

Article type: Research reports

Genomic alterations and transcriptional phenotypes in circulating tumor DNA and matched metastatic tumor

Nobuyuki Takahashi^{1,2, 3*}, Lorinc Pongor^{1*}, Shivam P. Agrawal⁴, Mariya Shtumpf⁴, Vinodh N. Rajapakse¹, Ahmad Shafiei⁵, Christopher W. Schultz¹, Sehyun Kim^{1,6}, Diana Roame⁷, Paula Carter⁷, Rasa Vilimas¹, Samantha Nichols¹, Parth Desai¹, William Douglas Figg Sr⁷, Mohammad Bagheri⁵, Vladimir B. Teif^{4**}, Anish Thomas^{1**}

¹Developmental Therapeutics Branch, Center for Cancer Research, National Cancer Institute, Bethesda, USA

²Medical Oncology Branch, Center Hospital, National Center for Global Health and Medicine, Tokyo, Japan

³Department of Medical Oncology, National Cancer Center East Hospital, Kashiwa, Japan

⁴School of Life Sciences, University of Essex, Colchester, UK

⁵Department of Radiology and Imaging Sciences, Center for Cancer Research, National Cancer Institute, Bethesda, USA

⁶Department of Internal Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Korea

⁷Genitourinary Malignancies Branch, Center for Cancer Research, National Cancer Institute, Bethesda, USA

23

24 *Equal contributions

25 **Corresponding authors

26 Lead contact: Anish Thomas, Center for Cancer Research, National Cancer Institute, Building 10

27 Room 4-5330, Bethesda, MD 20892; Ph: 240-760-7343; Fax: 954-827-0184; Email:

28 anish.thomas@nih.gov

29

30 Key words: circulating cell-free DNA, circulating tumor DNA; whole genome sequencing,

31 transcription factor binding site

32

Abstract

Background Profiling circulating cell-free DNA (cfDNA) has become a fundamental practice in cancer medicine, but the effectiveness of cfDNA at elucidating tumor-derived molecular features has not been systematically compared to standard single-lesion tumor biopsies in prospective cohorts of patients. The use of plasma instead of tissue to guide therapy is particularly attractive for patients with small cell lung cancer (SCLC), a cancer whose aggressive clinical course making it exceedingly challenging to obtain tumor biopsies.

Methods Here, a prospective cohort of 49 plasma samples obtained before, during, and after treatment from 20 patients with recurrent SCLC, we study cfDNA low pass whole genome (0.1X coverage) and exome (130X) sequencing in comparison with time-point matched tumor, characterized using exome and transcriptome sequencing.

Results Direct comparison of cfDNA versus tumor biopsy reveals that cfDNA not only mirrors the mutation and copy number landscape of the corresponding tumor but also identifies clinically relevant resistance mechanisms and cancer driver alterations not found in matched tumor biopsies. Longitudinal cfDNA analysis reliably tracks tumor response, progression, and clonal evolution. Genomic sequencing coverage of plasma DNA fragments around transcription start sites shows distinct treatment-related changes and captures the expression of key transcription factors such as NEUROD1 and REST in the corresponding SCLC tumors, allowing prediction of SCLC neuroendocrine phenotypes and treatment responses. **Conclusions** These findings have important implications for non-invasive stratification and subtype-specific therapies for patients with SCLC, now treated as a single disease.

55 Key words: circulating cell-free DNA, circulating tumor DNA; whole genome sequencing,
56 transcription factor binding site

Background

Circulating cell-free DNA (cfDNA) exists as fragmented DNA in the non-cellular fraction of the blood. In healthy individuals, death of normal cells of the hematopoietic lineage is the main contributor to plasma cfDNA. In cancer patients, plasma can carry circulating tumor DNA (ctDNA) fragments originating from tumor cells, providing non-invasive access to somatic genetic alterations in tumors. ctDNA can be differentiated from germline cfDNA through tumor-specific somatic genomic alterations (1). ctDNA is being investigated for a broad array of high-impact clinical applications in cancer but informs current clinical practice in two ways. First, ctDNA profiles actionable mutations to guide therapy in patients in whom a biopsy is not feasible or the amount of tissue is insufficient. Plasma-based detection of mutations and translocations in genes including *BRCA1*, *BRCA2*, *ATM*, *EGFR*, *PIK3CA*, and *ALK* are now approved by the United States Food and Drug Administration to identify patients eligible for treatment with targeted therapies (2). Second, ctDNA can track the trajectory of a patient's response to therapy, allowing early detection of impending disease progression, and identification of molecular mechanisms driving resistance (3, 4). ctDNA is also being investigated to assess molecular residual disease after surgery to identify patients who are at risk for relapse (5-8), and for early detection of cancer (9, 10).

Although ctDNA profiling is now a fundamental practice in cancer medicine, several key questions remain. A basic question is whether the mutational profile established through ctDNA testing reliably reproduces the mutational profile derived from a tumor biopsy. Early studies, based on small numbers of patient samples, suggested low concordance between DNA alterations detected in tumor and plasma samples from the same patient (11, 12). Subsequent

studies, mostly case reports or small cohorts querying single genes, or a panel of genes, suggested high concordance between ctDNA and tumor genomic alterations (13-17). A recent study used high-depth whole genome sequencing (WGS, median read depth 187X) to find concordant clonally expanded cancer driver alterations in both ctDNA and metastatic tumor, but ctDNA additionally harbored the genomes of multiple tumor subclones (18). However, cfDNA samples in this study were pre-selected based on high ctDNA fraction using targeted sequencing approaches and deep WGS of ctDNA is clinically not feasible at present. A second question is whether ctDNA testing can be applied to settings where tumor mutations are not known *a priori*. Next-generation sequencing-based assays afford the opportunity to broadly examine ctDNA, but most studies identify mutations in the tumor and then determine whether the same mutation is detectable in the plasma (5) or perform targeted sequencing of recurrently mutated cancer genes in ctDNA (9). Although these approaches have substantially advanced ctDNA as a diagnostic tool, they have limited utility when tumor mutations are not known *a priori*, for example due to difficulties in getting tumor biopsies, and in tumors driven by dysregulated transcriptional programs.

Small cell lung cancer (SCLC) represents a paradigm to study of ctDNA. SCLC represents about 15% of all lung cancers and is marked by an exceptionally high proliferative rate, strong predilection for early metastasis and poor prognosis (19). Obtaining SCLC tumors for molecular testing is exceedingly difficult as few patients undergo surgery and the cancer is usually extensively disseminated by the time it is diagnosed (20). At relapse, rapid disease progression generally precludes biopsies. As a result, SCLC is not included in large scale sequencing initiatives such as the Cancer Genome Atlas (TCGA). Mirroring its high metastatic predilection..

Detection of ctDNA is well validated in patients with SCLC and prior studies have found that ctDNA can track the disease course and identify recurrent gene alterations in *TP53* and *RBI* (21-29); but these studies are limited by lack of corresponding tumor samples, and the approaches used do not interrogate the transcription programs which underlie SCLC heterogeneity. Indeed, aberrations in transcription regulators are the primary genetic cause of SCLC (30). Cell cycle regulators, transcription factors (TFs) and chromatin modifiers including *RBI* and *TP53*, members of *MYC* family, *SOX2*, *MLL1/2*, *CREBBP-EP300*, *RBL2*, and *TP73* are frequently altered in SCLC, causing aberrant expression of a broad range of genes related to neuronal and neuroendocrine differentiation and proliferation (31). Importantly, SCLC transcriptional subtypes defined by differential expression of key transcription regulators (32) have therapeutic implications (33-36), but are not associated with specific mutational patterns.

Here we perform longitudinal profiling of ctDNA and time-point matched tumor from a molecularly defined prospective cohort of patients with relapsed SCLC, asking whether ctDNA reliably reproduces the genome-wide copy number aberrations and exome-wide tumor mutational profile, and going beyond mutations, whether ctDNA can be used to infer the expression of genes in the corresponding tumors (**Figure 1**).

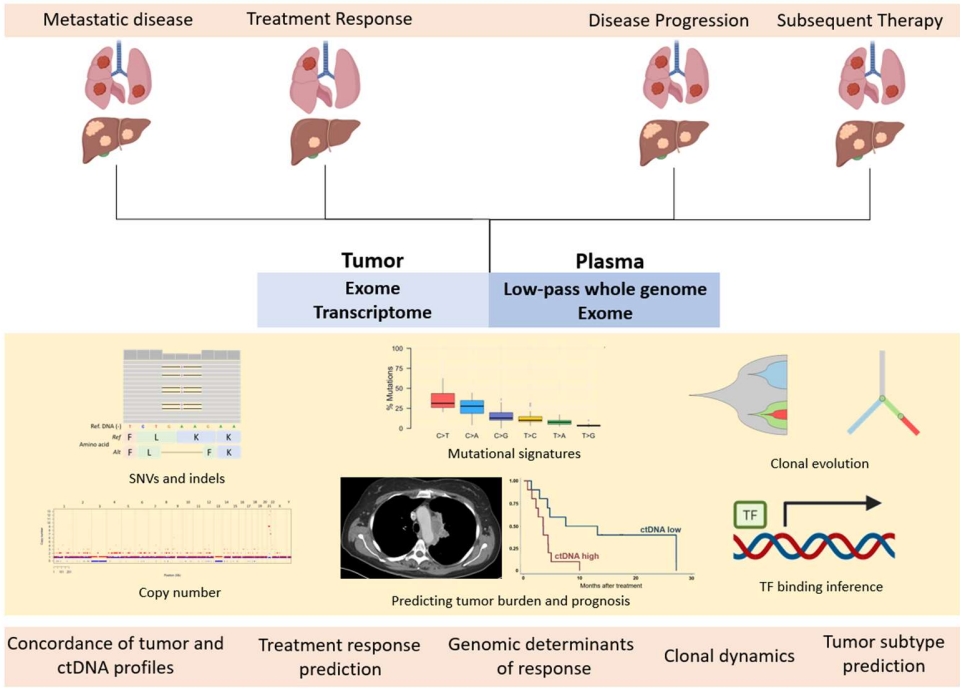


Figure 1. Study schema

Abbreviations: ctDNA: circulating tumor DNA; SNV; single nucleotide variant; indels: insertions and deletions; TF: transcription factor

