Genomic characterisation of pulmonary subsolid nodules: mutational landscape and radiological features

Yanmeng Li[1,6], Xiao Li[1,6], Hao Li [1,6], Yifan Zhao[1,6], Ziyang Liu [1,6], Kunkun Sun[2], Xiang Zhu[3], Qingyi Qi[4], Bei An[4], Danhua Shen[2], Ruoyan Li[1], Taorui Liu[1], Jiahui Mi [1], Lientu Wang[5], Fan Yang[1,7], Fan Bai[1,7] and Jun Wang[1,7]

Affiliations: [1]Biomedical Pioneering Innovation Center (BIOPIC), School of Life Sciences and Dept of Thoracic Surgery, People's Hospital, Peking University, Beijing, China. [2]Dept of Pathology, Peking University People's Hospital, Beijing, China. [3]Dept of Pathology, Peking University Third Hospital, Beijing, China. [4]Dept of Radiology, Peking University People's Hospital, Beijing, China. [5]Berry Oncology Co., Ltd, Fujian, China. [6]These authors contributed equally to the study. [7]These authors contributed equally to the study.

Correspondence: Jun Wang, Dept of Thoracic Surgery, Peking University People's Hospital, No. 11 Xizhimen South Street, Beijing, 100044, China. E-mail: wangjun@pkuph.edu.cn

ABSTRACT
Background: Lung adenocarcinomas (LUADs) that display radiologically as subsolid nodules (SSNs) exhibit more indolent biological behaviour than solid LUADs. SSNs, commonly encompassing pre-invasive and invasive yet early-stage adenocarcinomas, can be categorised as pure ground-glass nodules and part-solid nodules. The genomic characteristics of SSNs remain poorly understood.

Methods: We subjected 154 SSN samples from 120 treatment-naïve Chinese patients to whole-exome sequencing. Clinical parameters and radiological features of these SSNs were collected. The genomic landscape of SSNs and differences from that of advanced-stage LUADs were defined. In addition, we investigated the intratumour heterogeneity and clonal relationship of multifocal SSNs and conducted radiogenomic analysis to link imaging and molecular characteristics of SSNs. Fisher's exact and Wilcoxon rank sum tests were used in the statistical analysis.

Results: The median somatic mutation rate across the SSN cohort was 1.12 mutations per Mb. Mutations in *EGFR* were the most prominent and significant variation, followed by those in *RBM10*, *TP53*, *STK11* and *KRAS*. The differences between SSNs and advanced-stage LUADs at a genomic level were unravelled. Branched evolution and remarkable genomic heterogeneity were demonstrated in SSNs. Although multicentric origin was predominant, we also detected early metastatic events among multifocal SSNs. Using radiogenomic analysis, we found that higher ratios of solid components in SSNs were accompanied by significantly higher mutation frequencies in *EGFR*, *TP53*, *RBM10* and *ARID1B*, suggesting that these genes play roles in the progression of LUADs.

Conclusions: Our study provides the first comprehensive description of the mutational landscape and radiogenomic mapping of SSNs.

This article has supplementary material available from erj.ersjournals.com

Introduction

Early-stage lung adenocarcinomas (LUADs) can present as a spectrum of radiological appearances, ranging from pure ground-glass nodules (pGGNs) generally thought to be indolent, to more aggressive solid LUADs. Radiological subsolid nodules (SSNs) can be further categorised as pGGNs, which are defined as nodules manifesting as an area of hazy increased lung opacity with preservation of bronchial and vascular margins, and part-solid nodules (PSNs) comprised of both ground-glass and solid components [1]. A proportion of pGGNs eventually develop into PSNs. Pathologically, malignant SSNs encompass pre-invasive lesions (atypical adenomatous hyperplasias and adenocarcinomas *in situ*), minimally invasive adenocarcinomas and invasive adenocarcinomas.

To date, comprehensive studies describing the genomic landscape [2–5] and clonal architecture [6, 7] of LUAD have focused mainly on relatively late-stage tumours. Genomic features of the very early stage of LUAD [8], especially the subgroup that displays radiologically as SSNs [9], remain poorly understood. Furthermore, the diverse clinical trajectories of pGGNs and PSNs pose significant management challenges, and there have been limited predictive and prognostic biomarkers to risk-stratify patients. Radiological features of SSNs, such as the subclassification (pGGN or PSN) [10], the solid size [11] and the consolidation tumour ratio (CTR) [12], can serve as important clinical references for diagnosis and treatment of this disease. However, the potential genomic grounds supporting the use of radiological parameters as important clinical references for diagnosis and treatment of SSNs were not deciphered.

In this study, we performed whole-exome sequencing of 154 surgical SSN specimens from 120 patients. We defined the genomic characteristics of SSNs and delineated potential driver somatic mutations underlying the development and progression to advanced-stage LUAD. We also revealed the intratumour heterogeneity between ground-glass and solid regions of the same SSN (n=11) and the clonal relationship among multifocal SSNs from the same patient (n=26). Finally, we provided radiogenomic mapping of SSNs. Our findings shed light on the evolution of LUAD and provide valuable biological and clinical insights into SSNs.

Methods

SSN cohort

Pathologically confirmed malignant tumour samples from 154 SSNs along with paired controls from 120 treatment-naïve Chinese patients were subjected to whole-exome sequencing. The study design is summarised in figure 1a and b, and detailed clinical features of the cohort are summarised in table 1 and supplementary table S1. Of note, 11 large PSN samples were divided into two to five regions for whole-exome sequencing according to the gross appearance of the resected tumour and the radiological characteristics (figure 1c, supplementary table S2). Pathological diagnoses were classified as atypical adenomatous hyperplasia, adenocarcinoma *in situ*, minimally invasive adenocarcinoma or invasive adenocarcinoma according to the 2015 World Health Organization classification system (figure 1d).

Radiological evaluation

The maximum diameters of whole SSNs and solid components on lung windows were measured (figure 2b). The CTR was calculated as the ratio of the maximum diameter of consolidation divided by the tumour size. Volumetric parameters, such as solid volume percentage, were obtained semi-automatically using the Lung VCAR workstation (GE Healthcare, Chicago, IL, USA). The SSN mass was calculated using the following equation: mass=volume × ((mean CT attenuation + 1000) × 0.01). Image analysis details are included in the supplementary material.

