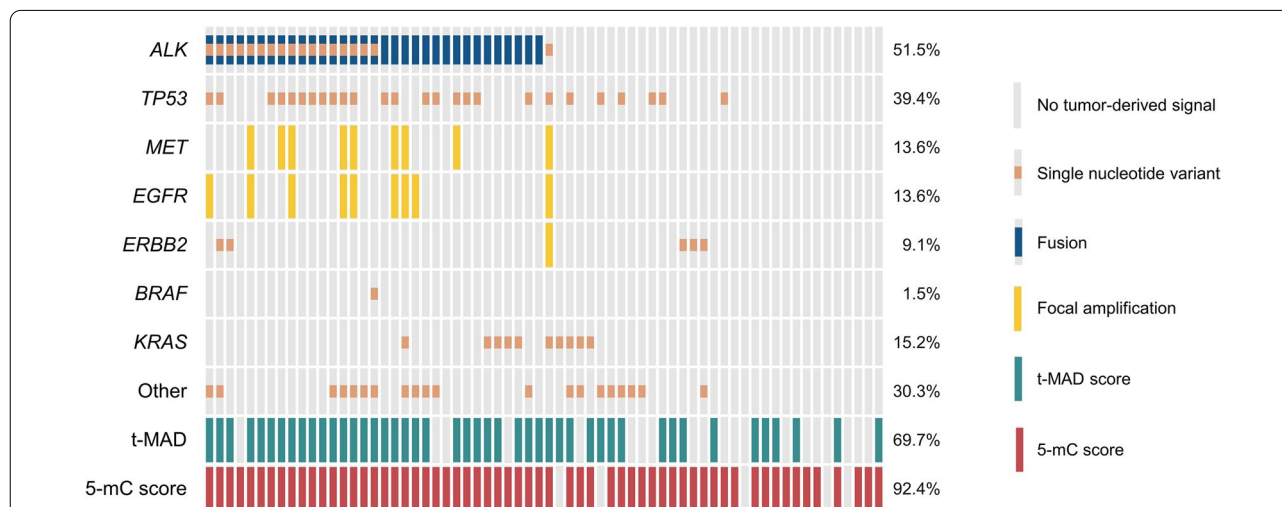
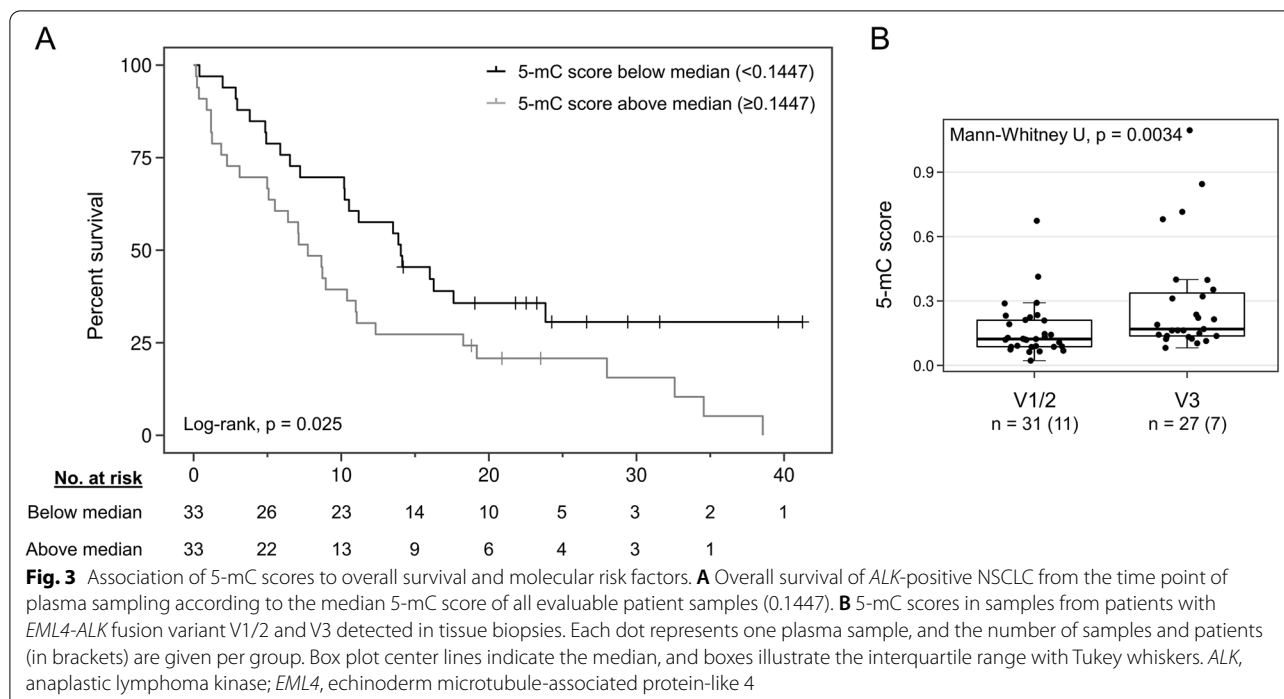