Methods

Patients

Patients with metastatic biopsy-proven SCLC enrolled on an interventional clinical trial (ClinicalTrials.gov identifier NCT02484404, n = 20) were included in this analysis (37). All patients provided written informed consent. Clinicopathologic data were abstracted from the medical record. The National Cancer Institute (NCI) Laboratory of Pathology confirmed the diagnoses of SCLC. All patients had received platinum-based chemotherapy for SCLC before enrolment. All patients were treated with durvalumab 1500 mg intravenously and olaparib 300 mg twice daily until disease progression or unacceptable toxicity. Plasma samples were collected

pre-treatment, post-treatment (2 to 4 weeks after treatment), and at disease progression. Tumor biopsies were performed by the Department of Radiology at National Institutes of Health at corresponding time points. Tumor response was assessed using Response Evaluation Criteria in Solid Tumors (RECIST) v 1.1. Patient survival was followed up every six months by phone until death or date of cut off.

cfDNA, tumor, and germline sequencing

The DSP Circulating DNA kit from Qiagen was utilized to extract cell-free DNA from aliquots of plasma which were eluted into 40-80 uL of re-suspension buffer using the Qiagen Circulating DNA kit on the QIA Symphony liquid handling system. Library preparation utilized the Kapa Hyper Prep kit with custom adapters (IDT and Broad Institute). Samples were sequenced to meet a goal of 0.1x mean coverage utilizing the Laboratory Picard bioinformatics pipeline and Illumina instruments were used for all of cfDNA sequencing with 150 bp and paired-end sequencing. Library construction was performed as previously described (38). Kapa HyperPrep reagents in 96-reaction kit format were used for end repair/A-tailing, adapter ligation, and library enrichment polymerase chain reaction (PCR). After library construction, hybridization and capture were performed using the relevant components of Illumina's Nextera Exome Kit and following the manufacturer's suggested protocol. Cluster amplification of DNA libraries was performed according to the manufacturer's protocol (Illumina) using exclusion amplification chemistry and flowcells. Flowcells were sequenced utilizing Sequencing-by-Synthesis chemistry. Each pool of whole genome libraries was sequenced on paired 76 cycle runs with two 8 cycle index reads across the number of lanes needed to meet coverage for all libraries in the pool. For tumor sequencing, formalin-fixed, paraffin-embedded (FFPE) tumor tissue samples were

prepared for whole exome sequencing (WES) and RNA sequencing. One hundred nanograms of DNA was shared to approximately 200 bp by sonication (Covaris, Woburn, MA). Exome enrichment was performed using SureSelect Clinical Research Exome Kits according to the manufacturer's instructions (Aglient, Santa Clara, CA). Paired-end sequencing (2 x 75bp) was performed on an Illumina NextSet 500 instrument. The sequences were compared to the human reference genome hg19 using internally developed somatic bioinformatic pipeline. In brief, raw sequencing data in FASTQ format were aligned against the reference human genome (hg19). RNA was extracted from FFPE tumor cores using RNeasy FFPE kits according to the manufacturer's protocol (Qiagen, Germantown, MD). For germline DNA sequencing, the patient's peripheral blood mononuclear cells were sequenced and genotyped with the HumanOmni2.5-8v1 array (Illumina) by Personal Genome Diagnostics. Briefly, 3 µg of genomic DNA per patient sample was sequenced using the Illumina Whole Genome Sequencing Service with the Illumina HiSeq 2000 (Illumina), generating 200 bp (2 × 100 bp reads) per fragment in the final library.

Variant calling, transcriptome expression, and mutational signature analysis

The Genome Analysis Toolkit (GATK) MuTect2 (39) and Strelka2 (40) were used for somatic SNV and small indel calling, respectively. ANNOVAR was used to functionally annotate genetic variants. The Sequenza software (41) was used to determine total and allele-specific DNA copy number from WES. RNA sequencing libraries were generated using TruSeq RNA Access Library Prep Kits (TruSeq RNA Exome kits; Illumina) and sequenced on NextSeq500 sequencers using 75bp paired-end sequencing method (Illumina, San Diego, CA). Each sample was processed through an RNA sequencing data analysis pipeline where reads were mapped to

the ENSEMBL human genome GRCh37 build 71 using TopHat2 (42). Read counts for each gene between samples were transformed to Trimmed Mean of M-values-normalized Fragments per kilobase per million mapped reads (TMM-FPKM). Mutational signatures of COSMIC version 3.2 (43) were identified using the deconstructSigs (44) package (version 1.8.0). Identified mutations were annotated using VEP (45). Mutations were processed using the *maftools* (46) package in R.

Somatic copy number alterations (SCNAs)

Somatic copy number calls were identified using *CNVkit* (47) (version 0.9.9) with default parameters. Tumor purity and ploidy were estimated by *sclust* (48) and *sequenza* (41). The *sclust* purity values were used for adjusted CNV calls using the *CNVkit* tool. Whole genome copy number representation was visualized using R (version 4.0.4) with the *rtracklayer* (49) (version 1.48.0), *ComplexHeatmap* (50) (version 2.4.3), and *ggplot2* (version 3.3.3) packages. The copy number heatmap summarizes the genome segmented into 1Mb sized regions, where the average CNV log2 ratio was calculated for each using the *CNVkit* .cns files.

Homologous recombination repair deficiency (HRD) score

Loss of heterozygosity (LOH), telomeric allelic imbalance (TAI) and large-scale state transition (LST) scores were calculated as described by Telli et al. (51). Allele intensities from CEL files were used to generate allelic imbalance profiles. A hidden Markov model (HMM) was used to define regions and breakpoints with these profiles. Allele specific copy number (ASCN) for each of the regions was determined. TAI (number of regions of allelic imbalance that extend to one of the subtelomeres but do not cross the centromere) and LST (number of break points between

regions longer than 10 Mb after filtering out regions shorter than 3 Mb) scores were calculated using the allelic imbalance profiles, while LOH (number of subchromosomal LOH regions longer than 15 Mb) was calculated using ASCN. To calculate the HRD score based on SNP data, noise to signal ratio (NSR) for SNP data was used as a quality metric. Noise was calculated as the standard deviation of allele dosage for informative SNPs (SNPs that are heterozygous in normal DNA). Signal was calculated as the weighted average of the difference in allele dosage between adjacent regions with weights defined as $1/S1+1/S2$, where S1 and S2 are sizes of the adjacent regions. By comparing HRD scores between samples run in duplicate, a cutoff of 0.85 for NSR was established. Samples with NSR below 0.85 were considered passing HRD scores.

Phylogenetic tumor evolution

Phylogenetic trees were inferred using *PyClone* (52) (version 0.13.1) using shared variants among samples for each patient, adjusted by tumor purities calculated by *sclust*. Phylogenetic trees were prepared using the *ClonEvol* (53) (version 0.99.11) package in R (version 4.0.3). In case of some patients, one or more samples had to be removed to complete the *ClonEvol* analysis.

TF-binding site analysis

Regions with differential cfDNA occupancy were called with NucTools (54), with 10,000-bp sliding window.

The NucTools-based analysis of differential cfDNA occupancy was based on the following steps:

1) cfDNA occupancy has been calculated for each sample, averaged within each 10-kb genomic window and normalized by the sequencing depth of that sample, taking into this analysis only DNA fragments with sizes 120-180 bp to account for the nucleosome protection.

2) The genome was scanned with 10-kb window to determine regions which have stable nucleosome occupancy in a given condition (pre-treatment, post-treatment, progression) based on the criterion that the standard deviation of the normalized cfDNA occupancies of all individual samples with this condition determined on step 1 is <0.5 . For these stable-nucleosome regions the normalized nucleosome occupancy of a given condition is defined as the average normalized nucleosome occupancy calculated across all samples with a given condition.

3) For genomic regions which have stable nucleosome occupancy both in pre-treatment and in post-treatment conditions (based on step 2 above), we performed pairwise comparisons of the averaged normalized occupancies. We defined regions where averaged normalized occupancy increased post-treatment versus pre-treatment as “gained-nucleosome regions” (requiring relative change of the average normalized occupancy >0.4). Similarly, we defined regions where averaged normalized occupancy decreased post-treatment versus pre-treatment as “lost-nucleosome regions” (again, requiring the relative change of the average normalized occupancy >0.4). This resulted in 267 and 342 regions where cfDNA occupancy correspondingly decreased or increased post-treatment.

These regions were then analyzed with MEME-ChIP (55) to determine names of TFs which show enrichment of binding sites. The locations of binding sites of these TFs inside regions with differential cfDNA occupancy were then determined with RSAT (56). Aggregate profiles of cfDNA occupancy around these binding sites were calculated with HOMER (57). Analysis of binding sites of TFs experimentally profiled by chromatin immunoprecipitation sequencing

(ChIP-seq) was done by finding motifs of the corresponding TFs inside ChIP-seq peaks with the help of *gimme* (58). ChIP-seq data for *CTCF* binding in A549 SCLC cell line was obtained from GEO accession GSE175135 (59); *CTCF* binding in healthy lung cells from GSE175135 (59). *REST* ChIP-seq dataset in A549 cells was obtained from GSM1010749 (60). Inference of TF activity from cfDNA was also performed using the TranscriptionFactorProfiling tool (61) with default settings based on the top 50% *NEUROD1* binding sites defined in the Gene Transcription Regulation Database (GTRD) (62) (available at <https://github.com/PeterUlz/TranscriptionFactorProfiling/tree/master/Ref/GTRD>). To quantify the activity of TFs, we used the calculated sequencing depth value at the center of the binding profile calculated by the tool.

Graph generation and statistical analysis

All figures were generated using Origin Pro 2021 (originlab.com), GraphPad PRISM software version 8.1.2 (GraphPad Software), R version 1.2.135 (R Foundation for Statistical Computing), and STATA software version 16.0 (Stata-Corp). All statistical tests were two-sided. P value < 0.05 was considered as statistically significant.

Results

Patients, plasma, and tumor samples

We evaluated a prospective cohort of 20 patients with relapsed SCLC. Patients were uniformly treated with a combination of programmed death-ligand 1 inhibitor durvalumab and poly (ADP-ribose) polymerase (PARP) inhibitor olaparib (ClinicalTrials.gov identifier NCT02484404), aimed at inducing tumor immunogenicity by DNA damage via activation of the cyclic GMP-

AMP synthase and stimulator of interferon genes pathway (37). Through a tumor-liquid biopsy program, all patients underwent systematic tumor and plasma sampling before starting treatment, early during (2 to 4 weeks) treatment, and at the time of disease progression (median [range]: 8.1 weeks [4.1–47.7]). A matched tumor biopsy was obtained before starting treatment, during treatment, at disease progression, and while on subsequent therapies. Forty-nine plasma samples were available, including 20 obtained pre-treatment, 17 on treatment, and 12 at disease progression. Twenty-nine tumor samples were available, 18 obtained pre-treatment, 6 on treatment, 2 at progression, and 3 at a subsequent timepoint. Plasma and tumors were time-point matched in 26 cases; in most cases, the plasma samples were obtained on the day of the biopsy, or within a 30-day window. Most of the 23 plasma samples without timepoint matched tumor were obtained on-treatment or at disease progression (**Figure S1**). Tumor exome and transcriptome, and germline exome sequencing were performed as described previously (63).

