doi:10.4149/neo_2021_201106N1189

A pilot prospective study of plasma cell-free DNA whole-genome sequencing identified chromosome 7p copy number gains as a specific biomarker for early lung cancer detection

Hong-Jie YU[1], Xiao-Jun YU[2,*], Jia TANG[1], Xun LU[1], Hai-Tao MA[2,*]

[1]Department of Thoracic Surgery, The First Affiliated Hospital of Soochow University, Suzhou, Jiangsu, China; [2]Dushu Lake Hospital Affiliated to Soochow University, Suzhou, Jiangsu, China

**Correspondence: 18362738721@163.com; mht7403@163.com*

Received November 6, 2020 / Accepted January 14, 2021

Chromosome 7 plays an important role in lung tumorigenesis. Chromosome 7 copy number changes might be an early event of lung cancer tumorigenesis. Here we investigate whether chromosome 7p copy number gain is a detectable genetic event with plasma cell-free DNA for early lung cancer detection. 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. 83.3% of patients were confirmed as malignancy post-operation, 12 patients with adenocarcinoma, and 3 with squamous-carcinoma. The other 16.7% were benign lesions. Cell-free DNA was successfully extracted from pre-surgical plasma samples, with a concentration range from 0.18 to 0.49 ng/μl. 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 carcinoembryonic antigen (CEA) and cytokeratin-19-fragment (CYFRA21-1) were only found in 1/18 (5.7%) and 4/18 (22.2%), respectively. Taking together, Ultrasensitive- Chromosomal Aneuploidy Detector (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 tend to be enriched in normal EGFR/KRAS patients (Fisher's test, p-value = 0.077). Chromosome 7p copy gain is a useful peripheral blood tumor biomarker from lung cancer detection.

Key words: chromosome 7, biomarker, lung cancer

Lung cancer is the leading cause of cancer-related death worldwide. Although there are improvements in lung cancer diagnosis and treatment in recent years, the 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 is diagnosed. A large prospective clinical trial of applying low-dose CT (LDCT) scan for early lung cancer discovery has shown great survival benefits. However, LDCT scan results in huge false positives [2]. About 94% of LDCT identified nodules are actually benign changes, which do not need further treatments. Usually, invasive diagnosis approaches will be applied to those patients. Hence, there is a need to identify non-invasive approaches to complement LDCT to reduce the over-diagnosis.

In recent years, chromosome aneuploidy detection by using plasma cell-free DNA (cfDNA) for prenatal tests, with minimal false positives and false negatives [3]. Similar to fetal tissues, tumors also keep shedding DNA into the peripheral bloodstream. 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 [4], hepatocellular carcinoma [5], and lung cancer [6]. 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 can be detected in early lung cancer.

Chromosome 7 aneuploidy was discovered as an early event in 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 healthy tissues, which support chromosome 7 copy number changes might be an early event of lung cancer tumorigenesis [7–9]. However, a tissue biopsy is an invasive and complicated procedure. And sometimes, it might cause tumor metastasis.

Patients and methods

Subjects. Eighteen lung cancer patients and eighteen benign lung tumor patients were admitted to the Department of Thoracic Surgery of the First Affiliated Hospital of Soochow University. All patients were chemotherapy-naive or radiotherapy-naive. Blood samples were collected for cfDNA extracting and CA125 measuring before surgery. 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. The Institutional Review Board (IRB) of Soochow University approved the study. All recruited subjects have signed written informed consent.

Table 1. Clinicopathological information.

	Ad n=12	Scc n=3	Benign n=3	Healthy n=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				
≥55 years	6	2	1	7
<55 years	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.

Notes: TM-tumor biomarker; Ad-Adenocarcinoma; Scc-Squamous cell carcinoma; WT-wildtype; del-deletion; NA-not available; Tis-carcinoma *in situ*, T1aN0M0-tumor ≤1cm, without lymph node or distant metastasis; T1b-1cm < tumor ≤2cm, T2a-3cm < tumor ≤5 cm

Next-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. DNA was fragmented into an average size of 300 bp (cfDNA without fragmentation), and then 100 ng of fragmented genomic DNA (cfDNA 10 ng) was used for the preparation of sequencing libraries (NEBnext Ultra II). 8 bp barcoded sequencing adaptors were then ligated with DNA fragments and amplified by PCR. Purified sequencing libraries were massively parallel sequenced by the 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 analyze copy number changes. A p-value <0.05 was considered as statistically significant binary segmentation. The 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 χ^2 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 were informed with signed consent. DNA was successfully extracted from all of these samples with DNA concentrations ranging from 0.19 to 0.49 ng/μl. The sequencing library was prepared with standard NEB protocol. It was then sent to Illumina X10 for sequencing. In average, 10 G data was collected for each sample.

Patient characterization. As shown in Table 1, 15 cancer patient samples, 3 benign lesions, and 18 healthy controls were collected. UCAD detection was performed in 12 adenocarcinomas and 3 squamous cell carcinomas before the surgery. Tumors show a significantly higher UCAD chr7 score compared to the control group. And advanced tumors show a higher UCAD chr7 score compared to stage T1abN0M0 patients. The 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.

