Multi-region exome sequencing reveals the intratumoral heterogeneity of surgically resected small cell lung cancer

Small cell lung cancer (SCLC) is a highly malignant tumor which is eventually refractory to any treatment. Intratumoral heterogeneity (ITH) may contribute to treatment failure. However, the extent of ITH in SCLC is still largely unknown. Here, we subject 120 tumor samples from 40 stage I-III SCLC patients to multi-regional whole-exome sequencing. The most common mutant genes are TP53 (88%) and RB1 (72%). We observe a medium level of mutational heterogeneity (0.30, range 0.0~0.98) and tumor mutational burden (TMB, 10.2 mutations/Mb, range 1.1~51.7). Our SCLC samples also exhibit somatic copy number variation (CNV) across all patients, with an average CNV ITH of 0.49 (range 0.02~0.99). In terms of mutation distribution, ITH, TMB, mutation clusters, and gene signatures, patients with combined SCLC behave roughly the same way as patients with pure SCLC. This condition also exists in smoking patients and patients with EGFR mutations. A higher TMB per cluster is associated with better disease-free survival while single-nucleotide variant ITH is linked to worse overall survival, and therefore these features may be used as prognostic biomarkers for SCLC. Together, these findings demonstrate the intratumoral genetic heterogeneity of surgically resected SCLC and provide insights into resistance to treatment.
lung cancer is the most prevalent cancer in the world, with 15% of patients diagnosed with the highly aggressive and metastatic malignancy small cell lung cancer (SCLC). About one-third of SCLC patients present with limited disease (LD) and the remaining patients are diagnosed with extensive disease (ED) SCLC at the time of initial diagnosis. The 5-year overall survival (OS) rate for ED SCLC is below 7%. For SCLC patients, there has been no significant progress in the treatment modalities over the past decade. While the vast majority of patients are sensitive to chemotherapy and radiotherapy at the time of the initial treatment, all patients inevitably face the dilemma of chemoresistance and disease progression. Recently, immunotherapy was approved for the comprehensive treatment of ED SCLC. Yet, recurrence, drug resistance, and cancer as the cause of death are still common in the course of SCLC. How to improve a patient’s prognosis remains an unmet need for this recalcitrant malignancy.

An important factor in the failure of anticancer treatment is intratumor heterogeneity (ITH), which refers to distinct tumor cell populations (with different molecular and phenotypic profiles) within the same tumor specimen, resulting in differences in the tumor growth rate, invasion ability, drug sensitivity, and prognosis. Next-generation sequencing (NGS) technology has been widely used for tumor genome research. Elucidating the heterogeneity associated with prognosis, which provides a reference for sub-sequent cancer genome research. The tumor growth rate, invasion ability, drug sensitivity, and prognosis. The 5-year OS was 81% (range 76–85%) at the time of initial diagnosis. The 5-year OS (range 76–85%) at the time of initial diagnosis. The 5-year OS was 81% (range 76–85%) at the time of initial diagnosis.

In this study, we aim to provide the intratumoral genetic heterogeneity landscape of surgically resected SCLC, by analyzing the whole-exome sequencing data of 120 samples from 40 patients with SCLC. We characterize their mutational burden, heterogeneity, evolution, and potential biomarkers. Considerable intratumoral genetic heterogeneity is present among SCLC. We further identify several heterogeneity-related prognostic biomarkers.

### Results

**Patients’ characteristics.** We included 40 surgically resected SCLC patients in this study, among them, 6 were diagnosed with combined SCLC (C-SCLC). Most SCLCs (34/40) were pure SCLC (P-SCLC). Table 1 shows the patients’ clinical characteristics. The median age was 62 years old. Most patients were male (35, 87.5%) and had a history of smoking (31, 77.5%). All patients underwent surgery, with a median tumor size of 22.5 mm. About 65% of patients received further treatment after surgery. Fifteen patients (15, 38%) died after a median follow-up time of 22.82 months.

**Mutation landscape of 40 SCLC patients using multiple-regional sequencing.** We subjected 120 formalin-fixed paraffin-embedded (FFPE) SCLC samples (3 regions per patient) to MRS. In total, 33,153 non-silent somatic mutations were identified with an average 252× sequencing depth (Supplementary Data 1). We found an average of 340 mutations (range 33–1552) from multi-region for each patient. The median multi-region based tumor mutation burden (TMB) of SCLC was similar with single-region based TMB in our cohort and The Cancer Genome Atlas (TCGA) cohort (Supplementary Fig. 1a, Mann–Whitney–Wilcoxon test, both p > 0.05). There was a positive correlation between TMB and tumor neoantigen burden (TNB) (Spearman’s correlation coefficient, r = 0.59, p < 0.001; Supplementary Fig. 1b). The most frequent mutant genes were TP53 (88%) and RB1 (72%), which were clonal mutations; while LRP1B (22%), PCLO (15%), and KMT2D (15%) were subclonal mutations (Fig. 1a, Supplementary Fig. 2c, Supplementary Data 2). The C > T and C > A transversions were enriched in these patients (Supplementary Fig. 1c, d). The age-associated, BRCA1/2-associated, tobacco-associated, and aflatoxin-associated signatures were also major mutational signatures in these patients (Fig. 1a). The age-associated, aflatoxin-associated, and DNA repair-associated signatures were the top signatures in the branch, while the age-associated and smoking-associated signatures were major ones in the trunk (Supplementary Fig. 1e, f).