SCLC patients have higher plasma ctDNA compared to patients with other solid tumors

cfDNA was extracted from the plasma collected at the pre-specified timepoints. The yield of cfDNA in our cohort (median [range] of plasma cfDNA concentration: 2.52 ng/μL [0.056 – 21.22]) was markedly higher than those of other epithelial cancers including non-small cell lung cancer (64, 65). We performed sparse WGS (~0.1× coverage) and used a previously validated analytical approach (64) to estimate tumor fraction based on somatic copy number alterations (SCNAs) in cfDNA while accounting for subclonality and tumor ploidy. Using this approach, cfDNA tumor fraction spanned a broad range (median [interquartile range] cfDNA tumor fraction: 37.0% [11.0%–49.0%]). However, all samples had detectable tumor DNA, and a high proportion of samples yielded tumor fraction of >10% (37/49, 75.5%). Most samples with lower

cfDNA tumor fraction (<10%) were collected after treatment from patients who achieved complete or partial responses (n=7), or from a patient whose tumor was found to have a component of atypical carcinoid (n=2). In comparison, a cohort of castration-resistant prostate cancer patients examined using a similar approach showed much lower tumor fraction (13.0% [4.0%-38.0%]) (66), and only 42% of patients with metastatic triple-negative breast cancer had tumor fraction of >10% (67) (**Figure 2A**).

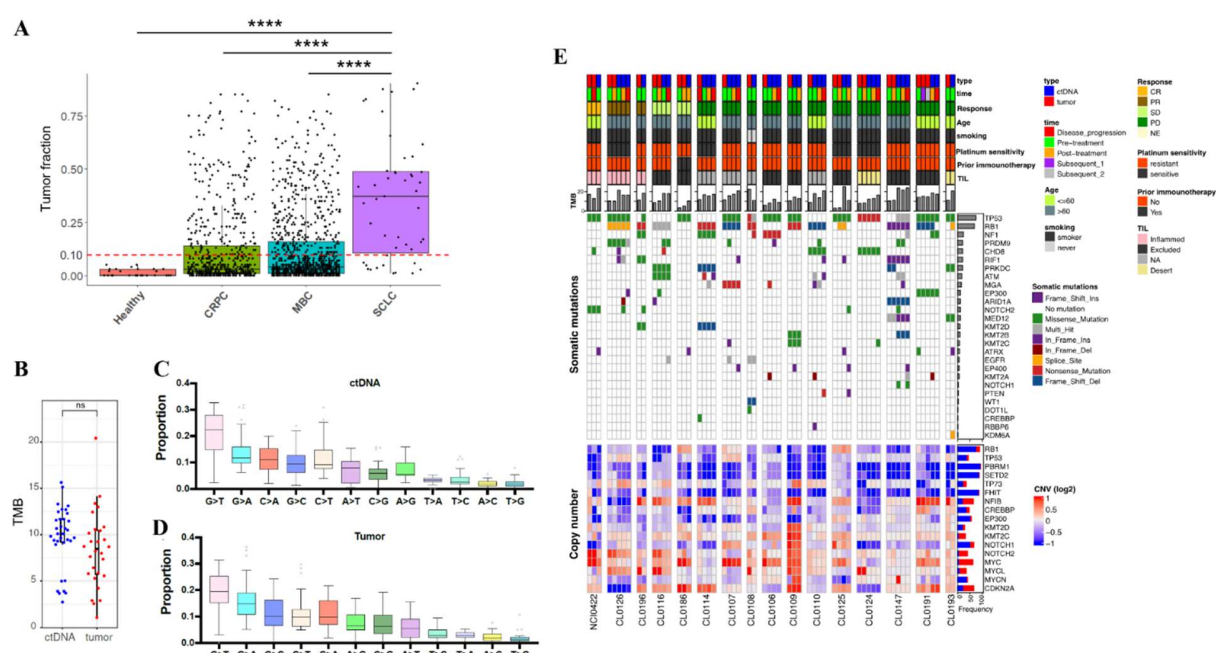


Figure 2. Mutation profiles are highly concordant between cfDNA and tumor

A: cfDNA tumor fraction in healthy donors and patients with CRPC, MBC, and SCLC. ****: $P < 0.0001$ by Kruskal-Wallis test followed by Dunn's multiple comparison test.

B: TMB between cfDNA and tumor samples. ns: $P > 0.05$ by Mann-Whitney U test.

C, D: Distributions of SNVs in cfDNA (C) and tumor (D)

E: Clinical characteristics, TMB, SNVs, and SCNAs in cfDNA and tumor.

Abbreviations: cfDNA: circulating cell-free DNA; CRPC: castration resistant prostate cancer; MBC: metastatic breast cancer; TMB: tumor mutational burden; ns: not significant; SNV: single nucleotide variant; CNV: copy number variant; TIL: tumor infiltrating lymphocytes; CR: complete response; PR: partial response; SD: stable disease; PD: progressive disease; NA: not assessed; SCNA: somatic copy number alteration; SCLC: small cell lung cancer; Ins: insertion; Del: deletion. The genes in the heatmap are recurrently altered genes in SCLC (30, 68). Platinum-sensitive is defined as disease progression ≥ 90 days after first-line platinum-based chemotherapy, and platinum-resistant as disease progression < 90 days or during first-line chemotherapy. TIL was evaluated by immunohistochemistry staining (37).

Mutation and copy number profiles are highly concordant between cfDNA and tumor

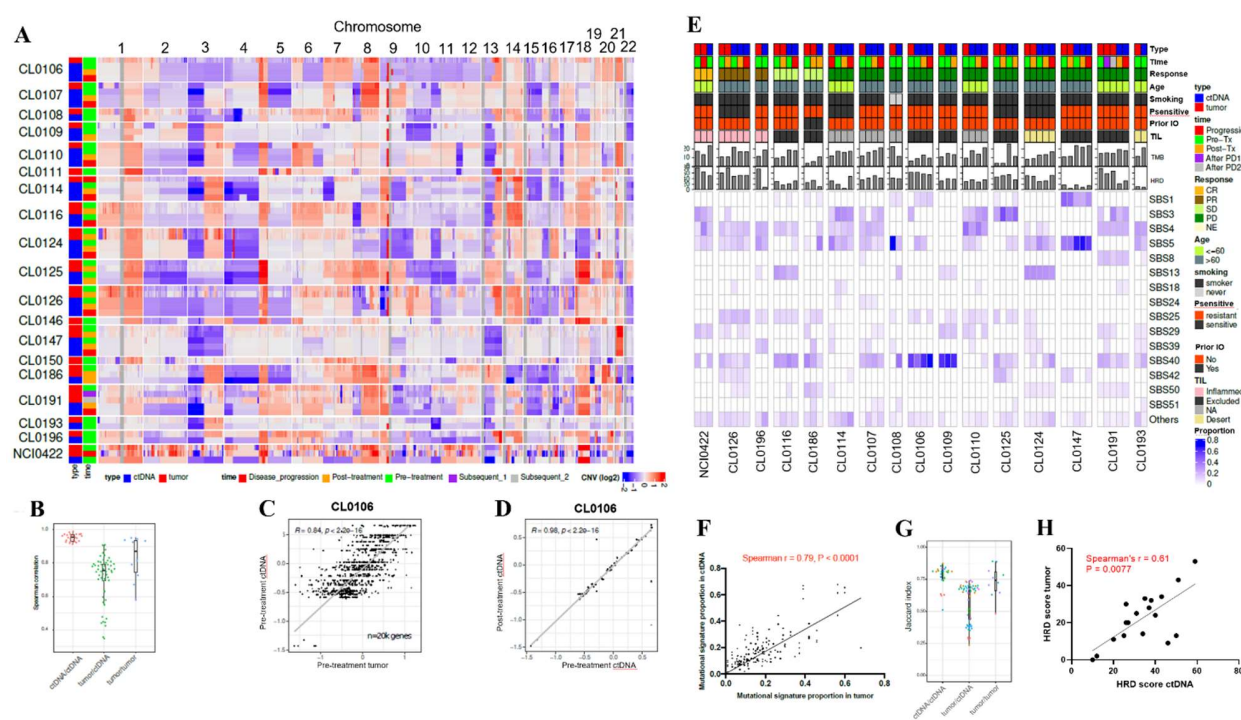
The observation that nearly every patient with SCLC had detectable ctDNA prompted us to examine how well the plasma ctDNA captured the tumor genomic features, a question particularly relevant in SCLCs which are difficult to biopsy. We performed WES of plasma samples with $>10\%$ tumor fraction (average depth of 131X). Tumor fraction estimates from WES and the initial low-pass WGS were highly correlated with each other (Spearman's $r = 0.80$, $P < 0.0001$; **Figure S2**). WES identified a total of 40,354 (median: 1229 per sample) somatic single nucleotide variants (SNVs) and 20,430 (median: 513 per sample) insertion or deletion (indels) among the 37 ctDNA samples with tumor fraction of $>10\%$. This amounted to an average mutation rate of 11.85 mutations per megabase (range: 1.36-21.3) which was similar to that of tumors (**Figure 2B**). The mutation burden of the cohort was higher than other cancer types (**Figure S3A**). G>T transversions were the most common mutations in both ctDNA and tumors, reflecting the mutagenic impact of tobacco smoking in SCLC tumorigenesis (69) (**Figure 2C, D**,

S3B). Mutations and SCNAs were highly concordant between ctDNA and tumors (**Figure 2E**, **S3C**). Mutations and/or copy loss of *TP53* and *RB1* were found in majority of plasma and tumor (55 of 61 [90.2%] and 51 of 61, [83.6%], respectively), consistent with the frequent inactivation of these genes in SCLC (30). Recurrent mutations and SCNAs of Myc paralogues (*MYC*, *MYCL*, *MYCN*), Notch genes (*NOTCH1*, *NOTCH2*), cell cycle regulators (*ATM*, *ATRX*, *CDH8*, *CDKN2A*), and chromatin modifiers (*ARID1A*, *KMT2B*, *KMT2C*, *KMT2D*, *EP300*) were identified in both ctDNA and tumors. Yet, there were several genes important for SCLC tumorigenesis and metastasis (68) that were recurrently mutated in ctDNA, but not detected in the corresponding tumor samples, such *TP53* and *KMT2B* (CL0147) and *NF1* (CL0191) (**Figure 2E**).