Cell-free DNA whole-genome copy number profiling. The reads were mapped to the human reference genome hg19. Genomic coverage was then counted by using software Samtools Mpileup. We then calculate average coverage for each 200 K bin. A circular binary segmentation algorithm was then used to detect significant genomic breakpoints.

As shown in Figures 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 the long arm, indicating short

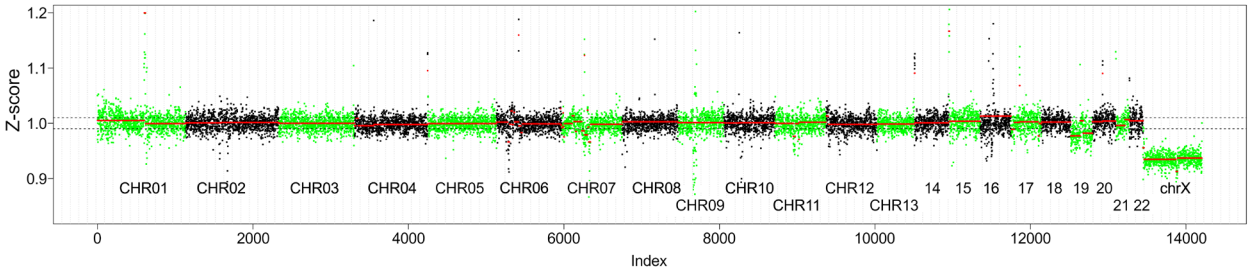

Figure 1. Chromosome 1 to 22 is a layout from left to right with green and black colors. Chromosomal segments are marked in red lines. Each dot indicates the normalized coverage value of a 200K bin.

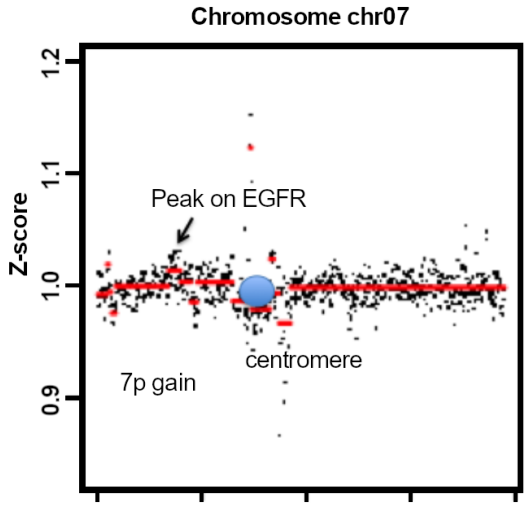

Figure 2. Zoom out view of chromosome 7, a breakpoint is found on chromosome 7 centromere (blue circle). Chr7p coverage is significantly higher than chr7q, which peak coverage around oncogene EGFR.

arm copy gains. Relatively much lower chromosomal imbalance was found in non-tumor controls.

UCAD chr7p gains in cancer patients of different clinical stages. A parameter set, alpha = 0.01, was used to detect genomic breakpoints in the DNA copy 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 ≥1.0005 and 2) Student t-test p-value <0.05. As shown in Table 2, 10 of 15 (66.7%) samples reached the significance, including 5/7 (71.43%) T1b/T2a, 4/6 (66.67%) T1a patients, and 1/2 (50%) Tis patients. All 3 benign controls did not reach the significance. The specificity was validated on another independent 'body-examination' volunteer cohort. 1 individual out of 18 (5.56%) was found to be chr7p positive, who need further follow-up.

UCAD chr7p+ is an independent predictor of lung malignancy. 13 serum tumor biomarkers were 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 were found with CYFRA21-1 mild increase. 1 of them (8.33%) was found with a mild CEA increase. The other 7 were negative for all the 13 tumor biomarkers. As shown in Table 3, UCAD chr7p gain is an independent predictor of lung malignancy.

Chr7p gain negatively correlated with EGFR/KRAS mutation status. EGFR and KRAS mutations were examined after the surgery. The 21 exon L858R and 19Del mutations were selected for EGFR detection because they account for the majority of EGFR mutations (87.85%), while KRAS mutations were realized through the widely studied exon 13 (G13). 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 a trend of negative correlation between UCAD+ and EGFR/KRAS mutations (Fisher test, p=0.067).

A model of plasma cell-free DNA UCAD chr7p and tumor biomarker to predict early-stage cancer. By combining 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.

Discussions

All 18 cancer samples were found with UCAD chr7 score >0.0020. However, 3 of them did not reach the statistical

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

Notes: Tis-carcinoma in situ, T1aN0M0-tumor ≤ 1cm, without lymph node or distant metastasis; T1b-1cm < tumor ≤ 2cm, T2a-3cm < tumor ≤ 5 cm

Table 3. UCAD chr7p+ is independent predictor of lung malignancy.