Whole-exome sequencing library preparation and data analysis

Protocols for genomic DNA extraction and whole-exome sequencing library preparation are detailed in the supplementary material. We used the Genome Analysis Toolkit (GATK, version 3.8.0) to pre-process the whole-exome sequencing data, and MuTect (version 1.1.4) and Strelka (version 2.8.4) to call single nucleotide variants and insertions and deletions, respectively. Further details of bioinformatic analyses arre provided in the supplementary material. We used Fisher's exact test and Wilcoxon rank sum test for statistical analysis.

Results

The mutational landscape of SSNs

The median (range) somatic mutation rate across the SSN cohort (n=154) was 1.12 (0.03–6.87) mutations per Mb, which is significantly lower than that of Chinese advanced-stage LUADs [5] (supplementary figure S1). The median sequencing depth per tumour was 217. The mutational landscape of SSNs is summarised in figure 3a. 12 significant mutated genes were identified by the MutSigCV algorithm. Notably, mutations in *EGFR* were the most prominent and significant variation (50%), followed by mutations in *RBM10* (16%), *TP53* (13%), *STK11* (8%) and *KRAS* (5%). Other frequent, but not
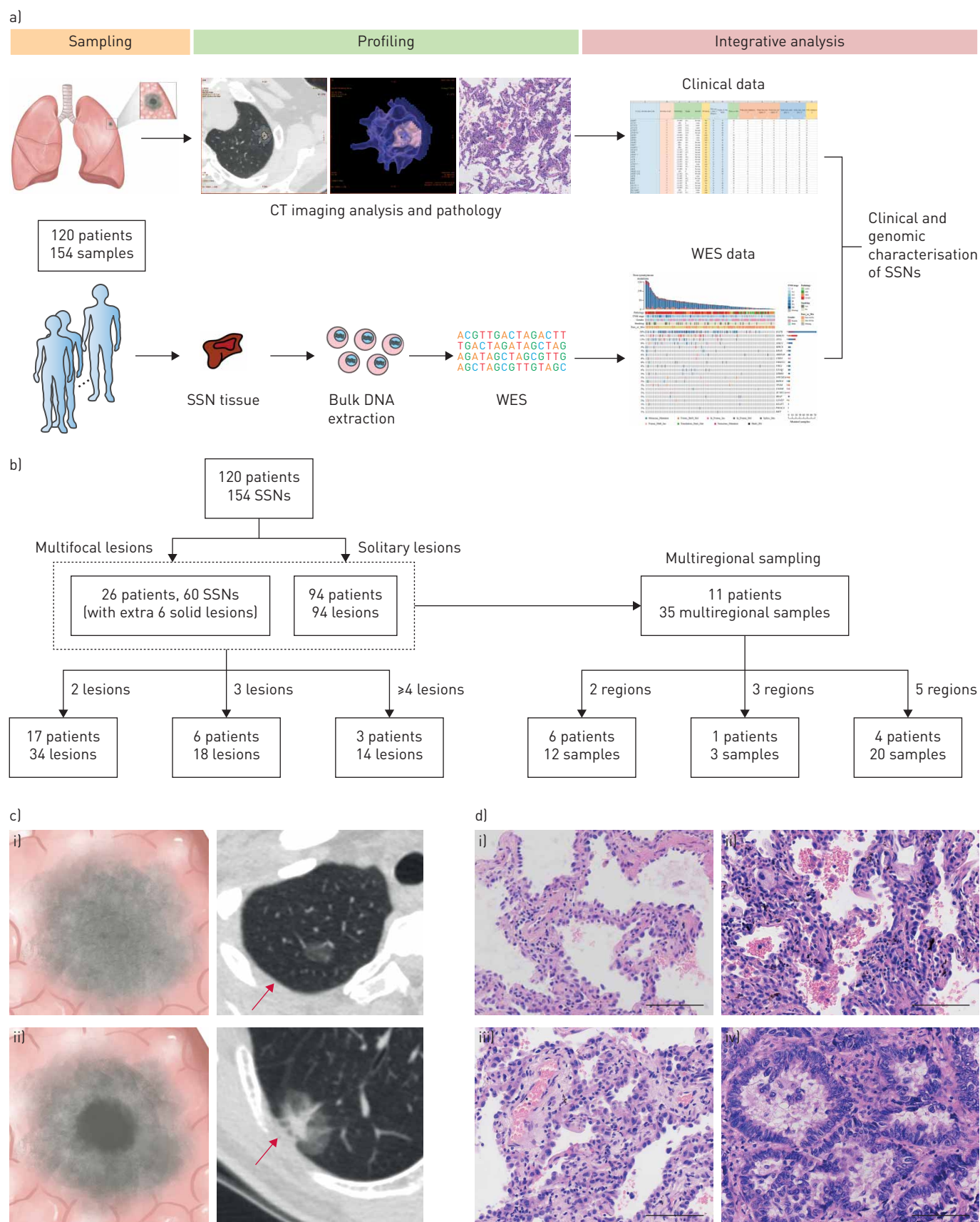
a)

b)

c)

d)

FIGURE 1 Study overview and sample information for 154 subsolid nodules (SSNs). a) Schematic of the study design; b) detailed information about the number of patients and samples assigned to different research strategies; c) schematic drawings of i) a pure ground-glass nodule and ii) a part-solid nodule with corresponding computed tomography (CT) images under the lung window setting (red arrows indicate sites of SSNs); d) histological features (haematoxylin and eosin staining) of SSNs including i) atypical adenomatous hyperplasia, ii) adenocarcinoma *in situ*, iii) minimally invasive adenocarcinoma iv) and invasive adenocarcinoma. WES: whole-exome sequencing. Scale bar=100 μm.