Consistent with the known SCNA landscape of SCLCs (30, 68), deletions in chromosome 3p and 10q, and gains in 1p, 2p, 8q were recurrently observed in both ctDNA and tumors (**Figure 3A**). SCNAs were highly concordant between ctDNA and tumors in timepoint-matched samples (median [range] Spearman's $r = 0.81$ [0.45-0.94], $P < 0.0001$ in all pairs) (**Figure 3B, C**), and samples from the same patients obtained at different time points (median [range] Spearman's $r = 0.79$ [0.37-0.94], $P < 0.0001$ in all pairs, **Figure 3B, D**), pointing to the limited evolution of SCNA with treatment. Finally, we assessed whether mutational signatures derived from genome-wide mutation and SCNA profiles might be comparable between plasma and tumor. These signatures hint at the causative origins of cancer, including infidelity of the DNA replication machinery, mutagen exposures, enzymatic modification of DNA, and defective DNA repair (43). The distribution of mutational signatures in the plasma and tumor were highly concordant, with enrichment of tobacco-related single base substitutions (SBS) 4 and 5 dominant in both. The

mutational signature profiles were also similar among ctDNA samples collected at different timepoints (**Figure 3E-G**). HRD score estimated by tumor loss of heterozygosity, telomeric allelic imbalance and large-scale state transition scores predicts efficacy of PARP inhibitors (51). HRD scores evaluated by ctDNA and tumor samples were highly concordant between the tumor and plasma (**Figure 3H**). Together, these observations show that plasma ctDNA recapitulates the tumor mutations and SCNAs, and genome-wide signatures that incorporate these features. In addition, ctDNA may capture key genomic alterations missed in small tumor biopsies.

Figure 3. Somatic copy number alterations (SCNAs) and mutational signature profiles are



highly concordant between circulating cell-free DNA (cfDNA) and tumor

A: Heatmap of SCNAs in cfDNA and tumor. In each row, samples from each patient are aligned tumor followed by ctDNA from top to bottom as indicated on the left.

B: Spearman's coefficients of correlations between cfDNA and tumor SCNA at matched or different time points

C: Representative correlation of SCNAs between pre-treatment tumor (x-axis) and pre-treatment ctDNA (y-axis) in patient CL0106

Spearman's correlation coefficient (R) and P value are indicated.

D: Representative correlation of SCNAs between pre-treatment ctDNA (x-axis) and post-treatment ctDNA (y-axis) in patient CL0106

Spearman's correlation coefficient (R) and P value are indicated.

E: Clinical characteristics, TMB, HRD score, and mutational signature profiles in ctDNA and tumor COSMIC mutational signature version 3.2 (43) is computed and shown in the heatmap.

Platinum-sensitive defined as disease progression ≥ 90 days after first-line platinum-based chemotherapy, and platinum-resistant disease progression < 90 days or during first-line chemotherapy. TIL was evaluated by immunohistochemistry staining (37)

F: Correlation of mutational signature proportions between tumor (x-axis) and ctDNA (y-axis)

G: Distribution of Jaccard index of mutational signatures between cfDNA and tumor at matched or different time points

H: Correlation of HRD scores in pre-treatment ctDNA (x-axis) and tumor (y-axis)

Abbreviations: SCNA: somatic copy number alteration; cfDNA: circulating cell-free DNA; ctDNA: circulating tumor DNA; SCNA: somatic copy number alteration; TMB: tumor mutational burden; HRD: homologous recombination repair deficiency; TIL: tumor infiltrating lymphocytes; CR: complete response; PR: partial response; SD: stable disease; PD: progressive disease; NE: not evaluable; NA: not assessed; COSMIC: Catalogue of Somatic Mutations in Cancer; Psensitive: platinum sensitivity; prior IO: prior immunotherapy.

cfDNA tracks the clinical course and reveal mechanisms of treatment response and resistance

Given the detectable levels of ctDNA in a high proportion of SCLC patients, with the broad dynamic range between patients, we assessed whether the ctDNA profiling could predict tumor burden. We performed volumetric segmentation, a three-dimensional assessment of computed tomography that may more accurately predict clinical outcomes than conventional evaluation by RECIST (64). The cfDNA tumor fraction was significantly positively correlated with volumetric measurements evaluated by timepoint-matched computed tomography (Spearman's $r = 0.66$ $P < 0.0001$, **Figure 4A**).

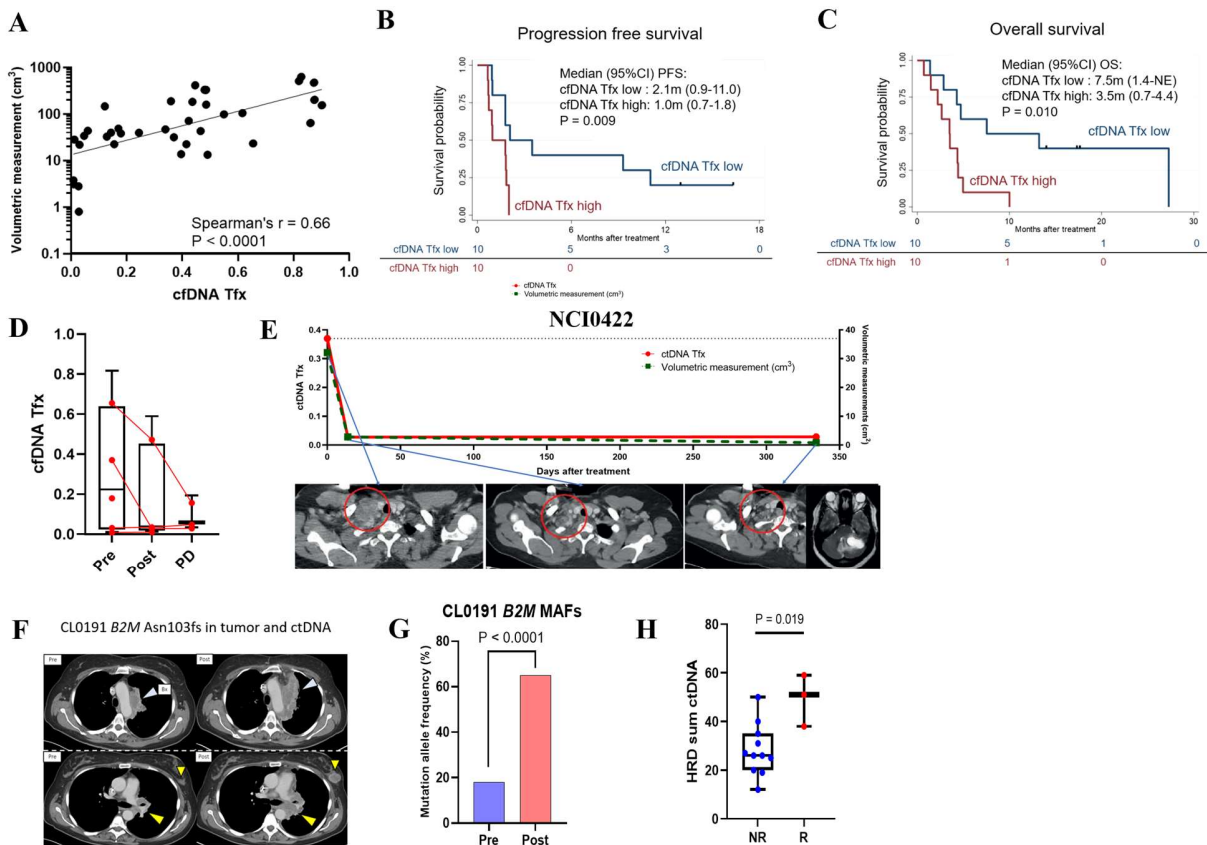


Figure 4. ctDNA tracks the clinical course and reveal mechanisms of treatment response and resistance

A: Correlation between cfDNA tumor fraction and volumetric measurements in time point-matched computed tomography. The y-axis (volumetric measurement) is logarithm transformed.

B, C: Kaplan-Meier curves of PFS (B) and OS (C) in patients with high vs. low cfDNA tumor fraction

High or low cfDNA tumor fraction is defined as patients whose cfDNA tumor fraction is higher or lower than the median of the cfDNA tumor fraction among all 20 samples pre-treatment. P values are evaluated by log-rank test.

D: Changes of cfDNA tumor fractions in patients who had CR or PR as best response

E: Changes of cfDNA tumor fraction (red solid line, left y-axis) and radiological volumetric tumor measurement (green dash line, right y-axis) through treatment time course in a patient who had CR followed by brain only progression (NCI0422). Red circles in CT images indicate right supraclavicular lymph node metastases.

F, G: CT images of para-aortic lymph node metastasis (top, light blue arrowheads), left breast metastasis (bottom, small yellow arrowheads), and left mediastinal lymph node metastasis (bottom, large yellow arrowheads) in a patient who had PD as best response and with *B2M*

Asn103fs variant in cfDNA and tumor (CL0191). “Bx” in panel F indicates the biopsy site for tumor sequencing. MAF changes of the *B2M Asn103fs* variant from pre-treatment to post-treatment cfDNA is indicated in panel G. P value is evaluated by Fisher’s exact test.

H: Comparison of HRD scores derived from pre-treatment ctDNA between patients with SD or PD (= Non-responder, NR) vs. those with CR or PR (= Responder, R). P value is evaluated by Mann-Whitney U test.

425
 426 Abbreviations: cfDNA: circulating cell-free DNA; Tfx: tumor fraction; PFS: progression free
 427 survival; OS: overall survival; CI: confidence interval; m: months; SCLC: small cell lung cancer;
 428 CR: complete response; PR: partial response; SD: stable disease; PD: progressive disease; HRD:
 429 homologous recombination repair deficiency; CT: computed tomography; MAF; mutation allele
 430 frequency; Bx: biopsy; ctDNA; circulating tumor DNA.

431
 432 Given previous studies reporting shortening of cfDNA fragments in cancer (70-72), we analyzed
 433 the distribution of DNA fragment lengths to predict tumor burden (**Figure S4**). We considered
 434 cases with available plasma at all the three time points (pre-treatment, post-treatment, and
 435 disease progression), excluded samples with low ctDNA fraction, and used predictive metrics
 436 based on peaks heights of the size distribution: DNA fragments protected by the chromosome
 437 (sizes around 165 bp), nucleosome core-particle (sizes around 150 bp) and TF binding (sizes
 438 around 50 bp) (**Figure S4A**). We calculated the ratio of the heights of
 439 chromosome/nucleosome and chromosome/TF peaks of the fragment size distribution. The
 440 ratio was shifted towards larger number of longer fragments after treatment, but this effect was
 441 not statistically significant over treatment time course for this patient cohort (**Figure S4B and**
 442 **C**).

