



Detecting Small Cell Transformation in Patients with Advanced *EGFR* Mutant Lung Adenocarcinoma through Epigenomic cfDNA Profiling

Talal El Zarif^{1,2,3}, Catherine B. Meador^{4,5}, Xintao Qiu³, Ji-Heui Seo^{1,2}, Matthew P. Davidsohn^{1,2}, Hunter Savignano^{1,2}, Gitanjali Lakshminarayanan^{1,2}, Heather M. McClure^{1,2}, John Canniff^{1,2}, Brad Fortunato^{1,2}, Rong Li², Mandeep K. Banwait^{4,5}, Karl Semaan^{1,2,6}, Marc Eid^{1,2}, Henry Long², Yin P. Hung⁷, Navin R. Mahadevan^{1,8}, David A. Barbie¹, Matthew G. Oser¹, Zofia Piotrowska^{4,5}, Toni K. Choueiri¹, Sylvan C. Baca^{1,2,6}, Aaron N. Hata^{4,5}, Matthew L. Freedman^{1,2,6}, and Jacob E. Berchuck¹

ABSTRACT

Purpose: Histologic transformation to small cell lung cancer (SCLC) is a mechanism of treatment resistance in patients with advanced oncogene-driven lung adenocarcinoma (LUAD) that currently requires histologic review for diagnosis. Herein, we sought to develop an epigenomic cell-free DNA (cfDNA)-based approach to noninvasively detect small cell transformation in patients with *EGFR* mutant (*EGFRm*) LUAD.

Experimental Design: To characterize the epigenomic landscape of transformed (t)SCLC relative to LUAD and *de novo* SCLC, we performed chromatin immunoprecipitation sequencing (ChIP-seq) to profile the histone modifications H3K27ac, H3K4me3, and H3K27me3; methylated DNA immunoprecipitation sequencing (MeDIP-seq); assay for transposase-accessible chromatin sequencing; and RNA sequencing on 26 lung cancer patient-derived xenograft (PDX) tumors. We then generated and analyzed H3K27ac ChIP-seq, MeDIP-seq, and whole genome sequencing cfDNA data from

1 mL aliquots of plasma from patients with *EGFRm* LUAD with or without tSCLC.

Results: Analysis of 126 epigenomic libraries from the lung cancer PDXs revealed widespread epigenomic reprogramming between LUAD and tSCLC, with a large number of differential H3K27ac ($n = 24,424$), DNA methylation ($n = 3,298$), and chromatin accessibility ($n = 16,352$) sites between the two histologies. Tumor-informed analysis of each of these three epigenomic features in cfDNA resulted in accurate noninvasive discrimination between patients with *EGFRm* LUAD versus tSCLC [area under the receiver operating characteristic curve (AUROC) = 0.82–0.87]. A multianalyte cfDNA-based classifier integrating these three epigenomic features discriminated between *EGFRm* LUAD versus tSCLC with an AUROC of 0.94.

Conclusions: These data demonstrate the feasibility of detecting small cell transformation in patients with *EGFRm* LUAD through epigenomic cfDNA profiling of 1 mL of patient plasma.

Introduction

Advanced *EGFR*-mutant (*EGFRm*) lung adenocarcinoma (LUAD) is an archetype of precision oncology, initially treated with selective *EGFR* tyrosine kinase inhibitors (TKI; ref. 1). While there are currently five *EGFR* TKIs approved in the United States, osimertinib, a highly selective third-generation *EGFR* TKI, is the preferred first-line choice based on improved progression-free and overall survival compared to earlier generation TKIs (2). Despite

initial robust responses to *EGFR* TKIs, acquired resistance inevitably occurs. Subsequent treatment decisions are ideally made following evaluation for potentially targetable resistance mechanisms on a post-progression tumor biopsy, which uncovers potential therapeutic approaches in about one-third of patients (3–6).

Histologic transformation from LUAD to small cell lung cancer (SCLC) is one well-characterized resistance mechanism that occurs in up to 15% of patients with *EGFRm* LUAD following progression on an *EGFR* TKI, and appears to be increasing in prevalence with

¹Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts. ²Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, Massachusetts. ³Department of Medicine, Yale School of Medicine, New Haven, Connecticut. ⁴Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts. ⁵Massachusetts General Hospital Cancer Center, Boston, Massachusetts. ⁶Eli and Edythe L. Broad Institute, Cambridge, Massachusetts. ⁷Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts. ⁸Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts.

T. El Zarif, C.B. Meador, and X. Qiu contributed equally to this article.

M.L. Freedman and J.E. Berchuck contributed equally to this article.

Corresponding Authors: Jacob E. Berchuck, Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Dana 9, Boston, MA 02215. E-mail: jacob_berchuck@dfci.harvard.edu; and Matthew L. Freedman, 450 Brookline Avenue, Smith 10, Boston, MA 02215. E-mail: matthew_freedman@dfci.harvard.edu

Clin Cancer Res 2024;30:3798–811

doi: 10.1158/1078-0432.CCR-24-0466

This open access article is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

©2024 The Authors; Published by the American Association for Cancer Research

Translational Relevance

Histologic transformation to small cell lung cancer (SCLC) is an increasingly common resistance mechanism to EGFR tyrosine kinase inhibitors in *EGFR* mutant lung adenocarcinoma (LUAD) that is underdiagnosed in clinical practice due to the requirement for tissue biopsy. Early and accurate detection of transformed (t)SCLC has important prognostic and therapeutic implications. To address this unmet need, we first comprehensively profiled the epigenomes of metastatic lung tumors finding widespread epigenomic reprogramming during histologic transformation from LUAD to SCLC. We then utilized a novel approach for epigenomic profiling of cell-free DNA (cfDNA), which discriminated patients with *EGFR*-mutant (*EGFR*m) tSCLC from patients with *EGFR*m LUAD with greater than 90% accuracy. This first demonstration of the ability to accurately and noninvasively detect small cell transformation in patients with *EGFR*m LUAD through epigenomic cfDNA profiling is a critical step toward a new paradigm of diagnostic and therapeutic precision for patients with advanced lung cancer.

the use of more selective EGFR TKIs, such as osimertinib (4–7). *EGFR*m LUADs that undergo SCLC transformation are aggressive cancers that portend poor prognosis and require a change in therapeutic regimen (8). Transformed (t)SCLC clinically mimics *de novo* SCLC—a distinct diagnosis which, unlike *EGFR*m LUAD, is typically associated with smoking history. Both *de novo* and tSCLC are treated with platinum-etoposide chemotherapy (1).

Current guidelines recommend that patients with *EGFR*m LUAD progressing on targeted therapy undergo tumor biopsy to evaluate for actionable mechanisms of TKI resistance, including acquired genomic alterations as well as histologic transformation (1). However, tumor biopsies pose risks to patients and are often not clinically feasible. Consequently, less than half of patients with metastatic *EGFR*m LUAD undergo tumor tissue biopsy at the time of TKI resistance (9). Further, due to inpatient tumor heterogeneity, targetable mechanisms of resistance (including histologic transformation) may be missed by sampling a single metastatic focus, resulting in lost opportunities to deliver guideline-recommended histology-directed therapy.

For these reasons, molecular profiling at the time of EGFR TKI resistance is increasingly performed via liquid biopsies, which can detect somatically acquired genomic alterations in circulating tumor cell-free DNA (cfDNA). Compared to tumor biopsies, liquid biopsies are minimally invasive, can easily be repeated at multiple timepoints, and may better capture inpatient tumor heterogeneity (10). However, limitations in current commercially available liquid biopsies preclude detection of certain clinically actionable resistance phenotypes that lack defining genomic alterations (6). In particular, the diagnosis of tSCLC cannot be made by currently available liquid biopsy assays.

The development of noninvasive diagnostic approaches to detect SCLC transformation has the potential to usher in a new paradigm of diagnostic and therapeutic precision for patients with advanced lung cancer (8, 11–14). In particular, histologic transformation is accompanied by widespread epigenomic reprogramming, which presents an opportunity to detect tSCLC through epigenomic analysis of cfDNA

(11). We, and others, have developed tools to profile tumor epigenomic features from patient plasma, including DNA methylation (15–18), chromatin accessibility (19, 20), and histone modifications (21, 22). Herein, we build upon this work, generating and analyzing 351 tissue and plasma epigenomic libraries to demonstrate the clinical utility of epigenomic cfDNA profiling to noninvasively detect small cell transformation in patients with *EGFR*m LUAD (Fig. 1).

Materials and Methods

Subjects and samples

Lung cancer patient-derived xenografts (PDX) were derived from patients with LUAD, *de novo* SCLC, and tSCLC as previously described (23–26). Tissue samples were obtained from donors who provided explicit written consent per the Declaration of Helsinki under an approved Institutional Review Board (IRB) protocol at Massachusetts General Hospital (MGH). All mouse studies were conducted through Institutional Animal Care and Use Committee–approved animal protocols in accordance with institutional guidelines (MGH Subcommittee on Research Animal Care, OLAW Assurance A3596-01). tSCLC PDXs were reviewed by a staff thoracic pathologist to confirm SCLC histology. Plasma samples were collected from patients with advanced lung cancer diagnosed and treated at the Dana-Farber Cancer Institute (DFCI) or MGH between November 2016 and March 2023. All patients provided written informed consent. The use of samples was approved by DFCI (01-045 and 09-171) and MGH (13-416) IRB. Studies were conducted in accordance with recognized ethical guidelines.

cfDNA processing and tumor content calculation

Peripheral blood was collected in EDTA Vacutainer tubes (BD) or Streck Cell-Free DNA tubes and processed within 3 hours of collection. Plasma was separated by centrifugation at 1,600 g for 10 minutes, transferred to microcentrifuge tubes, and centrifuged at 3,000 g at room temperature for 10 minutes. The supernatant was aliquoted and stored at -80°C until the time of DNA extraction. cfDNA was isolated from 1 mL of plasma, using the QIAGEN Circulating Nucleic Acids Kit (Qiagen), eluted in AE buffer, and stored at -80°C . Low-pass whole genome sequencing (LPWGS) was performed on all cfDNA samples. The ichorCNA R package was used to infer copy-number profiles and cfDNA tumor content from read abundance across bins spanning the genome using default parameters (27). Per the published limit of detection of ichorCNA, estimated cfDNA tumor content cut-off of greater than or less than 0.03 were used to characterize samples as having detectable or undetectable circulating tumor DNA, respectively.