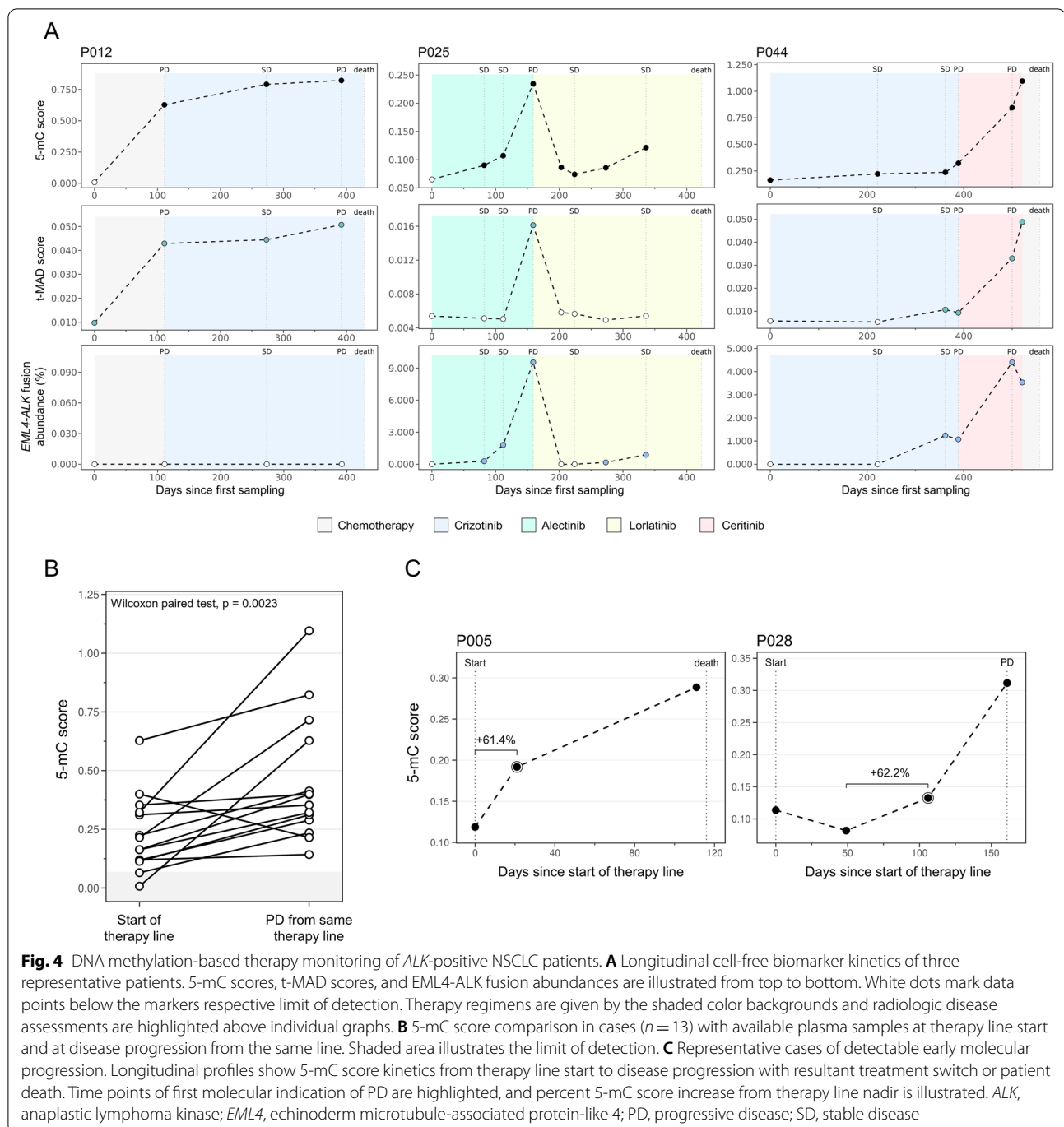
Fig. 2 Per-sample OncoPrint of genomic alterations and cancer-derived 5-mC signals. Plasma samples with 5-mC scores exceeding the maximum value in the healthy donor cohort (0.6870) were considered tumor DNA-positive. Positivity for copy number alterations was evaluated similarly (t-MAD score ≥ 0.0081). Mutations, fusions, and focal amplifications with a relative abundance or variant allele frequency $\geq 0.01\%$ were deemed tumor DNA-positive. Percent of tumor DNA-positive samples per biomarker are indicated on the right. *ALK*, anaplastic lymphoma kinase; *BRAF*, B-raf proto-oncogene; *EGFR*, epidermal growth factor receptor; *ERBB2*, Erb-B2 receptor tyrosine kinase 2; *KRAS*, v-Kir-Ras2 Kirsten rat sarcoma viral oncogene homolog; *MET*, MET proto-oncogene; *TP53*, tumor protein 53; t-MAD, trimmed median absolute deviation from copy number neutrality



cases of 5-mC-based therapy monitoring are illustrated in Fig. 4A and Additional file 1: Figure S5. 5-mC score kinetics reflected those found in the co-measured cell-free genomic biomarkers and recapitulated radiologic tumor progression. Patients P012, P025, and P044 exemplified cases indicating TKI failure by rising 5-mC scores, while decreasing 5-mC signals after administration of effective therapy regimens were observed in P007, P013, P025, and P028 (Fig. 4A and Additional file 1: Figure S5). Additionally, the cohort included cases without informative *EML4-ALK* fusion abundances (e.g., P012 and P007) or t-MAD scores (P007). 5-mC scores were detectable in both cases and indicated disease progression, highlighting the value of 5-mC profiling. Our cohort comprised 13 instances (in 7 patients) with available plasma at the start of a therapy line and at disease progression from the same line with consequential therapy switch or patient death. Compared to the therapy baseline, 5-mC scores were elevated at the progressive disease (PD) time point in 13/14 cases (Wilcoxon paired test, $p=0.0023$; Fig. 4B). *EML4-ALK* fusion abundances and t-MAD scores increased at PD in 10/14 and 11/14 cases, respectively (Wilcoxon paired test, $p=0.0367$ and $p=0.0419$; Additional file 1: Figure S6A). Interestingly, we observed rising 5-mC scores in samples taken prior to disease progression at radiologically stable disease (SD), potentially marking the development of drug resistance (Fig. 4C and Additional file 1: Figure S5). The plasma sampling scheme of this study allowed detecting early molecular signs of

PD at 8 instances (6 patients). Defining a $\geq 25\%$ increase from the therapy line nadir as an indication of molecular progression, we identified 7/8 instances in which 5-mC profiling preceded radiographic determination of TKI failure. The median lead time to radiological progression was 89 days (range 0 to 345 days), allowing significantly earlier relapse identification compared to imaging (Wilcoxon paired test, $p=0.0225$; Additional file 1: Figure S6B). *EML4-ALK* fusion abundances denoted lead times in 6/8 instances (median 66 days [range 0 to 150 days]; Wilcoxon paired test, $p=0.0360$) and t-MAD scores were not informative of early molecular progression (Additional file 1: Figure S6B).

