Lung Cancer 120 (2018) 113–121

Contents lists available at ScienceDirect

Lung Cancer

journal homepage: www.elsevier.com/locate/lungcan

# Genomic alterations of plasma cell-free DNAs in small cell lung cancer and their clinical relevance

Meijun Du[a], Jonathan Thompson[b], Hannah Fisher[a], Peng Zhang[a], Chiang-Ching Huang[c], Liang Wang[a,*]

[a] Department of Pathology and Cancer Center, Medical College of Wisconsin, Milwaukee, WI, USA
[b] Division of Hematology/Oncology, Department of Medicine, Medical College of Wisconsin, Milwaukee, WI, USA
[c] Joseph J. Zilber School of Public Health, University of Wisconsin, Milwaukee, WI, USA

ARTICLE INFO

*Keywords:*
Small cell lung cancer
Cell-free DNA
Copy number variation
Somatic mutation
Survival

ABSTRACT

*Objectives:* To identify genomic variations in cell-free DNA (cfDNA) and evaluate their clinical utility in small cell lung cancer (SCLC).

*Materials and methods:* We performed whole genome sequencing using plasma cfDNAs derived from 24 SCLC patients for copy number variation (CNV) analysis, and targeted sequencing using 17 pairs of plasma cfDNA and their matched gDNA for mutation analysis. We defined somatic mutations by comparing cfDNA to its matched gDNA with 5% variant alleles as the cutoff for mutation calls. We applied Kaplan-Meier to correlate the genomic alterations with overall survival (OS) and progression-free survival (PFS).

*Results:* We observed widespread somatic copy-number alterations and mutations, including amplification of *MYC* at 8q24, *FGF10* at 5p13, *SOX2* at 3q26 and *FGFR1* at 8p12, as well as deletion of *TP53* at 17p13, *RASSF1* at 3p21.3, *RB1* at 13q14.2, *FHIT* at 3p14, and *PTEN* at 10q23. The most frequent mutations were genes involved in chromatin regulation (*KMT2D*, *ARID1A*, *SETBP1* and *PBRM1)*, PI3K/MTOR pathway(*MTOR*,*PIK13G*), Notch1 signalling pathway (*NOTCH1*), and DNA repair related gene *ATRX*. Kaplan-Meier analysis revealed poor OS and PFS in patients with somatic mutations in gene *SETBP1* (P = 0.0061/0.0264, HR = 4.785/3.841, 95% CI = 2.014–28.25/1.286–16.58) and *PBRM1* (P = 0.0276/0.0286, HR = 3.532/3.506, 95% CI = 1.275 to 25.34/1.26–24.87). Poor OS was also associated with somatic mutations in *ATRX* (P = 0.0099, HR = 4.024, 95% CI = 1.926–42.95), *EP300* (P = 0.025/0.0622, HR = 3.382/2.891, 95% CI = 1.448–27.76/1.013-17.29), while poor PFS was associated with *ATM* mutation (P = 0.0038, HR = 4.604, 95% CI = 2.211–40.93). The mutation index produced by summing up the number of mutations in the five genes was significantly associated with the poor OS/PFS (P = 0.0185/0.0294) after adjusting the effect of the stage.

*Conclusions:* Our result supports blood plasma as a promising sample source for the genomic analysis in SCLC patients whose tumor tissues are scarcely available and demonstrates potential clinical utilities of cfDNA-based liquid biopsy for clinical management of this deadly disease.

## 1. Introduction

Small-cell lung cancer (SCLC) accounts for approximately 10–15% of all lung cancers[1,2]. The patients are often heavy smokers, and tumor cells express neuroendocrine markers. It is a highly aggressive malignancy frequently presenting with metastases at time of diagnosis. Although chemotherapy is initially effective in the treatment of SCLC, recurrence arises rapidly in the vast majority of cases and the survival outcome is poor. Five-year survival rates remain amongst the lowest of all solid tumours with only 6% of patients surviving 5 years from diagnosis [3]. Unfortunate outcomes are likely to be due to subsequent genetic events occurring during tumor progression, possibly under the direct selection pressure of the drugs employed. Furthermore, the most reproducible prognostic factor is stage of the disease. No histological or molecular features are prognostically useful [1,4]. Therefore, to improve treatment efficacy of the disease, it is essential to understand the genetic variations and identify prognostic biomarkers.

