A Pilot Prospective Study of Plasma Cell free DNA Whole Genome Sequencing Identified Chromosome 7p Copy Number Gains is a Specific Biomarker for Early Lung Cancer detection

Hongjie Yu
Department of Thoracic Surgery, the First Affiliated Hospital of Soochow University

Xiaojun Yu (✉ 18362738721@163.com )
The First Affiliated Hospital of Soochow University  https://orcid.org/0000-0002-7318-0220

Jia Tang
Department of Thoracic Surgery, First Affiliated Hospital of Soochow University

Xun Lu
Department of Thoracic Surgery, First Affiliated Hospital of Soochow University

Haitao Ma
Department of Thoracic Surgery, First Affiliated Hospital of Soochow University

Research

Keywords: Chromosome 7, biomarker, lung cancer

DOI: https://doi.org/10.21203/rs.3.rs-80693/v1

License: ☑️ ☁️ This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

**Purpose:** Chromosome 7 is playing an important role in lung tumorigenesis. Here we investigate whether chromosome 7p copy number gain as a detectable genetic event with plasma cell-free DNA for early lung cancer detection biomarker.

**Methods:** Eighteen surgical eligible lung cancer patients and eighteen non-cancer controls were recruited. Peripheral blood was collected before surgery. Cell-free DNA was profiled with low coverage whole genome sequencing. Chromosome 7 copy number gains were defined as chr7 normalized coverage ≥ 1.0005 and P value < 0.05. Plasma cell-free DNA chr7 copy gains were then compared to pathological examinations on surgical tissues.

**Results:** 83.3% patients were confirmed as malignancy post-operation, with 12 adenocarcinoma and 3 squamous-carcinoma. The other 16.7% were benign lesions. Cell-free DNA was successfully extracted from pre-surgical plasma samples, with concentration range from 0.18 to 0.49 ng/ul. Chromosome 7 short arm copy gains were found in 66.7% (10/15) patients, including 66.7% (4/6) T1aN0M0 and 50.0% (1/2) Tis patients, otherwise, chr7p gain was found in 0% (0/3) benign lesions. The specificity was further examined in 18 volunteers who undergoing routine body examinations. Meanwhile, positive CEA and CYFRA21-1 was only found in 1/18 (5.7%) and 4/18 (22.2%). Taking together, UCAD chr7p or UCAD chr7p and tumor biomarker positivity can predict 12/15 (80%, 95% CI: 49.0-94.3%) early lung cancers. Further analyses showed that chr7p copy number gains tends to be enriched in EGFR/KRAS silent patients (fisher test, P value = 0.077).

**Conclusions:** Chromosome 7p copy gains might be a useful peripheral blood tumor biomarker from lung cancer detection.

Introductions

Lung cancer is the leading cause of cancer-related death worldwide. Although there are improvements in lung cancer diagnosis and treatment recent years, 5-year survival rate is still less than 20% of all lung cancer patients[1]. The most important reason for poor survival is still relatively advanced stage when lung cancer was diagnosed. A large prospective clinical trial of applying low-dose CT (LDCT) scan for early lung cancer discovery has showed great survival benefits. However, LDCT scan results in huge false positives[2]. About 94% of LDCT identified nodules are actually benign changes, which dose need further treatments. Usually, invasive diagnosis approaches will be applied to those patients. Hence, there are needs to identify non-invasive approaches to complement LDCT to reduce over-diagnosis.

Chromosome 7 aneuploidy was discovered an early events on lung cancer development. Chromosome 7 amplifications were frequently reported in lung cancer tissue profiling studies. It is also reported in lung bronchial in-situ hybridization studies. Chromosome 7 aneuploidy could be found in both malignant tissues and in pre-cancerous lesions, but not health tissues, which supports chromosome 7 copy number...
changes might be an early event of lung cancer tumorigensis[3–5]. However, tissue biopsy is an invasive and complicated procedure. And sometimes, it might cause tumor metastasis.

Recently years, chromosome aneuploidy detection by using plasma cell free DNA for prenatal tests, with minimal false positives and false negatives[6]. Similar to fetal tissues, tumors also keep shedding DNA into peripheral blood stream. Cancer somatic mutations, such as EGFR mutations, were successfully detected as a biomarker indicating potential benefits from targeted therapies. In addition to somatic point mutations, chromosomal copy number changes were also detected in breast cancer[7], hepatocellular carcinoma [8] and lung cancer[9]. However, most of the complicated copy number changes happened in advanced stage tumors. It is showing limited power for early cancer detection. Here we investigate whether chromosome 7 aneuploidy detection in early lung cancer detection.