Copy number alteration estimation from cfMeDIP-seq data
The assessment of copy number changes by sWGS of cfDNA demonstrates a cost-effective method for minimally invasive estimation of tumor burden [2, 4, 29, 46]. Here, we evaluated whether cfMeDIP-seq data can be used to infer chromosomal instability, allowing simultaneous genomic and epigenomic tumor assessment from the same dataset. We generated global copy number profiles from cfMeDIP-seq data at 1-Mb bins [47] and compared them to CNAs detected by sWGS of matched plasma. The resulting CNA profiles were highly concordant between both datasets (Additional file 1: Figure S7A). For quantitative CNA detection comparison, we downsampled both datasets to a common read coverage of 5 M paired reads per sample and



calculated t-MAD scores for all patient samples. The resulting t-MAD scores were highly correlated (Pearson, $r = 0.9360$; $p < 2.2e-16$; Additional file 1: Figure S7B), showing that cfMeDIP-seq could be used for both 5-mC profiling and genome-wide copy number estimations. In this way, the utility of sequencing data could be increased without additional costs and using fewer resources (i.e., patient material).

Discussion

In this study, we showed that cancer-derived changes of the methylome can be detected in cfDNA samples of *ALK*-positive NSCLC patients. We implemented a workflow to identify cell-free 5-mC biomarkers, validated their tumor specificity using in-house and external reference datasets, and demonstrated the utility of these markers for prognosis and therapy monitoring.

Tumor-derived DNA can be detected in plasma cfDNA of cancer patients and allows minimally invasive disease assessment [1–7]. Previous reports showed that genome-wide 5-mC profiles can be derived from cfDNA of cancer patients and carry information about the tumor [10, 21–23, 25, 48]. However, a major challenge of liquid biopsies is the detection of minute amounts of tumor DNA molecules within a vast cfDNA background derived from hematopoietic cells. A pivotal element of our 5-mC biomarker identification approach was the initial reduction of 5-mC signals derived from blood cells (Fig. 1A). We found that this step increased the association between *ALK* cfDNA and *ALK* tumor tissue 5-mC signals and therefore reduced the number of genomic regions without tumor-informative DNA methylation. Additionally, the number of DMRs concordantly hypermethylated in cfDNA and tumor tissue increased after the exclusion of the signals derived from hematopoietic cells. This emphasized that the employed background suppression enriches genomic regions containing tumor-derived 5-mC signals and facilitates their detection from cfDNA. Similar observations were made by others showing that the selection of tumor DNA molecules in cfDNA via fragment characteristics (e.g., fragment length) allows the identification of genomic aberrations otherwise missed [29, 49]. Our approach differs from previously reported methods for 5-mC background exclusion [15, 21] in two major points. First, we used public WGBS reference data of individual cell types [18], rather than bulk peripheral blood mononuclear cells, to infer the cfDNA background. Thereby, we were able to account for the relative contribution of each cell type to the cfDNA composition, which has been described to deviate from their abundance in blood [18, 19]. Second, we segmented the genome based on methylation blocks (i.e., adjacent CpG sites with concordant methylation status) [18], instead of continuous 300-bp windows. The assessment of coordinated methylation blocks was reported to reflect fundamental functional 5-mC units and increased the robustness as well as the sensitivity of liquid biopsy assays [9, 50]. Additionally, the evaluation of methylation blocks is well suited to the resolution of cfMeDIP-seq, interrogating 5-mC signals at genomic regions rather than individual CpGs.

Another key advantage of the presented study was the availability of various in-house reference datasets for marker validation (i.e., cfMeDIP-seq data of *ALK* tumor tissue and cell-free genomic alterations determined in matched plasma). The concordance of hyper-DMRs found in both *ALK* cfDNA and cancer tissue confirmed their tumor specificity, which was further validated by the correlation of the 5-mC score to cancer-specific genomic alterations. External datasets (TCGA and GTEx) illustrated the biological plausibility of the identified 5-mC

markers and provided insights into transcriptional dysregulation of LUAD-specific genes. Of note, the comparison with TCGA data suggested that many of the cell-free 5-mC marker regions represent methylome alterations found in various molecular LUAD subtypes. Hence, these biomarkers might be applicable to a wider range of patients beyond the *ALK*-positive subtype addressed in this study. Furthermore, some of the 5-mC biomarkers were present in TCGA-LUAD patients with localized disease, indicating their potential applicability for early disease detection. This was corroborated by the identification of methylation-regulated genes (*PCDH10*, *TBX2*, and *CDO1*) recently described as biomarkers in localized lung cancers and premalignant lesions [32–34]. These observations are in line with the early occurrence of 5-mC changes during tumorigenesis and the pervasiveness of 5-mC patterns across tumor types [12, 51–53].

The combined analysis of 5-mC and genomic biomarkers within this study highlighted their complementarity for the detection of tumor-derived DNA in plasma cfDNA samples. We found that 5-mC markers, summarized as the 5-mC score, detected tumor DNA in more samples ($n=61$) compared to hybrid-capture sequencing ($n=49$) for mutation analysis or chromosomal instability assessment via the t-MAD score ($n=46$). A plausible explanation for this finding is the increased number of loci covered (245 tumor-informative 5-mC signals). It has previously been shown that the breadth can supplant the depth of sequencing and increases the sensitivity of liquid biopsy assays [10, 54]. However, 5-mC analysis failed to detect tumor DNA in some samples in which genomic markers were informative and vice versa. In addition, we demonstrated that the 5-mC score is indicative of OS in a per-sample survival analysis. We and others reported similar results for the detectability of tumor DNA by genomic markers (i.e., mutations, *EML4-ALK* fusion or t-MAD scores) [4, 55–57]. In this study, both *EML4-ALK* fusion abundances and t-MAD scores were superior in predicting OS compared to the 5-mC score. This might be explained by their association with *TP53* mutations, a well-described molecular risk factor in *ALK*-positive NSCLC [37, 39], which was not found for the 5-mC score. The high dynamic range of the 5-mC score in *TP53*-negative samples might render 5-mC-based disease assessment more suitable for this patient subgroup instead of the analysis of less abundant genomic markers [4].

Through 5-mC analysis in sequential plasma samples, we highlighted that the 5-mC score reflects tumor dynamics associated with TKI therapy response. We showed that 5-mC signals indicated disease progression and were informative in cases in which the *EML4-ALK* fusion abundance or the t-MAD score remained undetectable. Importantly, we observed several instances of

rising 5-mC scores prior to radiologically apparent disease progression which marked early signs of molecular progression. With the high incidence of disease relapse under *ALK* TKI therapy, early identification of disease progression is of particular importance for prolonged therapeutic benefit as we highlighted previously [2].