Since SCLC is rarely treated by surgery, surgical specimens of primary tumor are scarcely available for genomic characterization. Although three seminal genomic studies have recently been conducted

*Abbreviations:* SCLC, small cell lung cancer; cfDNA, cell-free DNA; gDNA, lymphocyte germline DNA; CNV, copy number variation; OS, overall survival; PFS, progression-free survival

* Corresponding author at: Department of Pathology and MCW Cancer Center, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226, USA.
*E-mail address:* liwang@mcw.edu (L. Wang).

https://doi.org/10.1016/j.lungcan.2018.04.008
Received 6 December 2017; Received in revised form 5 April 2018; Accepted 10 April 2018

[5–7] and the genomic features have been correlated with the clinical outcome[8], these studies were evaluated patients with surgically resectable tumors, a rare entity (< 1% of patients) termed peripheral SCLC. Moreover, due to the spatial and temporal genomic heterogeneity of cancers, genomic sequencing obtained from a single biopsy may not capture the complete genomic profile of tumors. To address these issues, liquid biopsy, characteristic of non-invasive sampling and the ability to capture the genetic heterogeneity of cancer in bodily fluids, has attracted a growing interst, especially in developing cell-free DNA (cfDNA)-based biomarkers for detecting the presence of malignancies, monitoring treatment response, judging prognosis, or evaluating recurrence [8]. Studies have shown that tumor-derived DNAs are detectable in peripheral blood and their genetic features are associated with clinical outcomes [9–12].

To test tumor-derived genetic and genomic changes in peripheral blood, we applied whole genome sequencing and targeted sequencing technologies in plasma cfDNAs and matched germline DNAs (gDNAs) from 24 SCLC patients. We provided the genetic and genomic profiles (both copy number variation and somatic mutations) of cfDNA in SCLC and evaluated the potential association of genetic alterations with the main clinical outcomes. The study result demonstrated the feasibility of using cfDNA for genomic profiling and prediction of survival in patients with SCLC.

## 2. Materials and methods

### 2.1. Patient samples

A total of 24 SCLC plasma and paired lymphocyte gDNA were provided by the Medical College of Wisconsin Tissue Bank. Plasma was collected, uniformly processed, and stored at −80 °C before use, as previously described [13,14]. This study was approved by Institutional Review Board at the Medical College of Wisconsin.

### 2.2. Plasma cfDNA extraction

The process of cfDNA extraction has previously been described [14,15]. In brief, blood plasma samples underwent a centrifugation at 3000 rpm for 10 min before DNA extraction. The cfDNAs from 250 µl of plasma were extracted using the DNA Blood Mini Kit (Qiagen, Valencia, CA,USA). The extracted cfDNA was eluted in 50 µl water and quantified using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA). Lymphocyte gDNAs were extracted using the DNA Blood Mini Kit according to the standard protocol. The extracted DNA was stored at −80 °C until preparation of sequencing libraries.

### 2.3. Whole genome sequencing

DNA libraries were prepared using a ThruPLEX DNA-seq Library Kit (Rubicon Genomics, Ann Arbor, MI) per manufacturer's instructions. 1 ng of cfDNA and 2 ng of lymphocyte gDNA [14–16]were used for library preparation including end-repair, addition of stem-loop adaptors, and 15 cycles of high fidelity amplification. Following amplification, the libraries were purified using a 1:1 ratio of DNA sample to Agencourt AMPure XP Beads (Beckman Coulter, Indianapolis, IN). The library quality and insert size were examined using High Sensitivity DNA Analysis Chip (Agilent Technologies, Santa Clara, CA). The 24 indexed libraries were pooled for 50 bp single-read sequencing on a HiSeq2500 Sequencing System (Illumina, San Diego, CA).