significantly mutated genes included *BIRC6* (5%), *ARID1B* (4%), *CHD2* (4%), *FBXO11* (4%) and *GNAQ* (4%). Concurrent and mutually exclusive mutation patterns in SSNs are shown in figure 3d. Similar to previous reports in LUADs [3, 5, 13], mutations in *KRAS* were mutually exclusive with those in *EGFR*. Both *RBM10* and *TP53* mutations commonly co-occurred with *EGFR* mutations. Interestingly, *TP53* was found to be co-altered in only 19% (15 out of 77) of the *EGFR* mutant population, lower than the

TABLE 1 Clinical characteristics of 120 patients and 154 subsolid nodules (SSNs)	
Age years	59 (29–82)
Sex	
Female	68 (56.7)
Male	52 (43.3)
Smoking status	
Smoker	32 (26.7)
Nonsmoker	88 (73.3)
Lesion maximum diameter mm	
⩽10.0	41 (26.6)
10.1–20.0	68 (44.2)
>20.0	45 (29.2)
SSN type	
pGGN	66 (42.9)
PSN	87 (56.5)
NA	1 (0.6)
Solid maximum diameter mm	
⩽5	79 (51.3)
>5	74 (48.1)
NA	1 (0.6)
CTR	
0–0.25	68 (44.2)
>0.25	85 (55.2)
NA	1 (0.6)
Solid volume percentage %	
⩽50	120 (77.9)
>50	22 (14.3)
NA	12 (7.8)
SSN mass mg	
⩽250	53 (34.4)
>250	89 (57.8)
NA	12 (7.8)
Histological type	
AAH	14 (9.1)
AIS	19 (12.3)
MIA	30 (19.5)
IAC	91 (59.1)
Tumour stage[#]	
NA (AAH)	14 (9.1)
0	19 (12.3)
IA1	74 (48.1)
IA2	33 (21.4)
IA3	3 (1.9)
IB	8 (5.2)
IIA	1 (0.6)
IIB	2 (1.3)

Data are presented as median (range) or n (%). pGGN: pure ground glass nodule; PSN: part-solid nodule; NA: not applicable; CTR: consolidation tumour ratio; AAH: atypical adenomatous hyperplasia; AIS: adenocarcinoma *in situ*; MIA: minimally invasive adenocarcinoma; IAC: invasive adenocarcinoma. [#]: according to the International Union Against Cancer eighth tumour, node, metastasis staging.
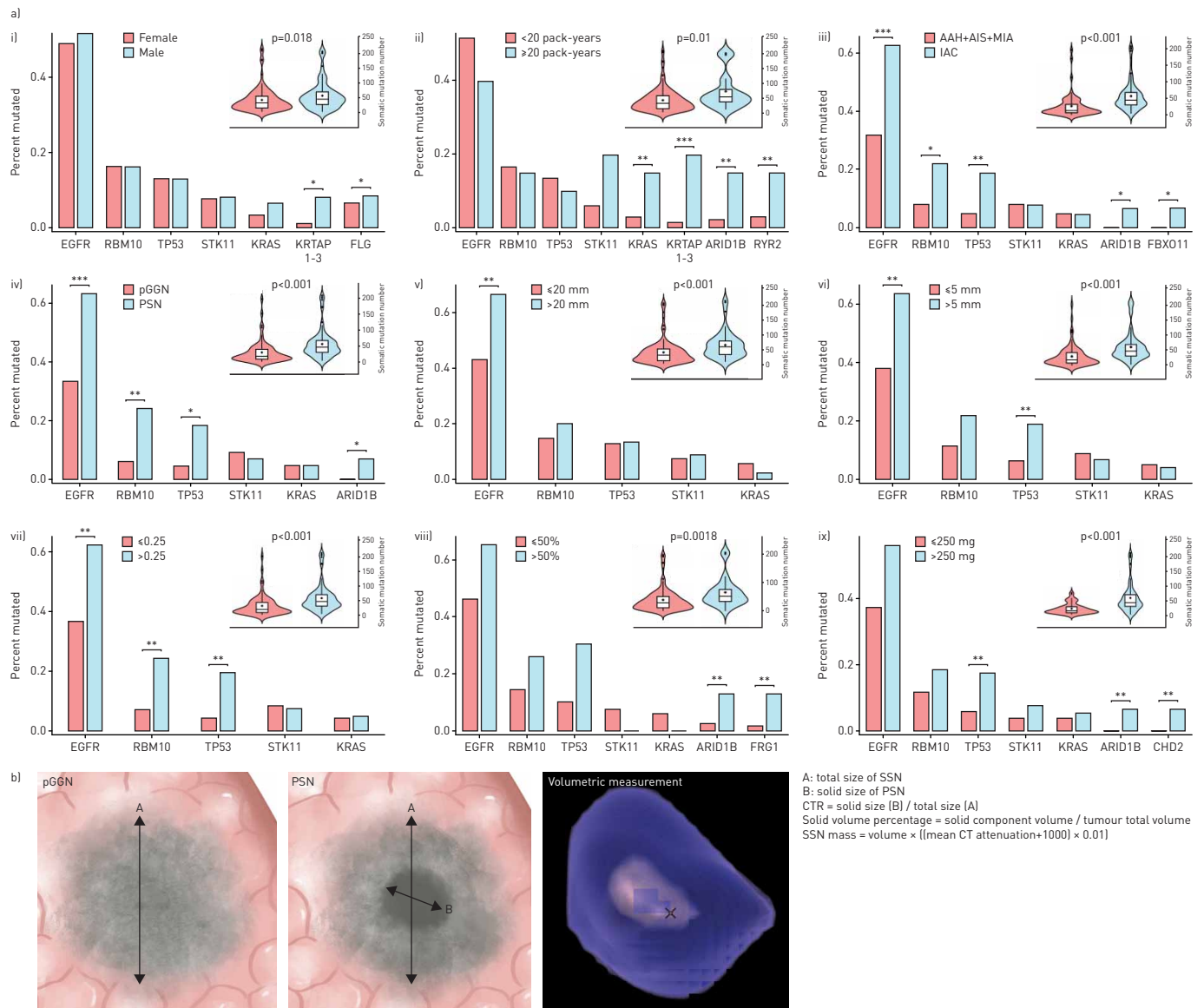
FIGURE 2 Clinical enrichment analysis of mutated genes by different clinical criterion groups. a) Both barplots, which indicate the mutation frequencies of significantly mutated genes as well as additional non-significantly enriched genes (for all genes with mutation frequencies >3%), and violin plots, which show the somatic mutation number, demonstrate the comparisons between sample groups classified by i) sex, ii) smoking condition with a cut-off of 20 pack-years, iii) pathology, iv) radiological subsolid nodule (SSN) classification (pure ground-glass nodule (pGGN) or part-solid nodule (PSN)), v) total size with a cut-off of 20 mm, vi) solid component size with a cut-off of 5 mm, vii) consolidation tumour ratio (CTR) with a cut-off of 0.25, vii) solid volume percentage with a cut-off of 50% and viii) SSN mass with a cut-off of 250 mg. AAH: atypical adenomatous hyperplasia; AIS: adenocarcinoma *in situ*; MIA: minimally invasive adenocarcinoma; IAC: invasive adenocarcinoma. For the barplots, significance was calculated by Fisher's exact test; for the violin plots, the p-value was calculated by Wilcoxon rank sum test. *: p<0.05, **: p<0.01, ***: p<0.001. b) Schematic diagrams and formulas for calculation of several of the clinical criteria mentioned in (a).