The main limitations of this study were its retrospective design with heterogeneous sampling time points and administered therapy lines. The varying number of plasma samples per patient might have introduced errors due to the overrepresentation of certain individuals. Corresponding findings should ideally be validated in a larger, prospective study with defined sampling intervals. Moreover, the limited number of samples impeded the definition of robust 5-mC score thresholds, both for the assignment of tumor DNA positivity and for the presence of early molecular progression, and precluded the application of machine learning for 5-mC biomarker identification. In addition, technical restraints of the cfMeDIP-seq method, such as its limited capability to assess DNA hypomethylation, precluded the evaluation of the entire methylome.

Conclusion

To our knowledge, this is the first study that analyzed 5-mC alterations in cfDNA of *ALK*-rearranged NSCLC and comprehensively monitored their targeted TKI therapy using 5-mC biomarkers. We demonstrated that the employed biomarker identification approach could reliably identify tumor-associated 5-mC signals and might be used as a blueprint for 5-mC marker detection in future studies. We established a quantitative measure for the assessment of cancer-derived 5-mC changes (5-mC score) and demonstrated its suitability for prognostication and longitudinal therapy monitoring.

Methods

Patients

All individuals provided written informed consent and the study was approved by the ethics committee at Heidelberg (S-270/2001, S-296/2016, S-435/2019) and Lübeck Universities (AZ 12-238). Patients were screened for *ALK* rearrangements in tissue using at least two of the following approaches: *ALK* immunohistochemistry (D5F3 clone, Roche, Mannheim, Germany), *ALK* fluorescence in situ hybridization (ZytoLight SPEC *ALK* probe, ZytoVision GmbH, Bremerhaven, Germany), and RNA-based next-generation sequencing (NGS, Thermo Fisher Lung Cancer Fusion Panel, Waltham MA, USA). Plasma samples from 21 *ALK*-positive metastatic NSCLC patients and 13 healthy donors (i.e., subjects without known current disease) were provided by the Lung Biobank Heidelberg, member of the Biomaterial Bank

Heidelberg (BMBH), and LungenClinic Grosshansdorf. Serial plasma throughout TKI therapy was available for eleven patients resulting in a total of 79 collected plasma specimens (patient samples, $n=66$; healthy donor samples, $n=13$). Additionally, 15 tissue samples from *ALK*-positive metastatic NSCLC patients and 10 distant normal lung tissue specimens (>5 cm) from NSCLC patients who underwent resection of primary lung cancer at the Thoraxklinik at the University Hospital Heidelberg were provided by the Lung Biobank Heidelberg. For 6 patients, matched plasma and tumor tissue samples were available. The remaining 9 tumor tissue samples were taken from patients not included in the *ALK*-positive cfDNA cohort. All diagnoses were made according to the 2004 WHO classification for lung cancer by at least two experienced pathologists. Tumor histology was classified according to the third edition of the World Health Organization classification system. Clinical data, relevant molecular information (i.e., information about *ALK* fusion variants and *TP53* mutation positivity), and radiographic assessments by chest/abdominal computed tomography and brain magnetic resonance imaging were collected based on patient records with a cutoff on March 3, 2020.

Blood processing and cfDNA isolation

Peripheral blood was collected in K₂EDTA tubes and subjected to plasma isolation within one hour of venipuncture employing the previously described centrifugation protocol [4]. Plasma samples were stored at -80 °C in the Lung Biobank Heidelberg until further processing. cfDNA isolation was performed from 2 mL of plasma using the QIAamp MinElute ccfDNA Kit (Qiagen, Hilden, Germany). The concentration and integrity of cfDNA were assessed by the Qubit dsDNA High Sensitivity Kit (Thermo Fisher Scientific, Waltham MA, USA) and the Bioanalyzer 2100 System with DNA High Sensitivity reagents (Agilent Technologies, Santa Clara CA, USA), respectively.

Tissue collection and DNA extraction

Tissues were snap-frozen within 30 min after resection and stored at -80 °C until the time of analysis. For nucleic acid isolation, 10 to 15 tissue cryosections (10 to 15 μ m each) were prepared for each patient. The first and the last sections in each series were stained with hematoxylin and eosin (H&E) and tumor samples were reviewed by an experienced lung pathologist to determine the proportions of viable tumor cells, stromal cells, normal lung cell cells, infiltrating lymphocytes and necrotic areas. Only samples with a viable tumor content of $\geq 50\%$ were used for subsequent analyses. Frozen tumor cryosections and matched normal lung

tissue pieces were homogenized using the TissueLyser mixer-mill disruptor (2 × 2 min at 25 Hz, Qiagen, Hilden, Germany). Total DNA was isolated with the AllPrep DNA/RNA Universal Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. DNA was stored at − 80 °C until further use.

cfMeDIP-seq library preparation and sequencing

5-mC-enriched sequencing libraries were generated employing a previously published protocol designed for small DNA input quantities (cfMeDIP-seq; [20]). In brief, 2 to 10 ng cfDNA was subjected to end-repair and A-tailing followed by sequencing adapter ligation at 16 °C for 15 h using the KAPA HyperPrep Kit with KAPA Dual-Indexed Adapters (Roche, Mannheim, Germany). Prior to immunoprecipitation, libraries were purified by bead-based double-sided size selection and spiked with methylated and unmethylated control DNA fragments (prepared as described in Song et al. [58]) for assessment of 5-mC enrichment efficiency. Additionally, λ phage filler DNA was added to bring the DNA amount to a total of 100 ng. The MagMeDIP qPCR and iPure v2 Kits (Diagenode, Seraing, Belgium) were used for methylation immunoprecipitation and DNA purification, respectively. Enrichment efficiency was assessed by means of qPCR quantification of methylated and unmethylated spike-ins and library amplification was carried out using 12 PCR cycles, followed by another bead clean-up. The final libraries were quantified using the Qubit DNA High Sensitivity Kit and checked for appropriate adapter ligation with the Bioanalyzer 2100 System. Libraries were pooled equimolarly and sequenced in 8-plexes on an Illumina NextSeq550 instrument with high-output reagents (75-bp paired-end reads). Sheared genomic DNA from tissue specimens was processed using the same protocol with two exceptions: (1) 250 ng of DNA input was used per sample and (2) library preparation was performed without the addition of λ phage filler DNA.