	CYFRA21-1 and/or CEA	
	TM+	TM–
UCAD chr7p+	3	7
UCAD chr7p-	2	5
	χ^2 test, p-value = 1.00	

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	

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

	UCAD chr7p+ TM+	UCAD chr7p+ TM–	UCAD chr7p– TM+	UCAD chr7p– TM–
Count	3	7	2	3
%	20.00	46.67	13.33	20.00

significance. The potential reason could be the coverage variations across the 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 the Illumina sequencer, GC bias has been considered to be the major cause of coverage bias. In addition to GC content, the secondary structure of cfDNA fragments might be another major factor contributing to coverage bias. The size of the cfDNA fragment is around 170 base pairs, which enables a complicated secondary structure. Hence, PCR reaction conditions need further be optimized for chromosome 7 copy number detections.

Copy number variations detected in the non-cancer patient. There are minor chromosome changes, which were found in non-cancer patients, which might indicate the background copy number variations among non-cancer individuals. In a 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 needed.

EGFR, KRAS, and other gene mutations are often found in patients with lung adenocarcinoma, and relevant targeted therapy is feasible for gene mutations [10, 11]. In our study, we found that chr7p amplification could be an independent predictor of lung cancer, but whether this is related to different pathological types of lung cancer has not been studied. Therefore, chr7p amplification was negatively correlated with EGFR and KRAS mutations, which may be related to the inclusion of 3 squamous cell carcinoma patients in the study. Whether CHR7 amplification can predict lung cancer of different pathological types in the future remains to be further studied. We also compared the UCAD chr7 score to traditional lung cancer tumor biomarkers, such as CEA, SCC and etc. Traditional tumor biomarker is rarely positive in the current dataset. In comparison, all tumor patients showed 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.

Chromosomal abnormalities have been widely studied for tumor diagnosis, such as chr1 and thyroid cancer, CHR9 and liver cancer, etc., but the association between CHR7 and lung cancer has not been found [12, 13]. In this research, the discovery was found in a small-scale prospective study, including 8 cancer patients and 4 healthy controls. Although the preliminary data was encouraging, we still need a large prospective clinical trial to further confirm the discovery.

Acknowledgments: 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 2008; 359: 1367–1380. https://doi.org/10.1056/NEJMra0802714

[2] NATIONAL LUNG SCREENING TRIAL RESEARCH TEAM, ABERLE DR, ADAMS AM, BERG CD, BLACK WC et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. N Engl J Med 2011; 365: 395–409. https://doi.org/10.1056/NEJMoa1102873

[3] NORTON ME, JACOBSSON B, SWAMY GK, LAURENT LC, RANZINI AC et al. Cell-free DNA Analysis for Noninvasive Examination of Trisomy. N Engl J Med 2015; 373: 2582. https://doi.org/10.1056/NEJMc1509344

[4] STOVER DG, PARSONS HA, HA G, FREEMAN SS, BARRY WT 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 2018; 36: 543–553. https://doi.org/10.1200/JCO.2017.76.0033

[5] JIANG P, CHAN CWM, CHAN KCA, CHENG SH, WONG J et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. Proc Natl Acad Sci U S A 2015; 112: E1317–1325. https://doi.org/10.1073/pnas.1500076112

[6] XIA S, HUANG CC, LE M, DITTMAR R, DU M et al. Genomic variations in plasma cell free DNA differentiate early stage lung cancers from normal controls. Lung Cancer 2015; 90: 78–84. https://doi.org/10.1016/j.lungcan.2015.07.002

[7] BEROUKHIM R, MERMEL CH, PORTER D, WEI G, RAYCHAUDHURI S et al. The landscape of somatic copy-number alteration across human cancers. Nature 2010; 463: 899–905. https://doi.org/10.1038/nature08822

[8] ZOJER N, DEKAN G, ACKERMANN J, FIEGL M, KAUFMANN H 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 2000; 28: 225–235. https://doi.org/10.1016/s0169-5002(00)00097-0

[9] RICKE RM, VAN DEURSEN JM. Aneuploidy in health, disease, and aging. J Cell Biol 2013; 201: 11–21. https://doi.org/10.1083/jcb.201301061

[10] HA SY, CHOI SJ, CHO JH, CHOI HJ, LEE J et al. Lung cancer in never-smoker Asian females is driven by oncogenic mutations, most often involving EGFR. Oncotarget 2015; 6: 5465–5474. https://doi.org/10.18632/oncotarget.2925

[11] EL OSTA B, BEHERA M, KIM S, BERRY LD, SICA G et al. Characteristics and Outcomes of Patients With Metastatic KRAS-Mutant Lung Adenocarcinomas: The Lung Cancer Mutation Consortium Experience. J Thorac Oncol 2019; 14: 876–889. https://doi.org/10.1016/j.jtho.2019.01.020

[12] JEONG S, LEE J, KIM D, LEE WK, JEONG JJ et al. Relationship of Focally Amplified Long Noncoding on Chromosome 1 (FAL1) lncRNA with E2F Transcription Factors in Thyroid Cancer. Medicine (Baltimore) 2016; 95: e2592. https://doi.org/10.1097/MD.0000000000002592

[13] ZUCMAN-ROSSI J, VILLANUEVA A, NAULT JC, LLOVET JM. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. Gastroenterology 2015; 149: 1226–1239. e4. https://doi.org/10.1053/j.gastro.2015.05.061