anticipated 50–60% in early-stage solid tumours or metastatic disease [5, 14], suggesting that these combinations may facilitate the transformation of early stage lung neoplasm to malignancy [15].

We further explored the aetiological factors that underlie mutagenesis in Chinese SSNs. Mutational spectrum analysis revealed a strong enrichment of C>T transversions and C>A transitions, which are associated with a history of smoking (figure 3b). From 154 SSNs, we confidently identified three mutational signatures, which displayed high similarity to COSMIC signatures 4, 6 and 5 (figure 3c) [16]. COSMIC signature 4 is associated with smoking and has been frequently reported in LUADs [4, 5]. Intriguingly, COSMIC signature 6, which is often identified in colorectal cancers but less in LUADs [2], is characteristic of tumours with defective DNA mismatch repair. These results suggested that smoking and dysregulation of DNA repair contribute to SSN tumourigenesis.

Comparing the genomic features of SSNs to advanced-stage LUADs

To unravel the differences between SSNs and advanced-stage LUADs at a genomic level, we compared the mutation frequencies of driver genes in 10 oncogenic pathways [17] and other previously reported mutations related to LUAD [3, 5]. The mutation frequencies of these genes in advanced-stage LUADs were calculated by reanalysing data from the Cancer Genome Atlas (106 samples of stage II–IV LUADs [3]) and published Chinese LUADs (75 samples of stage II–IV LUADs [5]) (figure 3e, supplementary figure S2).

First, we compared the mutation frequencies of oncogenic pathways in SSNs and stage II–IV LUADs and found the differences between the three groups (supplementary table S3). Next, we examined the frequencies of driver mutations in SSNs and Chinese stage II–IV LUADs (figure 3e, supplementary figure S2). The smoking status of these two cohorts were similar (73.3% nonsmokers in Chinese SSNs and 61.3% in Chinese stage II–IV LUADs). Of note, *EGFR* (50.0% *versus* 30.7%) and *RBM10* (16.2% *versus* 1.3%) mutations exhibited significantly higher frequencies in SSNs. Of all the 83 *EGFR* mutations from 77 SSNs, 60% (50 out of 83) were L858R, and 20% (17 out of 83) were exon 19 deletions, while the remainder (16 out of 83) were other rare mutations (T790M, L858M, L861Q, L833F, and so on) (supplementary figure S3). This was consistent with a previous study which reported that *EGFR* mutations were found in 64% of Japanese SSNs [9]. Interestingly, *RBM10* exhibited a significantly higher mutation frequency in SSNs with specific mutational types (nonsense, frameshift and splice site) (supplementary figure S3) [4] when compared to Chinese advanced-stage LUADs. *RBM10*, which encodes an RNA-binding protein, was found to be frequently mutated (7%) in Caucasian LUADs [4] and highly mutated in pre-invasive and early-stage LUADs [18]. We speculated that alterations in RNA splicing alterations are a hallmark of SSNs, but do not confer a sustained evolutionary advantage during progression to advanced stage adenocarcinoma.

A number of genes displayed notably lower mutation frequencies in SSNs when compared with Chinese stage II–IV LUADs (figure 3e). These included tumour suppressor genes in key oncogenic pathways, including *TP53*, *APC*, *FAT1/2/3/4*, *CRB1/2*, *NOTCH1/2/3/4*, *SPEN* and *ARID1A* and oncogenes including *KRAS*, *ERBB4*, *ROS1*, *PIK3CA/B*, *TEAD1/2* and *NFE2L2*, as well as genes associated with cytoskeleton remodelling including *PHPN2* and *TRIO*. Above all, we speculated that *EGFR* and *RBM10* mutations are critical for SSN tumourigenesis, but progression of SSNs to advanced-stage LUADs is driven by mutations in other important cancer-related genes.

Genomic profile of intrapatient multifocal SSNs

Multiple synchronous occurrence is a feature of SSNs. To explore the clonal relationship between multifocal SSNs and solid tumours from the same patient, we sequenced 66 samples from 26 patients (two to five per patient). The interlesional genomic profile (the median sequencing depth of these regions was 186.9) was distinct in 23 out of 26 patients and generally confirmed the multicentric origin of SSNs (supplementary figure S4).

However, of note, P119 with four pGGNs (T1, T3–T5) and one PSN (T2) was found to have a metastatic event (between T4 and T5). P119-T4 and P119-T5, both of which were located in the apical-posterior segment of left upper lobe and histologically presented as minimally invasive adenocarcinomas, were demonstrated to share 10 nonsynonymous and eight synonymous mutations (figure 4a). The potential driver *MLLT1* mutation was shared by these two pGGNs, while *EGFR(L858R)*, *TP53* and *PASK* mutations were exclusive to P119-T4. Taken together, multicentric tumours and intrapulmonary metastatic lesions with a common clonal origin co-existed in this patient, which can be visualised in the regional distribution of mutations and phylogenetic tree.