Fastq raw reads were adapter trimmed by Cutadapt v3.7 [59] and aligned to the human reference genome GRCh37/hg19 using bowtie2 v2.3.5.1 [60] in paired mode. Aligned reads were indexed, sorted, and filtered by samtools v1.9 [61], retaining only properly paired reads with MAPQ > 10. Duplicate reads were marked with Picard v2.25.1 (MarkDuplicates) and collapsed to allow one read per alignment position. Sequencing data quality was assessed with fastqc v0.11.5 and the MEDIPS R package [62], evaluating coverage saturation (MEDIPS.saturation), CpG enrichment (MEDIPS.CpGenrich), and CpG coverage (MEDIPS.seqCoverage). Per-sample quality metrics are summarized in Additional file 1: Table S5.

Identification of ALK-associated DMRs

Paired fragments were counted by Subread v1.5.3 (featureCounts) [63] at 7,264,350 non-overlapping windows previously described to span CpGs with concordant 5-mC signals (methylation blocks) [18]. Windows covering < 3 CpG sites, mapping to chromosomes X, Y or the mitochondrial genome were excluded. To enrich for windows with cancer-informative 5-mC signals, we inferred regions frequently hypermethylated in plasma cfDNA of healthy individuals and excluded those from further analyses. In brief, whole-genome bisulfite sequencing data of cell types composing healthy donor cfDNA was downloaded (GSE186458 [18] and BLUEPRINT [64]) and processed with the wgbstools suite (https://github.com/nloyfer/wgbs_tools), averaging beta values falling into the same window. Beta values of each cell type were weighted according to their predicted abundance in healthy cfDNA [18] and summed to yield total DNA methylation. Windows with β values > 0.15 in the combined dataset were excluded.

To identify *ALK*-associated DMRs, we performed differential methylation analysis between cfDNA from healthy donors ($n=13$) and *ALK*-positive NSCLC patients ($n=21$). For patients with longitudinal plasma available, we only considered the sample with the highest t-MAD score, reasoning that these contain elevated quantities of cancer-derived cfDNA. Differential analysis was limited to cancer-informative genomic windows remaining after the filtering steps described before. Additionally, windows with low read counts across all samples were excluded (i.e., < 20% of the total number of samples). Trimmed mean of M values (TMM)-normalized counts [65] were subjected to differential methylation analysis using the limma package in R [66]. Following variance smoothing, a linear model using weighted least squares was fit for each genomic region. *P* values between cancer and control conditions were calculated by empirical Bayes smoothing. Significantly hyper- or hypomethylated regions were called at adjusted *p* values (Benjamini-Hochberg) < 0.1 and $|\log_2FC| > 1$. *ALK* tissue DMRs (i.e., *ALK* tissues vs. normal lung tissues) were identified likewise, omitting the exclusion of genomic regions hypermethylated in healthy plasma cfDNA.

5-mC score calculation

Aligned bam files of 13 healthy control samples were downsampled to a common read coverage and merged, yielding a combined read depth of 28 million paired-reads (median read depth across all cfDNA cfMeDIP-seq datasets). The resulting normal reference file was used as a baseline to quantitatively assess the extent of cancer-derived 5-mC changes in our patient plasma samples.

The median absolute RPKM (reads per kilobase per million mapped reads) deviation from this baseline at relevant hyper-DMRs was calculated per sample and defined as ‘5-mC score.’

Processing of publicly available DNA methylation and gene expression data of tissue samples

Illumina 450 k methylation array and RNA sequencing data of primary tumor tissues from lung adenocarcinoma patients—alongside various adjacent normal tissues (breast, bladder, colon, endometrium, head and neck, kidney, liver, lung, prostate, and thyroid gland)—was obtained from TCGA [30, 31]. Additional gene expression data of normal lung tissues ($n=288$) were retrieved from the GTEx [67]. All datasets, alongside clinical and molecular annotations, were downloaded from the Xena platform [68]. DNA methylation array data were adjusted to the genomic regions used for DMR calling from cfMeDIP-seq datasets by averaging β values of CpGs falling into the same region. Differential methylation analysis between LUAD ($n=455$) and normal lung tissue samples ($n=75$; taken from LUAD and LUSC cohorts) was performed using limma [66], only considering genomic regions with healthy cfDNA 5-mC signals $\beta \leq 0.15$. Regions with $|\Delta\beta| \geq 0.25$ and adjusted p value < 0.01 were deemed significantly hyper- or hypomethylated. Further differential methylation analyses, stratifying patient samples by pathologic stage and molecular driver event, are summarized in Additional file 1: Table S3. *ALK*-positive patients within the TCGA-LUAD cohort were determined using the TumorFusions data portal [69]. To identify genomic regions with gene regulatory 5-mC signals, we correlated matched DNA methylation and gene expression data of adjacent normal tissues ($n=15$ per tissue type). 5-mC signals demonstrating a significant negative correlation (Spearman, $p < 0.05$) to the expression level of its associated gene were considered to be involved in transcriptional regulation. Genomic feature annotation was performed using the annotatr R package [70].

Genomic cfDNA biomarkers

Information on cancer-specific genomic alterations of the patient plasma samples profiled in this study was obtained from previously published work (for detailed descriptions see [2, 4]). Somatic mutations and *EML4-ALK* fusion abundances were determined by hybrid-capture sequencing using the AVENIO ctDNA Library Preparation Kit followed by sequencing with the Targeted or Surveillance Panel (Roche, Mannheim, Germany). Mutations with variant allele frequencies (VAFs) $\geq 30\%$ were considered germline mutations and consequentially excluded from further analyses. VAFs $< 0.01\%$ were

deemed undetectable. Genome-wide copy number profiles and t-MAD scores were estimated from sWGS data using the ichorCNA algorithm [47] and t-MAD score calculation documentation (<https://github.com/sdchandra/tMAD>) [29], respectively. CNA calling was carried out at 1-Mb bin sizes using sWGS data of 16 healthy control samples as copy number neutral references. The sequencing data was downsampled to 5 M paired reads prior to t-MAD score calculation. The maximal t-MAD score across all healthy control samples (0.0081) was set as the detection threshold. Chromosomal instabilities were similarly assessed from cfMeDIP-seq data with 5-mC-enriched sequencing data of healthy controls ($n=13$) as copy number neutral reference for normalization.