443
 444 Next, we assessed whether ctDNA profiles were associated with clinical characteristics and
 445 prognosis. Using the median pre-treatment cfDNA tumor fraction within the cohort, we divided
 446 the patients into low (n=10) or high (n=10) cfDNA tumor fraction groups. No significant
 447 differences were found in age, sex, smoking status, stage at diagnosis, and platinum-sensitivity

between the two groups (**Table S1**). However, progression free survival (PFS) and overall survival (OS) durations were significantly longer in patients with low cfDNA tumor fraction than those with high cfDNA tumor fraction (median [95% confidence interval, CI] PFS and OS in low vs. high cfDNA tumor fraction groups: 2.1 months [0.9–11.0] vs. 1.0 months [0.7–1.8], $p=0.009$; 7.5 months [1.4–not estimable] vs. 3.5 months [0.7–4.4], $p=0.010$, respectively, **Figure 4B, C**). After adjusting for covariates (age, sex, platinum-sensitivity) using a multivariate Cox proportional hazards model, cfDNA tumor fraction at baseline was an independent determinant of PFS and OS (**Table S2, 3**).

We then examined whether cfDNA tumor fraction can track tumor response or progression. As indicated in **Figure 4D**, cfDNA tumor fraction declined or stabilized at low levels in patients who achieved complete or partial response to treatment. Marked reduction of cfDNA tumor fraction was observed in patients who achieved complete (NCI0422, **Figure 4E**) or partial (CL0126, **Figure S5A**) tumor responses. Both patients developed brain-only disease progression, but notably showed no detectable signals in the plasma (**Figure 4E, S5A**), suggesting that this approach may not be sensitive enough to detect intracranial tumor shedding. In contrast, cfDNA tumor fraction significantly increased over time in patients with non-responding tumors (**Figure S5B**). In a patient who had a minor radiographic response of a pleural lesion followed by rapid progression in liver (total volumetric measurements: from 47.2 cm³ to 44.2 cm³ to 205.1 cm³ pre-treatment, post-treatment, and disease progression, respectively), the cfDNA tumor fraction tracked these radiographic changes (CL0116, **Figure S5C**). The cfDNA tumor fraction in a patient who had disease progression as best response steadily increased through treatment time course (CL0124, **Figure S5D**).

471

472 Given that SCLC responds poorly to immunotherapies despite its highly mutated genome (73),

473 we sought to identify potential mechanisms of immunotherapy resistance from ctDNA.

474 Truncating mutation of *B2M* (*B2M* Asn103fs), a known resistance mechanism to immune

475 checkpoint blockade (74, 75), was identified in the ctDNA of a patient who did not respond to

476 treatment (CL0191, **Figure 4F**). Notably, the ctDNA mutation allele frequency (MAF) of the

477 *B2M* variant significantly increased at disease progression compared with baseline (from 17.6%

478 to 64.9%, $p < 0.0001$ by Fisher exact test, **Figure 4G**), suggesting clonal expansion of *B2M* as a

479 potential resistant mechanism. Consistent with the synthetic lethality of PARP inhibition in HRD

480 tumors (76), a ctDNA-derived HRD score (51) was significantly higher in patients who achieved

481 tumor responses compared with those who did not (median [range] HRD score in responder vs.

482 non-responder: 58 [38-59] vs. 26 [12-50], respectively, $P = 0.019$, **Figure 4H**). Together,

483 profiling of ctDNA can predict SCLC tumor burden, non-invasively track the disease course, and

484 discover mechanisms of treatment response and resistance.

485

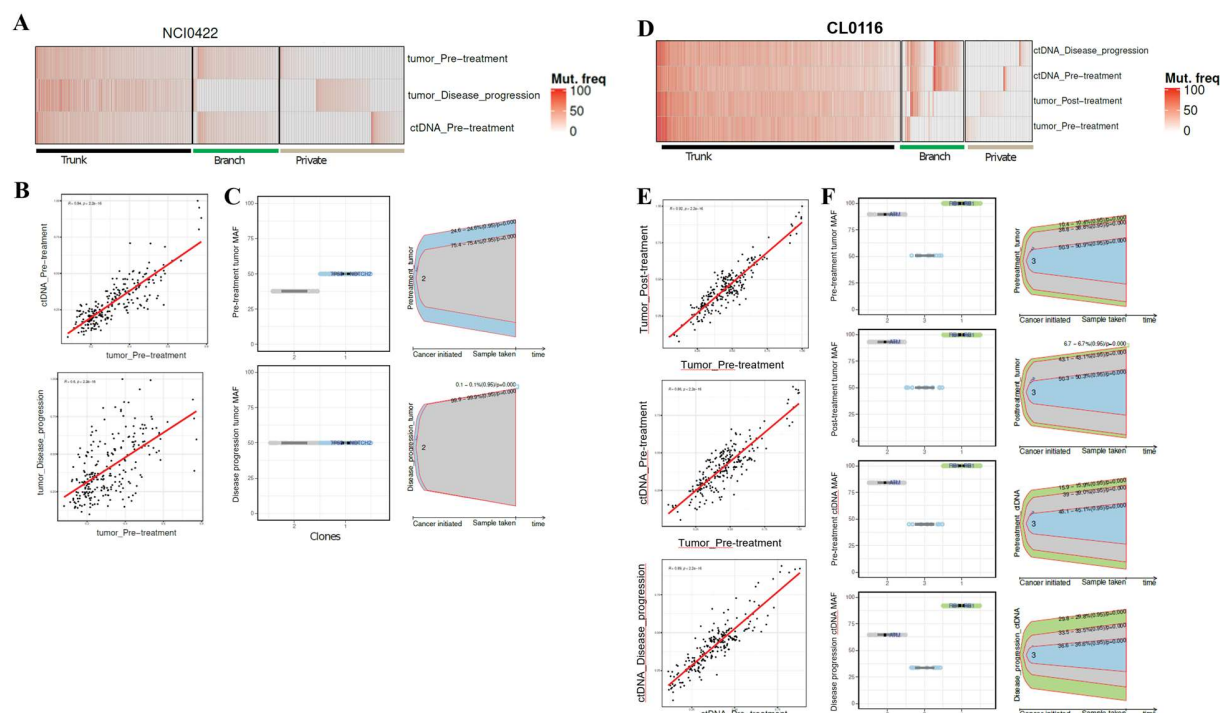


Figure 5. Longitudinal profiling of cfDNA reveal SCLC clonal architecture and track treatment responses

A: Mutation frequencies of variants at different time points in cfDNA and tumor from a patient who achieved complete response followed by brain only progression (NCI0422)

B: Correlations of mutation frequencies between cfDNA vs. tumor pre-treatment (top) and pre-treatment vs at disease progression tumors (bottom) in a patient who achieved complete response followed by brain only progression (NCI0422). Spearman's coefficients (R) and P values are indicated.

C: Visualization of genomic clones through treatment time course in a patient who achieved complete response followed by brain only progression (NCI0422)

D: Mutation frequencies of variants at different time points in cfDNA and tumor in a patient who had disease progression as the best response (CL0116)

E: Correlations of mutation frequencies between pre-treatment vs. post-treatment tumors (top), pre-treatment tumor vs. cfDNA (middle), and cfDNA pre-treatment vs at disease progression (bottom) in a patient who had disease progression as the best response (CL0116)

F: Visualization of genomic clones through treatment time course in a patient who had disease progression as the best response (CL0116)

Abbreviations: cfDNA: circulating cell-free DNA; ctDNA: circulating tumor DNA; SCLC: small cell lung cancer; Mut. freq: mutation frequency; MAF: mutation allele frequency.

Longitudinal profiling of cfDNA reveal SCLC clonal architecture and track treatment responses

Since longitudinal tumor biopsies are not practical in the setting of the generally rapid progression of SCLC, the evolutionary patterns of SCLC under treatment pressure remain poorly understood. By sampling ctDNA and tumors at multiple timepoints, we evaluated the clonal evolution of SCLC through the treatment course. Phylogenic analyses revealed linear evolution in 18 of 19 patients whose longitudinal samples were successfully processed, indicating a genetic landscape that does not change markedly over treatment time course and is dominated by truncal clones harboring alterations of *TP53* and *RBI*. Prior studies have also reported low subclonal diversity in SCLC (30, 77). However, there were few notable exceptions. In a patient who achieved complete response followed by recurrence in the brain one year after the treatment initiation (NCI0422), a relatively higher proportion of unique mutations were identified in the relapsed tumor compared with either pre-treatment ctDNA or tumor, whereas the pre-treatment samples harbored clonally similar cell populations (**Figure 5A, B**). The median MAFs among

mutations in the subclone increased at disease progression, indicating the expansion of the clone through acquired resistant mechanisms with the treatment (**Figure 5C**). On the other hand, the majority of mutations were shared between ctDNA and tumor obtained at multiple time points in a patient who did not respond to treatment (CL0116, **Figure 5D, E**). Three genomic clones were identified and their MAFs did not change over treatment time course (**Figure 5F**). Together, the subclonal architecture of SCLC profiled non-invasively using ctDNA demonstrate a generally linear evolution with treatment, pointing to non-genetic mechanisms such as transcriptional plasticity (32, 78) rather than genetic mechanisms underlying treatment resistance.

cfDNA occupancy at transcription factor binding sites (TFBS) predicts SCLC phenotypes and treatment response

DNA is protected from nuclease digestion through its association with a nucleosome core particle and other chromatin proteins (**Figure S4A**) and therefore it may be possible to infer differences in TF binding between different medical conditions from cfDNA (61, 70, 79). We analyzed differential occupancy of cfDNA at TF binding sites (TFBS). First, we identified loci with largest changes of cfDNA occupancy by scanning the genome with 10,000 base-pair (bp) sliding window and identifying regions which have similar cfDNA occupancy across all samples with the same condition (pre- or post-treatment respectively), but significant change in occupancy post-treatment versus pre-treatment. Following these criteria (detailed in Materials and methods), we identified 267 and 342 regions where cfDNA occupancy decreased or increased respectively post-treatment.