We identified three PSNs (T1–T3) and one solid nodule (T4) of P114 that originated from the same clonal origin (figure 4b). P114-T1 showed predominant solid features by computed tomography scan (CTR 0.87, solid volume percentage 68%), mainly acinar in pathology. P114-T2 (CTR 0.39, solid volume percentage 30%) and P114-T3 (CTR 0.66, solid volume percentage 41%) were ground-glass-predominant lesions with adenocarcinoma *in situ* and mainly acinar in their pathology, respectively. P114-T1 was very large and therefore underwent multiregion sequencing (R1–R5). The phylogenetic tree of this patient revealed that P114-T1-R1~R5, P114-T2, P114-T3, and P114-T4 all shared a substantial amount of mutations. Interestingly, P114-T1 displayed remarkable regional heterogeneity and the most central part (R5) of P114-T1 was the nearest to the putative metastases (P114-T2, P114-T3 and P114-T4) in evolutionary distance, suggesting that metastatic lesions of SSNs can either stay as SSNs or progress to solid tumours.

Intratumour heterogeneity between the ground-glass and solid components of SSNs

To investigate the evolution and intratumour heterogeneity of SSNs, we performed multiregion sequencing on 35 spatially separate regions sampled from 11 large SSNs. The regional mutation distributions and
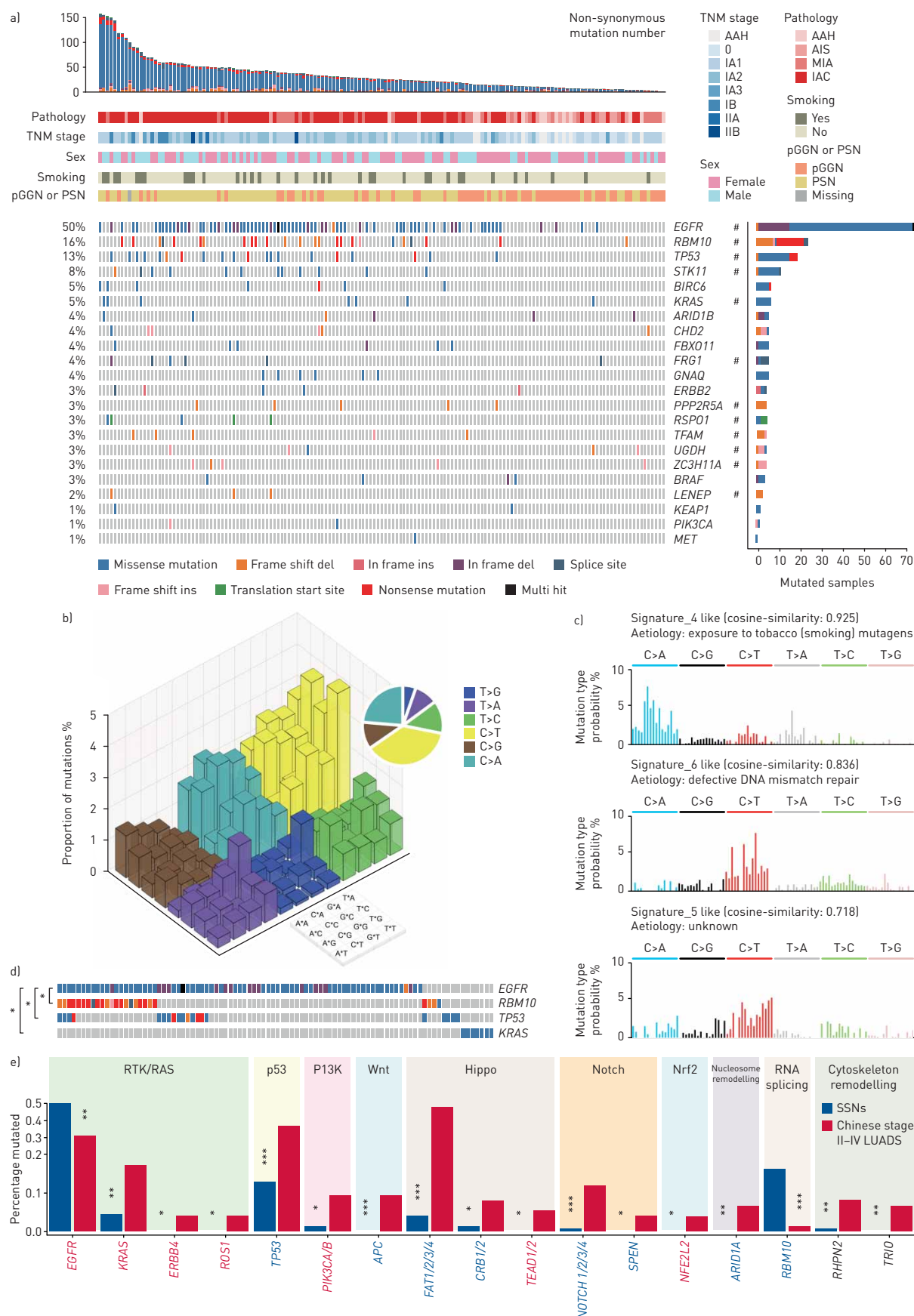
FIGURE 3 The mutational landscape and signatures of Chinese subsolid nodules (SSNs). a) Mutational landscape of nonsynonymous somatic mutations, including single nucleotide variants and insertions (ins) and deletions (del). Samples displayed as columns are ordered by the number of nonsynonymous somatic mutations, as shown in the upper panel. Clinical features are annotated in the middle panel. The heatmap below displays the somatic mutated genes ordered by their respective mutation frequencies. TNM: tumour, node, metastasis; pGGN: pure ground-glass nodule; PSN: part-solid nodule; AAH: atypical adenomatous hyperplasia; AIS: adenocarcinoma *in situ*; MIA: minimally invasive adenocarcinoma; IAC: invasive adenocarcinoma. #: 12 significant genes with a corrected q-value <0.1 were identified by the MutSigCV algorithm. b) "Lego" plots displaying the frequency of base substitutions within specific trinucleotide mutational contexts. c) Mutational signatures of Chinese SSNs. d) Concurrent and mutually exclusive somatic mutation patterns of significantly mutated genes. Significance was calculated using Fisher's exact test. e) Comparison of the mutation frequencies of important genes in cancer-associated pathways between Chinese SSNs and Chinese advanced-stage lung adenocardinomas (LUADS; stage II–IV) [5]. Oncogenes labelled in red and tumour suppressor genes labelled in blue are listed on the horizontal axis ordered by the corresponding pathways. The vertical axis represents the mutation frequencies obtained from different cohorts. Significance was calculated using Fisher's exact test. *: p<0.05, **: p<0.01, ***: p<0.001.