**Materials And Methods**

**Funding**

This work is supported by the science and technology fund social development project of Jiangsu Province (No. BE2015640).

**Ethics statement**

The design and methods of the current study involving human subjects were clearly described in a research protocol. Informed consent was obtained from all patients. Institutional Review Board (IRB) of Soochow University approved the study. All recruited subjects have signed a written informed consent.

**Subjects**

Twenty-six ovarian cancer patients and thirty-four benign ovarian tumor patients were admitted to the Department of Thoracic of the First Affiliated Hospital of Suchoow University. All patients were chemotherapy-naive or radiotherapy-naive. Blood samples were collected for cfDNA extracting and CA125 measuring before surgery.

**Declarations**

The authors have no conflicts of interest to declare.

**generation sequencing**

Total genomic DNA and cfDNA were isolated from tissue samples and plasma using the Amp Genomic DNA Kit (TIANGEN) and QIAseq cfDNA Extraction kit (Qiagen) respectively. Next generation sequencing was performed as previously described[18,19]. DNA was fragmented into an average size of 300bp (cfDNA without fragmentation), and then 100 ng of fragmented genomic DNA (cfDNA 10ng) was used
for preparation of sequencing libraries (NEBnext Ultra II). 8bp barcoded sequencing adaptors were then ligated with DNA fragments and amplified by PCR. Purified sequencing libraries were massively parallel sequenced by Illumina HiSeq Xten platform. 4G sequencing raw data per sample were filtered and aligned to the human reference genome.

Statistical Analysis.

R package ‘DNACopy’ was used to analysis copy number changes. A P value of <0.05 was considered as statistically significant binary segmentation. Absolute segment value is used for further analysis. The sensitivity and specificity of UCAD were estimated by Receiver Operating Characteristic (ROC) curves. For categorical variables, the chi-square test was used as appropriate. All statistical analyses were performed using SPSS17.0.

Results

Fresh plasma samples were collected from 8 lung cancer patients. All the patients are informed with signed consent. DNA was successfully extracted from all of these samples with DNA concentration range from 0.19 to 0.49 ng/ul. Sequencing library was prepared with standard NEB protocol. It was then send to Illumina X10 for sequencing. In average, 10G data was collected for each sample.

1. Patient characterization

As shown in table 1, 16 cancer patient samples, 3 benign lesions and 18 health controls was collected. UCAD detection was performed in 12 adenocarcinoma and 4 squamous cell carcinoma before surgery. Tumors show a significant higher UCAD chr7 score compared to control group. And advanced tumors shows higher UCAD chr7 score compared to stage T1abN0M0 patients. Tumor biomarker was almost negative for all 8 samples. There is no obvious correlation between UCAD chr7 score and patient age, histological subtype, molecular subtype (EGFR/KRAS/ALK/ROS1/PIK3CA status). And there is also no obvious correlation between UCAD chr7 score and smoking status.

Table 1. Clinical pathological information
|                | Ad  | Scc | Benign | Healthy |
|----------------|-----|-----|--------|---------|
| **N**          | 12  | 3   | 3      | 18      |
| **Stage**      |     |     |        |         |
| T1b, T2a       | 5   | 2   | n.a.   |         |
| T1aN0M0        | 5   | 1   | n.a.   |         |
| Tis            | 2   | 0   | n.a.   |         |
| **Sex**        |     |     |        |         |
| M              | 4   | 3   | 3      | 8       |
| F              | 9   | 0   | 0      | 10      |
| **Age**        |     |     |        |         |
| >=55y          | 6   | 2   | 1      | 7       |
| <55y           | 6   | 1   | 2      | 11      |
| **EGFR mut**   |     |     |        |         |
| L858R          | 1   | 0   | 0      | n.a.    |
| 19del          | 2   | 0   | 0      | n.a.    |
| WT             | 1   | 4   | 0      | n.a.    |
| NA             | 9   | 0   | 3      |         |
| **Serum TM**   |     |     |        |         |
| CEA            | 1   | 0   | 0      | n.a.    |
| CYFRA21-1      | 3   | 1   | 0      | n.a.    |

Note: TM, tumor biomarker; Ad, Adenocarcinoma; Scc, Squamous cell carcinoma; WT, wildtype; del: deletion; NA, not available

**2 cell-free DNA whole genome copy number profiling**

The reads were mapped to human reference genome hg19. Genomic coverage was then counted by using software samtools mpileup. We then calculate average coverage for each 200k bin. Circular binary segmentation algorithm was then used to detect significant genomic breakpoints.

