

Targeted next-generation sequencing reveals recurrence-associated genomic alterations in early-stage non-small cell lung cancer

William C.S. Cho¹, Kien Thiam Tan², Victor W.S. Ma¹, Jacky Y.C. Li¹, Roger K.C. Ngan³, Wah Cheuk⁴, Timothy T.C. Yip⁵, Yi-Ting Yang² and Shu-Jen Chen²

¹Department of Clinical Oncology, Queen Elizabeth Hospital, Kowloon, Hong Kong

²ACT Genomics, Co. Ltd., Taipei, Taiwan

³Department of Clinical Oncology, The University of Hong Kong, Gleneagles Hong Kong Hospital, Wong Chuk Hang, Hong Kong

⁴Department of Pathology, Queen Elizabeth Hospital, Kowloon, Hong Kong

⁵ACT Genomics, Co. Ltd., Kowloon, Hong Kong

Correspondence to: William C.S. Cho, **email:** chocs@ha.org.hk or williamcscho@gmail.com
Yi-Ting Yang, **email:** yitingyang@actgenomics.com
Shu-Jen Chen, **email:** sjchen@actgenomics.com

Keywords: biomarker; early-stage; lung cancer; next-generation sequencing; relapse

Received: August 20, 2018

Accepted: November 01, 2018

Published: November 20, 2018

Copyright: Cho et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Purpose: The identification of genomic alterations related to recurrence in early-stage non-small cell lung cancer (NSCLC) patients may help better stratify high-risk individuals and guide treatment strategies. This study aimed to identify the molecular biomarkers of recurrence in early-stage NSCLC.

Results: Of the 42 tumors evaluable for genomic alterations, TP53 and EGFR were the most frequent alterations with population frequency 52.4% and 50.0%, respectively. Fusion genes were detected in four patients, which had lower mutational burden and relatively better genomic stability. EGFR mutation and fusion gene were mutually exclusive in this study. CDKN2A, FAS, SUFU and SMARCA4 genomic alterations were only observed in the relapsed patients. Increased copy number alteration index was observed in early relapsed patients. Among these genomic alterations, early-stage NSCLCs harboring CDKN2A, FAS, SUFU and SMARCA4 genomic alterations were found to be significantly associated with recurrence. Some of these new findings were validated using The Cancer Genome Atlas (TCGA) dataset.

Conclusions: The genomic alterations of CDKN2A, FAS, SUFU and SMARCA4 in early-stage NSCLC are found to be associated with recurrence, but confirmation in a larger independent cohort is required to define the clinical impact.

Materials and Methods: Paired primary tumor and normal lung tissue samples were collected for targeted next-generation sequencing analysis. A panel targets exons for 440 genes was used to assess the mutational and copy number status of selected genes in three clinically relevant groups of stage I/II NSCLC patients: 1) Early relapse; 2) Late relapse; and 3) No relapse.

INTRODUCTION

Lung cancer is the leading cause of cancer-related mortality in the world, and non-small cell lung cancer (NSCLC) accounts for 80–85% of all lung cancers [1]. Current treatment for NSCLC is mainly based on the clinical staging systems in addition to some factors, such as smoking history, gender and genetic mutation [2, 3]. The 5-year survival rate of NSCLC patients is about 30–60%, with local recurrence and metastasis being the leading causes of death [4]. Stage I/II NSCLC patients are typically treated with complete surgical resection of the tumor, over half of the patients with early-stage NSCLC will never recur after surgical resections. However, individual clinical outcome varies widely. Even after the entire resection of tumor, 30–55% of patients will develop disease recurrence within the first five years of surgery and ultimately die of the disease [4–8]. Unfortunately, the molecular mechanisms underlying recurrence remain unclear. Identification of patients with high risk of recurrence may help to guide more tailored treatment plan for the individual patient, in particular for those who may receive additional benefit from adjuvant therapy and targeted therapy.

In recent years, high-throughput genomic analysis has been used to obtain genomic information of NSCLC. With the advancement of next-generation sequencing (NGS), it is now possible to identify the oncogenic alterations that will previously be missed by conventional pathological diagnosis. Rather than sequencing the entire genome or exome, a clinical cancer gene test (which includes a panel of genes that show frequent mutations in cancer) can reduce the amount of specimen, time and effort required to perform sequencing. These tests usually use a polymerase chain reaction capture-based or amplicon-based NGS assay for the deep targeted sequencing of the genes of interest in limited biological specimens [9–11]. Our study performed this kind of targeted NGS to identify the genomic alterations associated with recurrence in early-stage NSCLC patients. The identification of these genomic alterations may help to distinguish patients at higher risk and guide treatment strategies for early-stage NSCLC.

RESULTS

We identified 45 pathologically confirmed stage I/II NSCLC patients with available formalin-fixed paraffin-embedded (FFPE) tissues. Of these, three tumor specimens were excluded due to the insufficiency of tumor cells or tumor purity was lower than acceptance criteria. 42 patients were included in NGS analysis. At the time of analysis, 16 patients were reported having early relapse (ER), 15 patients having late relapse (LR) and 11 patients having no relapse (NR). The general characteristics of the study population are summarized in Table 1. The median follow-up time of NR patients from the date of surgery until the last follow-up or the end of the study was 50.4 months.

Genomic alterations

Genomic alterations were identified in 324 (73.6%) of 440 targeted genes, including 180 (40.9%) of single nucleotide variants (SNVs), 193 (43.9%) of gene deletions and 92 (20.9%) of gene amplifications (Supplementary Table 1). The genomic alterations with the highest mutated frequencies were shown in Figure 1. Gene deletion was more common event than gene amplification in this study (Figure 2, copy number variant (CNV) index). *TP53* and *EGFR* were the most frequent alterations detected in this study, with a genomic alteration frequency of 52.4% and 50%, respectively (Figure 2). *EGFR* exon 19 in-frame deletion was the most common *EGFR* mutation (11 cases), followed by activating mutation L858R in exon 21 (5 cases), wild-type *EGFR* amplification (2 cases), and exon 20 insertion (1 case). Other *EGFR* mutations were also detected, including K860I, A871G, L861Q, G719S, V592I and T302H. Interestingly, two coexisting *EGFR* alterations were found in 4 ER patients (D00807, D00825, D01207 and D01303). The L858R mutation occurred in *cis* with A871G and K860I, whereas the allelic relationship of G719S and L861Q mutations were unknown (D01303) due to the limited read length of NGS platform. No T790M point mutation was observed (Table 2).

KRAS G12C, amplification and G12V were detected in 3 relapsed patients. *ERBB2* V659E, E507K and exon 20 insertion (A775_G776insYVMA) were also found in 3 relapsed patients. *ERBB2* mutations were mutated exclusively from *EGFR*. *PIK3CA* genomic alterations were only found in ER patients, including mutation type I459V and H1047R, and also amplification. *BRAF* mutation was identified in 4 cases, including mutation type K601E, G219A and G469A, as well as amplification and deletion.

Fusion genes (*SDC4-ROS1*, *EML4-ALK*, *SDC4-ROS1*, *EZR-ROS1*) were detected in 4 patients, whose tumors had lower mutation burden and demonstrated relatively better genome stability with extremely low copy number alteration (CNA) index (Figure 2). *EGFR* mutation and fusion gene were mutually exclusive in this study.

Genomic alterations absent in the NR patients

CDKN2A, *FAS* and *SUFU* were the most frequently altered genes detected in the relapsed patients, with a population frequency of 31.0% (13 cases), 16.7% (7 cases) and 14.3% (6 cases), respectively. A total of five loss-of-function (LOF) mutations (E61*, R80*, R58*, E61 and P114L) and one variant of unknown significance (VUS) in the splice region of *CDKN2A* were detected in the relapsed patients, whereas the deletions of *CDKN2A* were found in seven relapsed patients. A *FAS* deletion was detected in 7 relapsed (16.7%) cases, and no nucleotide variant was detected. *SUFU* genomic alterations were found in 6 relapsed patients (14.2%), including 5 deletions and a VUS missense mutation (D159E). *SMARCA4* variants were detected in 4 relapsed (9.5%) cases, including a VUS (G1159L) and deletions (Figure 2).