phylogenetic trees of P118, P117 and P116 are shown in figure 4c, d and e, respectively (results for the remaining eight patients are included in supplementary figure S5). Branched evolution was evident in SSNs, with potential driver mutations presenting both on the trunks and branches in the phylogenetic trees (figure 4c, d and e). This demonstrated that genomic heterogeneity and branched evolution are salient features of LUADs, even in their early stages when they radiologically display as SSNs.

Next, we focused on the difference between the ground-glass and solid components of PSNs in the hopes of uncovering genomic factors that shape the distinct radiological phenotypes within the same lesion. Interestingly, we found that solid components had comparable or even fewer somatic mutation numbers than the ground-glass components, and no significant enrichment of specific mutations was identified in ground-glass or solid components. The evolutionary distance between regions of the same radiological phenotype varied among patients, as demonstrated in the phylogenetic trees (figure 4b, c, d and e). For example, R4 and R5 sampled from the solid component of P114-T1 were remarkably heterogeneous. The same result was observed among R1, R2 and R3, which were sampled from the ground-glass component of P114-T1 (figure 4b). In contrast, R2 and R3 from the ground-glass component of P117-T displayed high genomic similarity, whereas R1, R4 and R5 from the solid component grouped together on the phylogenetic tree (figure 4d), consistent with their separation in the radiological phenotype. In this regard, mutational features are not deterministically correlated with radiological features of different regions in the same SSN.

Radiogenomic analysis links imaging and molecular characteristics of SSNs

Radiological features of SSNs can serve as important clinical references for diagnosis and treatment. Given the clinical significance of these radiological parameters of SSNs, we investigated the differences in the number of somatic mutations and patterns of mutated genes (for all genes with mutation frequencies >3%) between any two groups separated by thresholds of each parameter with relevant clinical implications.

The mutation patterns were first correlated with some basic clinical features such as sex, smoking condition and pathology (figure 2a(i, ii and iii), respectively). Driver mutations of SSNs, including *EGFR*, *RBM10*, *TP53* and *STK11* displayed roughly the same mutation frequencies between sexes, an observation which differed from previous studies [3, 5]. *KRAS* mutations were more commonly observed in males, although the difference was not statistically significant. In addition, we found that *KTRAP1-3* mutations were significantly enriched in males, although *FLG* mutations were enriched in female patients (p<0.05, Fisher's exact test). Compared with nonsmokers and light smokers (<20 pack-years), medium–heavy smokers (⩾20 pack-years) had significantly greater numbers of *KRAS*, *KTRAP1-3* and *ARID1B* mutations (p<0.01 or p<0.001, Fisher's exact test) along with a higher somatic mutation burden (p=0.01, Wilcoxon rank sum test). When considering the pathological characteristics of the SSNs (supplementary table S4), the mutation frequencies of *EGFR*, *RBM10*, *TP53*, *ARID1B* and *FBXO11* were significantly greater in SSNs that pathologically presented as invasive adenocarcinoma compared to those that presented as atypical adenomatous hyperplasia, adenocarcinoma *in situ* and minimally invasive adenocarcinoma (p<0.05, Fisher's exact test). In addition, we found that invasive adenocarcinoma SSNs possessed a much higher somatic mutation load (p<0.001, Wilcoxon rank sum test).