The regions with altered cfDNA occupancy were analyzed for enrichment of TFBS (**Table S4**) and aggregate profiles of cfDNA occupancy calculated around these TFBS (**Figure S6A, B**). Interestingly, for the class of regions where cfDNA occupancy increased post-treatment, it usually declined at the time of disease progression but remained higher than at pre-treatment levels. Among the individual TFs with marked differences in cfDNA occupancy over the treatment course, the most prominent were *NRF1* and *REST* (**Figure 6A, B**), which are known to co-localize on DNA binding sites, with REST facilitating NRF1 occupancy by promoting local DNA hypomethylation (80). Plotting cfDNA occupancy profiles for patients from our cohort around sites bound by *REST* in SCLC cell line A549 showed clear distinction of the pre-treatment profile from post-treatment and disease progression (**Figure 6C**). Binding sites of the chromatin organizer protein CTCF and its close paralog BORIS also showed distinct changes pre- and post-treatment (**Figure 6D, E, S6C, D**), suggesting the potential impact of treatment on three-dimensional genome organization.

We also separated our cohort into two groups based on platinum sensitivity and analyzed cfDNA occupancy at TFBS. Higher REST occupancy was seen at disease progression in platinum-resistant compared with platinum-sensitive cases (**Figure 6F**). *NEUROD1* occupancy also changed over the treatment time-course, with numerically higher occupancy at disease progression in platinum-resistant compared with platinum-sensitive cases (**Figure 6G, S6E**). Of note, average GC content was not different genome-wide and in regions with differential cfDNA occupancy, suggesting that GC content did not confound the analysis of cfDNA occupancy (**Figure S6F**). The regions with differential cfDNA occupancy defined above also had very limited overlap with CNVs (only ~11% overlap with regions undergoing amplifications with

log₂fold change >1). Thus, CNV-based and cfDNA occupancy-based analyses are complementary. The average cfDNA occupancy profiles were also not different between samples with high or low cfDNA tumor fraction (**Figure S6G, H**). Together, cfDNA TFBS occupancy analysis showed distinct treatment related changes, which may reflect the impact of treatment on the transcriptional landscape.

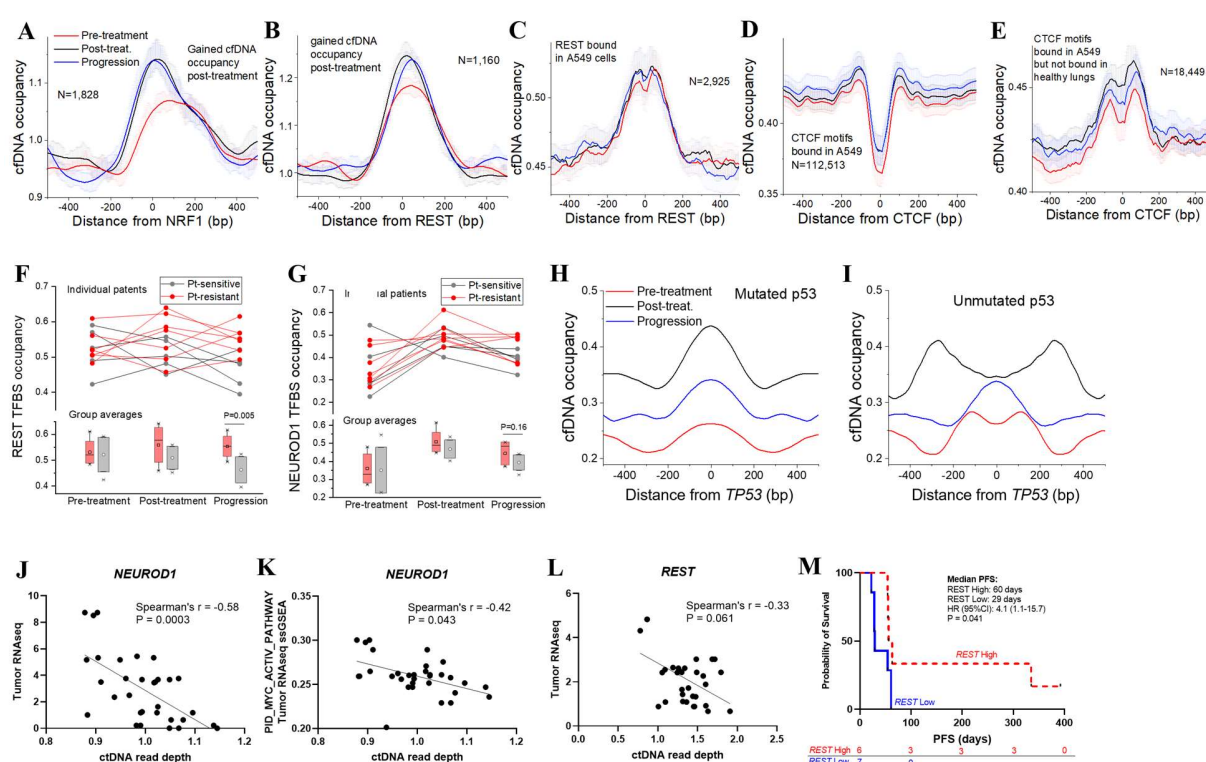


Figure 6. cfDNA transcription factor occupancy predicts SCLC phenotypes and treatment response

A-E: Aggregate cfDNA occupancy profiles around binding sites of *NRF1*, *REST*, and *CTCF* pre-treatment (red), post-treatment (black) and at disease progression (blue)

A: Computationally predicted *NRF1* binding sites inside regions which where cfDNA occupancy increases post-treatment

B: cfDNA occupancy profiles around computationally predicted *REST* binding sites inside regions where cfDNA occupancy increases post-treatment

C: cfDNA occupancy profiles around experimentally determined *REST* binding sites in A549 SCLC cells

D: cfDNA occupancy profiles around *CTCF* binding motifs inside experimentally determined *CTCF* binding sites in A549 SCLC cells.

E: cfDNA occupancy profiles around a subset of *CTCF* binding sites from (D), which do not overlap with *CTCF* sites bound in healthy lungs.

F, G: cfDNA occupancy at TFBSs of *REST* and *NEUROD1*. TFBSs occupancies are shown for the three time points of individual patients, along with group averages. Patients were split into two groups which were sensitive (grey circles and connecting lines) and resistant (red circles and lines) to platinum-based chemotherapy. Top: individual TFBS activity in *REST* (F) or *NEUROD1* (G) are shown. Bottom: averages within the groups of platinum-sensitive (Pt-sensitive) and resistant (Pt-resistant) in *REST* (F) or *NEUROD1* (G) are shown. *REST* sites are defined based on chromatin immunoprecipitation sequencing in A549 cells. *NEUROD1* motifs are defined computationally inside regions with increased cfDNA occupancy post-treatment vs pre-treatment. Platinum-sensitive defined as disease progression ≥ 90 days after first-line platinum-based chemotherapy, and platinum-resistant disease progression < 90 days or during first-line chemotherapy. P values are evaluated by Mann-Whitney U test.

H, I: cfDNA occupancy profiles at different timepoints around computationally predicted *TP53* binding sites inside regions which have increased cfDNA occupancy post-treatment, in samples with (H) vs. without (I) mutations in *TP53*.

J: Correlation between ctDNA read depth of *NEUROD1* binding sites (x-axis) and *NEUROD1* gene expression (TMM-FPKM) in timepoint-matched tumors (y-axis)
Higher read depth indicates less TF binding, predicting less gene expression.

K: Correlation between *NEUROD1* ctDNA read depth at TFBS (x-axis) and the PID_MYC_ACTIV_PATHWAY scores by ssGSEA in timepoint-matched tumors (y-axis)
Higher read depth indicates less TF binding, predicting less gene expression.

L: Correlation between ctDNA read depth at TFBS (x-axis) and gene expression of tumor RNA sequencing (TMM-FPKM, y-axis) in the gene *REST*.
Higher read depth indicates less TF binding, predicting less gene expression.

M: A Kaplan-Meier curve of PFS in patients with high vs. low predicted *REST* expression
High vs. low predicted *REST* expression is defined as higher or lower than median predicted *REST* expression by cfDNA read depth among 13 patients whose pre-treatment cfDNA was successfully processed for the TFBS analysis. Higher predicted *REST* expression was defined as lower read depth and vice versa, given that higher read depth indicates less TF binding, predicting less gene expression. P value is evaluated by Log-rank test.

Abbreviations: cfDNA: circulating cell-free DNA; bp: base pair; TFBS: transcriptional factor binding site; Pt: platinum-based chemotherapy; TMM-FPKM: Trimmed Mean of M-values-normalized Fragments per kilo base per million mapped reads; ssGSEA; single sample gene set enrichment analysis; PFS; progression free survival; HR: hazard ratio; CI: confidence interval.

The analysis performed above gave equal weights to binding sites of a given TF even if some sites contained weaker DNA binding motifs. To check whether TFBS strength influences this analysis, we have also profiled SCLC-specific TFs using a recently proposed nucleosome

footprint analysis method (61) which takes into account top 50% strongest TFBS from the Gene Transcription Regulation Database (GTRD) (62). Using this approach, binding site accessibility of *NEUROD1*, an SCLC lineage defining TF (32), was significantly correlated with *NEUROD1* gene expression in the corresponding time point matched tumor (Spearman's $r = -0.58$, $P = 0.0003$; higher ctDNA read depth indicates less binding of TFs predicting less gene expression) (**Figure 6J**). Consistent with MYC driving the *NEUROD1*-high SCLC subtype (32, 81, 82), increased accessibility of *NEUROD1* in cfDNA was associated with upregulation of targets of MYC transcriptional activation in the corresponding tumors (Spearman's $r = -0.51$, $P = 0.0030$) (**Figure 6K**). Binding site accessibility of *REST*, a transcriptional repressor of neuroendocrine differentiation (32, 82) was higher in cfDNA corresponding to tumor samples with increased *REST* expression (Spearman's $r = -0.33$, $P = 0.061$, **Figure 6L**). Importantly, patients with higher predicted *REST* expression based on lower *REST* read depth in ctDNA had prolonged PFS following immunotherapy than patients with lower predicted expression (**Figure 6M**). *REST* binding site accessibility was not associated with OS (**Figure S7**). These results are consistent with recent observations of low neuroendocrine SCLC differentiation driven by Notch signaling which targets *REST* being predictive of benefit from immunotherapy-based approaches, but not a prognostic marker (33, 34). Thus, TFBS occupancy/accessibility estimation derived from cfDNA can inform prediction of SCLC neuroendocrine phenotypes and treatment response.

Discussion

There are few systematic comparisons of the effectiveness of cfDNA at elucidating tumor-derived molecular features relative to standard single-lesion tumor biopsies in prospective cohorts of patients. The use of plasma instead of tissue to guide therapy is a particularly

attractive alternative for patients with SCLC, a cancer driven by transcription addiction, and whose clinical course makes it exceedingly challenging to obtain tumor biopsies. Here, in a prospective cohort of molecularly defined patients with recurrent SCLC, treated uniformly with an immunotherapy-based combination, we find that cfDNA not only mirrored the mutation and copy number landscape of the tumor, but also its genomic signatures, while also detecting clinically relevant resistance mechanisms, and cancer driver alterations not found in matched tumor biopsies. Our study also confirmed previous observations of high cfDNA tumor fraction in patients with SCLC compared with other solid tumors (22), as well as the utility of longitudinal cfDNA analysis to reliably track tumor response and progression, and reveal mechanisms of treatment response and resistance (83).