Table 1: Patient characteristics

Characteristics	Number	%
Patients (n)	42	
Sex		
Male	26	61.9
Female	16	38.1
Age, median (range, year)	65	(46–85)
Stage		
T1	38	90.5
T2	4	9.5
Smoking status		
Non-smoker	19	45.2
Ex-smoker	17	40.5
Current smoker	5	11.9
Not mention	1	2.4
Recurrence		
Early relapse	16	38.1
Late relapse	15	35.7
No relapse	11	26.2

Tumor mutational burden (TMB) and CNA index

The median TMB was 3.86 mutations per megabase (Mb), with a range of 0.86–46.31 mutations/Mb in this subset of patients (Table 3). The average values of TMB were 9.33, 8.98 and 6.00 mutations/Mb in ER, LR, and NR patients, respectively. The median of TMB were 7.29, 3.43 and 3.43 mutations/Mb in ER, LR and NR patients, respectively. The range of CNA index was between 0–22.79% and the average index values were 6.81%, 5.89% and 3.34% in the ER, LR and NR patients, respectively. The TMB and CNA index were not significantly associated with relapse status, while both TMB and CNA indices were increased in the ER patients (Figure 3A and 3B). Furthermore, TMB was highly associated with smoking status ($^{***}p = 0.0004$), and borderline significance was associated with *EGFR* alterations ($p = 0.0564$) (Figure 3C and 3D). Four *EGFR* mutated patients with high TMB were all smokers. Most of the high TMB patients were male smokers in this study. Several genes were found to cluster with smoking status and TMB, including *TP53*, *USH2A*, *LRP1B*, *MUC16* and *SYNE1* (Figure 4).

Genomic alterations associated with the risk of recurrence

Further analysis found that the mutations of *CDKN2A*, *FAS*, *SUFU* and *SMARCA4* were significantly associated with an increased risk of recurrence in early-stage NSCLC (Figure 5). The median time to recurrence was 14.5 months in the *CDKN2A*-mutated patients and 55.0 months in the *CDKN2A* normal patients ($p = 0.0134$).

The median time to recurrence was 10.0 months in the *FAS* loss patients and 35.0 months in the *FAS* normal patients ($p = 0.0138$). The median time to recurrence was 9.0 months in the *SUFU*-mutated patients and 28.5 months in the *SUFU* normal patients ($p = 0.0210$). The median time to recurrence was 4.5 months in the *SMARCA4*-mutated patients and 28.5 months in the *SMARCA4* normal patients ($p = 0.0007$). In order to evaluate if the genomic alterations identified in *CDKN2A*, *FAS*, *SUFU* and *SMARCA4* genes are significantly associated with early relapse, multivariate Cox regression model was tested. Our results revealed that the genomic alterations of *CDKN2A*, *FAS*, *SUFU* and *SMARCA4* are independent risk factors significantly associated with the risk of recurrence regardless of age, gender, stage and smoking status (Table 4).

Validation of findings in The Cancer Genome Atlas (TCGA) dataset

We aimed to validate our findings in a publically available TCGA dataset. In the TCGA dataset (TCGA, Provisional), 177 out of 283 (62.5%) stage I/II NSCLC patients showed *CDKN2A* mutation, and the median relapse-free survival (RFS) was 37.65 months in the *CDKN2A*-altered patients and 38.37 months in the *CDKN2A* normal patients ($p = 0.5676$). We also analyzed the TCGA dataset for the presence of *FAS*, *SUFU* and *SMARCA4* alterations in recurrent NSCLC. A total of 95 out of 268 (35.4%) showed a deletion of *FAS*, 97 out of 269 (36.1%) showed an alteration of *SUFU*, and 161 out of 300 (53.7%) showed a mutation of *SMARCA4*, with Gehan-Breslow-Wilcoxon test p value of 0.0314, 0.0221 and 0.5265, respectively. RFS was shorter in patients with genomic alterations in *CDKN2A*, *FAS*, *SUFU*

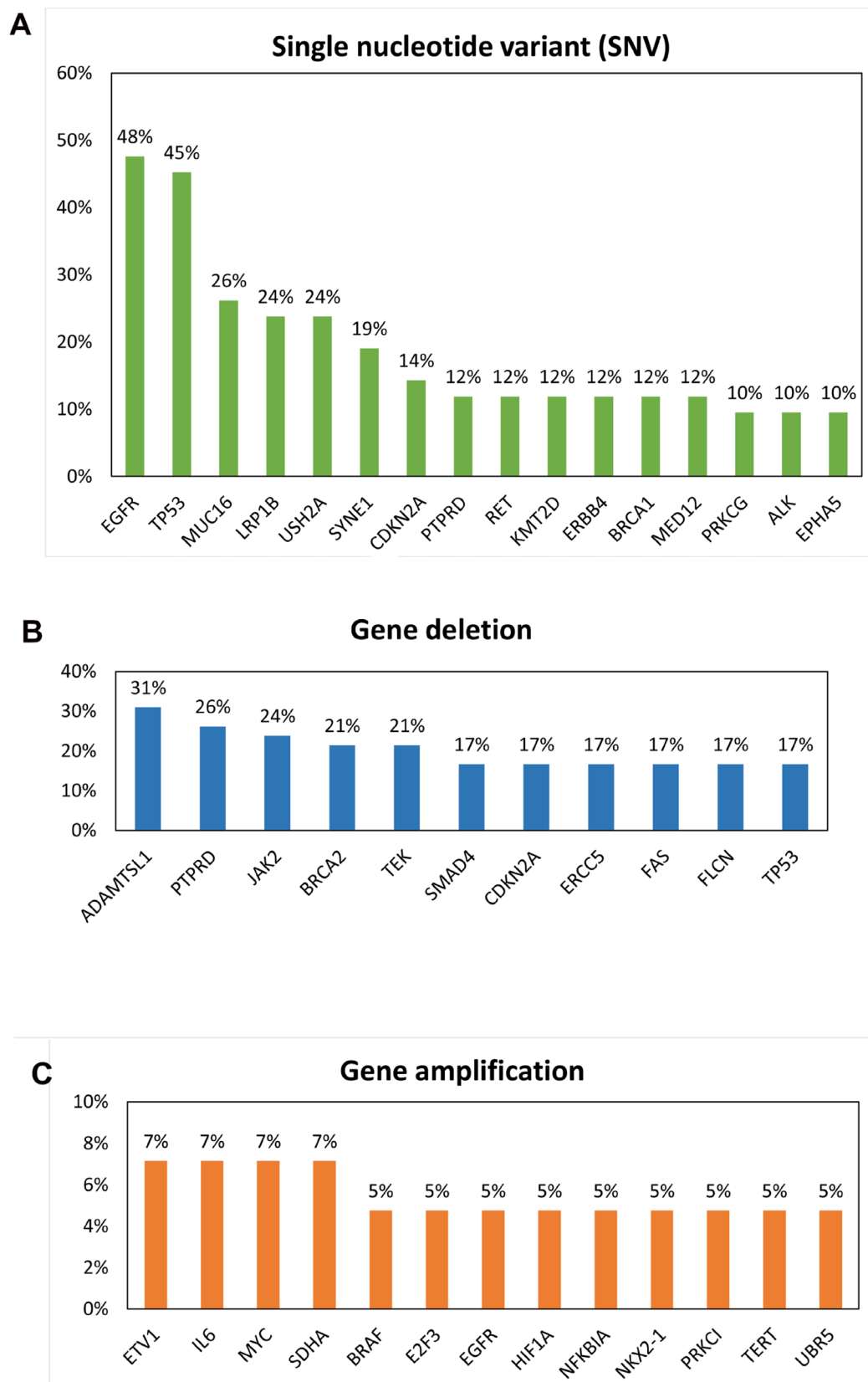


Figure 1: Frequency of genomic alterations in each gene. Top genes of (A) single nucleotide variant, copy number alteration of (B) deletion and (C) amplification are sorted by the total frequency of alteration events on the x-axis.

and *SMARCA4* genes (Figure 6). Using Cox regression model for multivariate analysis, we found that only *FAS* and *SUFU* mutants were significantly associated with the risk of recurrence (Table 4).