Tissue chromatin immunoprecipitation sequencing

Frozen tissue was pulverized using the Covaris CryoPrep system and fixed with 2 mmol/L disuccinimidyl glutarate for 10 minutes followed by 1% formaldehyde buffer for 10 minutes and quenched with glycine. Chromatin was sheared to 300 to 500 bp using the Covaris E220 ultrasonicator and then incubated overnight with the following antibodies coupled with 40 μL protein A and protein G beads (Invitrogen) at 4°C overnight: H3K27ac (Abcam #ab4729, Lot GR3442890-1), H3K4me3 (Thermo Fisher Scientific #PA5-27029, Lot XI3696063), and H3K27me3 (Cell Signaling #9733S, Lot 19). Five percent of the sample was not exposed to antibody and was used as a control. Beads were washed three times each with Low-Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2 mmol/L EDTA, 20

mmol/L Tris-HCl pH 7.5, 150 mmol/L NaCl), High-Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2 mmol/L EDTA, 20 mmol/L Tris-HCl pH 7.5, 500 mmol/L NaCl), and LiCl Wash Buffer (10 mmol/L Tris pH 7.5, 250 mmol/L LiCl, 1% NP-40, 1% Na-Doc, 1 mmol/L EDTA) and rinsed with TE buffer (pH 8.0) once. Samples were then de-cross-linked, treated with RNase and proteinase K, and DNA was extracted (Qiagen). DNA sequencing libraries were prepared from the purified immunoprecipitated and non-immunoprecipitated DNA using the ThruPLEX DNA-seq Kit (TakaraBio). Libraries were sequenced on an Illumina HiSeq 4000 to generate 150 bp paired-end reads (Novogene Corporation).

Chromatin immunoprecipitation sequencing (ChIP-seq) reads were aligned to the human genome build hg19 using the Burrows-Wheeler Aligner version 0.7.17 (RRID:SCR_010910; ref. 28). Non-uniquely mapped and redundant reads were discarded. MACS v2.1.1.20140616 (RRID:SCR_013291) was used for ChIP-seq peak calling with a q-value threshold of 0.01 (29). IGV v2.8.2 was used to visualize normalized ChIP-seq read counts at specific genomic loci (30). ChIP-seq heatmaps were generated with deepTools v3.3.1 (RRID:SCR_016366) and show normalized read counts at the peak center ± 2 kb unless otherwise noted (31). Overlap of ChIP-seq peaks was assessed using BEDTools v2.26.0. Peaks were considered overlapping if they shared one or more base pairs.

Tissue assay for transposase-accessible chromatin sequencing

Frozen tissue was resuspended and dounce homogenized in 1,000 μ L of homogenization buffer. Nuclei were filtered using a 70- μ m Flowmi strainer, isolated using iodixanol density-gradient centrifugation method, and washed with RSB buffer (10 mmol/L Tris-HCl pH 7.4, 10 mmol/L NaCl, and 3 mmol/L $MgCl_2$ in water). Fifty thousand nuclei were resuspended in 50 μ L of transposition mix [2.5 μ L transposase (100 nmol/L), 16.5 μ L PBS, 0.5 μ L 1% digitonin, 0.5 μ L 10% Tween-20, and 5 μ L water; ref. 32]. Transposition reactions were incubated at 37°C for 30 minutes in a thermomixer shaking at 1,000 rpm. Reactions were cleaned with Qiagen columns. Libraries were amplified using the Omni-ATAC protocol and sequenced on an Illumina platform (Novogene Corporation) using 150-base paired-end reads (33).

Identification and annotation of histology-specific ChIP-seq and assay for transposase-accessible chromatin sequencing peaks

Sample clustering, principal component analysis, and identification of lineage-enriched peaks were performed using Cobra v2.0 (RRID:SCR_005677), a ChIP-seq analysis pipeline implemented with Snakemake (34, 35). ChIP-seq data from LUAD, *de novo* SCLC, and tSCLC PDXs were compared to identify H3K27ac, H3K4me3, and H3K27me3 peaks with significant enrichment in the three tumor subtypes. A union set of peaks for each histone modification was created using BEDTools (RRID:SCR_006646). narrowPeak calls from MACS were used for H3K27ac and H3K4me3 while broadPeak calls were used for H3K27me3. The number of unique aligned reads overlapping each peak in each sample was calculated from BAM files using BEDtools. Quantile normalization was applied to this matrix of normalized read counts for clustering and PCA analysis. Unsupervised hierarchical clustering was performed based on Spearman correlation between samples. Principal component analysis was performed using the prcomp R function. Raw read counts for each peak were normalized to the total number of mapped reads for each sample. Then using DESeq2 v1.14.1 (RRID:SCR_015687), histology-enriched peaks were identified at

the indicated FDR-adjusted P value (p_{adj}) < 0.001 and \log_2 fold-change > 2 (36).

RNA sequencing and differential expression analysis

RNA was extracted from frozen tumor samples using the Qia-gen RNeasy Mini Kit (Cat No./ID: 74104). RNA sequencing (RNA-seq) libraries were constructed from 1 μ g RNA using the Illumina TruSeq Stranded mRNA LT Sample Prep Kit. Barcoded libraries were pooled and sequenced on the Illumina HiSeq 2,500 generating 50 bp paired-end reads. FASTQ files were processed using the VIPER workflow (37). Read alignment to human genome build hg19 was performed with STAR (RRID:SCR_004463; ref. 38). Cufflinks (RRID:SCR_014597) was used to assemble transcript-level expression data from filtered alignments (39). Differential gene expression analysis was conducted using DESeq2 (36).

Methylated DNA immunoprecipitation sequencing

Methylated DNA immunoprecipitation sequencing (MeDIP-seq) was performed on tissue and plasma following published methods (15–18). Library preparation was performed on 10 ng of DNA using the KAPA HyperPrep Kit (KAPA Biosystems). We then performed end-repair, A-tailing, and ligation of NEBNext adaptors (NEBNext Multiplex Oligos for Illumina kit, New England BioLabs). Libraries were digested using the USER enzyme (New England BioLabs). λ DNA, consisting of unmethylated and *in vitro* methylated DNA, was added to prepared libraries to achieve a total amount of 100 ng DNA. Methylated and unmethylated *Arabidopsis thaliana* DNA (Diagenode) was added for quality control. MeDIP was performed using the MagMeDIP Kit (Diagenode) following the manufacturer's protocol. Samples were purified using the iPure Kit v2 (Diagenode). Success of the immunoprecipitation was confirmed using qPCR to detect recovery of the spiked-in *Arabidopsis thaliana* methylated and unmethylated DNA. KAPA HiFi Hotstart ReadyMix (KAPA Biosystems) and NEBNext Multiplex Oligos for Illumina (New England Biolabs) were added to a final concentration of 0.3 μ mol/L and libraries were amplified. Samples were pooled and sequenced (Novogene Corporation) on Illumina HiSeq 4,000 to generate 150 bp paired-end reads.

Quality and quantity of raw MeDIP-seq reads were examined using FastQC version 0.11.5 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) and MultiQC version 1.7 (40). Raw reads were quality and adapter trimmed using Trim Galore! version 0.6.0 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) using default settings in paired-end mode. The trimmed reads then were aligned to hg19 using Bowtie2 version 2.3.5.1 in paired-end mode and all other settings default (41). The SAMtools version 1.10 software suite was used to convert SAM alignment files to BAM format, sort and index reads, and remove duplicates (42). The R package RSamtools version 2.2.1 was used to calculate the number of unique mapped reads. Saturation analyses to evaluate reproducibility of each library were carried out using the R Bioconductor package MEDIPS version 1.38.0 (43).

Cell-free ChIP-seq

One microgram of antibody was coupled with 10 μ L protein A (Invitrogen, cat #10002D) and 10 μ L protein G (Invitrogen, cat #10004D) for at least 6 hours at 4°C with rotation in 0.5% BSA (Jackson Immunology, cat #001-000-161) in PBS (Gibco, cat #14190250), followed by blocking with 1% BSA in PBS for 1 hour at 4°C with rotation. The following antibodies were used: H3K27ac (Abcam #ab4729) and H3K4me3 (Thermo Fisher

Scientific #PA5-27029). A total of 800 μ L of thawed plasma was centrifuged at 3,000 g for 15 minutes at 4°C. The supernatant was precleared with the magnetic beads with 20 μ L protein A and 20 μ L protein G for 2 hours at 4°C. Then, the precleared and conditioned plasma was subjected to antibody-coupled magnetic beads overnight with rotation at 4°C. The reclaimed magnetic beads were washed with 1 mL of each washing buffer twice. Three washing buffers were used in following order: low salt washing buffer (0.1% SDS, 1% Triton X-100, 2 mmol/L EDTA, 150 mmol/L NaCl, 20 mmol/L Tris-HCl pH 7.5), high salt buffer (0.1% SDS, 1% Triton X-100, 2 mmol/L EDTA, 500 mmol/L NaCl, 20 mmol/L Tris-HCl pH 7.5), and LiCl washing buffer (250 mmol/L LiCl, 1% NP-40, 1% Na Deoxycholate, 1 mmol/L EDTA, 10 mmol/L Tris-HCl pH 7.5). Subsequently, the beads were rinsed with TE buffer (Thermo Fisher Scientific, cat #BP2473500) and resuspended and incubated in 100 μ L of DNA extraction buffer containing 0.1 mol/L NaHCO₃, 1% SDS, and 0.6 mg/mL Proteinase K (Qiagen, cat #19131) and 0.4 mg/mL RNaseA (Thermo Fisher Scientific, cat #12091021) for 10 minutes at 37°C, for 1 hour at 50°C and for 90 minutes at 65°C. DNA was purified through phenol extraction (Invitrogen, cat #15593031) and ethanol precipitation was performed with 3 mol/L NaOAc (Ambion, cat #AM9740) and glycogen (Ambion, cat #AM9510). Cell-free ChIP-seq (cfChIP-seq) libraries were prepared with ThruPLEX

DNA-Seq Kit (Takara Bio, cat #R400675) following the manufacturer's instructions. After library amplification, the DNA was purified by AMPure XP (Beckman coulter, cat# A63880). The size distribution of the purified libraries was examined using Agilent 2100 Bioanalyzer with a high sensitivity DNA Chip (Agilent, cat #5067-4626). The library was submitted for the 150 base-pair paired-end sequencing on an Illumina Nova-Seq6000 system (Novogene Corporation).