### 2.4. Targeted sequencing

Targeted sequencing was carried out in 17 plasma cfDNA samples and 17 matched lymphocyte gDNA from the same patient. The xGen Pan-Cancer Panel (IDT, Coralville, IA) was used for targeted sequencing. The panel covers 800 kb of the human genomic sequences targeting 127 significantly mutated genes implicated across 12 tumor types. The target genes were captured from whole genome sequencing libraries following manufacturer's instructions with some modifications. In brief, 500 ng of plasma cf DNA libraries, 5ug Cot-1 DNA, 2 ul xGen® Universal Blockers-TS Mix were mixed and dried in a 1.5 ml tube. Then, 8.5 ul of 2x hybridization buffer, 2.7 ul hybridization buffer enhancer and 1.8 ul nuclease-free water were added to suspend the DNA mixture. After hybridization at 65 °C for 4 h, the targets were captured by streptavidin-coated magnetic beads (Dynabeads M-270 Streptavidin; Life Technologies, Oslo Cat #2014-08) followed by removing unbound DNA using washing buffer, the captured DNA fragments were amplified for an additional 12 amplification cycles and purified by Agencourt AMPure beads (Beckman Coulter, Inc., Cat # A63881). Final enriched libraries were subjected to 100 bp PE sequencing on a HiSeq2500 Sequencing System.

### 2.5. Copy number variation (CNV) calculation

CNV analysis was performed as previously described [14,16]. In brief, Raw reads were alligned to the human reference genome (NCBI37/hg19) by using DNASTAR (Madison, WI). The mapped reads were then binned into either 1Mb (for overall copy number analysis) or 60Kb (for locus-specific copy number analysis) and re-scaled to 10 million reads after excluding sex chromosomes. The read count ratio in each genomic bin was calculated by dividing cfDNA by lymphocyte gDNA from the same patient. For the four cfDNA without gDNA samples, the read count ratio in each genomic bin was calculated by dividing cfDNA by average read count of the 20 gDNAs. The resulting ratios were further transformed with log2 and corrected for GC content [17]. The fully normalized log2 ratios in genomic bins were subjected to segmentation using the copy number analysis method algorithm (Golden Helix, Bozeman, MT).

### 2.6. Mutation analysis

Muational analysis was performed using DNASTAR software (Madison, WI). In brief, raw sequencing data (fastq files) were assembled using an Exome and Gene Panel pipeline with the reference genome of hg19 and target region file: Xgen- Pan-Cancer-targets, downloaded from IDT website. Assembly options included somatic/cancer/heterogenous and PCR duplicate removal. Somatic mutations were identified by comparing cfDNA to gDNA in the same patient with 5% variant alleles as the cutoff for variant calls. Any variant call that matched to dbSNPs was excluded. Variations present in a plasma sample but absent in a matched gDNA were defined as somatic mutations. The minimun unique sequence depth was 50 for cfDNA. To ensure correct mutation calls, we visually checked all sequence alignments at detected mutation sites. Finally, all reported mutations met the following criteria: 1). variants in coding regions only, 2). variant allele frequency > 5%, 3). variants not in dbSNPs, and 4). variants not in gDNA.

### 2.7. Statistical analysis of genomic data

Overall survival (OS) was measured from the date of diagnosis to the date of death and censored at the date of last follow-up for survivors. Progression-free survival was measured from the date of diagnosis to the date of disease progression or the date of death, whichever occurred first and censored at the date of last follow-up for survivors without disease progression. Survival distribution was estimated using Kaplan–Meier methods and statistical signficance in survival between groups were examined by the log-rank test. The effect of clinical factors, including age, sex and stage on survival were further estimated using Cox model. The genes with mutations in at least four patients were included for statistical analysis.