DISCUSSION

Local recurrence or metastasis is generally believed to account for the failure of therapy and recurrence after initial surgical resection. Although early-stage NSCLC patients have better prognosis, nearly 30–35% of them will relapse [12–14]. Owing to the fact that NSCLC has heterogeneous histopathological features, conventional classification systems (such as tumor-nodes-metastasis staging) cannot fully explain its clinical behaviors [15]. Regarding the unsatisfactory survival rates, many researchers try to look for possible methods that can help to predict the outcome of early-stage patients. The tumor size of the invasive component has been reported to be an important prognostic factor for early-stage adenocarcinoma [16, 17]. In addition to low-dose computed tomography, many biomarkers (especially molecular biomarkers) have been developed to supplement clinical diagnosis, such as carcinoembryonic antigen level [18–20], epigenetic modifications [21, 22], gene-expression profiles [23–25] and the detection of genomic alterations [26–29]. However, few studies have explored the genomic alterations of early-

stage NSCLC patients [20, 26, 30–32]. Contemporary lung cancer research has distinguished itself from traditional study with an unprecedented large amount of data and tremendous diagnostic and therapeutic innovations. Data are currently generated in high-throughput fashions with the integration and application of sequencing analysis. The application of NGS on mutational analysis gives us a more comprehensive genomic landscape in NSCLC. If we can identify early-stage NSCLC patient with a high risk of recurrence, we are able to select a more appropriate candidate treatment (such as adjuvant therapy or targeted therapy). The ability to identify patients with high-risk of early recurrence following surgical resection can also help formulating more intensive surveillance schedule, performing more extensive surgery including lymph node dissection, instituting adjuvant postoperative chemotherapy and other adjuvant personalized or targeted therapy for these patients [12, 33].

In this study, we aimed to investigate the impact of genomic alterations on the recurrence of early-stage NSCLC. Oncogenic mutations were detected more frequently in the ER patients than in the LR and NR patients. Two coexisting *EGFR* mutations representing two oncogenic drivers were found in some ER patients, but not in the LR and NR patients. Some genomic alterations were absent in the NR patients, whereas some mutations were found only in the ER patients. These findings add

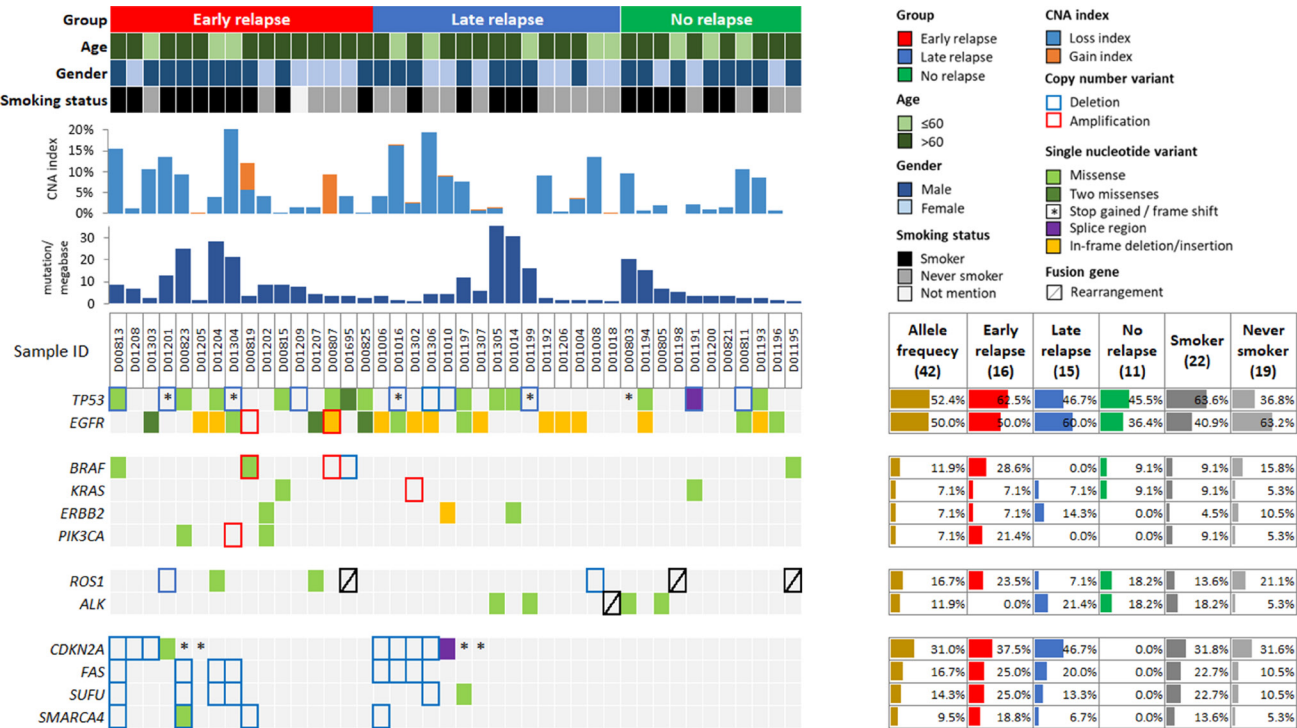


Figure 2: Genomic alterations identified in early-stage non-small cell lung cancer. Tumor samples of early relapse, late relapse, and no relapse are arranged from left to right. Patient information is displayed on the top panel, including group, age, gender, and smoking status. Tumor mutational number and copy number alteration (CNA) index are shown below the patient information. Genomic alterations are annotated according to the color panel on the top right corner of the image. Mutation rates of each gene are plotted on the right of the genomic alterations.

Table 2: EGFR alterations in the patients

Sample ID	Genomic alteration	Gender	Age	Smoking status	Stage	Relapse
D00807	Amplification, 767_V769dup	F	61	Non-smoker	IA (T ₁ N ₀)	Early
D00819	Amplification	M	69	Ex-smoker	IB (T _{2a} N ₀)	Early
D00825	L858R, K860I	M	79	Ex-smoker	IB (T _{2a} N ₀)	Early
D01204	E746_A750del	M	43	Ex-smoker	IB (T _{2a} N ₀)	Early
D01205	L747_A750delinsP	M	62	Current smoker	IA (pT _{1a} N ₀)	Early
D01207	L858R, A871G	F	64	Non-smoker	IIA (pT _{2a} N ₁)	Early
D01303	G719S, L861Q	M	47	Non-smoker	IB (pT _{2a} N ₀)	Early
D01304	V592I	M	57	Ex-smoker	IB (T _{2a} N ₀)	Early
D01004	E746_A750del	M	77	Non-smoker	IB (T _{2a} N ₀)	Late
D01006	E746_A750del	M	74	Non-smoker	IIB (T ₂ N ₁)	Late
D01016	L858R	M	52	Non-smoker	IIB (T ₂ N ₁)	Late
D01192	E746_A750del	F	70	Non-smoker	IA (T ₁ N ₀)	Late
D01197	T302H	M	63	Ex-smoker	IA (T _{1b} N ₀)	Late
D01206	E746_A750del	F	62	Non-smoker	IA (pT _{1b} N ₀)	Late
D01302	E746_A750del	M	65	Ex-smoker	IA (pT _{1b} N ₀)	Late
D01306	L747_P753delinsS	F	61	Non-smoker	IB (pT ₂ N ₀)	Late
D01307	E746_A750del	F	71	Non-smoker	IA (pT _{1b} N ₀)	Late
D00811	L858R	M	58	Non-smoker	IB (T _{2a} N ₀)	No
D01193	E746_A750del	M	81	Ex-smoker	IA (pT _{1b})	No
D01194	L747_T751del	M	65	Current smoker	IA (T ₁ N ₀)	No
D01196	L858R	F	76	Non-smoker	IA (T _{1b} N ₀)	No

Table 3: Distribution of tumor mutational burden (TMB)

Mutation/Megabase	All patients	%	Early relapse	%	Late relapse	%	No relapse	%
0.86	3	7.1	0	0	2	4.8	1	2.4
1.72	6	14.3	1	2.4	4	9.5	1	2.4
2.57	5	11.9	2	4.8	1	2.4	2	4.8
3.43	7	16.7	3	7.1	1	2.4	3	7.1
4.29	3	7.1	1	2.4	2	4.8	0	0
5.15	1	2.4	0	0	0	0	1	2.4
6.00	1	2.4	0	0	1	2.4	0	0
6.86	2	4.8	1	2.4	0	0	1	2.4
7.72	1	2.4	1	2.4	0	0	0	0
8.58	3	7.1	3	7.1	0	0	0	0
12.01	1	2.4	0	0	1	2.4	0	0
12.86	1	2.4	1	2.4	0	0	0	0
15.44	1	2.4	0	0	0	0	1	2.4
16.30	1	2.4	0	0	1	2.4	0	0
20.58	1	2.4	0	0	0	0	1	2.4
21.44	1	2.4	1	2.4	0	0	0	0
24.87	1	2.4	1	2.4	0	0	0	0
28.30	1	2.4	1	2.4	0	0	0	0
30.87	1	2.4	0	0	1	2.4	0	0
46.31	1	2.4	0	0	1	2.4	0	0
Total patients	42	100	16	38.1	15	35.7	11	26.2