Generation of cfDNA SCLC risk scores

For cfDNA *histone* data, we computed reads per kilobase per million mapped reads for each cfChIP-seq library using bed files of the SCLC-enriched and LUAD-enriched peaks in the PDXs (FDR-adjusted $P < 0.001$ and \log_2 fold-change > 2). We excluded the top and bottom 0.5% of sites based on the average signal across all cfChIP-seq samples. For cfDNA *methylation* data, we used the MeDIPs R package to calculate CpG-normalized relative methylation scores across 300 bp windows across the genome for each cell-free methylated DNA immunoprecipitation and sequencing (cfMeDIP-seq) library (43, 44). We then summed relative methylation scores in cfDNA at SCLC-enriched and LUAD-enriched differentially methylated regions from the PDXs (FDR-adjusted $P < 1.0 \times 10^{-6}$ and \log_2 fold-change > 3) for each cfDNA sample and

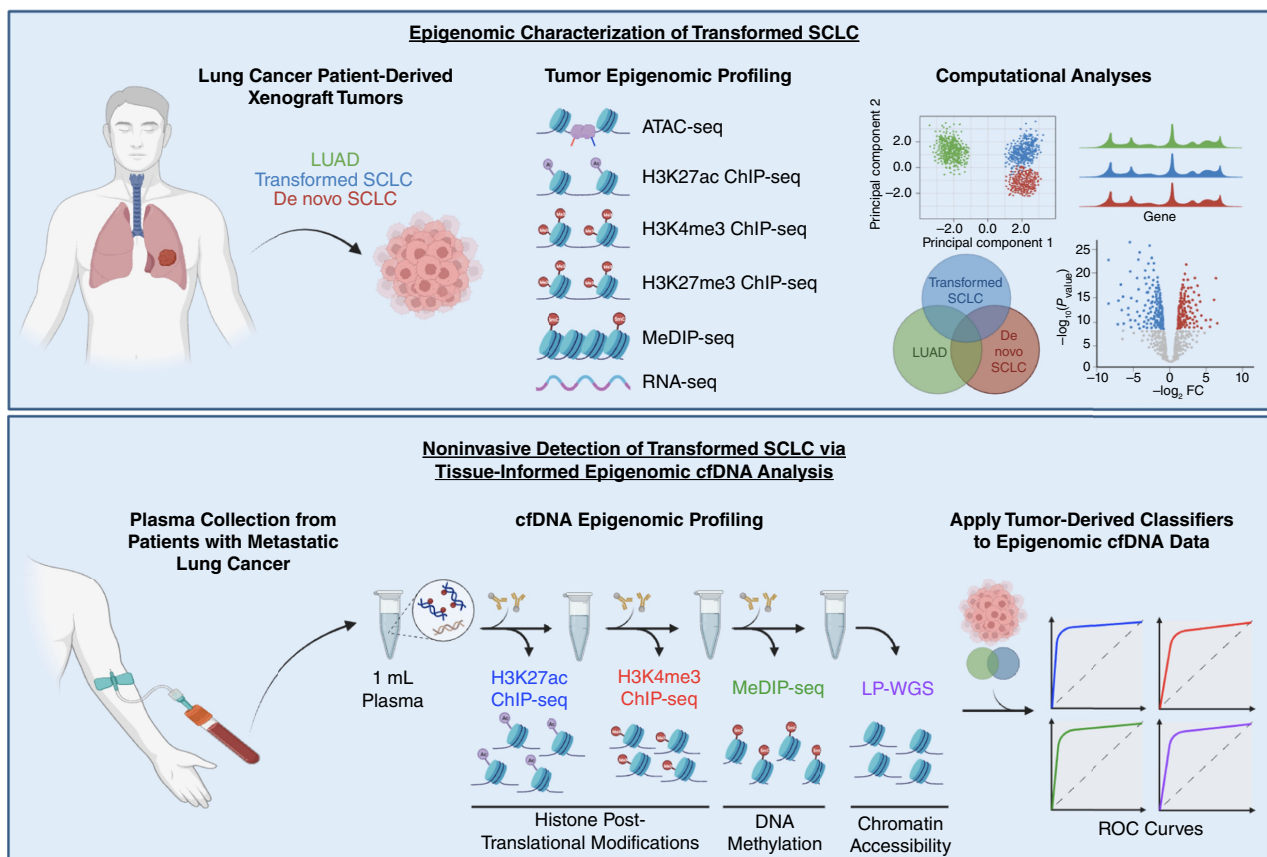


Figure 1.

Overview of the experimental approach to perform comprehensive epigenomic profiling of lung cancer PDXs and multi-analyte epigenomic profiling of cfDNA from 1 mL of patient plasma and noninvasively detect SCLC transformation in patients with *EGFR*^m LUAD. (Created with BioRender.com).

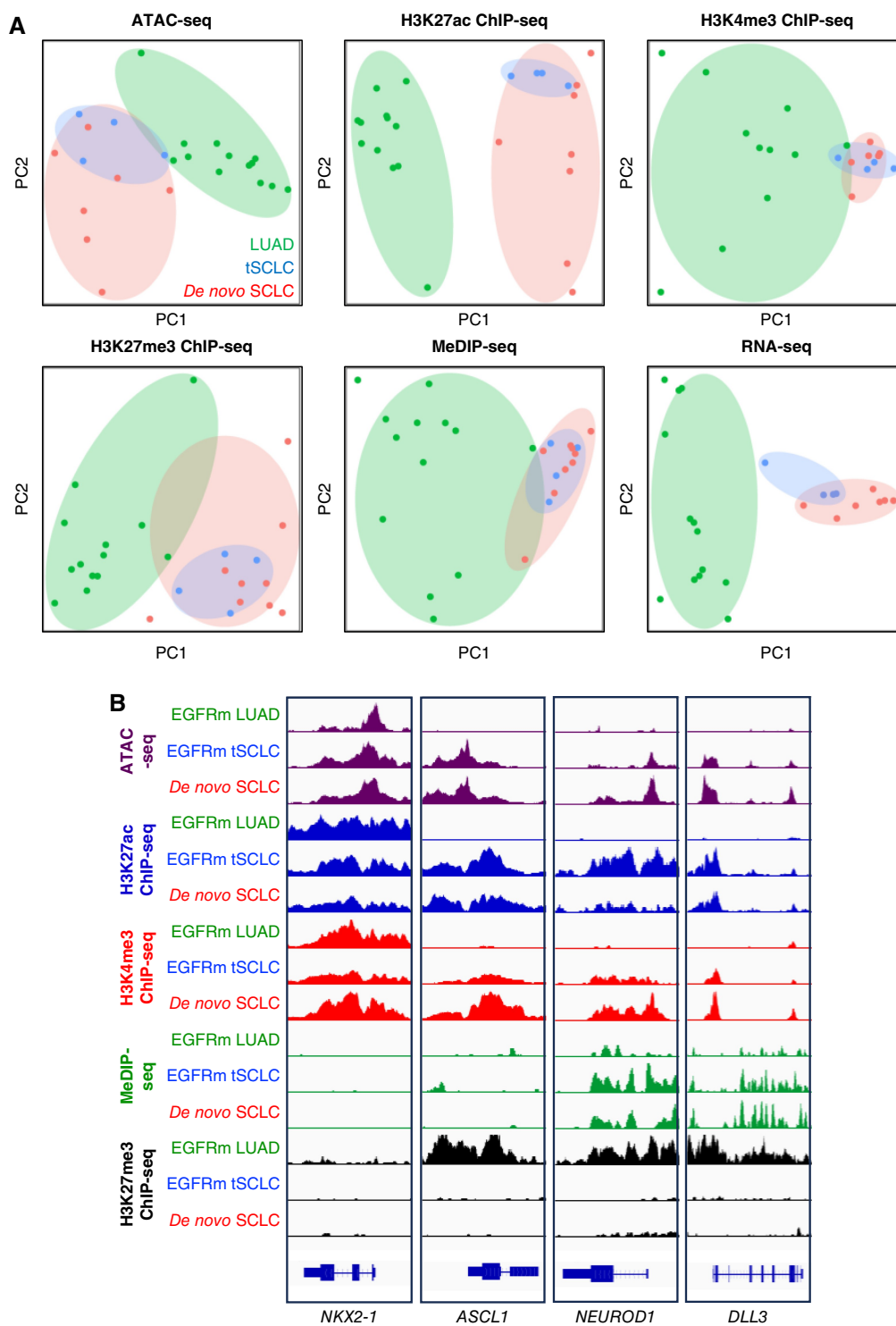


Figure 2.

Comprehensive epigenomic profiling of LUAD, tSCLC, and *de novo* SCLC reveals widespread epigenomic reprogramming in small transformation. **A**, Principal component analysis (PCA) plots of the ATAC, H3K27ac ChIP, H3K4me3 ChIP, H3K27me3 ChIP, MeDIP, and RNA-seq data reveal clustering of tSCLC tumors from *de novo* SCLC tumors and their distinction from LUAD tumors. **B**, Representative epigenomic data from an *EGFRm* tSCLC, *EGFRm* LUAD, and *de novo* SCLC PDX, showing gain of signal in markers of active gene transcription (ATAC-seq, H3K27ac ChIP-seq, HK4me3 ChIP-seq, gene body DNA methylation) and loss of signal with repressive marks (H3K27me3 ChIP-seq) at neural lineage-defining genes in tSCLC. Each track depicts signal intensity for the indicated epigenetic mark in the indicated sample.

normalized this value to the sum of rms values across all 300 bp windows (18). For the cfDNA *chromatin accessibility* analysis, we applied Griffin (v0.1.0), a computational tool that profiles nucleosome protection and accessibility from cfDNA, to all LPWGS libraries (19). Fragments aligning to the region within $\pm 5,000$ bp from the site of interest [defined as differential assay for transposase-accessible chromatin sequencing (ATAC-seq) peaks between SCLC and LUAD PDXs] were extracted. Duplicate and low-quality alignments (mapping quality < 20) were filtered out. Fragments within the nucleosome size range (140–250 bp) were preserved. We then extracted the “mean window coverage” at the SCLC-enriched and the LUAD-enriched regions of open chromatin in the PDXs (FDR-adjusted $P < 0.001$ and \log_2 fold-change > 2). For each data type, the SCLC risk score was calculated as the ratio of signal at the SCLC-enriched sites over the LUAD-enriched sites. This value was normalized to the median score across all samples for that data type and the \log_2 value was calculated, which was reported as the cfDNA SCLC risk score. To generate the *integrated epigenomic* SCLC risk score, we first calculated Z-scores for each sample within each individual data type (DNA methylation, H3K27ac, and chromatin accessibility). The integrated epigenomic cfDNA SCLC risk score was calculated as the sum of the Z-scores for each of the individual cfDNA SCLC risk scores for each sample.

Statistical tests

χ^2 test was used to calculate the P value for the correlation of epigenomic features with differentially expressed genes. All statistical tests were two-sided except where otherwise indicated.

Data availability

The PDX epigenomic data have been deposited in Gene Expression Omnibus (GSE269746). The cfDNA data generated from patient samples that support the findings of this study are available upon request from the corresponding authors (J.E. Berchuck, M.L. Freedman) to comply with the DFCI ethics regulations to protect patient privacy. All requests for raw and analyzed data will be promptly reviewed by the Belfer Office for Dana-Farber Innovations to verify if the request is subject to any intellectual property or confidentiality obligations. Any data and materials that can be shared will be released via a Data Transfer Agreement.

Results

Epigenomic characterization of tSCLC

We first sought to comprehensively characterize the epigenomic landscape of tSCLC relative to LUAD and *de novo* SCLC. We performed this analysis on PDXs developed from biopsies obtained from patients with LUAD, *de novo* SCLC, and tSCLC. We chose this model system for several reasons. First, compared to primary tumor biopsies whose cellular content often contains a high proportion of stromal cells, PDXs facilitate analysis of pure tumor cell populations to more cleanly study tumor-intrinsic molecular features. Second, PDXs can produce large tumor volumes to facilitate molecular analyses that would be infeasible to perform on primary tumor biopsies, which often produce insufficient material for the epigenomic profiling studies performed herein. Importantly, we previously demonstrated that the PDXs in this study recapitulate their original tumor molecular profiles (24). Our analysis focused on 26 lung cancer PDXs, comprising LUAD ($n = 13$), tSCLC ($n = 4$), and *de novo* SCLC ($n = 9$) tumors (Supplementary Table S1). The SCLC tumors were predominantly *ASCL1* (SCLC-A) and/or *NEUROD1*

(SCLC-N) subtype, with one sample each of the *POU2F3* (SCLC-P) and *YAP1* [SCLC-Y, also referred to as triple negative or inflamed (SCLC-I)] subtype (Supplementary Fig. S1; refs. 45–47). Consistent with prior reports, all tSCLC PDXs exhibited loss of *TP53* and/or *RBI* (Supplementary Table S2; refs. 11–14). For each PDX, we performed ChIP-seq to profile the histone modifications H3K27ac (a mark of active gene promoters and enhancers), H3K4me3 (a mark of active gene promoters), and H3K27me3 (a mark of repressed regulatory elements), MeDIP-seq to profile DNA methylation, and ATAC-seq to profile chromatin accessibility, resulting in 126 epigenome-wide libraries (Fig. 1; Supplementary Fig. S2; refs. 23, 24). We also performed bulk RNA-seq on PDXs for which this data had not already been generated (23, 24).