Statistical analyses and data visualization

A comparison between independent and paired data was made using the Mann–Whitney U test and Wilcoxon’s paired test, respectively (as labeled in graphs). Spearman’s correlation was used to test the association between *ALK* cfDNA and *ALK* tissue 5-mC signals as well as 5-mC scores and genomic biomarker abundances. T-MAD scores inferred from sWGS and cfMeDIP-seq were compared by Pearson’s correlation. Permutation testing to estimate the significance of the overlap between tissue and cfDNA DMRs was performed with the regioneR R package [71], comparing the observed overlap to a null distribution of 10,000 random samplings. The permutation test p value represents the number of random samplings with overlaps greater or equal to the observed overlap divided by the number of random permutations. Survival data were analyzed according to Kaplan–Meier using the log-rank test for OS comparison. Statistical analyses were performed in R (version 3.6.2) [72] and relevant graphs were generated using the ggplot2 R package [73].

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13148-022-01387-4>.

Additional file 1: Supplementary results.

Acknowledgements

We thank Simon Ogronnik and Sabrina Gerhardt for their excellent technical assistance throughout this project. We thank Ingrid Heinzmann-Groth, Karin Schnorr-Teichert, Saskia Oestinger, Christa Stolp, and Martin Fallenbuechel for clinical sample collection and processing.

Author contributions

F.J. contributed to conceptualization, methodology, data acquisition, formal analysis, writing—original draft, writing—review and editing, and visualization; A.K.A., A.L.R., S.B., M.R., A.S., M.A.S., T.M., M.T., contributed to methodology, formal analysis, visualization, and writing—review and editing; P.C. contributed to conceptualization, methodology, data acquisition, formal analysis, writing—original draft, writing—review and editing, visualization, and supervision;

H.S. contributed to conceptualization, methodology, writing—original draft, writing—review and editing, supervision, project administration, and funding acquisition. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. The study was funded by The German Center for Lung Research (DZL; Grant ID: 82ZDL004A4).

Availability of data and materials

All sequencing data generated in this study are deposited in the European Genome-Phenome Archive (EGA) under accession number EGAS00001006573. TCGA Illumina 450 k methylation array and RNA sequencing data are publicly available through the Genomics Data Commons (<https://portal.gdc.cancer.gov>) and the GTEx portal (<https://gtexportal.org>). For this study, data were downloaded as uniformly preprocessed datasets from the Xena platform (<https://xenabrowser.net>). WGBS data of hematopoietic cell types, vascular endothelial cells, and hepatocytes were downloaded from the Gene Expression Omnibus (GSE186458). The code used to process cfMeDIP-seq data, identify 5-mC biomarkers, and calculate 5-mC scores can be obtained from the authors upon reasonable request.

Declarations

Ethics approval and consent to participate

All individuals provided written informed consent, and the study was approved by the ethics committee at Heidelberg (S-270/2001, S-296/2016, S-435/2019) and Lübeck Universities (AZ 12-238).

Consent for publication

Not applicable.

Competing interests

MR reports personal fees from Amgen, AstraZeneca, BMS, Boehringer-Ingelheim, Lilly, Merck, MSD, Novartis, Pfizer, Roche, and Samsung, outside the submitted work. AS reports advisory board honoraria from BMS, Bayer, AstraZeneca, Thermo Fisher, Novartis, Seattle Genomics speaker's honoraria from BMS, Bayer, Illumina, AstraZeneca, Novartis, Thermo Fisher, MSD, Roche, as well as research funding from Chugai, Bayer, and BMS, outside the submitted work. MAS reports personal fees from the German Center for Lung Research (DZL), outside the submitted work. TM reports grants and non-financial support from Roche Diagnostics GmbH, Penzberg, Germany, outside the submitted work; In addition, TM has patents WO2019158460, WO2019211418, WO2019215223, and EP3365679 pending and EP3391053 issued. PC reports research funding from AstraZeneca, Novartis, Roche, and Takeda as well as an advisory board and/or lecture fees from Boehringer Ingelheim, Chugai, Lilly, Pfizer, and Takeda, outside the submitted work. HS reports grants from Roche Sequencing Solutions and personal fees from Roche, outside the submitted work.

Author details

¹Division of Cancer Genome Research, German Cancer Research Center, National Center for Tumor Diseases, Heidelberg, Germany. ²German Center for Lung Research (DZL), TLRC Heidelberg, Heidelberg, Germany. ³Helmholtz Young Investigator Group, Multiparametric Methods for Early Detection of Prostate Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁴Department of Urology, Heidelberg University Hospital, Heidelberg, Germany. ⁵Faculty of Biosciences, Heidelberg University, Heidelberg, Germany. ⁶Lung Clinic Grosshansdorf, Airway Research Center North, German Center for Lung Research, Grosshansdorf, Germany. ⁷Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany. ⁸German Cancer Consortium (DKTK), Heidelberg, Germany. ⁹Translational Research Unit, Thoraxklinik at University Hospital Heidelberg, Heidelberg, Germany. ¹⁰Department of Oncology, Thoraxklinik and National Center for Tumor Disease (NCT) at Heidelberg University Hospital, Heidelberg, Germany.