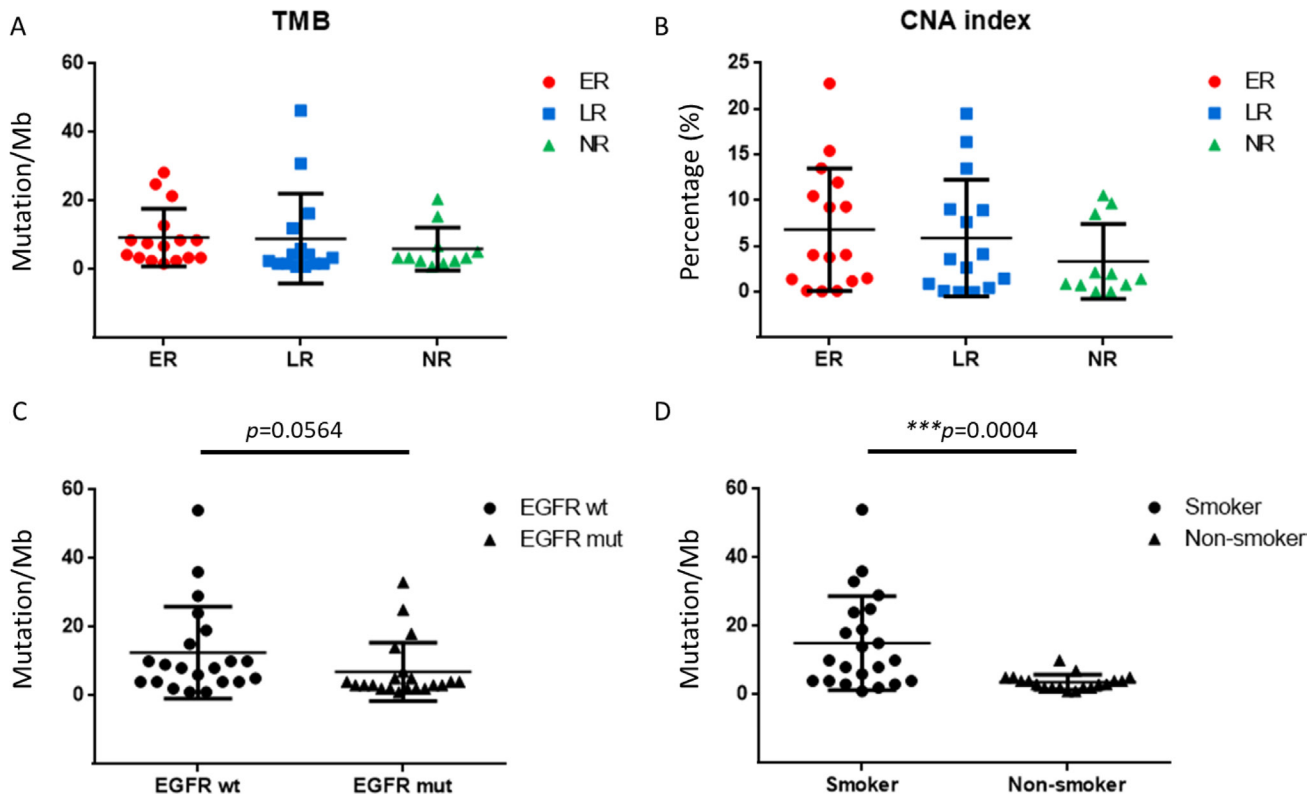


Figure 3: Tumor mutational burden (TMB) was highly correlated with smoking status. (A) TMB did not show significant differences among early relapse (ER), late relapse (LR), and no relapse (NR) patients. (B) Copy number alteration (CNA) index was not correlated with relapse status, whereas CNA index was increased in ER patients. (C) TMB is of borderline significance with *EGFR* mutations ($p = 0.0564$). (D) TMB was highly associated with smoking status ($***p = 0.0004$).

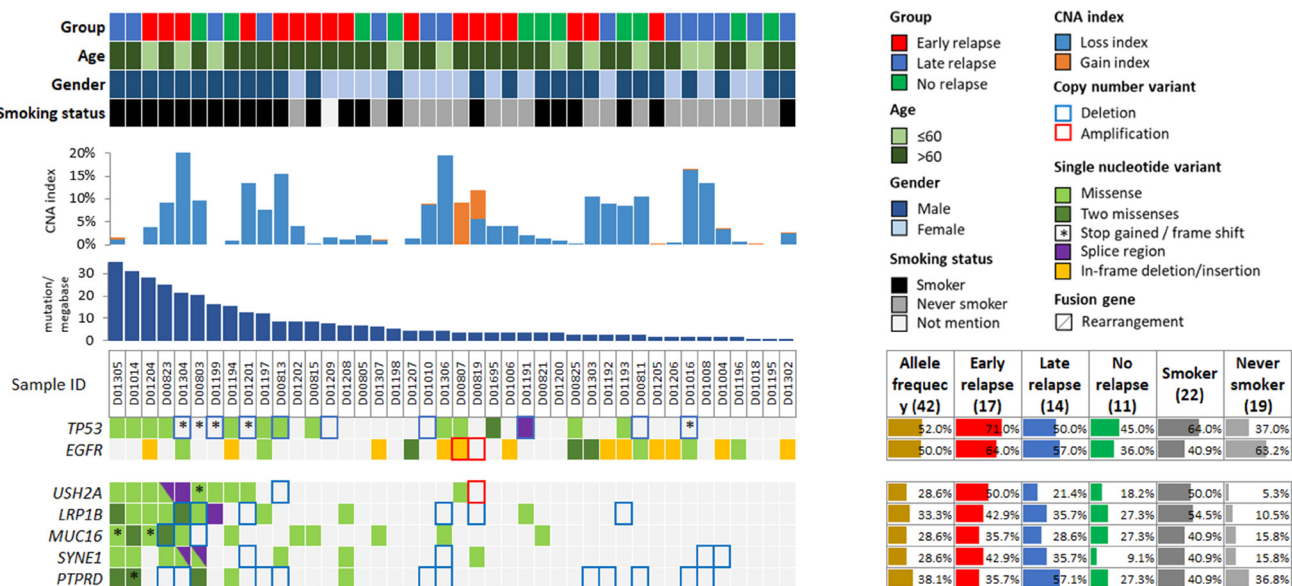


Figure 4: Genomic alterations identified in high tumor mutational burden (TMB) patients. Tumor samples of high TMB are arranged from left to right. Patient information is displayed on the top panel, including group, age, gender, and smoking status. TMB and copy number alteration (CNA) index are shown below the patient information. Genomic alterations are annotated according to the color panel on the top right corner of the image. Alteration rates of each gene are plotted on the right of the genomic alterations.

to the growing evidence that genomic alterations may be a significant factor for recurrence in early-stage NSCLC. Several rare *EGFR* mutations were detected in the extracellular domain (V592I and T302H) and tyrosine kinase domain (A871G, K860I, G719S, and L861Q), suggesting NGS is a promising and comprehensive platform to detect multiple alterations simultaneously. Our results have provided a genome-wide view on the genomic landscape of early-stage NSCLC, including TMB and CNA burden, which allows us to correlate genomic profiles to the risk of recurrence. Consistent with a previous study [34], we observed genomic alterations in several known prognostic biomarkers. In addition, we identified some uncommon but potentially actionable mutations in early-stage NSCLC.

This study is unique in that it identified 4 (*CDKN2A*, *FAS*, *SUFU*, and *SMARCA4*) potential genomic biomarkers of recurrence in early-stage NSCLC. We found *CDKN2A* to be the most frequently altered genes detected in the relapsed patients. Indeed, *CDKN2A* status was found to be associated with various cancers [35–40]. A recent study identified *CDKN2A* as a tumor suppressor whose inactivation promoted homotypic cell-in-cell formation in human cancer cells [41]. Our analysis also suggested that

CDKN2A might be a potential biomarker of recurrence in early-stage NSCLC.

FAS gene has been reported to be associated with tumor progression. The previous study has provided evidence that Fas was important for natural killer cell-mediated immune surveillance and chemosensitivity. Their model for Fas LOF in tumor progression showed that Fas and FasL interactions were important in the control of malignant disease and that changes in the level of Fas expression could determine immune escape and therapeutic responses [42]. We have identified *FAS* as a potential biomarker of recurrence, as our results showed that *FAS* mutation was significantly associated with recurrence in early-stage NSCLC.