We next investigated the relationship between the radiological predictors of invasiveness and the genomic profiles of SSNs. Similar to the results found for pathology, PSNs in our cohort were enriched for *EGFR*, *RBM10*, *TP53* and *ARID1B* mutations (p<0.05, Fisher's exact test) and displayed significantly higher numbers of somatic mutations (p<0.001, Wilcoxon rank sum test) than did pGGNs (figure 2a(iv)). We then compared the patterns of mutation of the subgroups separated by total size and solid size (figure 2a(v and vi)). Only mutation in *EGFR* was significantly enriched in SSNs with a total size >20 mm when compared to smaller SSNs (p<0.05, Fisher's exact test). However, when considering the size of the solid
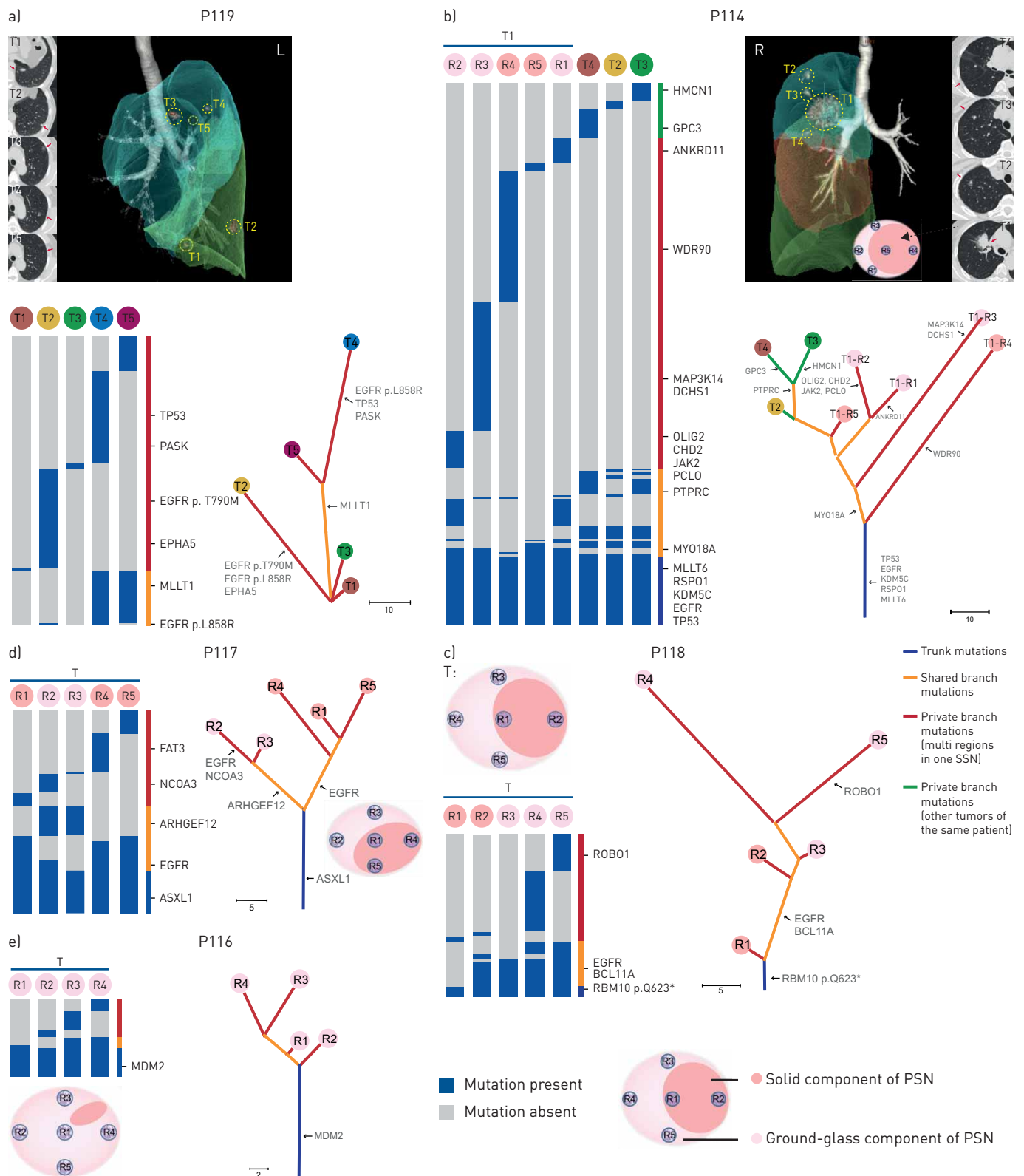
FIGURE 4 Intertumour and intratumour heterogeneity of subsolid nodules (SSNs). Heatmaps on the left of each panel depict the regional distribution of all of the somatic mutations among different sequencing sites of multiregion SSNs and multifocal SSNs. Phylogenetic trees on the right of each panel were constructed for each patient using the maximum parsimony algorithm. The colour of each line corresponds to the categories of mutations shown in the heatmaps, and the lengths of the trunks and branches are proportional to the number of mutations in each sample. Spatial location information of multifocal samples in b) P114 and a) P119 is represented in the three-dimensional reconstructions with the corresponding computed tomography images on the side. L: left lobe. R: right lobe.

component, in addition to *EGFR*, *TP53* was also significantly enriched in SSNs with a solid size >5 mm (p<0.01, Fisher's exact test). This finding underscored the importance of mutation of *TP53* in driving the growth of the invasive tumour component (solid size) rather than the total SSN growth (total size). In contrast, mutation in *EGFR* may play an important role in both SSN growth and SSN invasiveness. Both groups (total size >20 mm and solid size >5 mm) exhibited much higher somatic mutation burdens than other groups (p<0.01, Wilcoxon rank sum test) (figure 2a(v and vi)).

The CTR, solid volume percentage and SSN mass are additional predictive radiological parameters that are used to evaluate the tumour solid components (figure 2a(vii, viii and ix)). We found that the number of somatic mutations was significantly higher in groups with a CTR >0.25, a solid volume percentage >50% and an SSN mass >250 mg, which suggested that increased solid component in SSNs is correlated with increased mutation events. *EGFR*, *RBM10* and *TP53* mutations were significantly enriched in SSNs with a CTR ⩾0.25 (p<0.01, Fisher's exact test). Similar enrichments were found when comparing subgroups classified by solid volume percentage and mass, although without statistical significance. However, these two classification groups both displayed a higher frequency of *ARID1B* mutations (p<0.01, Fisher's exact test).

Discussion

In this study, we provided a comprehensive genomic landscape of Chinese SSNs. We highlighted that mutations in genes including *EGFR*, *RBM10*, *TP53*, *STK11*, *BIRC6* and *KRAS* play important roles in driving SSN tumourigenesis and discovered mutational signatures associated with smoking and defective DNA mismatch repair, the latter of which is often identified in colorectal cancer [16], but is rare in LUADs [19]. Consistent with this observation, genes associated with DNA repair, such as *ATM* [20], *ASXL1* [21], *ATRX* [22] and *ARID1B* [23], were also mutated in SSNs as well as most early lesions of LUADs [15]. By comparing the genomic features of SSNs to those of advanced-stage LUADs, we were able to delineate the potential order of driver mutations during the progression of SSNs to invasive LUADs. These findings may explain why SSNs are able to remain indolent for many years.

Our study confirmed that most multifocal SSNs are tumours of multiple synchronous occurrence. However, we identified two cases with metastatic multifocal SSNs. Together with the two cases we reported previously [24], we found that SSNs can disseminate metastatic lesions while the metastatic lesions can remain as SSNs. The exact metastatic routes remain unknown, but are possibly through haematogenous, lymphatic or airway metastasis. We demonstrated that genomic heterogeneity and branched evolution are present even in the early stages of LUADs, as shown in a recent study [25], highlighting the complex evolutionary history of lung cancers and the therapeutic challenges we face.