Across every epigenomic feature assessed, unsupervised analyses demonstrated that tSCLC tumors clustered with *de novo* SCLC tumors and were distinct from LUAD tumors (Fig. 2A; Supplementary Fig. S3). tSCLC tumors exhibited a gain of signal associated with active gene transcription—e.g., open chromatin, promoter H3K27ac and H3K4me3, gene body methylation, and loss of gene body H3K27me3—at neural lineage-defining genes, such as *ASCL1*, *NEUROD1*, and *DLL3* (Fig. 2B). Likewise, more of the H3K27ac ChIP-seq and ATAC-seq peaks in tSCLC PDXs were also peaks in *de novo* SCLC than in LUAD PDXs (Supplementary Fig. S4). Collectively, these data affirm a prior report of shifts in DNA methylation and gene expression (11) and demonstrate that histologic transformation from LUAD to SCLC is characterized by widespread epigenomic reprogramming, converging on a profile that resembles *de novo* SCLC.

Development of a tissue-informed epigenomic liquid biopsy approach to detect tSCLC

We sought to leverage these divergent epigenomic profiles to develop a diagnostic test to detect SCLC transformation in patients with *EGFRm* LUAD through epigenomic analysis of cfDNA. Building upon our prior work that tissue-informed analysis improves the detection of tumor-specific epigenomic features in cfDNA (18), we first identified epigenomic features enriched in tSCLC compared to *EGFRm* LUAD. Given the similarity in epigenomic profiles of *EGFRm* and non-*EGFRm* LUAD, and tSCLC and *de novo* SCLC, we included all 26 PDXs in this analysis (Supplementary Fig. S3). Comparative analysis of the LUAD and SCLC tumors resulted in a set of 24,424 H3K27ac, 2,272 H3K4me3, 3,298 DNA methylation, and 16,352 accessible chromatin sites with significantly greater signal in one or the other histologic subtypes. To identify features that would be clinically informative for a liquid biopsy, we removed sites with peaks in white blood cells (WBC), the primary source of background contamination. This step removed 50% of all sites across features, resulting in 10,907 H3K27ac, 833 H3K4me3, 1,210 DNA methylation, and 9,995 accessible chromatin sites enriched in SCLC or LUAD without signal in WBCs (Fig. 3A; Supplementary Fig. S5). Importantly, these epigenomic features strongly correlated with transcriptional activity. There was a strong association of upregulated genes in one histologic subtype having a nearby H3K27ac, H3K4me3, or open chromatin peak in that subtype (Fig. 3B; Supplementary Fig. S6).

We next performed multianalyte epigenomic cfDNA profiling on 48 plasma samples collected from 32 patients with metastatic lung cancer. This cohort comprised 20 patients with *EGFRm* LUAD who never developed tSCLC (*EGFRm* LUAD) and 12 patients with *EGFRm* LUAD following diagnosis of biopsy-proven tSCLC (tSCLC), 6 of whom also had plasma collected prior to being

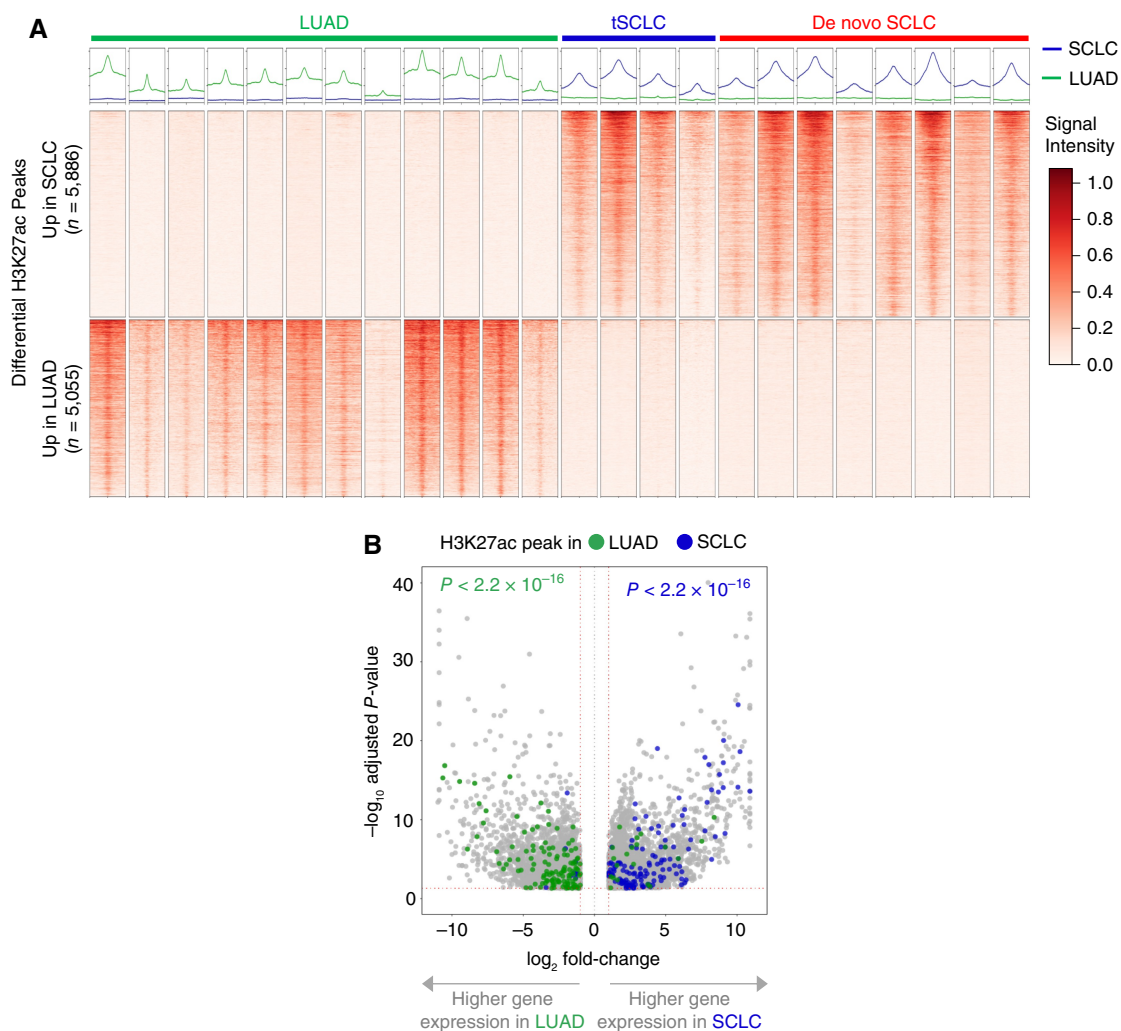


Figure 3.

Comparative analysis identifies a robust set of highly differential epigenomic features between LUAD and SCLC. **A**, Heatmap of normalized H3K27ac tag densities at differential H3K27ac sites between LUAD and SCLC tumors (FDR-adjusted $P < 0.001$ and \log_2 fold-change > 2) located ± 2 kb from peak center. **B**, Volcano plot showing overlap of the \log_2 fold-change differentially expressed genes between LUAD and SCLC PDXs with respective differential H3K27ac peaks enriched in LUAD (blue) and SCLC (red). Two-sided P values were corrected for multiple hypothesis testing (FDR-adjusted $P < 0.05$).

diagnosed with tSCLC (Supplementary Table S3). Consistent with prior reports, all of the tumors from patients with tSCLC included in this cohort exhibited loss of *TP53* and/or *RBI* (Supplementary Table S2; refs. 11–14). On all plasma samples, we generated epigenome-wide data for four analytes—H3K27ac, H3K4me3, DNA methylation, and chromatin accessibility—from 1 mL of plasma (Fig. 1). In brief, we utilized a novel assay (cfChIP-seq) to identify nucleosomal cfDNA fragments bound to H3K27ac and H3K4me3, respectively (22). We then performed cfMeDIP-seq to profile cfDNA methylation (15–18). Finally, we performed LPWGS and utilized Griffin to infer regions of open chromatin based on cfDNA coverage patterns (19, 20). All subsequent analyses include samples with detectable circulating tumor DNA, defined as estimated tumor fraction $> 3\%$ based on the lower limit of detection of ichorCNA applied to LPWGS data (27). More than 95% of plasma samples collected from patients with *EGFRm* LUAD progressing on osimertinib have cfDNA tumor fraction $> 3\%$ using this method,

supporting the clinical relevance of this criteria (48). Notably, there was no difference in cfDNA tumor fraction between samples from patients with tSCLC and *EGFRm* LUAD ($P = 0.10$; Supplementary Table S3). Filtering based on this cut-off and quality control metrics resulted in a final set of 105 epigenomic cfDNA libraries, which were included in subsequent analyses (Supplementary Fig. S7).

To evaluate the ability to noninvasively detect tSCLC in patients with *EGFRm* LUAD, we developed individual SCLC risk scores for each epigenomic analyte based on cfDNA signals at the SCLC- and LUAD-enriched sites in the PDXs for that feature (Fig. 1; Supplementary Fig. S5). The following SCLC risk scores were calculated as the normalized ratio of signals at the SCLC- versus LUAD-enriched sites (see “Materials and Methods”).

cfDNA nucleosome analysis to detect tSCLC

We first assessed the ability to accurately detect tSCLC through cfDNA nucleosome profiling. We observed significantly higher