SUFU (suppressor of fused) is an important negative regulator of the Hedgehog (HH) pathway [43]. Activation of HH pathway signaling has been reported in various cancers, including cancers of the lung, skin, colon and stomach, which is involved in cancer cell proliferation and metastasis [44]. Moreover, several studies have demonstrated that the HH pathway is activated in NSCLC [45–48]. However, the prognostic roles of *SUFU* in lung cancer have not been addressed. Further study to clarify whether *SUFU* is indeed involved

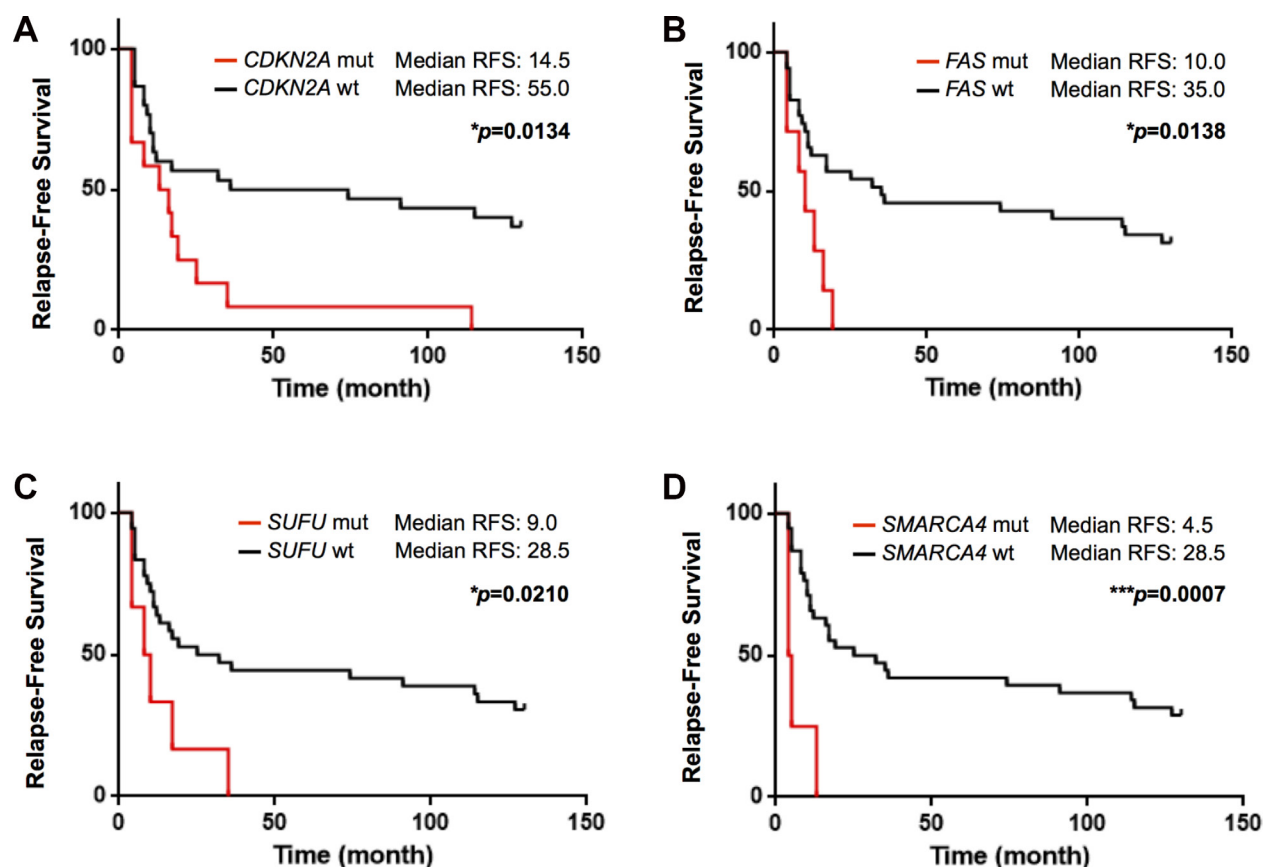


Figure 5: Kaplan–Meier curve of relapse-free survival according to mutational status in this study. The alterations of (A) *CDKN2A*, (B) *FAS*, (C) *SUFU* and (D) *SMARCA4* showed shorter median relapse-free survival and all the alterations were significantly associated with an increased risk of recurrence in early-stage NSCLC (Gehan-Breslow-Wilcoxon test). * $p < 0.05$, *** $p < 0.001$.

Table 4: Univariate and multivariate analysis of the risk factors associated with disease recurrence

Mutation	Total (%)	Median time to recurrence (months)	Univariate test	Multivariate test			
			<i>p</i> value ^a	This cohort	TCGA cohort		
				HR ^b (95% CI)	<i>p</i> value ^c	HR ^b (95% CI)	<i>p</i> value ^c
<i>CDKN2A</i>							
Wild-type	29 (69.0)	55.0	0.0134	1	0.011	1	0.931
Mutant	13 (31.0)	14.5		2.927 (1,278–6.702)		0.985 (0.701–1.384)	
<i>FAS</i>							
Wild-type	35 (83.3)	35.0	0.0138	1	0.010	1	0.033
Mutant	7 (16.7)	10.0		4.835 (1.462–15.992)		1.461 (1.031–2.070)	
<i>SUFU</i>							
Wild-type	38 (90.5)	28.5	0.0210	1	0.010	1	0.017
Mutant	4 (9.5)	9.0		3.505 (1.342–9.154)		1.528 (1.080–2.163)	
<i>SMARCA4</i>							
Wild-type	38 (90.5)	28.5	0.0007	1	0.009	1	0.681
Mutant	4 (9.5)	4.5		6.844 (1.630–28.726)		1.074 (0.764–1.509)	

^a Gehan-Breslow-Wilcoxon test.

^b Cox proportional hazards regression model based on age, gender, stage and smoking status.

^c *P* value for multivariate Cox proportional hazard regression.

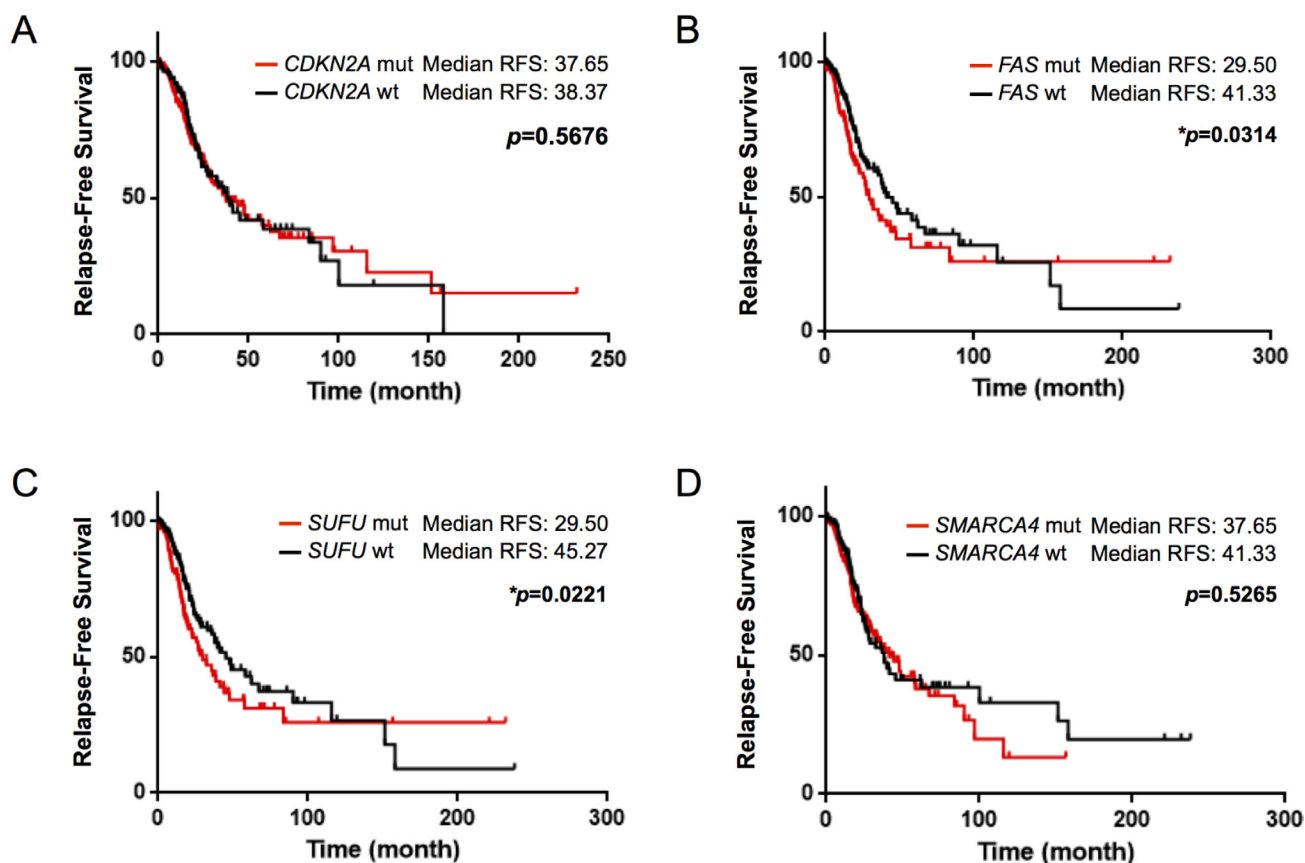


Figure 6: Kaplan-Meier curve of relapse-free survival according to mutational status in the TCGA cohort (TCGA, Provisional). The alterations of (A) *CDKN2A*, (B) *FAS*, (C) *SUFU* and (D) *SMARCA4* showed shorter median relapse-free survival. Only *FAS* and *SUFU* genes were significantly associated with an increased risk of recurrence in early-stage NSCLC (Gehan-Breslow-Wilcoxon test). **p* < 0.05.

in HH pathway activation or ER in lung cancer would be interesting.