Received: 29 September 2022 Accepted: 25 November 2022

Published online: 02 December 2022

References

- Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature*. 2017;545(7655):446–51.
- Angeles AK, Christopoulos P, Yuan Z, Bauer S, Janke F, Ogdodnik SJ, et al. Early identification of disease progression in ALK-rearranged lung cancer using circulating tumor DNA analysis. *npj Precision Oncol*. 2021;5(1):100.
- Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science*. 2018;359(6378):926–30.
- Dietz S, Christopoulos P, Yuan Z, Angeles AK, Gu L, Volckmar AL, et al. Longitudinal therapy monitoring of ALK-positive lung cancer by combined copy number and targeted mutation profiling of cell-free DNA. *EBioMedicine*. 2020;62:103103.
- Kurtz DM, Soo J, Co Ting Keh L, Alig S, Chabon JJ, Sworder BJ, et al. Enhanced detection of minimal residual disease by targeted sequencing of phased variants in circulating tumor DNA. *Nat Biotechnol*. 2021;39(12):1537–47.
- Newman AM, Bratman SV, To J, Wynne JF, Eclov NCW, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med*. 2014;20(5):548–54.
- Rothwell DG, Ayub M, Cook N, Thistlethwaite F, Carter L, Dean E, et al. Utility of ctDNA to support patient selection for early phase clinical trials: the TARGET study. *Nat Med*. 2019;25(5):738–43.
- Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med*. 2014;6(224):224ra24.
- Lehmann-Werman R, Neiman D, Zemmour H, Moss J, Magenheimer J, Vaknin-Dembinsky A, et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc Natl Acad Sci*. 2016;113(13):E1826–34.
- Shen SY, Singhanian R, Fehringer G, Chakravarthy A, Roehrl MHA, Chadwick D, et al. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature*. 2018;563(7732):579–83.
- Sun K, Jiang P, Chan KCA, Wong J, Cheng YKY, Liang RHS, et al. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. *Proc Natl Acad Sci*. 2015;112(40):E5503–12.
- Dor Y, Cedar H. Principles of DNA methylation and their implications for biology and medicine. *Lancet*. 2018;392(10149):777–86.
- Duruiseaux M, Esteller M. Lung cancer epigenetics: from knowledge to applications. *Semin Cancer Biol*. 2018;51:116–28.
- Esteller M. Epigenetics in cancer. *N Engl J Med*. 2008;358(11):1148–59.
- Cheng N, Skead K, Ouellette T, Bratman S, De Carvalho D, Soave D, et al. Early signatures of breast cancer up to seven years prior to clinical diagnosis in plasma cell-free DNA methylomes. *Research Square*; 2022.
- Klein EA, Richards D, Cohn A, Tummala M, Lapham R, Cosgrove D, et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann Oncol*. 2021;32(9):1167–77.
- Li W, Li Q, Kang S, Same M, Zhou Y, Sun C, et al. CancerDetector: ultrasensitive and non-invasive cancer detection at the resolution of individual reads using cell-free DNA methylation sequencing data. *Nucleic Acids Res*. 2018;46(15):e89.
- Loyfer N, Magenheimer J, Peretz A, Cann G, Bredno J, Klochendler A, et al. A human DNA methylation atlas reveals principles of cell type-specific methylation and identifies thousands of cell type-specific regulatory elements. *bioRxiv*. 2022:2022.01.24.477547.
- Moss J, Magenheimer J, Neiman D, Zemmour H, Loyfer N, Korach A, et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat Commun*. 2018;9(1):5068.
- Shen SY, Burgener JM, Bratman SV, De Carvalho DD. Preparation of cfMeDIP-seq libraries for methylation profiling of plasma cell-free DNA. *Nat Protoc*. 2019;14(10):2749–80.
- Burgener JM, Zou J, Zhao Z, Zheng Y, Shen SY, Huang SH, et al. Tumor-Naïve multimodal profiling of circulating tumor DNA in head and neck squamous cell carcinoma. *Clin Cancer Res*. 2021;27(15):4230–44.
- Nassiri F, Chakravarthy A, Feng S, Shen SY, Nejad R, Zuccato JA, et al. Detection and discrimination of intracranial tumors using plasma cell-free DNA methylomes. *Nat Med*. 2020;26(7):1044–7.