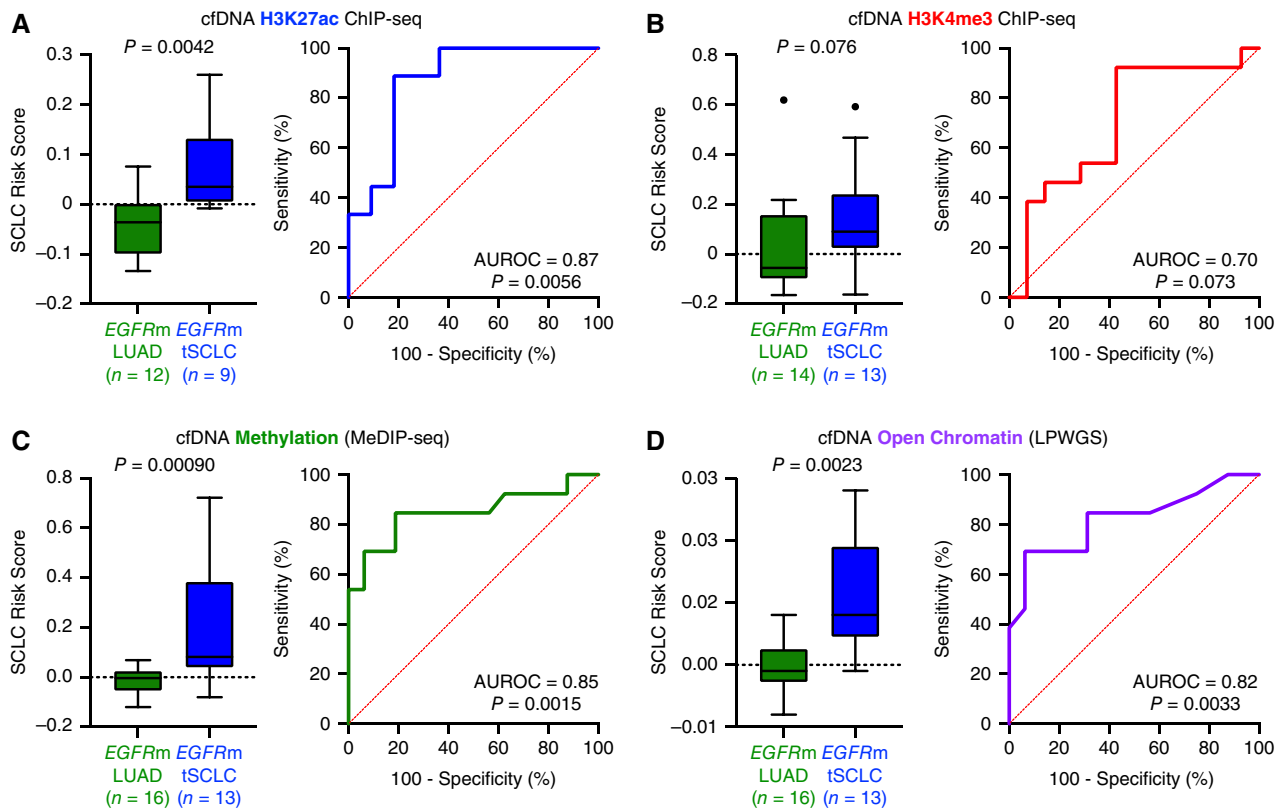


Figure 4.

Noninvasive detection of tSCLC via tissue-informed epigenomic cfDNA analysis. Box plots show cfDNA SCLC risk scores for plasma samples from patients with *EGFRm* LUAD and *EGFRm* tSCLC based on H3K27ac cfChIP-seq analysis (A), H3K4me3 cfChIP-seq analysis (B), cfDNA methylation analysis (C), or cfDNA chromatin accessibility analysis (D). Box plots show the interquartile range and median value for each dataset with whiskers representing the larger value of 1.5 times the interquartile range or the largest value in the dataset. *P* values were calculated using Mann-Whitney test. Corresponding ROC curves with the AUROC are included.

H3K27ac SCLC risk scores in cfDNA samples from patients with tSCLC than those with *EGFRm* LUAD ($P = 0.0042$; Fig. 4A). H3K27ac SCLC risk scores discriminated between plasma samples from patients with tSCLC versus *EGFRm* LUAD with an area under the receiver operating characteristic curve (AUROC) of 0.87 ($P = 0.0056$). H3K4me3 SCLC risk scores trended in the same direction, but did not achieve statistical significance (AUROC = 0.70; $P = 0.073$; Fig. 4B). We hypothesized that the superior performance of H3K27ac relative to H3K4me3 was due in large part to the presence of more than 13 times the number of differential H3K27ac sites ($n = 10,907$) between SCLC and LUAD than H3K4me3 sites ($n = 833$), since both features mark active promoters, but only H3K27ac marks active enhancers as well. This is illustrated by the distribution of these two histone modifications near *INSM1*, a neural lineage-defining gene, wherein an H3K4me3 peak marks the gene promoter and H3K27ac peaks mark the promoter as well as the two most proximal enhancers in tSCLC (GH20J020396 and GH20J020399; Supplementary Fig. S8). In cfDNA, H3K27ac not only marked neural-lineage genes but also potential SCLC therapeutic targets such as *DLL3* (Supplementary Fig. S9). These data demonstrate the ability to noninvasively detect tSCLC and potentially therapeutic target expression, through cfDNA nucleosome profiling and highlight the value of profiling gene enhancers in addition to promoters.

To further validate the ability of the PDX-derived H3K27ac SCLC risk score to discriminate between histology subtypes, we performed H3K27ac ChIP-seq on 20 LUAD cell lines and 13 SCLC cell lines (Supplementary Table S4). Notably, all four molecular subtypes of SCLC were represented, including SCLC-A ($n = 5$), SCLC-N ($n = 3$), SCLC-P ($n = 3$), and SCLC-Y ($n = 2$; ref. 49). H3K27ac SCLC risk scores were significantly higher in the SCLC than the LUAD cell lines ($P = 9.5 \times 10^{-7}$) and achieved an AUROC of 0.95 ($P = 1.4 \times 10^{-5}$) for accurate discrimination of the two histologic subtypes (Supplementary Fig. S10A and S10B). Notably, the SCLC-A ($P = 3.8 \times 10^{-5}$), SCLC-N ($P = 0.0011$), and the SCLC-P (0.0023) cell lines all individually exhibited higher SCLC risk scores than the LUAD cell lines. In contrast, the SCLC risk scores for the SCLC-Y cell lines were not different from the LUAD cell lines ($P = 0.36$) and were significantly lower than the scores for the other three SCLC subtypes ($P = 0.026$; Supplementary Fig. S10A). When excluding the SCLC-Y cell lines, the SCLC risk score achieved an AUROC of 0.995 ($P = 6.8 \times 10^{-6}$) for accurately discriminating SCLC-A, SCLC-N, and SCLC-P cell lines from LUAD cell lines (Supplementary Fig. S10C).

cfDNA methylation and chromatin accessibility analysis to detect tSCLC

cfDNA methylation and chromatin accessibility analysis also discriminated between patients with tSCLC and those with *EGFRm*

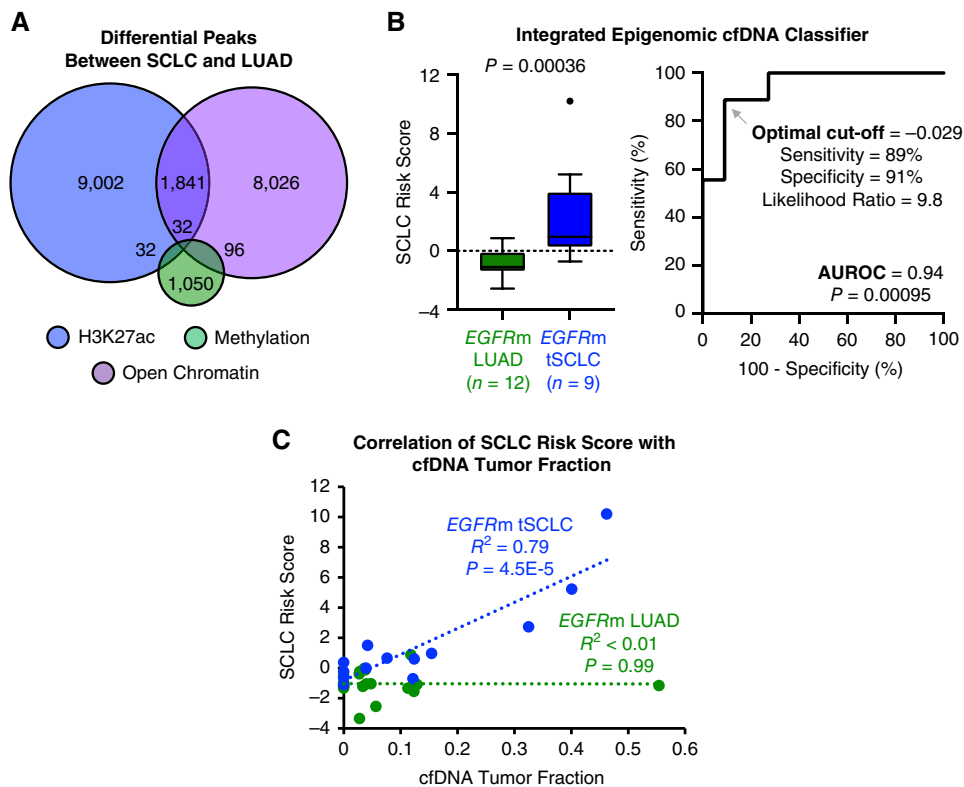


Figure 5.

Integrating multiple epigenomic cfDNA analytes improves noninvasive detection of tSCLC. **A**, Venn diagram showing the overlap of differential H3K27ac, DNA methylation, and open chromatin sites between SCLC and LUAD PDXs. **B**, Box plot shows cfDNA SCLC risk scores for plasma samples from patients with EGFRm LUAD and EGFRm tSCLC based on an integrated epigenomic classifier incorporating H3K27ac, DNA methylation, and chromatin accessibility analysis. Box plot shows the interquartile range and median value for each dataset with whiskers representing the larger value of 1.5 times the interquartile range or the largest value in the dataset. P value was calculated using Mann-Whitney test. ROC curve with the AUROC is included. Optimal cut-off was calculated using Youden index. **C**, Correlation of SCLC risk score with estimated cfDNA tumor fraction in patients with EGFRm tSCLC and EGFRm LUAD.

LUAD. We again observed significantly higher SCLC risk scores in cfDNA samples from patients with tSCLC than those with EGFRm LUAD for both DNA methylation ($P = 0.00090$; Fig. 4C) and chromatin accessibility ($P = 0.0023$; Fig. 4D) data. Classifiers based on signal in cfDNA at differential methylation and chromatin accessibility sites between SCLC and LUAD tumors discriminated between patients with tSCLC and EGFRm LUAD with AUROCs of 0.85 ($P = 0.0015$) and 0.82 ($P = 0.0033$), respectively. These data add to a growing body of literature demonstrating that cfDNA methylation and chromatin accessibility analysis can detect clinically actionable tumor biology (15–17, 19, 20, 50, 51).

Multianalyte epigenomic cfDNA classifier to detect tSCLC

We hypothesized that integrated cfDNA analysis of multiple epigenomic features would improve our ability to discriminate between patients with EGFRm LUAD and tSCLC. To characterize the extent to which multiple epigenomic features would provide additive versus redundant information, we assessed the overlap of differential sites between SCLC and LUAD PDXs for H3K27ac, methylation, and open chromatin. Of the 20,079 differential sites across the three features without background signal in WBCs, the vast majority (90%; $n = 18,078$) were unique to one epigenomic data type (Fig. 5A). The number of combined sites was 1.8, 2.0, and 16.6 times greater than the number of H3K27ac, accessible chromatin, and DNA methylation sites, respectively, and spanned more than 60 Mb—approximately two times the number of base pairs in the protein-coding genome. Given the largely non-overlapping information from these three epigenomic features, we tested a multi-analyte classifier integrating H3K27ac, DNA methylation, and chromatin accessibility data. This integrated epigenomic SCLC risk score discriminated between cfDNA samples from patients with

tSCLC and EGFRm LUAD with greater accuracy than any of the individual analytes (AUROC = 0.94; $P = 0.0095$; Fig. 5B). The optimal diagnostic cut-off (-0.029) demonstrated 89% sensitivity and 91% specificity to identify cfDNA samples from patients with tSCLC versus EGFRm LUAD (likelihood ratio of 9.8).