Alteration in *SMARCA4* was detected only in the relapsed patients. Low expression of *SMARCA4* has been reported to be associated with worse prognosis and is supposed to be a predictive biomarker for increased sensitivity to platinum-based chemotherapy in NSCLC [49]. In this study, we found that the loss of *SMARCA4* was only detected in the relapsed patients and thus might serve as a potential biomarker for recurrence in early-stage NSCLC.

On the other hand, *USH2A*, *LRP1B*, *MUC16*, and *SYNE1* genes have been reported to have frequently mutated in various cancer types, particularly lung adenocarcinoma and lung squamous cell carcinoma [50, 51]. However, the functional roles of these genes in tumorigenesis remain unclear. We noted that the length of proteins coded by *USH2A*, *LRP1B*, *MUC16* and *SYNE1* were of 5202, 4599, 14507 and 8797 amino acids, respectively. A recent investigation has addressed that mutation rate was higher in the genes coding for proteins of longer length. For example, extremely long genes such as *TTN* have a high mutational frequency (52%) in lung squamous TCGA data [52]. Therefore, further study is required to clarify the biological meaning of these alterations.

There are a number of advantages of using targeted NGS panel for sequencing analysis [53]. It can reduce the amount of specimen, time, and effort required to perform deep targeted sequencing. However, it also had some limitations. Targeted NGS may have missed smaller region copy number alterations and other mutations in regions not covered by the panel. We tried to avoid this by using a panel consisted of as many as 440 genes that showed frequent mutations in cancer and thus could detect the genomic alterations of interest. The detection of genomic alterations in some known prognostic biomarkers confirmed the reliability of our panel.

In conclusion, our new findings revealed that genomic mutations, single nucleotide variant and CNV might play a role in the clinical outcomes of early-stage NSCLC patients. In addition, our results also indicated that the mutated genes might serve as potential biomarkers of recurrence. Notably, *CDKN2A*, *FAS*, *SUFU* and *SMARCA4* mutations were significantly associated with an increased risk of recurrence in early-stage NSCLC. Our results suggest that utilizing these genomic alterations to guide additional adjuvant therapies after surgery may improve outcomes in selected patients with high-risk of recurrent disease. Although some of our findings have been validated using TCGA dataset, confirmation in a larger independent cohort of the Asian population is warranted.

MATERIALS AND METHODS

Patient selection

We used a cohort of patients with pathologically confirmed stage I/II (AJCC 7th edition) NSCLC who underwent lobectomy and/or thoracotomy at Queen

Elizabeth Hospital (Hong Kong) between March 2004 and March 2015 (Supplementary Table 2). Patients with early-stage NSCLC who developed recurrence ≤ 1 year after treatment will be included in this study. In addition, the tissues of patients with early-stage NSCLC who did not develop recurrence ≤ 1 year after treatment will be included as a control. We defined ER as patients who relapsed within one year of treatment. LR were those patients who did not develop recurrence within one year of treatment. NR were those patients without evidence of recurrence after treatment on the date of last follow-up or before the end of the study.

Sample collection

The archival FFPE tissue samples of patients with pathologically confirmed stage I/II NSCLC were collected for NGS analysis. Paired primary tumor and normal lung tissue samples were included to identify somatic mutations. This study was approved by the Kowloon Central/Kowloon East Cluster Research Ethics Committee (Hospital Authority, Hong Kong).

Sample preparation

One 5 μ m hematoxylin and eosin stained slide and ten 10 μ m unstained slides were prepared. Samples needed to be > 25 mm² and tumor cell content $> 30\%$ tumor cells were considered eligible for targeted NGS analysis.

DNA and RNA extraction

Ten 10 μ m lung tissue sections were obtained from each patient. Genomic DNA and RNA were isolated from FFPE tissue sections with the RecoverAll™ Total Nucleic Acid Isolation Kit (Invitrogen, MA, USA) according to manufacturer's protocol. Extracted DNA and RNA concentration were measured by Qubit-iT™ dsDNA HS Assay Kit (Invitrogen) and Qubit RNA HS Assay Kit (Invitrogen), respectively. The integrity of DNA and RNA were assessed using the Fragment Analyzer (Advanced Analytical Technologies, Inc, IA, USA).

Next-generation sequencing of genomic DNA

Targeted deep NGS was used to assess the mutational status, single nucleotide variant, small insertions and deletions and copy number variant of 440 cancer-related genes (Supplementary Table 3) (ACT Genomics, Taiwan). Extracted genomic DNA from FFPE was amplified using 18,700 primer pairs targeting selected genes. Amplicons were constructed with barcoded libraries using the Ion AmpliSeq Library Kit (Life Technologies, MA, USA). Sequencing was performed on the Ion Proton sequencer (Life Technologies) according to the manufacturer's protocol. The mean sequencing depth was more than 700x, and the mean uniformity was more than 75%.

NGS data analysis

For the 440-gene panel, sequencing raw reads were mapped to the hg19 human reference genome using Torrent Suite Server version 5.2, base calling and variant calling were performed with the Torrent Suite Variant Caller plug-in version 5.2. The Ion Torrent default pipeline and parameters were used for data analysis. Variants reported in 1000 Genomes Project Phase 3 with > 1% minor allele frequency (Asian populations) were considered as polymorphisms and excluded from further analysis. Variants detected in 25 peripheral blood mononuclear cell (PBMC) in-house samples from healthy volunteers were also disregarded as SNPs or technical errors. Variants with coverage ≥ 25 , allele frequency $\geq 5\%$ were retained. CNV was analyzed using ONCOCNV (<https://github.com/BoevaLab/ONCOCNV>) [54]. The diploid reference baseline was established according to our in-house PBMC samples from healthy volunteers. ADTEX was applied for estimating tumor purity and correcting baseline shifts based on SNP information [55]. Copy number amplification was defined as an observed copy number ≥ 10 , whereas copy number loss was defined as an observed copy number ≤ 1 . Paired primary tumor and normal lung tissue samples were compared to identify tumor somatic mutations. TMB was calculated according to the number of detected mutations and the number of analyzed base pairs (1.166 Mb). CNA index was calculated by the percentage of the regions of genes altered in a tumor and the total regions of the genes that covered in the test on the chromosome to measure degree of genomic instability across the entire genome of a tumor. Statistical analysis of TMB and CNA index distribution with differences between each group was assessed as unpaired *t*-test using GraphPad Prism (v.6.0; GraphPad Inc., CA, USA).

Fusion gene test

Fusion transcripts for *ALK*, *ROS1*, *RET* and *NTRK* genes were tested for genetic rearrangement. Extracted RNA was reverse transcribed using SuperScript VILO cDNA Synthesis Kit (Invitrogen) according to the manufacturer's instructions. The library was constructed using the Ion AmpliSeq™ RNA Fusion Lung Cancer Research Panel (Life Technologies). Sequencing was performed on the Ion Proton sequencer (Life Technologies) according to the manufacturer's protocol. Raw reads were mapped to the targeted fusion transcripts using BWA (Burrows-Wheeler Aligner) software and using the in-house script to identify reads covered the fusion breakpoint.

Analysis of TCGA public data

Cases from TCGA were selected from the lung adenocarcinoma (TCGA, Provisional). A total of 313 stage I/II patients were selected (Supplementary Table 4). Alteration types of single nucleotide variant, heterozygous

and homozygous deletions were all included in a mutant group.