23. Nuzzo PV, Berchuck JE, Korthauer K, Spisak S, Nassar AH, Abou Alaiwi S, et al. Detection of renal cell carcinoma using plasma and urine cell-free DNA methylomes. *Nat Med*. 2020;26(7):1041–3.
24. Lasseter K, Nassar AH, Hamieh L, Berchuck JE, Nuzzo PV, Korthauer K, et al. Plasma cell-free DNA variant analysis compared with methylated DNA analysis in renal cell carcinoma. *Genet Med*. 2020;22(8):1366–73.
25. Peter MR, Bilenyk M, Isserlin R, Bader GD, Shen SY, De Carvalho DD, et al. Dynamics of the cell-free DNA methylome of metastatic prostate cancer during androgen-targeting treatment. *Epigenomics*. 2020;12(15):1317–32.
26. Elsayed M, Christopoulos P. Therapeutic sequencing in ALK(+) NSCLC. *Pharmaceuticals (Basel)*. 2021;14(2).
27. Lin JJ, Riely GJ, Shaw AT. Targeting ALK: precision medicine takes on drug resistance. *Cancer Discov*. 2017;7(2):137–55.
28. Mok T, Camidge DR, Gadgeel SM, Rosell R, Dziadziuszko R, Kim DW, et al. Updated overall survival and final progression-free survival data for patients with treatment-naïve advanced ALK-positive non-small-cell lung cancer in the ALEX study. *Ann Oncol*. 2020;31(8):1056–64.
29. Moliere F, Chandrananda D, Piskorz AM, Moore EK, Morris J, Ahlborn LB, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med*. 2018;10(466).
30. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a shared vision for cancer genomic data. *N Engl J Med*. 2016;375(12):1109–12.
31. Weisenberger DJ. Characterizing DNA methylation alterations from The Cancer Genome Atlas. *J Clin Invest*. 2014;124(1):17–23.
32. Chen C, Huang X, Yin W, Peng M, Wu F, Wu X, et al. Ultrasensitive DNA hypermethylation detection using plasma for early detection of NSCLC: a study in Chinese patients with very small nodules. *Clin Epigenetics*. 2020;12(1):39.
33. Harada H, Miyamoto K, Yamashita Y, Taniyama K, Mihara K, Nishimura M, et al. Prognostic signature of protocadherin 10 methylation in curatively resected pathological stage I non-small-cell lung cancer. *Cancer Med*. 2015;4(10):1536–46.
34. Khalil AA, Sivakumar S, Lucas FAS, McDowell T, Lang W, Tabata K, et al. TBX2 subfamily suppression in lung cancer pathogenesis: a high-potential marker for early detection. *Oncotarget*. 2017;8(40):68230–41.
35. Wen SWC, Andersen RF, Petersen LMS, Hager H, Hillberg O, Jakobsen A, et al. Comparison of mutated KRAS and methylated HOXA9 tumor-specific DNA in advanced lung adenocarcinoma. *Cancers (Basel)*. 2020;12(12).
36. Li W, Zhou XJ. Methylation extends the reach of liquid biopsy in cancer detection. *Nat Rev Clin Oncol*. 2020;17(11):655–6.
37. Aisner DL, Sholl LM, Berry LD, Rossi MR, Chen H, Fujimoto J, et al. The impact of smoking and TP53 mutations in lung adenocarcinoma patients with targetable mutations—the lung cancer mutation consortium (LCMC2). *Clin Cancer Res*. 2018;24(5):1038–47.
38. Christopoulos P, Endris V, Bozorgmehr F, Elsayed M, Kirchner M, Ristau J, et al. 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*. 2018;142(12):2589–98.
39. Christopoulos P, Kirchner M, Bozorgmehr F, Endris V, Elsayed M, Budczies J, et al. Identification of a highly lethal V3(+) TP53(+) subset in ALK(+) lung adenocarcinoma. *Int J Cancer*. 2019;144(1):190–9.
40. Woo CG, Seo S, Kim SW, Jang SJ, Park KS, Song JY, et al. 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*. 2017;28(4):791–7.
41. Shaw AT, Kim DW, Nakagawa K, Seto T, Crinó L, Ahn MJ, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med*. 2013;368(25):2385–94.
42. Solomon BJ, Mok T, Kim DW, Wu YL, Nakagawa K, Mekhail T, et al. First-line crizotinib versus chemotherapy in ALK-positive lung cancer. *N Engl J Med*. 2014;371(23):2167–77.
43. Dagogo-Jack I, Brannon AR, Ferris LA, Campbell CD, Lin JJ, Schultz KR, et al. Tracking the evolution of resistance to ALK tyrosine kinase inhibitors through longitudinal analysis of circulating tumor DNA. *JCO Precis Oncol*. 2018;2018.
44. Dagogo-Jack I, Rooney M, Lin JJ, Nagy RJ, Yeap BY, Hubbeling H, et al. Treatment with next-generation ALK inhibitors fuels plasma ALK mutation diversity. *Clin Cancer Res*. 2019;25(22):6662–70.
45. McCoach CE, Blakely CM, Banks KC, Levy B, Chue BM, Raymond VM, et al. Clinical utility of cell-free DNA for the detection of ALK fusions and genomic mechanisms of ALK inhibitor resistance in non-small cell lung cancer. *Clin Cancer Res*. 2018;24(12):2758–70.
46. Smith CG, Moser T, Moliere F, Field-Rayner J, Eldridge M, Riediger AL, et al. Comprehensive characterization of cell-free tumor DNA in plasma and urine of patients with renal tumors. *Genome Medicine*. 2020;12(1):23.
47. Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun*. 2017;8(1):1324.
48. Chemi F, Pearce SP, Clipson A, Hill SM, Conway A-M, Richardson SA, et al. cfDNA methylome profiling for detection and subtyping of small cell lung cancers. *Nature Cancer*. 2022.
49. Moliere F, Piskorz AM, Chandrananda D, Moore E, Morris J, Smith CG, et al. Selecting short DNA fragments in plasma improves detection of circulating tumour DNA. *bioRxiv*. 2017:134437.
50. Guo S, Diep D, Plongthongkum N, Fung H-L, Zhang K, Zhang K. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat Genet*. 2017;49(4):635–42.
51. Alvarez H, Opalinska J, Zhou L, Sohal D, Fazzari MJ, Yu Y, et al. Widespread hypomethylation occurs early and synergizes with gene amplification during esophageal carcinogenesis. *PLoS Genet*. 2011;7(3):e1001356.
52. Yang X, Gao L, Zhang S. Comparative pan-cancer DNA methylation analysis reveals cancer common and specific patterns. *Brief Bioinform*. 2017;18(5):761–73.
53. Zhang J, Huang K. Pan-cancer analysis of frequent DNA co-methylation patterns reveals consistent epigenetic landscape changes in multiple cancers. *BMC Genomics*. 2017;18(Suppl 1):1045.
54. Zviran A, Schulman RC, Shah M, Hill STK, Deochand S, Khamnei CC, et al. Genome-wide cell-free DNA mutational integration enables ultra-sensitive cancer monitoring. *Nat Med*. 2020;26(7):1114–24.
55. Christopoulos P, Dietz S, Angeles AK, Rheinheimer S, Kazdal D, Volckmar AL, et al. Earlier extracranial progression and shorter survival in ALK-rearranged lung cancer with positive liquid biopsies. *Transl Lung Cancer Res*. 2021;10(5):2118–31.
56. Kwon M, Ku BM, Olsen S, Park S, Lefterova M, Odegaard J, et al. Longitudinal monitoring by next-generation sequencing of plasma cell-free DNA in ALK rearranged NSCLC patients treated with ALK tyrosine kinase inhibitors. *Cancer Med*. 2022.
57. Madsen AT, Winther-Larsen A, McCulloch T, Meldgaard P, Sorensen BS. Genomic Profiling of Circulating Tumor DNA Predicts Outcome and Demonstrates Tumor Evolution in ALK-Positive Non-Small Cell Lung Cancer Patients. *Cancers (Basel)*. 2020;12(4).
58. Song C-X, Yin S, Ma L, Wheeler A, Chen Y, Zhang Y, et al. 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Res*. 2017;27(10):1231–42.
59. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011. 2011;17(1):3.
60. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–9.
61. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
62. Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics*. 2013;30(2):284–6.
63. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–30.
64. Fernández JM, de la Torre V, Richardson D, Royo R, Puiggròs M, Moncunill V, et al. The BLUEPRINT data analysis portal. *Cell Syst*. 2016;3(5):491–5.e5.
65. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40.
66. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.

67. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580–5.
68. Goldman MJ, Craft B, Hastie M, Repčička K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol.* 2020;38(6):675–8.
69. Hu X, Wang Q, Tang M, Barthel F, Amin S, Yoshihara K, et al. TumorFusions: an integrative resource for cancer-associated transcript fusions. *Nucleic Acids Res.* 2018;46(D1):D1144–9.
70. Cavalcante RG, Sartor MA. annotatr: genomic regions in context. *Bioinformatics.* 2017;33(15):2381–3.
71. Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics.* 2016;32(2):289–91.
72. Team RC. R: A language and environment for statistical computing. 2013.
73. Wilkinson L. ggplot2: Elegant graphics for data analysis by WICKHAM. H *Biometrics.* 2011;67(2):678–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