It is widely accepted that radiological features of SSNs can serve as important clinical parameters for the diagnosis and treatment of SSNs. The total size and solid size of SSNs can predict their clinical behaviours. A more aggressive surveillance strategy should be applied to patients with pGGNs with a total size >20 mm [26]. CTR, which takes both total size and solid size into account, also plays a critical role in the evaluation and treatment of SSNs [27]. Meanwhile, tumour volume and mass are novel indexes that reflect the invasiveness of SSNs [28]. Using radiogenomic analysis, we revealed the potential genomic grounds supporting the use of these parameters as important clinical references for diagnosis and treatment of SSNs. To summarise, when considering the total size of SSNs, only mutation in *EGFR* was remarkably different between subgroups stratified by this parameter. However, when we took the presence (pGGN or PSN), the size (solid size), and the weight (CTR, solid volume percentage, mass) of solid components into consideration, we found that more solid components were accompanied by significantly higher mutation frequencies in *EGFR*, *TP53*, *RBM10* and *ARID1B*. We revealed that the formation of the solid components of SSNs is potentially driven by additional mutations in cancer-related genes. Together with the relationship between pathological characteristics and genomic profiles of SSNs, higher mutation frequencies in *EGFR*, *TP53*, *RBM10* and *ARID1B* were proven to relate to a more invasive behaviour radiologically and pathologically.

References

1 Naidich DP, Bankier AA, MacMahon H, *et al.* Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society. *Radiology* 2013; 266: 304–317.

2 Campbell JD, Alexandrov A, Kim J, *et al.* Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat Genet* 2016; 48: 607–616.

3 Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014; 511: 543–550.

4 Imielinski M, Berger AH, Hammerman PS, *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 2012; 150: 1107–1120.

5 Wu K, Zhang X, Li F, *et al.* Frequent alterations in cytoskeleton remodelling genes in primary and metastatic lung adenocarcinomas. *Nat Commun* 2015; 6: 10131.

6 de Bruin EC, McGranahan N, Mitter R, *et al.* Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* 2014; 346: 251–256.

7 de Bruin EC, McGranahan N, Swanton C. Analysis of intratumor heterogeneity unravels lung cancer evolution. *Mol Cell Oncol* 2015; 2: e985549.

8 Inamura K. Clinicopathological characteristics and mutations driving development of early lung adenocarcinoma: tumor initiation and progression. *Int J Mol Sci* 2018; 19: E1259.

9 Kobayashi Y, Mitsudomi T, Sakao Y, *et al.* Genetic features of pulmonary adenocarcinoma presenting with ground-glass nodules: the differences between nodules with and without growth. *Ann Oncol* 2015; 26: 156–161.

10 Kakinuma R, Noguchi M, Ashizawa K, *et al.* Natural history of pulmonary subsolid nodules: a prospective multicenter study. *J Thorac Oncol* 2016; 11: 1012–1028.

11 MacMahon H, Naidich DP, Goo JM, *et al.* Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017. *Radiology* 2017; 284: 228–243.

12 Sun F, Xi J, Zhan C, *et al.* Ground glass opacities: imaging, pathology, and gene mutations. *J Thorac Cardiovasc Surg* 2018; 156: 808–813.

13 Kadara H, Choi M, Zhang J, *et al.* Whole-exome sequencing and immune profiling of early-stage lung adenocarcinoma with fully annotated clinical follow-up. *Ann Oncol* 2017; 28: 75–82.

14 Nahar R, Zhai W, Zhang T, *et al.* Elucidating the genomic architecture of Asian EGFR-mutant lung adenocarcinoma through multi-region exome sequencing. *Nat Commun* 2018; 9: 216.

15 Izumchenko E, Chang X, Brait M, *et al.* Targeted sequencing reveals clonal genetic changes in the progression of early lung neoplasms and paired circulating DNA. *Nat Commun* 2015; 6: 8258.

16 Alexandrov LB, Nik-Zainal S, Wedge DC, *et al.* Signatures of mutational processes in human cancer. *Nature* 2013; 500: 415–421.

17 Sanchez-Vega F, Mina M, Armenia J, *et al.* Oncogenic signaling pathways in the Cancer Genome Atlas. *Cell* 2018; 173: 321–337.

18 Vinayanuwattikun C, Le Calvez-Kelm F, Abedi-Ardekani B, *et al.* Elucidating genomic characteristics of lung cancer progression from *in situ* to invasive adenocarcinoma. *Sci Rep* 2016; 6: 31628.

19 Takamochi K, Takahashi F, Suehara Y, *et al.* DNA mismatch repair deficiency in surgically resected lung adenocarcinoma: microsatellite instability analysis using the Promega panel. *Lung Cancer* 2017; 110: 26–31.

20 Li GM. Mechanisms and functions of DNA mismatch repair. *Cell Res* 2008; 18: 85–98.

21 Ismail IH, Davidson R, Gagné JP, *et al.* Germline mutations in BAP1 impair its function in DNA double-strand break repair. *Cancer Res* 2014; 74: 4282–4294.

22 Clynes D, Higgs DR, Gibbons RJ. The chromatin remodeller ATRX: a repeat offender in human disease. *Trends Biochem Sci* 2013; 38: 461–466.

23 Watanabe R, Ui A, Kanno S, *et al.* SWI/SNF factors required for cellular resistance to DNA damage include ARID1A and ARID1B and show interdependent protein stability. *Cancer Res* 2014; 74: 2465–2475.

24 Li R, Li X, Xue R, *et al.* Early metastasis detected in patients with multifocal pulmonary ground-glass opacities (GGOs). *Thorax* 2018; 73: 290–292.

25 Zhang C, Zhang J, Xu FP, *et al.* Genomic landscape and immune microenvironment features of preinvasive and early-invasive lung adenocarcinoma. *J Thorac Oncol* 2019; 14: 1912–1923.

26 National Comprehensive Cancer Network. NCCN guide-lines for NSCLC. www.nccn.org/professionals/physician_gls Date last accessed: April 2, 2019.

27 Yano M, Yoshida J, Koike T, *et al.* Survival of 1737 lobectomy-tolerable patients who underwent limited resection for cStage IA non-small-cell lung cancer. *Eur J Cardiothorac Surg* 2015; 47: 135–142.

28 Wang XW, Chen WF, He WJ, *et al.* CT features differentiating pre- and minimally invasive from invasive adenocarcinoma appearing as mixed ground-glass nodules: mass is a potential imaging biomarker. *Clin Radiol* 2018; 73: 549–554.