Statistical analysis

RFS was defined from the date of first surgery until tumor progression, death, the end of follow-up or the end of the study. Survival analysis was conducted to correlate genomic alterations with time to NSCLC relapse using the Kaplan–Meier curve and Gehan-Breslow-Wilcoxon test (v.6.0; GraphPad Inc.). Multivariate Cox regression modeling was performed using potential risk factors (age, gender, tumor stage, smoking and mutation status of genes of interest) by SPSS Version 23.0 (SPSS Inc., Chicago, IL, USA). All calculations were two-sided tests, with a *p* value < 0.05 considered as statistically significant.

Abbreviations

CNA: Copy number alteration; CNV: Copy number variant; ER: Early relapse; FFPE: Formalin-fixed paraffin-embedded; HH: Hedgehog; LOF: Loss-of-function; LR: Late relapse; Mb: Megabase; NGS: Next-generation sequencing; NR: No relapse; NSCLC: Non-small cell lung cancer; PBMC: Peripheral blood mononuclear cell; RFS: Relapse-free survival; TCGA: The Cancer Genome Atlas; TMB: Tumor mutational burden; VUS: Variant of unknown significance.

Author contributions

W.C.C. conceptualized the project, analyzed the data and wrote the manuscript; Y.T.Y. performed sequencing, computational analyses and wrote the manuscript; K.T.T. (experimental design, data analysis), V.M. (sample preparation), J.Y.L. and R.K.N. (clinical case identification), W.C. (histopathology), T.T.Y. (funding), S.J.C. (experimental design and consultation).

CONFLICTS OF INTEREST

Y.T.Y., K.T.T., T.T.Y. and S.J.C. are employees of ACT Genomics, Co., Ltd.

REFERENCES

1. Islami F, Torre LA, Jemal A. Global trends of lung cancer mortality and smoking prevalence. *Transl Lung Cancer Res*. 2015; 4:327–38.
2. Tanoue LT. Staging of non-small cell lung cancer. *Semin Respir Crit Care Med*. 2008; 29:248–260.
3. Ettinger DS, Wood DE, Akerley W, Bazhenova LA, Borghaei H, Camidge DR, Cheney RT, Chirieac LR, D'Amico TA, Dilling TJ, Dobelbower MC, Govindan R, Hennon M, et al. NCCN guidelines insights: non-small

- cell lung cancer, Version 4.2016. *J Natl Compr Canc Netw*. 2016; 14:255–264.
4. Lou F, Huang J, Sima CS, Dycoco J, Rusch V, Bach PB. Patterns of recurrence and second primary lung cancer in early-stage lung cancer survivors followed with routine computed tomography surveillance. *J Thorac Cardiovasc Surg*. 2013; 145:75–81.
 5. al-Kattan K, Sepsas E, Fountain SW, Townsend ER. Disease recurrence after resection for stage I lung cancer. *Eur J Cardiothorac Surg*. 1997; 12:380–384.
 6. Hoffman PC, Mauer AM, Vokes EE. Lung cancer. *Lancet*. 2000; 355:479–485.
 7. Carnio S, Novello S, Papotti M, Loiacono M, Scagliotti GV. Prognostic and predictive biomarkers in early stage non small-cell lung cancer: tumor based approaches including gene signatures. *Transl Lung Cancer Res*. 2013; 2:372–381.
 8. Uramoto H, Tanaka F. Recurrence after surgery in patients with NSCLC. *Transl Lung Cancer Res*. 2014; 3:242–49.
 9. Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, Schnall-Levin M, White J, Sanford EM, An P, Sun J, Juhn F, Brennan K, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol*. 2013; 31:1023–1031.
 10. Chang F, Li MM. Clinical application of amplicon-based next-generation sequencing in cancer. *Cancer Genet*. 2013; 206:413–419.
 11. Takeda M, Sakai K, Terashima M, Kaneda H, Hayashi H, Tanaka K, Okamoto K, Takahama T, Yoshida T, Iwasa T, Shimizu T, Nonagase Y, Kudo K, et al. Clinical application of amplicon-based next-generation sequencing to therapeutic decision making in lung cancer. *Ann Oncol*. 2015; 26:2477–2282.
 12. Arriagada R, Dunant A, Pignon JP, Bergman B, Chabowski M, Grunenwald D, Kozlowski M, Le Péchoux C, Pirker R, Pinel MI, Tarayre M, Le Chevalier T. Long-term results of the international adjuvant lung cancer trial evaluating adjuvant Cisplatin-based chemotherapy in resected lung cancer. *J Clin Oncol*. 2010; 28:35–42.
 13. Maeda R, Yoshida J, Ishii G, Hishida T, Nishimura M, Nagai K. Risk factors for tumor recurrence in patients with early-stage (stage I and II) non-small cell lung cancer: patient selection criteria for adjuvant chemotherapy according to the seventh edition TNM classification. *Chest*. 2011; 140:1494–1502.
 14. Kadota K, Colovos C, Suzuki K, Rizk NP, Dunphy MP, Zabor EC, Sima CS, Yoshizawa A, Travis WD, Rusch VW, Adusumilli PS. FDG-PET SUVmax combined with IASLC/ATS/ERS histologic classification improves the prognostic stratification of patients with stage I lung adenocarcinoma. *Ann Surg Oncol*. 2012; 19:3598–3605.
 15. Colby T, Noguchi M, Henschke C. Tumors of the lung. Pathology and Genetic of Tumors of the Lung, Pleura, Thymus and Heart. WHO Classification of Tumours. 2004; 10:35–8.
 16. Tsutani Y, Miyata Y, Mimae T, Kushitani K, Takeshima Y, Yoshimura M, Okada M. The prognostic role of pathologic invasive component size, excluding lepidic growth, in stage I lung adenocarcinoma. *J Thorac Cardiovasc Surg*. 2013; 146:580–585.
 17. Yanagawa N, Shiono S, Abiko M, Ogata SY, Sato T, Tamura G. New IASLC/ATS/ERS classification and invasive tumor size are predictive of disease recurrence in stage I lung adenocarcinoma. *J Thorac Oncol*. 2013; 8:612–618.
 18. Kawachi R, Tsukada H, Nakazato Y, Takei H, Furuyashiki G, Koshi-ishi Y, Goya T. Early recurrence after surgical resection in patients with pathological stage I non-small cell lung cancer. *Thorac Cardiovasc Surg*. 2009; 57:472–475.
 19. Shiono S, Abiko M, Sato T. Positron emission tomography/computed tomography and lymphovascular invasion predict recurrence in stage I lung cancers. *J Thorac Oncol*. 2011; 6:43–47.
 20. Cho WC. Potentially useful biomarkers for the diagnosis, treatment and prognosis of lung cancer. *Biomed Pharmacother*. 2007; 61:515–519.
 21. Guo M, House MG, Hooker C, Han Y, Heath E, Gabrielson E, Yang SC, Baylin SB, Herman JG, Brock MV. Promoter hypermethylation of resected bronchial margins: a field defect of changes? *Clin Cancer Res*. 2004; 10:5131–5136.
 22. Sandoval J, Mendez-Gonzalez J, Nadal E, Chen G, Carmona FJ, Sayols S, Moran S, Heyn H, Vizoso M, Gomez A, Sanchez-Cespedes M, Assenov Y, Müller F, et al. A prognostic DNA methylation signature for stage I non-small-cell lung cancer. *J Clin Oncol*. 2013; 31:4140–4147.
 23. Pio R, Agorreta J, Montuenga LM. Prognostic signature of early lung adenocarcinoma based on the expression of ribonucleic acid metabolism-related genes. *J Thorac Cardiovasc Surg*. 2015; 150:986–92.e1, 11.
 24. Ludovini V, Bianconi F, Siggillino A, Piobbico D, Vannucci J, Metro G, Chiari R, Bellezza G, Puma F, Della Fazio MA, Servillo G, Crinò L. Gene identification for risk of relapse in stage I lung adenocarcinoma patients: a combined methodology of gene expression profiling and computational gene network analysis. *Oncotarget*. 2016; 7:30561–74. <https://doi.org/10.18632/oncotarget.8723>.
 25. Xu W, Jia G, Davie JR, Murphy L, Kratzke R, Banerji S. A 10-Gene Yin Yang expression ratio signature for stage IA and IB non-small cell lung cancer. *J Thorac Oncol*. 2016; 11:2150–2160.
 26. Urban T, Ricci S, Danel C, Antoine M, Kambouchner M, Godard V, Lacave R, Bernaudin JF. Detection of codon 12 K-ras mutations in non-neoplastic mucosa from bronchial carina in patients with lung adenocarcinomas. *Br J Cancer*. 2000; 82:412–417.
 27. Jassem J, Jassem E, Jakóbkiewicz-Banecka J, Rzyman W, Badzio A, Dziadziuszko R, Kobińska-Gulida G,