As shown in Figure 1 and 2, chromosomal breakpoints were commonly found on centromere regions (vertical dash lines). Chromosome 7 short-arm generally was found with higher coverage compared to
long arm, indicating short arm copy gains. Relatively much lower chromosomal imbalance was found in non-tumor controls.

Figure 1, chromosome 1 to 22 is layout from left to right with green and black colors. Chromosomal segments are marked in red lines.

Figure 2, zoom out view of chromosome 7, a breakpoint is found on chromosome 7 centromere (blue circle). Chr7p's coverage is significantly higher than chr7q, which peak coverage around oncogene EGFR.

3 UCAD chr7p gains in cancer patients of different clinical stages

A parameter set, alpha=0.01, was used to detect genomic breakpoints in DNAcopy software package. Student test was then performed to test the differences between chromosome short arm (chr7p) and normalized genome-wide average value 1. A ‘significant’ chr7p gain was defined by: 1) absolute value large or equal than 1.0005 and 2) Student t-test P value less than 0.05. As shown in table 2, 10 of 15 (66.7%) samples reached significance, including 5/7 (71.43%) T1b/T2a, 4/6 (66.67%) T1a patients and 1/2 (50%) Tis patients. All 3 benign controls does not reach significance. The specificity was validated on another independent ‘body-examination’ volunteer cohort. 1 of 18 (5.56%) individual was found to be chr7p positive, who need further follow ups.

Table 2. statistical analyses of chr7 aneuploidy by T stages

|                  | T1b,T2a (N=7) | T1aN0M0 (N=6) | Tis (N=2) | Not specified (N=1) | Benign Lesion (N=2) | Body-examination Volunteer (N=18) |
|------------------|---------------|---------------|-----------|--------------------|---------------------|----------------------------------|
| UCAD+            | 5             | 4             | 1         | 0                  | 0                   | 1                                |
| UCAD-            | 2             | 2             | 1         | 1                  | 2                   | 17                               |
| %                | 71.43%        | 66.67%        | 50.00%    | 0.00%              | 0.00%               | 5.56%                            |

4 UCAD chr7p+ is independent predictor of lung malignancy

13 serum tumor biomarkers was tested, including AFP,CEA,NSE,PROGRP,β2-MG, Serum ferritin, CA242, CA153, CYFRA21-1,CA50 ,CA199 , CA72-4 and SCC. 4 of 12 (33.3%) patients was found with CYFRA21-1 mild increase. 1 of them (8.33%) was found with mild CEA increase. The other 7 was negative for all the 13 tumor biomarkers. As shown in table 3, UCAD chr7p gains is a independent predictor of lung malignancy.

Table 3. UCAD chr7p+ is independent predictor of lung malignancy
5 Chr7p gain negatively correlated with EGFR/KRAS mutation status

EGFR and KRAS mutations were examined post surgery. 2 EGFR exon 19 deletions, 1 EGFR L858R and 1 KRAS G13D mutation were found. Interestingly, as shown in table 4, 3 of 4 (75%) of the EGFR/KRAS mutant patients were UCAD negative. And 9 of 10 (90%) UCAD+ patients were EGFR/KRAS wildtype. There is tend of negative correlation between UCAD+ and EGFR/KRAS mutations (Fisher.test, p=0.067).

Table 4. Chr7p gain negatively correlated with EGFR/KRAS mutation status

|               | EGFR/KRAS + (N=4) | EGFR/KRAS – (N=11) |
|---------------|-------------------|--------------------|
| UCAD chr7p+   | 1                 | 9                  |
| UCAD chr7p-   | 3                 | 2                  |
|                 | Fisher.test P value = 0.067 |

6 A model of plasma cell free DNA UCAD Chr7p and tumor biomarker to predict early stage cancer

By combinding biomarker and independent of UCAD chr7p positivity, a prediction power of 12/15 (80%, 95% CI:49.0-94.3%) was reached in table 5.