Association of cfDNA SCLC risk score with tumor fraction

Circulating tumor DNA levels are a critical factor in the ability of liquid biopsies to accurately detect and characterize cancers. As such, we assessed the relationship of the integrated epigenomic cfDNA SCLC risk score with cfDNA tumor content. We observed that SCLC risk scores strongly correlated with cfDNA tumor fraction in patients with tSCLC ($R^2 = 0.79$; $P = 4.5E-5$) but not in patients with EGFRm LUAD ($R^2 < 0.01$; $P = 0.99$; Fig. 5C). This finding suggests that the SCLC risk score reflects tumor biology (i.e., the amount of signal in cfDNA at the site of epigenomic features enriched in a histologic subtype) rather than solely reflecting levels of circulating tumor DNA.

Longitudinal epigenomic cfDNA profiling in patients with EGFRm tSCLC

A unique feature of our cohort is plasma samples collected prior to and at the time of small cell transformation in 2 patients with EGFRm LUAD, providing the opportunity to correlate longitudinal assessment of the integrated epigenomic cfDNA SCLC risk scores with the emergence of tSCLC. The first patient (Fig. 6A) was a 73-year-old woman diagnosed with metastatic LUAD with EGFR exon 19 deletion involving the brain. She was started on first-line erlotinib and remained in remission 36 months later when the first plasma timepoint was collected showing an SCLC risk score of -0.4. Four months later, she experienced radiographic progression with

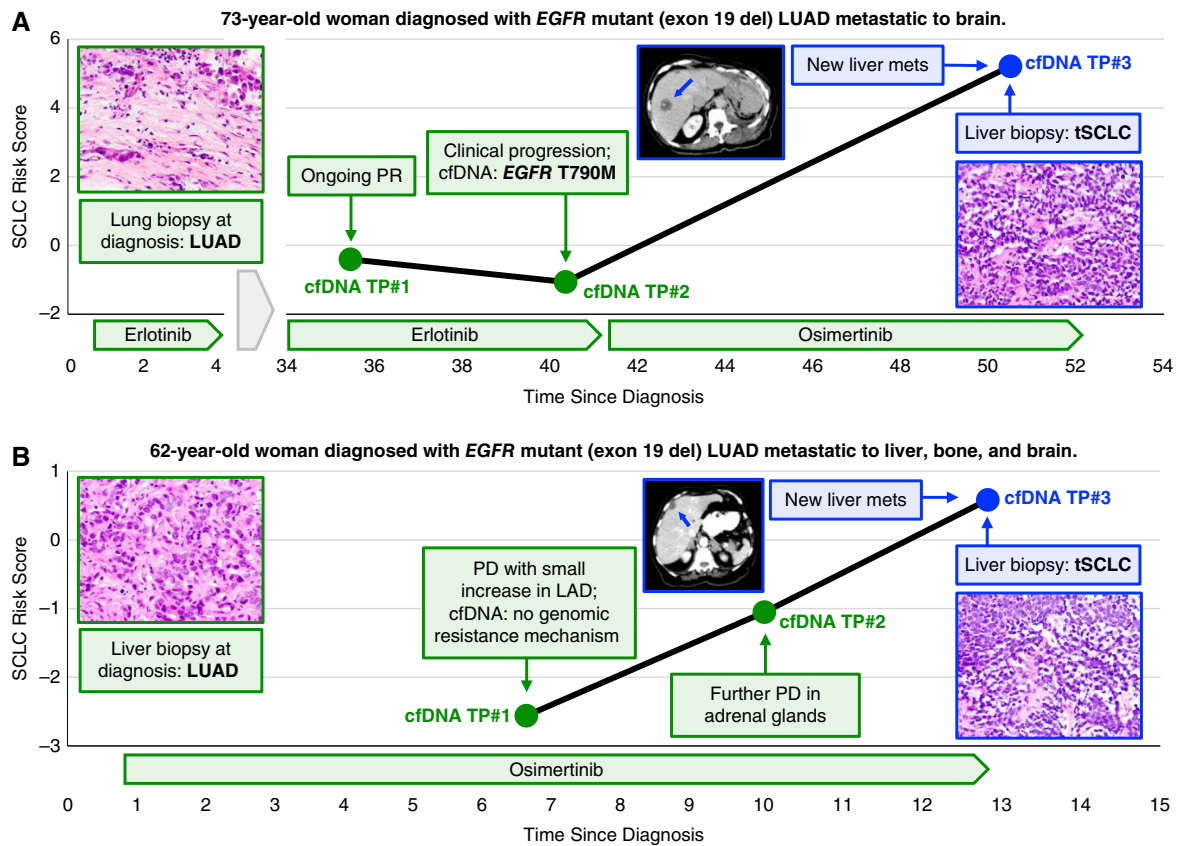


Figure 6.

A–B. Patient vignettes highlight the ability to noninvasively detect small cell transformation in patients with *EGFR*^m LUAD through cfDNA epigenomic profiling. Longitudinal assessment of the integrated epigenomic cfDNA SCLC risk score in 2 patients with *EGFR*^m LUAD who experienced biopsy-proven SCLC.

cfDNA analysis showing an *EGFR* T790M mutation. A second plasma sample at this timepoint showed an SCLC risk score of -1.0 . Treatment was switched from erlotinib to osimertinib. A third plasma timepoint collected 10 months later showed a marked increase in the SCLC risk score to 5.2. Scans at that time revealed new liver metastases and a liver biopsy showed tSCLC. The second patient (Fig. 6B), a 62-year-old woman, was diagnosed with metastatic LUAD with *EGFR* exon 19 deletion involving the liver, bone, and brain for which she was started on first-line osimertinib. The first two plasma samples collected 3 months apart at the time of minor radiographic progression on osimertinib showed SCLC risk scores of -2.6 and -1.1 , respectively. A third plasma sample 3 months later showed an increase in the SCLC risk score to 0.6. Scans at that time revealed new liver metastases and a liver biopsy showed tSCLC. Notably, the rise in the SCLC risk score preceded clinical diagnosis of tSCLC by 92 days in this patient. These patient vignettes illustrate that longitudinal epigenomic cfDNA analysis reflects the emergence of tSCLC in patients with *EGFR*^m LUAD.

Discussion

Histologic transformation to SCLC is an aggressive, clinically actionable resistance phenotype that emerges in a subset of patients with *EGFR*^m LUAD. Limitations of current clinical tools result in delays in diagnosis and underdiagnosis of tSCLC, and consequently,

missed opportunities to deliver optimal guideline-recommended histology-directed systemic therapy. Herein, in the largest study to date of cfDNA samples from patients with tSCLC, we demonstrate for the first time the ability to non-invasively detect small cell transformation in patients with *EGFR*^m LUAD through epigenomic profiling of 1 mL of plasma. With the limitations of tissue biopsy resulting in fewer than half of patients with metastatic *EGFR*^m LUAD undergoing histologic tumor assessment at the time of *EGFR* TKI resistance, these data highlight the potential to advance diagnostic and therapeutic precision for patients with advanced lung cancer by augmenting the current diagnostic paradigm with epigenomic cfDNA analysis.

Liquid biopsies are now widely utilized in clinical oncology to detect cancer recurrence and inform therapeutic decisions. However, most commercially-available cfDNA assays only detect tumor genomic alterations. The lack of genomic alterations exclusive to tSCLC limits the utility of these genomic-based cfDNA approaches to detect small cell transformation in patients with *EGFR*^m LUAD. To address this limitation, we and others have developed tools to analyze several tumor epigenomic features from patient plasma, including DNA methylation (15–18), chromatin accessibility (19, 20), and histone modifications (21, 22). Several recent studies highlight the potential of epigenomic cfDNA profiling to provide dynamic insights into tumor biology in patients with advanced lung cancer. Haq and colleagues analyzed cfDNA methylation profiles

from patients with advanced *de novo* SCLC, identifying two methylation-defined subsets that associate with distinct tumor biology and clinical prognosis (52). Heeke and colleagues (50) and Chemi and colleagues (51) demonstrated that SCLC subtypes (based upon predominant transcription factor expression, i.e., ASCL1, NEUROD1, POU2F3) harbor distinct methylation profiles that can be detected through cfDNA methylation analysis. Our study is distinct in that we demonstrate for the first time that epigenomic cfDNA profiling can be used to detect small cell transformation in patients with *EGFRm* LUAD progressing on EGFR TKIs. Diagnosing tSCLC by cfDNA profiling would be immediately clinically actionable, as guidelines recommend that tSCLC be treated with a *de novo* SCLC regimen of platinum-etoposide chemotherapy which would not otherwise be used in patients with LUAD (1, 53).

An accurate and easily implementable diagnostic test to detect tSCLC in the clinic, would not only facilitate timely delivery of standard of care therapy but could also potentially identify patients for tSCLC-directed therapeutic clinical trials. The first cohort of clinical trials designed specifically for patients with tSCLC are currently enrolling, testing addition of programmed death-ligand 1 inhibitors in combination with either platinum-etoposide chemotherapy or poly-ADP ribose polymerase 1 inhibitors (NCT04538378; NCT05957510; NCT03944772). Notably, we present data supporting the ability of epigenomic cfDNA analysis to noninvasively detect expression of potential therapeutic targets for tSCLC, such as *DLL3*, the target of tarlatamab, a bispecific antibody that recently demonstrated a 40% response rate in patients with previously treated *de novo* SCLC (53). Finally, the potential feasibility of detecting an emerging signal of tSCLC histology with our assay lends itself to the possibility of early adaptive treatment strategies to prevent or delay outright SCLC transformation, similar to the concept behind an ongoing trial (NCT03567642) treating patients with *EGFRm* LUAD at higher risk for SCLC transformation (concurrent *TP53/RB1* mutations) with four cycles of platinum/etoposide chemotherapy in addition to osimertinib.

A major strength of our study is the breadth and depth of epigenomic data generated from tSCLC tumors. To our knowledge, only one study has previously investigated the epigenomic landscape of tSCLC (11). This multiomic analysis of tSCLC tumors by Quintanal-Villalonga and colleagues found that small cell transformation is primarily driven by transcriptional reprogramming, with integrated methylation analysis providing insights into epigenomic changes that underlie histologic transformation. Our work builds upon this observation, generating epigenome-wide data on DNA methylation, chromatin accessibility, and three histone modifications involved in gene regulation on a series of 26 LUAD, *de novo* SCLC, and tSCLC PDXs. These data affirmed that the observed transcriptional reprogramming that accompanies histologic transformation is characterized by widespread epigenomic reprogramming, converging on an epigenomic profile that resembles *de novo* SCLC. While the focus of this manuscript was to develop a tumor-informed cfDNA-based epigenomic classifier to detect tSCLC, we believe that this publicly available epigenomic-transcriptomic dataset will be a valuable resource to further our understanding of the biology that underlies histologic transformation in lung cancer and clinically relevant differences between *de novo* and transformed SCLC.