- Szymanowska A, Skrzypski M, Zylicz M. P53 and K-ras mutations are frequent events in microscopically negative surgical margins from patients with nonsmall cell lung carcinoma. *Cancer*. 2004; 100:1951–1960.
28. Tang X, Shigematsu H, Bekele BN, Roth JA, Minna JD, Hong WK, Gazdar AF, Wistuba II. EGFR tyrosine kinase domain mutations are detected in histologically normal respiratory epithelium in lung cancer patients. *Cancer Res*. 2005; 65:7568–7572.
 29. Wu K, Huang RS, House L, Cho WC. Next-generation sequencing for lung cancer. *Future Oncol*. 2013; 9:1323–1336.
 30. Masasysva BG, Tong BC, Brock MV, Pilkington T, Goldenberg D, Sidransky D, Harden S, Westra WH, Califano J. Molecular margin analysis predicts local recurrence after sublobar resection of lung cancer. *Int J Cancer*. 2005; 113:1022–1025.
 31. Vatan O, Bilaloglu R, Tunca B, Cecener G, Gebitekin C, Egeli U, Yakut T, Urer N. Low frequency of p53 and k-ras codon 12 mutations in non-small cell lung carcinoma (NSCLC) tumors and surgical margins. *Tumori*. 2007; 93:473–477.
 32. Cao B, Feng L, Lu D, Liu Y, Liu Y, Guo S, Han N, Liu X, Mao Y, He J, Cheng S, Gao Y, Zhang K. Prognostic value of molecular events from negative surgical margin of non-small-cell lung cancer. *Oncotarget*. 2016; 8:53642–53. <https://doi.org/10.18632/oncotarget.10949>.
 33. Liu CH, Peng YJ, Wang HH, Chen YC, Tsai CL, Chian CF, Huang TW. Heterogeneous prognosis and adjuvant chemotherapy in pathological stage I non-small cell lung cancer patients. *Thorac Cancer*. 2015; 6:620–628.
 34. Burotto M, Thomas A, Subramaniam D, Giaccone G, Rajan A. Biomarkers in non-small cell lung cancer: current concepts and future directions. *J Thorac Oncol*. 2014; 9:1609–1617.
 35. Bartels S, van Luttikhuisen JL, Christgen M, Mägel L, Luft A, Hänzelmann S, Lehmann U, Schlegelberger B, Leo F, Steinemann D, Kreipe H. CDKN2A loss and PIK3CA mutation in myoepithelial-like metaplastic breast cancer. *J Pathol*. 2018; 245:373–383.
 36. Brandt LP, Albers J, Hejhal T, Pfundstein S, Gonçalves AF, Catalano A, Wild PJ, Frew IJ. Mouse genetic background influences whether *Hras*^{G12V} expression plus *Cdkn2a* knockdown causes angiosarcoma or undifferentiated pleomorphic sarcoma. *Oncotarget*. 2018; 9:19753–66. <https://doi.org/10.18632/oncotarget.24831>.
 37. Chen WS, Bindra RS, Mo A, Hayman T, Husain Z, Contessa JN, Gaffney SG, Townsend JP, Yu JB. CDKN2A copy number loss is an independent prognostic factor in HPV-negative head and neck squamous cell carcinoma. *Front Oncol*. 2018; 8:95.
 38. Ibrahim I, Sibinga Mulder BG, Bonsing B, Morreau H, Farina Sarasqueta A, Inderson A, Luelmo S, Feshtali S, Potjer TP, de Vos Tot Nederveen Cappel W, Wasser M, Vasen HF. Risk of multiple pancreatic cancers in CDKN2A-p16-Leiden mutation carriers. *Eur J Hum Genet*. 2018; 26:1227–29. <https://doi.org/10.1038/s41431-018-0170-y>.
 39. Karagianni F, Njauw CN, Kypreou KP, Stergiopoulou A, Plaka M, Polydorou D, Chasapi V, Pappas L, Stratigos I, Champsas G, Panagiotou P, Gogas H, Evangelou E, et al. CDKN2A/CDK4 status in Greek patients with familial melanoma and association with clinico-epidemiological parameters. *Acta Derm Venereol*. 2018; 98:862–66. <https://doi.org/10.2340/00015555-2969>.
 40. Siref A, Patel V, Reith JD, Balzer BL, Shon W. Evaluation of p16 protein expression and CDKN2A deletion in conventional and fibrosarcomatous dermatofibrosarcoma protuberans. *Pathology*. 2018; 50:474–475.
 41. Liang J, Fan J, Wang M, Niu Z, Zhang Z, Yuan L, Tai Y, Chen Z, Song S, Wang X, Liu X, Huang H, Sun Q. CDKN2A inhibits formation of homotypic cell-in-cell structures. *Oncogenesis*. 2018; 7:50.
 42. Maecker HL, Yun Z, Maecker HT, Giaccia AJ. Epigenetic changes in tumor Fas levels determine immune escape and response to therapy. *Cancer Cell*. 2002; 2:139–148.
 43. Lee Y, Kawagoe R, Sasai K, Li Y, Russell HR, Curran T, McKinnon PJ. Loss of suppressor-of-fused function promotes tumorigenesis. *Oncogene*. 2007; 26:6442–6447.
 44. Rubin LL, de Sauvage FJ. Targeting the Hedgehog pathway in cancer. *Nat Rev Drug Discov*. 2006; 5:1026–1033.
 45. Mizuarai S, Kawagishi A, Kotani H. Inhibition of p70S6K2 down-regulates Hedgehog/GLI pathway in non-small cell lung cancer cell lines. *Mol Cancer*. 2009; 8:44.
 46. Taipale J, Chen JK, Cooper MK, Wang B, Mann RK, Milenkovic L, Scott MP, Beachy PA. Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine. *Nature*. 2000; 406:1005–1009.
 47. Gialmanidis IP, Bravou V, Amanetopoulou SG, Varakis J, Kourea H, Papadaki H. Overexpression of hedgehog pathway molecules and FOXM1 in non-small cell lung carcinomas. *Lung Cancer*. 2009; 66:64–74.
 48. Gialmanidis IP, Bravou V, Petrou I, Kourea H, Mathioudakis A, Lilis I, Papadaki H. Expression of Bmi1, FoxF1, Nanog, and γ -catenin in relation to hedgehog signaling pathway in human non-small-cell lung cancer. *Lung*. 2013; 191:511–521.
 49. Bell EH, Chakraborty AR, Mo X, Liu Z, Shilo K, Kirste S, Stegmaier P, McNulty M, Karachaliou N, Rosell R, Bepler G, Carbone DP, Chakravarti A. SMARCA4/BRG1 is a novel prognostic biomarker predictive of cisplatin-based chemotherapy outcomes in resected non-small cell lung cancer. *Clin Cancer Res*. 2016; 22:2396–2404.
 50. Kim N, Hong Y, Kwon D, Yoon S. Somatic mutational profile in human cancer tissues. *Genomics Inform*. 2013; 11:239–244.
 51. Tan Q, Li F, Wang G, Xia W, Li Z, Niu X, Ji W, Yuan H, Xu Q, Luo Q, Zhang J, Lu S. Identification of FGF19 as a prognostic marker and potential driver gene of lung

- squamous cell carcinomas in Chinese smoking patients. *Oncotarget*. 2016; 7:18394–402. <https://doi.org/10.18632/oncotarget.7817>.
52. Hudson AM, Wirth C, Stephenson NL, Fawdar S, Brognard J, Miller CJ. Using large-scale genomics data to identify driver mutations in lung cancer: methods and challenges. *Pharmacogenomics*. 2015; 16:1149–1160.
53. Lianos GD, Mangano A, Cho WC, Roukos DH. From standard to new genome-based therapy of gastric cancer. *Expert Rev Gastroenterol Hepatol*. 2015; 9:1023-1026.
54. Boeva V, Popova T, Lienard M, Toffoli S, Kamal M, Le Tourneau C, Gentien D, Servant N, Gestraud P, Rio Frio T, Hupé P, Barillot E, Laes JF. Multi-factor data normalization enables the detection of copy number aberrations in amplicon sequencing data. *Bioinformatics*. 2014; 30:3443–3450.
55. Amarasinghe KC, Li J, Hunter SM, Ryland GL, Cowin PA, Campbell IG, Halgamuge SK. Inferring copy number and genotype in tumour exome data. *BMC Genomics*. 2014; 15:732.