Table 5. Combination of UCAD chr7p and tumor biomarker to predict early stage lung cancers

|             | UCAD chr7p+ | UCAD chr7p+ | UCAD chr7p- | UCAD chr7p- |
|-------------|-------------|-------------|-------------|-------------|
| TM+         | 3           | 7           | 2           | 3           |
| Count       |             |             |             |             |
| %           | 20.00%      | 46.67%      | 13.33%      | 20.00%      |
Discussions

All 8 cancer samples was found with UCAD chr7 score larger than 0.0020. However, 3 of them did not reach statistical significance. The potential reason could be the coverage variations across chromosome due to PCR amplification bias, sequencer bias and etc. PCR efficiency is different for different sequence composition, for example, high GC content regions might be preferred by specific PCR reactions. Similarly for Illumina sequencer, GC bias has been considered to be the major cause of coverage bias. In addition to GC content, secondary structure of cfDNA fragments might be another major factor contribute to coverage bias. The size of cfDNA fragment is around 170 base pair which enable a complicated secondary structure. Hence, PCR reaction conditions need further be optimized for chromosome 7 copy number detections.

Copy number variations detected in non-cancer patient. There is minor chromosome changes which was found in non-cancer patients, which might indicates the background copy number variations among non-cancer individuals. In previous report, slight chromosome aneuploidy might be found in pre-cancerous lesions. To find the baseline across non-tumor individuals, a large scale baseline study might be need.

We also compared UCAD chr7 score to traditional lung cancer tumor biomarkers, such as CEA, SCC and etc. Traditional tumor biomarker is rarely positive in current dataset. In comparison, all tumor patient show higher UCAD chr7 score compared to control samples. The results suggest UCAD ch7 score might be another used non-invasive peripheral biomarker to monitor lung cancer. We even can use UCAD chr7 score to screen cancer by competence LDCT and tumor biomarkers.

In this research, the discovery was found in a small scale prospective study, including 8 cancer patients and 4 health controls. Although the preliminary data was encouraging, we still need a large prospective clinical trial to further confirm the discovery.

Declarations

Ethics statement

The design and methods of the current study involving human subjects were clearly described in a research protocol. Informed consent was obtained from all patients. Institutional Review Board (IRB) of Soochow University approved the study. All recruited subjects have signed a written informed consent.

Funding

This work is supported by the science and technology fund social development project of Jiangsu Province (No. BE2015640).
References

1. Herbst RS, Heymach JV, Lippman SM. Lung Cancer. N Engl J Med [Internet] Massachusetts Medical Society. 2008;359:1367–80.

2. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. N Engl J Med [Internet] Massachusetts Medical Society. 2011;365:395–409.

3. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The landscape of somatic copy-number alteration across human cancers. Nature [Internet]. 2010;463:899–905.

4. Zojer N, Dekan G, Ackermann J, Fiegl M, Kaufmann H, Drach J, et al. Aneuploidy of chromosome 7 can be detected in invasive lung cancer and associated premalignant lesions of the lung by fluorescence in situ hybridisation. Lung Cancer [Internet]. 2000;28:225–35.

5. Ricke RM, van Deursen JM. Aneuploidy in health, disease, and aging. J. Cell Biol. [Internet]. 2013;201:11 LP-21.

6. Norton ME, Jacobsson B, Swamy GK, Laurent LC, Ranzini AC, Brar H, et al. Cell-free DNA Analysis for Noninvasive Examination of Trisomy. N Engl J Med [Internet] Massachusetts Medical Society. 2015;372:1589–97.

7. 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. J. Clin. Oncol. [Internet]. American Society of Clinical Oncology; 2018;JCO.2017.76.0033.

8. Jiang P, Chan CWM, Chan KCA, Cheng SH, Wong J, Wong VW-S, et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. Proc. Natl. Acad. Sci. [Internet]. 2015;112:E1317–25.

9. Xia S, Huang C-C, Le M, Dittmar R, Du M, Yuan T, et al. Genomic variations in plasma cell free DNA differentiate early stage lung cancers from normal controls. Lung Cancer [Internet] Elsevier. 2017;90:78–84.