Most applications of cfDNA to date are gene-centric focusing on somatic variants, and are of limited utility when tumor mutations are not known a priori and in tumors driven by dysregulated transcriptional programs. We find that cfDNA TF binding profiles are reflective of the altered transcriptional landscape in response to treatment (before vs. after treatment vs. tumor progression) and chemo-sensitivity (platinum sensitive vs. resistant). We find a striking association between cfDNA accessibility of *NEUROD1* inferred from nucleosome footprint analysis and expression of *NEUROD1* in the corresponding tumor. A similar trend was observed between *REST* accessibility and expression, which defined tumors with low neuroendocrine differentiation and higher likelihood of response to immunotherapy (33, 34). SCLC tumors exhibit distinct inter-tumor heterogeneity with respect to expression of neuroendocrine features, driven by expression of lineage TFs (32, 33). Whether the subtypes engender specific therapeutic vulnerabilities is an area of active investigation (33-36). A major barrier to clinical validation of

the proposed subtypes is the limited availability of high-quality tumors for molecular analyses.

The number of available sequenced cfDNA datasets from different cancer subtypes reported by different labs continues to increase exponentially (84), thus the clinical validation of this study can be expected quite fast. Our findings – the distinction between SCLC neuroendocrine phenotypes based on cfDNA REST and NEUROD1 accessibility – could allow for noninvasive characterization and treatment for patients.

Our cohort was limited by the small sample size, of whom only few patients had clinical benefit from the treatment (37). Future studies are needed to validate these findings in a general SCLC population and in other tumor types. Nevertheless, our cohort represents a prospective population, and the collection and processing of all samples was performed in a systematic fashion, ensuring homogeneity of pre-analytical characteristics and careful control of experimental and analytical variables. Targeted cfDNA sequencing approaches enriching cfDNA fragments covering known transcription start sites across the genome or binding sites of SCLC-specific TFs might make this approach more amenable to clinical application.

Conclusions

By direct comparisons of cfDNA versus tumor biopsy, we offers insights into non-invasive stratification and subtype-specific therapies for SCLC, now treated as a single disease, and has broad implications for mapping tumor-specific transcription factor binding on blood samples. Further studies with large cohort in a prospective manner is warranted.

693 **List of abbreviations**

- 694 cfDNA: circulating cell-free DNA
- 695 SCLC: small cell lung cancer
- 696 ctDNA: circulating tumor DNA
- 697 TCGA: the Cancer Genome Atlas
- 698 NCI: National Cancer Institute
- 699 PCR: polymerase chain reaction
- 700 FFPE: formalin-fixed, paraffin-embedded
- 701 TMM-FPKM: Trimmed Mean of M-values-normalized Fragments per kilobase per million
- 702 mapped reads
- 703 LOH: Loss of heterozygosity
- 704 TAI: telomeric allelic imbalance
- 705 LST: large-scale state transition
- 706 HMM: hidden Markov model
- 707 ASCN: Allele specific copy number
- 708 NSR: noise to signal ratio
- 709 PARP: poly (ADP-ribose) polymerase
- 710 WES: whole exome sequencing
- 711 SNV: single nucleotide variants
- 712 Indel: insertion or deletion
- 713 SBS: single base substitutions
- 714 HRD: Homologous recombination repair deficiency
- 715 SCNA: Somatic copy number alteration

716 RECIST: Response Evaluation Criteria in Solid Tumors

717 PFS: progression free survival

718 OS: overall survival

719 MAF: mutation allele frequency

720 TFBS: transcription factor binding sites

721 GTRD: transcription factor binding sites

722

Declarations

Ethics approval and consent to participate

The trial was conducted under a NCI Center for Cancer Research–sponsored investigational new drug application with institutional review board approval. Written informed consent was obtained from all patients as well as to use and share data and specimens collected for the study.

Consent for publication

All patients' information was anonymized and informed consent for publication is obtained for all datasets.

Availability of data and materials

Sequencing data (ctDNA WGS, tumor WES, RNA sequencing) used in this study will be deposited in dbGaP . Other information in this study are available upon request from the corresponding author.

Competing Interests

We have no declaration of interests to report.

Funding

This study was supported by the Center for Cancer Research, the Intramural Program of the National Cancer Institute (ZIA BC 011793). AT reports research funding from AstraZeneca, Tarveda, EMD Serono, and Prolynx.

Authors' contributions

NT, LP, and VNR performed data analysis and interpretation. SPA, MS, and VT performed analyses of cfDNA occupancy at TFBS. AS and MB performed volumetric segmentation of CT scan. CWS, SK, and PD provided supports for interpretations and discussion about results. DR, PC, and WDF corresponded specimen storage. NT, RV, SN, PD, and AT provided clinical supports during the clinical trial. NT, LP, VNR, SK, DP, and AT contributed to manuscript design and interpretation of data. NT, LP and AT wrote the manuscript with comments and contributions from all authors. All authors read and approved the final manuscript.

Acknowledgements

The authors gratefully acknowledge the contributions of the patients who participated in the study.

References

1. Corcoran RB. Circulating Tumor DNA: Clinical Monitoring and Early Detection. *Annu Rev Canc Biol.* 2019;3:187-201.
2. Ignatiadis M, Sledge GW, Jeffrey SS. Liquid biopsy enters the clinic — implementation issues and future challenges. *Nature Reviews Clinical Oncology.* 2021;18(5):297-312.
3. Dawson SJ, Tsui DWY, Murtaza M, Biggs H, Rueda OM, Chin SF, et al. Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer. *New Engl J Med.* 2013;368(13):1199-209.
4. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. *Nat Med.* 2008;14(9):985-90.
5. Powles T, Assaf ZJ, Davarpanah N, Banchereau R, Szabados BE, Yuen KC, et al. ctDNA guiding adjuvant immunotherapy in urothelial carcinoma. *Nature.* 2021;595(7867):432-+.
6. Chaudhuri AA, Chabon JJ, Lovejoy AF, Newman AM, Stehr H, Azad TD, et al. Early Detection of Molecular Residual Disease in Localized Lung Cancer by Circulating Tumor DNA Profiling. *Cancer Discov.* 2017;7(12):1394-403.
7. Garcia-Murillas I, Schiavon G, Weigelt B, Ng C, Hrebien S, Cutts RJ, et al. Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Science Translational Medicine.* 2015;7(302).
8. Beaver JA, Jelovac D, Balukrishna S, Cochran RL, Croessmann S, Zabransky DJ, et al. Detection of Cancer DNA in Plasma of Patients with Early-Stage Breast Cancer. *Clin Cancer Res.* 2014;20(10):2643-50.

9. Cohen JD, Li L, Wang YX, 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-+.
10. Chan KCA, Woo JKS, King A, Zee BCY, Lam WKJ, Chan SL, et al. Analysis of Plasma Epstein-Barr Virus DNA to Screen for Nasopharyngeal Cancer. *N Engl J Med*. 2017;377(6):513-22.
11. Kuderer NM, Burton KA, Blau S, Rose AL, Parker S, Lyman GH, et al. Comparison of 2 Commercially Available Next-Generation Sequencing Platforms in Oncology. *Jama Oncol*. 2017;3(7):996-8.
12. Chae YK, Davis AA, Carneiro BA, Chandra S, Mohindra N, Kalyan A, et al. Concordance between genomic alterations assessed by next-generation sequencing in tumor tissue or circulating cell-free DNA. *Oncotarget*. 2016;7(40):65364-73.
13. Thierry AR, Pastor B, Jiang ZQ, Katsiampoura AD, Parseghian C, Loree JM, et al. Circulating DNA Demonstrates Convergent Evolution and Common Resistance Mechanisms during Treatment of Colorectal Cancer. *Clin Cancer Res*. 2017;23(16):4578-91.
14. Bettgowda 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.
15. Goyal L, Saha SK, Liu LY, Siravegna G, Leshchiner I, Ahronian LG, et al. Polyclonal Secondary FGFR2 Mutations Drive Acquired Resistance to FGFR Inhibition in Patients with FGFR2 Fusion-Positive Cholangiocarcinoma. *Cancer Discov*. 2017;7(3):252-63.

16. Russo M, Siravegna G, Blaszkowsky LS, Corti G, Crisafulli G, Ahronian LG, et al. Tumor Heterogeneity and Lesion-Specific Response to Targeted Therapy in Colorectal Cancer. *Cancer Discov.* 2016;6(2):147-53.
17. Parikh AR, Leshchiner I, Elagina L, Goyal L, Levovitz C, Siravegna G, et al. Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers. *Nat Med.* 2019;25(9):1415-21.
18. Herberts C, Annala M, Sipola J, Ng SWS, Chen XE, Nurminen A, et al. Deep whole-genome ctDNA chronology of treatment-resistant prostate cancer. *Nature.* 2022;608(7921):199-208.
19. Rudin CM, Brambilla E, Faivre-Finn C, Sage J. Small-cell lung cancer. *Nat Rev Dis Primers.* 2021;7(1).
20. Thomas A, Pattanayak P, Szabo E, Pinsky P. Characteristics and Outcomes of Small Cell Lung Cancer Detected by CT Screening. *Chest.* 2018;154(6):1284-90.
21. Almodovar K, Iams WT, Meador CB, Zhao Z, York S, Horn L, et al. Longitudinal Cell-Free DNA Analysis in Patients with Small Cell Lung Cancer Reveals Dynamic Insights into Treatment Efficacy and Disease Relapse. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer.* 2018;13(1):112-23.
22. Nong J, Gong Y, Guan Y, Yi X, Yi Y, Chang L, et al. Circulating tumor DNA analysis depicts subclonal architecture and genomic evolution of small cell lung cancer. *Nature communications.* 2018;9(1):3114.
23. Mohan S, Foy V, Ayub M, Leong HS, Schofield P, Sahoo S, et al. Profiling of Circulating Free DNA Using Targeted and Genome-wide Sequencing in Patients with SCLC.

Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer. 2020;15(2):216-30.

24. Mondelo-Macía P, García-González J, León-Mateos L, Castillo-García A, López-López R, Muínelo-Romay L, et al. Current Status and Future Perspectives of Liquid Biopsy in Small Cell Lung Cancer. *Biomedicines*. 2021;9(1).