The platform developed and employed for cfDNA analysis in this study generated new insights into the feasibility and clinical utility of profiling multiple epigenomic features in cfDNA. Because epigenomic cfDNA profiling approaches have been developed independently, it is not known whether integrating multiple epigenomic

analytes is feasible in real-world patient samples or whether this results in better ability to detect clinically relevant tumor biology. We demonstrate for the first time the ability to generate high-quality epigenome-wide cfDNA data on DNA methylation, histone modifications, and chromatin accessibility from 1 mL of patient plasma. Notably, some of the plasma samples included in this study were collected as long as 8 years prior to analysis. Combined with the minimal sample requirement, these technologies potentially unlock the ability to obtain multianalyte epigenome-wide cfDNA data from real-world plasma samples. Additionally, to our knowledge, this is the first study to suggest that combining multiple cfDNA epigenomic features may improve the ability to non-invasively detect a clinically actionable resistance phenotype. The observation that histology-specific H3K27ac, DNA methylation, and open chromatin sites were largely nonoverlapping led us to develop a multi-analyte classifier integrating these three epigenomic analytes that resulted in better diagnostic accuracy (AUROC of 0.94) than any of the individual features (AUROC of 0.82–0.87). While further studies are needed to understand the value of multianalyte versus single-analyte epigenomic analysis, we demonstrate for the first time the feasibility of profiling multiple epigenomic features in cfDNA from 1 mL of real-world plasma samples with data suggesting that integrated multianalyte cfDNA profiling may improve the ability to non-invasively detect clinically relevant tumor biology.

We acknowledge important limitations of this study. First is the modest number of patient samples in the cfDNA cohort. While small in absolute terms, this represents the largest study to date of plasma samples from patients with pathologically confirmed tSCLC and the first to demonstrate the ability to noninvasively detect small cell transformation in patients with *EGFRm* LUAD through cfDNA analysis. Given the rarity of these specimens, plasma samples in this study were collected across 8 years at two institutions under different conditions, e.g., EDTA tubes for some samples and Streck tubes for others, and different plasma extraction methods. The results of the cfDNA analysis in this real-world cohort—despite the variability in the several pre-analytical conditions of the samples—demonstrate the robustness of the epigenomic assays deployed in this study. A second limitation is the lack of an independent validation cohort. Given the rarity of plasma samples from patients with biopsy-proven tSCLC, this was not feasible. We would like to highlight, however, that the epigenomic classifiers were developed solely from an independent cohort of tumors (i.e., comparative analysis of unrelated SCLC and LUAD PDXs) and applied to the cfDNA samples, thus the results are less subject to diagnostic biases, such as overfitting. Nevertheless, the performance of the described epigenomic classifiers, as well as the optimal cut-offs that maximize diagnostic accuracy for identifying patients with *EGFRm* LUAD who have undergone small cell transformation, need to be validated in independent cohorts. Additionally, further validation in cohorts with representation of all molecular subtypes of SCLC will be important (45–47). The majority of the SCLC PDXs from which the classifier was derived, and the tumors from patients with tSCLC in the cfDNA cohort, were of the SCLC-A or SCLC-N subtypes. Encouragingly, the PDX-derived H3K27ac SCLC risk score demonstrated excellent diagnostic accuracy for not only SCLC-A and SCLC-N, but also the SCLC-P in our cell line validation experiment. However, the inability of this classifier to distinguish SCLC-Y from LUAD is an important limitation. For now, tumor tissue biopsy remains the gold standard to diagnose histologic transformation in patients with advanced LUAD with liquid biopsy representing a complementary approach that may better assess intratumoral spatial

heterogeneity, can provide diagnostic information when tissue is unavailable, and allows for serial sampling over a patient's disease course. Evaluation of the performance of the SCLC risk score in cfDNA cohorts from patients with annotated tSCLC tumor molecular subtypes is needed. Finally, the retrospective nature of our cohort resulted in variability in timing of collection of plasma samples relative to biopsy-proven tSCLC diagnosis, so we were unable to methodically determine how far epigenomic cfDNA-based identification of SCLC histology precedes clinical tissue diagnosis. Intriguingly, we observed in one patient that a rise in the cfDNA SCLC risk score preceded clinical diagnosis of tSCLC by more than 3 months. To more formally evaluate this and other clinically relevant questions around the clinical utility of epigenomic cfDNA-based diagnostics, we strongly encourage incorporation of plasma collection into prospective clinical trials.

In summary, we observed widespread epigenomic reprogramming in tSCLC tumors relative to *EGFR*^R LUAD tumors and leveraged these divergent molecular profiles to demonstrate for the first time the ability to non-invasively detect tSCLC in patients with *EGFR*^R LUAD progressing on an EGFR TKI through epigenomic cfDNA analysis. With clinical validation, this epigenomic cfDNA-based approach to detect small cell transformation could usher in a new paradigm of diagnostic and therapeutic precision for patients with advanced lung cancer.

Authors' Disclosures

Y.P. Hung reports honoraria from Elsevier and American Society of Clinical Pathology on textbook writing and continuing medical education-related activity, both of which are unrelated to this study. N.R. Mahadevan reports stock ownership in AstraZeneca and Roche. D.A. Barbie reports personal fees from Qiagen/N of One and other support from Xspera Biosciences outside the submitted work. Z. Piotrowska reports grants from NIH during the conduct of the study. Z. Piotrowska also reports personal fees from Eli Lilly, Boehringer Ingelheim, Bayer, Sanofi, C4 Therapeutics, and Taiho Pharmaceuticals; grants, personal fees, and other support from Janssen and AstraZeneca; grants and personal fees from Takeda, Cullinan Oncology, Daiichi Sankyo, and Blueprint Medicines; grants from Novartis, Spectrum Pharmaceuticals, AbbVie, GlaxoSmithKline/Tesaro, and Phanes Therapeutics; and grants and other support from Genentech/Roche outside the submitted work. T.K. Choueiri reports personal fees and other support from Precede Bio during the conduct of the study, as well as grants and other support from Precede Bio outside the submitted work; in addition, T.K. Choueiri has a patent for Precede Bio with royalties paid. T.K. Choueiri also reports institutional and/or personal, paid and/or unpaid support for research, advisory boards, consultancy, and/or honoraria past 5 years, ongoing or not, from Alkermes, Arcus Bio, AstraZeneca, Aravive, Aveo, Bayer, Bristol Myers Squibb, Calithera, Circle Pharma, Deciphera Pharmaceuticals, Eisai, EMD Serono, Exelixis, GlaxoSmithKline, Gilead, HiberCell, IQVA, Infinity, Ipsen, Janssen, Kanaph, Lilly, Merck, Nikang, Neomorph, Nuscan/Precede Bio, Novartis, Oncohost, Pfizer, Roche, Sanofi/Aventis, Scholar Rock, Surface Oncology, Takeda, Tempest, Up-To-Date, CME events (Peerview, OnLive, MJH, CCO and others), outside the submitted work; institutional patents filed on molecular alterations and immunotherapy response/toxicity, and ctDNA; equity from Tempest, Pionyr, Osel, Precede Bio, CureResponse, InnDura Therapeutics, and Primium; committees for NCCN, GU Steering Committee, ASCO (BOD 6-2024+), ESMO, ACCRU, and KidneyCan; medical writing and editorial assistance support may have been funded by communications companies in part; mentored several non-US citizens on research projects with potential funding (in part) from non-US sources/foreign components; and the institution (Dana-Farber Cancer Institute) may have received additional independent funding of drug companies and/or royalties potentially involved in research around the subject matter. T.K. Choueiri is supported in part by the Dana-Farber/Harvard Cancer Center Kidney SPORE (2P50CA101942-16) and Program 5P30CA006516-56, the Kohlberg Chair at Harvard Medical School and the Trust Family, Michael Brigham, Pan Mass Challenge, Hinda and Arthur Marcus Fund, and Loker Pinarid Funds for Kidney Cancer Research at DFCI. S.C. Baca reports personal fees and other support

from Precede Biosciences outside the submitted work. A.N. Hata reports grants and personal fees from Amgen, Nuvalent, and Pfizer; grants from BridgeBio, Bristol-Myers Squibb, C4 Therapeutics, Eli Lilly, Novartis, and Scorpion Therapeutics; and personal fees from Engine Biosciences, Oncovalent, Tigenix, and Tolremo outside the submitted work. M.L. Freedman reports personal fees and other support from Precede Biosciences outside the submitted work; in addition, M.L. Freedman has a patent for 'Methods, kits and systems for determining the status of lung cancer and methods for treating lung cancer based on same' pending. J.E. Berchuck reports non-financial support and other support from Precede Biosciences during the conduct of the study. J.E. Berchuck also reports grants, personal fees, and non-financial support from Guardant Health; personal fees and other support from Genome Medical; and other support from Oncotect, TracerDx, and Musculo outside the submitted work. In addition, J.E. Berchuck has an institutional patent on methods to detect neuroendocrine prostate cancer through tissue-informed cell-free DNA methylation analysis issued, licensed, and with royalties paid from Precede Biosciences and an institutional patent on methods to detect small cell lung cancer through epigenomic cfDNA analysis pending. No disclosures were reported by the other authors.

Authors' Contributions

T. El Zarif: Data curation, formal analysis, investigation, writing-review and editing. **C.B. Meador:** Data curation, formal analysis, investigation, writing-original draft, project administration. **X. Qiu:** Software, formal analysis, investigation, methodology, writing-review and editing. **J.-H. Seo:** Formal analysis, investigation, methodology, writing-review and editing. **M.P. Davidsohn:** Investigation. **H. Savignano:** Investigation. **G. Lakshminarayanan:** Investigation. **H.M. McClure:** Investigation. **J. Canniff:** Investigation. **B. Fortunato:** Formal analysis. **R. Li:** Formal analysis. **M.K. Banwait:** Data curation. **K. Semaan:** Formal analysis. **M. Eid:** Investigation. **H. Long:** Supervision, methodology. **Y.P. Hung:** Data curation, writing-review and editing. **N.R. Mahadevan:** Writing-review and editing. **D.A. Barbie:** Conceptualization, resources, writing-review and editing. **M.G. Oser:** Conceptualization, resources, writing-review and editing. **Z. Piotrowska:** Resources, writing-review and editing. **T.K. Choueiri:** Resources, writing-review and editing. **S.C. Baca:** Software, methodology, writing-review and editing. **A.N. Hata:** Conceptualization, resources, supervision, funding acquisition, writing-review and editing. **M.L. Freedman:** Conceptualization, resources, formal analysis, supervision, methodology, writing-review and editing. **J.E. Berchuck:** Conceptualization, resources, data curation, formal analysis, supervision, funding acquisition, investigation, methodology, writing-original draft, project administration, writing-review and editing.

Acknowledgments

J.E. Berchuck is supported by the Dana-Farber/Harvard Cancer Center Lung Cancer SPORE Research Development Award, the Department of Defense (W81XWH-20-1-0118, HT9425-23-1-0048), and the Dave Page Cancer Research Fund. C.B. Meador is supported by an Institutional Research Grant from the American Cancer Society (2022A019104). N.R. Mahadevan is supported by NCI/NIH (K08CA270077). J.E. Berchuck, C.B. Meador, D.A. Barbie, M.G. Oser, Z. Piotrowska, and A.N. Hata are supported by the LUNGSTRONG Foundation and Pan-Mass Challenge Team 3G. Z. Piotrowska and A.N. Hata are supported by the NIH (R01CA137008). S.C. Baca is supported by the Department of Defense (W81XWH-21-1-0358), the Damon Runyon Cancer Research Foundation, the Fund for Innovation in Cancer Informatics, and the Kure It Cancer Research Foundation. M.L. Freedman is supported by the National Institute of Health (R01CA262577, R01CA251555), the Claudia Adams Barr Program for Innovative Cancer Research, the Dana-Farber Cancer Institute Presidential Initiatives Fund, the H.L. Snyder Medical Research Foundation, the Cutler Family Fund for Prevention and Early Detection, the Donahue Family Fund, the Department of Defense (W81XWH-21-1-0339, W81XWH-22-1-0951), and the Movember PCF Challenge Award.

Note

Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

Received February 9, 2024; revised April 4, 2024; accepted June 20, 2024; published first June 24, 2024.

References

- National Comprehensive Cancer Network Practice Guidelines in Oncology (NCCN Guidelines). Non-small cell lung cancer version 1.2024. National Comprehensive Cancer Network.
- Soria J-C, Ohe Y, Vansteenkiste J, Reungwetwattana T, Chewaskulyong B, Lee KH, et al. Osimertinib in untreated EGFR-mutated advanced non-small-cell lung cancer. *N Engl J Med* 2018;378:113–25.
- Leonetti A, Sharma S, Minari R, Perego P, Giovannetti E, Tiseo M. Resistance mechanisms to osimertinib in EGFR-mutated non-small cell lung cancer. *Br J Cancer* 2019;121:725–37.
- Schoenfeld AJ, Chan JM, Kubota D, Sato H, Rizvi H, Daneshbod Y, et al. Tumor analyses reveal squamous transformation and off-target alterations as early resistance mechanisms to first-line osimertinib in EGFR-mutant lung cancer. *Clin Cancer Res* 2020;26:2654–63.
- Piotrowska Z, Isozaki H, Lennerz JK, Gainor JF, Lennes IT, Zhu VW, et al. Landscape of acquired resistance to osimertinib in EGFR-mutant NSCLC and clinical validation of combined EGFR and RET inhibition with osimertinib and BLU-667 for acquired *RET* fusion. *Cancer Discov* 2018;8:1529–39.
- Piotrowska Z, Chen LN, Shum E, Park C, Guo M, Liu Y, et al. Tissue and plasma-based mechanisms of resistance to first-line osimertinib in EGFR-mutant NSCLC: a multi-institutional cohort. *J Clin Oncol* 2023;41:9108.
- Sequist LV, Waltman BA, Dias-Santagata D, Digumarthy S, Turke AB, Fidias P, et al. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci Transl Med* 2011;3:75ra26.
- Marcoux N, Gettinger SN, O’Kane G, Arbour KC, Neal JW, Husain H, et al. EGFR-mutant adenocarcinomas that transform to small-cell lung cancer and other neuroendocrine carcinomas: clinical outcomes. *J Clin Oncol* 2019;37:278–85.
- Piotrowska Z, Ahn M-J, Pang YK, How SH, Kim S-W, Voon PJ, et al. LBA53 ELIOS: a multicentre, molecular profiling study of patients (pts) with epidermal growth factor receptor-mutated (EGFRm) advanced NSCLC treated with first-line (1L) osimertinib. *Ann Oncol* 2022;33(Suppl 7):S1420-21.
- Chabon JJ, Simmons AD, Lovejoy AF, Esfahani MS, Newman AM, Haringsma HJ, et al. Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients. *Nat Commun* 2016;7:11815.
- Quintanal-Villalonga A, Taniguchi H, Zhan YA, Hasan MM, Chavan SS, Meng F, et al. Multiomic analysis of lung tumors defines pathways activated in neuroendocrine transformation. *Cancer Discov* 2021;11:3028–47.
- Lee J-K, Lee J, Kim S, Kim S, Youk J, Park S, et al. Clonal history and genetic predictors of transformation into small-cell carcinomas from lung adenocarcinomas. *J Clin Oncol* 2017;35:3065–74.
- Offin M, Chan JM, Tenet M, Rizvi HA, Shen R, Riely GJ, et al. Concurrent RB1 and TP53 alterations define a subset of EGFR-mutant lung cancers at risk for histologic transformation and inferior clinical outcomes. *J Thorac Oncol* 2019;14:1784–93.
- Niederst MJ, Sequist LV, Poirier JT, Mermel CH, Lockerman EL, Garcia AR, et al. RB loss in resistant EGFR mutant lung adenocarcinomas that transform to small-cell lung cancer. *Nat Commun* 2015;6:6377.
- Shen SY, Singhanian R, Fehrer G, Chakravarthy A, Roehrl MHA, Chadwick D, et al. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature* 2018;563:579–83.
- Shen SY, Burgener JM, Bratman SV, De Carvalho DD. Preparation of cfMeDIP-seq libraries for methylome profiling of plasma cell-free DNA. *Nat Protoc* 2019;14:2749–80.
- 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;26:1041–3.
- Berchuck JE, Baca SC, McClure HM, Korthauer K, Tsai HK, Nuzzo PV, et al. Detecting neuroendocrine prostate cancer through tissue-informed cell-free DNA methylation analysis. *Clin Cancer Res* 2022;28:928–38.
- Doebley A-L, Ko M, Liao H, Cruikshank AE, Santos K, Kikawa C, et al. A framework for clinical cancer subtyping from nucleosome profiling of cell-free DNA. *Nat Commun* 2022;13:7475.
- De Sarkar N, Patton RD, Doebley A-L, Hanratty B, Adil M, Kreitzman AJ, et al. Nucleosome patterns in circulating tumor DNA reveal transcriptional regulation of advanced prostate cancer phenotypes. *Cancer Discov* 2023;13:632–53.
- Sadeh R, Sharkia I, Fialkoff G, Rahat A, Gutin J, Chappleboim A, et al. ChIP-seq of plasma cell-free nucleosomes identifies gene expression programs of the cells of origin. *Nat Biotechnol* 2021;39:586–98.
- Baca SC, Seo J-H, Davidsohn MP, Fortunato B, Semaan K, Sotudian S, et al. Author correction: liquid biopsy epigenomic profiling for cancer subtyping. *Nat Med* 2024;30:907.
- Farago AF, Yeap BY, Stanzione M, Hung YP, Heist RS, Marcoux JP, et al. Combination olaparib and temozolomide in relapsed small-cell lung cancer. *Cancer Discov* 2019;9:1372–87.
- Drapkin BJ, George J, Christensen CL, Mino-Kenudson M, Dries R, Sundaresan T, et al. Genomic and functional fidelity of small cell lung cancer patient-derived xenografts. *Cancer Discov* 2018;8:600–15.
- Lin JJ, Langenbucher A, Gupta P, Yoda S, Fetter IJ, Rooney M, et al. Small cell transformation of *ROS1* fusion-positive lung cancer resistant to *ROS1* inhibition. *NPJ Precis Oncol* 2020;4:21.
- Nangia V, Siddiqui FM, Caenepeel S, Timonina D, Bilton SJ, Phan N, et al. Exploiting *MCL1* dependency with combination MEK + *MCL1* inhibitors leads to induction of apoptosis and tumor regression in *KRAS*-mutant non-small cell lung cancer. *Cancer Discov* 2018;8:1598–613.
- 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;8:1324.
- Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;10:R25.
- Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol* 2008;9:R137.
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol* 2011;29:24–6.
- Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 2014;42:W187–191.
- Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* 2017;14:959–62.
- Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol* 2015;109:21.29.1–21.29.9.
- Qiu X, Feit AS, Feiglin A, Xie Y, Kesten N, Taing L, et al. CoBRA: containerized bioinformatics workflow for reproducible ChIP/ATAC-seq analysis. *Genomics Proteomics Bioinformatics* 2021;19:652–61.
- Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* 2012;28:2520–2.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
- Cornwell M, Vangala M, Taing L, Herbert Z, Köster J, Li B, et al. VIPER: visualization Pipeline for RNA-seq, a Snakemake workflow for efficient and complete RNA-seq analysis. *BMC Bioinformatics* 2018;19:135.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21.
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 2012;7:562–78.
- Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016;32:3047–8.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–9.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078–9.
- 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* 2014;30:284–6.
- Pelizzola M, Koga Y, Urban AE, Krauthammer M, Weissman S, Halaban R, et al. MEDME: an experimental and analytical methodology for the estimation of DNA methylation levels based on microarray derived MeDIP-enrichment. *Genome Res* 2008;18:1652–9.

45. Hwang S, Hong TH, Park S, Jung H-A, Sun J-M, Ahn JS, et al. Molecular subtypes of small cell lung cancer transformed from adenocarcinoma after EGFR tyrosine kinase inhibitor treatment. *Transl Lung Cancer Res* 2021;10:4209–20.
46. Gay CM, Stewart CA, Park EM, Diao L, Groves SM, Heeke S, et al. Patterns of transcription factor programs and immune pathway activation define four major subtypes of SCLC with distinct therapeutic vulnerabilities. *Cancer Cell* 2021;39:346–60.e7.
47. Mahadevan NR, Knelson EH, Wolff JO, Vajdi A, Saigí M, Campisi M, et al. Intrinsic immunogenicity of small cell lung carcinoma revealed by its cellular plasticity. *Cancer Discov* 2021;11:1952–69.
48. Buder A, Heitzer E, Waldspühl-Geigl J, Weber S, Moser T, Hochmair MJ, et al. Somatic copy-number alterations in plasma circulating tumor DNA from advanced EGFR-mutated lung adenocarcinoma patients. *Biomolecules* 2021;11:618.
49. Tlemsani C, Pongor L, Elloumi F, Girard L, Huffman KE, Roper N, et al. SCLC-CellMiner: a resource for small cell lung cancer cell line genomics and pharmacology based on genomic signatures. *Cell Rep* 2020;33:108296.
50. Heeke S, Gay CM, Estecio MR, Tran H, Morris BB, Zhang B, et al. Tumor- and circulating-free DNA methylation identifies clinically relevant small cell lung cancer subtypes. *Cancer Cell* 2024;42:225–37.e5.
51. 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. *Nat Cancer* 2022;3:1260–70.
52. Ul Haq S, Schmid S, Aparnathi MK, Hueniken K, Zhan LJ, Sacdalan D, et al. Cell-free DNA methylation-defined prognostic subgroups in small-cell lung cancer identified by leukocyte methylation subtraction. *iScience* 2022;25:105487.
53. National Comprehensive Cancer Network Practice Guidelines in Oncology (NCCN Guidelines). Small cell lung Cancer version 2.2024. National Comprehensive Cancer Network.
54. Ahn M-J, Cho BC, Felip E, Korantzis I, Ohashi K, Majem M, et al. Tarlatamab for patients with previously treated small-cell lung cancer. *N Engl J Med* 2023; 389:2063–75.