25. Fernandez-Cuesta L, Perdomo S, Avogbe PH, Leblay N, Delhomme TM, Gaborieau V, et al. Identification of Circulating Tumor DNA for the Early Detection of Small-cell Lung Cancer. *EBioMedicine*. 2016;10:117-23.

26. Devarakonda S, Sankararaman S, Herzog BH, Gold KA, Waqar SN, Ward JP, et al. Circulating Tumor DNA Profiling in Small-Cell Lung Cancer Identifies Potentially Targetable Alterations. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2019;25(20):6119-26.

27. Board RE, Williams VS, Knight L, Shaw J, Greystoke A, Ranson M, et al. Isolation and extraction of circulating tumor DNA from patients with small cell lung cancer. *Annals of the New York Academy of Sciences*. 2008;1137:98-107.

28. Du M, Thompson J, Fisher H, Zhang P, Huang CC, Wang L. Genomic alterations of plasma cell-free DNAs in small cell lung cancer and their clinical relevance. *Lung cancer (Amsterdam, Netherlands)*. 2018;120:113-21.

29. Iams WT, Almodovar K, Meador C, York SJ, Horn L, Raymond C, et al. Longitudinal monitoring of circulating tumor DNA and peripheral T cell repertoire in patients with small cell lung cancer. *Journal of Clinical Oncology*. 2017;35(15_suppl):8571-.

30. George J, Lim JS, Jang SJ, Cun Y, Ozretić L, Kong G, et al. Comprehensive genomic profiles of small cell lung cancer. *Nature*. 2015;524(7563):47-53.

31. Kim DW, Kim KC, Kim KB, Dunn CT, Park KS. Transcriptional deregulation underlying the pathogenesis of small cell lung cancer. *Transl Lung Cancer Res.* 2018;7(1):4-20.
32. Rudin CM, Poirier JT, Byers LA, Dive C, Dowlati A, George J, et al. Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. *Nature reviews Cancer.* 2019;19(5):289-97.
33. 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(3):346-60.e7.
34. Roper N, Velez MJ, Chiappori A, Kim YS, Wei JS, Sindiri S, et al. Notch signaling and efficacy of PD-1/PD-L1 blockade in relapsed small cell lung cancer. *Nature communications.* 2021;12(1):3880.
35. Thomas A, Takahashi N, Rajapakse VN, Zhang X, Sun Y, Ceribelli M, et al. Therapeutic targeting of ATR yields durable regressions in small cell lung cancers with high replication stress. *Cancer cell.* 2021;39(4):566-79.e7.
36. Lissa D, Takahashi N, Desai P, Manukyan I, Schultz CW, Rajapakse V, et al. Heterogeneity of neuroendocrine transcriptional states in metastatic small cell lung cancers and patient-derived models. *Nature communications.* 2022;13(1):2023.
37. Thomas A, Vilimas R, Trindade C, Erwin-Cohen R, Roper N, Xi L, et al. Durvalumab in Combination with Olaparib in Patients with Relapsed SCLC: Results from a Phase II Study. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer.* 2019.

38. Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey TM, et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome biology*. 2011;12(1):R1.
39. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature biotechnology*. 2013;31(3):213-9.
40. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nature methods*. 2018;15(8):591-4.
41. Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q, et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Annals of oncology : official journal of the European Society for Medical Oncology*. 2015;26(1):64-70.
42. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*. 2013;14(4):R36.
43. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578(7793):94-101.
44. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome biology*. 2016;17:31.
45. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome biology*. 2016;17(1):122.
46. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome research*. 2018;28(11):1747-56.

47. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS computational biology*. 2016;12(4):e1004873.
48. Cun Y, Yang TP, Achter V, Lang U, Peifer M. Copy-number analysis and inference of subclonal populations in cancer genomes using ScIust. *Nature protocols*. 2018;13(6):1488-501.
49. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics (Oxford, England)*. 2009;25(14):1841-2.
50. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics (Oxford, England)*. 2016;32(18):2847-9.
51. Telli ML, Timms KM, Reid J, Hennessy B, Mills GB, Jensen KC, et al. Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2016;22(15):3764-73.
52. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, et al. PyClone: statistical inference of clonal population structure in cancer. *Nature methods*. 2014;11(4):396-8.
53. Dang HX, White BS, Foltz SM, Miller CA, Luo J, Fields RC, et al. ClonEvol: clonal ordering and visualization in cancer sequencing. *Annals of oncology : official journal of the European Society for Medical Oncology*. 2017;28(12):3076-82.
54. Vainshtein Y, Rippe K, Teif VB. NucTools: analysis of chromatin feature occupancy profiles from high-throughput sequencing data. *BMC Genomics*. 2017;18(1):158.
55. Machanick P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*. 2011;27(12):1696-7.

56. Nguyen NTT, Contreras-Moreira B, Castro-Mondragon JA, Santana-Garcia W, Ossio R, Robles-Espinoza CD, et al. RSAT 2018: regulatory sequence analysis tools 20th anniversary. *Nucleic Acids Res.* 2018;46(W1):W209-W14.
57. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010;38(4):576-89.
58. van Heeringen SJ, Veenstra GJ. GimmeMotifs: a de novo motif prediction pipeline for ChIP-sequencing experiments. *Bioinformatics.* 2011;27(2):270-1.
59. ENCODE_Project_Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57-74.
60. Gertz J, Savic D, Varley KE, Partridge EC, Safi A, Jain P, et al. Distinct properties of cell-type-specific and shared transcription factor binding sites. *Molecular cell.* 2013;52(1):25-36.
61. Ulz P, Perakis S, Zhou Q, Moser T, Belic J, Lazzeri I, et al. Inference of transcription factor binding from cell-free DNA enables tumor subtype prediction and early detection. *Nature communications.* 2019;10(1):4666.
62. Kolmykov S, Yevshin I, Kulyashov M, Sharipov R, Kondrakhin Y, Makeev VJ, et al. GTRD: an integrated view of transcription regulation. *Nucleic acids research.* 2021;49(D1):D104-d11.
63. Tlemsani C, Takahashi N, Pongor L, Rajapakse VN, Tyagi M, Wen X, et al. Whole-exome sequencing reveals germline-mutated small cell lung cancer subtype with favorable response to DNA repair-targeted therapies. *Science translational medicine.* 2021;13(578).

64. 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. *Nature communications*. 2017;8(1):1324.
65. Szpechcinski A, Chorostowska-Wynimko J, Struniawski R, Kupis W, Rudzinski P, Langfort R, et al. Cell-free DNA levels in plasma of patients with non-small-cell lung cancer and inflammatory lung disease. *Br J Cancer*. 2015;113(3):476-83.
66. Choudhury AD, Werner L, Francini E, Wei XX, Ha G, Freeman SS, et al. Tumor fraction in cell-free DNA as a biomarker in prostate cancer. *JCI insight*. 2018;3(21).
67. Stover DG, Parsons HA, Ha G, Freeman SS, Barry WT, Guo H, et al. Association of Cell-Free DNA Tumor Fraction and Somatic Copy Number Alterations With Survival in Metastatic Triple-Negative Breast Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2018;36(6):543-53.
68. Kim KB, Dunn CT, Park KS. Recent progress in mapping the emerging landscape of the small-cell lung cancer genome. *Experimental & molecular medicine*. 2019;51(12):1-13.
69. Pfeifer GP, Hainaut P. On the origin of G --> T transversions in lung cancer. *Mutat Res*. 2003;526(1-2):39-43.
70. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell*. 2016;164(1-2):57-68.
71. Udomruk S, Orrapin S, Pruksakorn D, Chaipayat P. Size distribution of cell-free DNA in oncology. *Critical reviews in oncology/hematology*. 2021;166:103455.
72. Underhill HR, Kitzman JO, Hellwig S, Welker NC, Daza R, Baker DN, et al. Fragment Length of Circulating Tumor DNA. *PLoS genetics*. 2016;12(7):e1006162.

73. Yarchoan M, Hopkins A, Jaffee EM. Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *New England Journal of Medicine*. 2017;377(25):2500-1.
74. Gettinger S, Choi J, Hastings K, Truini A, Datar I, Sowell R, et al. Impaired HLA Class I Antigen Processing and Presentation as a Mechanism of Acquired Resistance to Immune Checkpoint Inhibitors in Lung Cancer. *Cancer discovery*. 2017;7(12):1420-35.
75. Sade-Feldman M, Jiao YJ, Chen JH, Rooney MS, Barzily-Rokni M, Eliane JP, et al. Resistance to checkpoint blockade therapy through inactivation of antigen presentation. 2017;8(1):1136.
76. Lord CJ, Ashworth A. PARP inhibitors: Synthetic lethality in the clinic. *Science*. 2017;355(6330):1152-8.
77. Zhou H, Hu Y, Luo R, Zhao Y, Pan H, Ji L, et al. Multi-region exome sequencing reveals the intratumoral heterogeneity of surgically resected small cell lung cancer. *Nature communications*. 2021;12(1):5431.
78. Wagner AH, Devarakonda S, Skidmore ZL, Krysiak K, Ramu A, Trani L, et al. Recurrent WNT pathway alterations are frequent in relapsed small cell lung cancer. *Nat Commun*. 2018;9.
79. Ulz P, Thallinger GG, Auer M, Graf R, Kashofer K, Jahn SW, et al. Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nature genetics*. 2016;48(10):1273-8.
80. Domcke S, Bardet AF, Adrian Ginno P, Hartl D, Burger L, Schubeler D. Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature*. 2015;528(7583):575-9.

81. Mollaoglu G, Guthrie MR, Böhm S, Brägelmann J, Can I, Ballieu PM, et al. MYC Drives Progression of Small Cell Lung Cancer to a Variant Neuroendocrine Subtype with Vulnerability to Aurora Kinase Inhibition. *Cancer cell*. 2017;31(2):270-85.
82. Ireland AS, Micinski AM, Kastner DW, Guo B, Wait SJ, Spainhower KB, et al. MYC Drives Temporal Evolution of Small Cell Lung Cancer Subtypes by Reprogramming Neuroendocrine Fate. *Cancer cell*. 2020;38(1):60-78.e12.
83. Hench IB, Hench J, Tolnay M. Liquid Biopsy in Clinical Management of Breast, Lung, and Colorectal Cancer. *Frontiers in medicine*. 2018;5:9.
84. Shtumpf M, Piroeva KV, Agrawal SP, Jacob DR, Teif VB. NucPosDB: a database of nucleosome positioning in vivo and nucleosomics of cell-free DNA. *Chromosoma*. 2022;131(1-2):19-28.