Non-silent mutation distribution showed ITH in patients with SCLC varied significantly (Fig. 1b). Percentages ranged from 17 to 100% (Fig. 1c). We found a medium mutational heterogeneity (0.30, quartile 0.12–0.56) in our SCLC cohort, and the SNV ITH of P-SCLC and C-SCLC were not significantly different with NSCLC of TRACERx study (p = 0.065 and p = 0.32) (Fig. 1c and Supplementary Fig. 2b). We also showed the distribution of mutations in ten common oncogenic signaling pathways (Supplementary Fig. 2g) and identified that mutations in the TP53 and RTK-Ras-ERK signaling pathways were predominantly clonal mutations.

**Intratumoral heterogeneity in CNV.** SCLC exhibited somatic arm-level CNV alterations including amplification at chromosomes 1, 12, 18, 19, 20, 3q, 5p, 6p, and 8q, and deletions at chromosomes 4, 10, 3p, 5q, 15q, 16q, 17q, 21p, and 11q (Fig. 2a, Supplementary Data 3–5). Significantly amplified regions

### Table 1 Clinical characterization of our SCLC cohort.

| Characteristics          | Total (n = 40) | P-SCLC (n = 34) | C-SCLC (n = 6) |
|--------------------------|---------------|----------------|---------------|
| Median age in years (range) | 62 (23–76)    | 64 (23–75)     | 63 (50–76)    |
| Sex (%)                  |               |                |               |
| Male                     | 35 (87.5%)    | 31 (91%)       | 4 (66%)       |
| Female                   | 5 (12.5%)     | 3 (9%)         | 2 (33%)       |
| Smoking (%)              |               |                |               |
| Non-smoker               | 8 (20%)       | 5 (15%)        | 3 (50%)       |
| Smoker                   | 31 (77.5%)    | 28 (82%)       | 3 (50%)       |
| NA                       | 1 (2.5%)      | 1 (3%)         | 0 (0%)        |
| Drinking (%)             |               |                |               |
| Never                    | 17 (42.5%)    | 13 (38%)       | 4 (66%)       |
| Drinking                 | 22 (55%)      | 20 (56%)       | 2 (33%)       |
| NA                       | 1 (2.5%)      | 1 (3%)         | 0 (0%)        |
| Median tumor size (range) | 22.5 (1.6–420) | 22.5 (1.6–420) | 22.83 (2.4–54) |
| Stage (%)                |               |                |               |
| I                        | 15 (37.5%)    | 12 (35%)       | 3 (50%)       |
| II                       | 7 (17.5%)     | 5 (15%)        | 2 (33%)       |
| III                      | 18 (45%)      | 17 (50%)       | 1 (16%)       |
| IV                       | 0 (0%)        | 0 (0%)         | 0 (0%)        |
| Treatment after surgery (%) |             |                |               |
| Yes                      | 26 (65%)      | 23 (68%)       | 3 (50%)       |
| No                       | 14 (35%)      | 11 (32%)       | 3 (50%)       |
| Status (%)               |               |                |               |
| Alive                    | 25 (62%)      | 20 (59%)       | 5 (83%)       |
| Dead                     | 15 (38%)      | 14 (41%)       | 1 (16%)       |
included 1p34.2 (HEYL), 1q21.3 (APH1A), 2p24.3 (MYCN), 3q29 (PIK3CA), 5p13.2 (IL7R), 6p22.3 (E2F3), 8q24.21 (MYC), and 9p24.1 (CD274, PDCD1LG2) as well as deleted regions 3p12.1, 4q13.2, 5q35.3, 9q21.11 (CBWD3), 10q23.31 (PTEN), 13q14.2 (RB1), 14q11.2, 15q25.3 (NTRK3), 19p12 (ZNF429), and 22q11.1 (Fig. 2b, c). Using CNV ITH, a median of 0.485 (range 0.02–0.99 per sector) was found in SCLC (Fig. 2d). Among them, IL7R, PIK3CA, SETDB1, TERT, SEPT9, MYC, CEBPA, and CD274 genes were amplified as frequently recurring clonal genes, while the clonal depleted genes like CBWD3, RB1, and PTEN were identified in our patients (Supplementary Fig. 2e).

**Clonal evolution and pathway enrichment.** We also constructed phylogenetic trees based on somatic mutations detected in multiple regions. Figure 3a shows the phylogenetic tree for each patient according to their disease stage. In particular, TP53, EGFR, and CREBBP mutations were common early clonal events involved in the evolution of SCLC (Fig. 3b), while RB1 and other mutations were late clonal events. Generally, among clonal and subclonal mutations, passenger mutations were proportionally higher than driver mutations (oncogene and TSG, Fig. 4e).

**Correlation between genetic alterations and clinical characterization.** No significant relationship was observed between ITH and other clinical variables, including pathology, smoking history, EGFR mutation status, and tumor stage (Fig. 4a, b, Supplementary Fig. 2d). Among the EGFR mutations, three patients carried non-classic EGFR mutations (p.G652W, p.E114Q, p.Q701L/p.R108K; Supplementary Data 6) and four had classic mutations (p.L858R and EX19del). Classic EGFR mutations were found in two (5.9%, 2/34) P-SCLC and two (33%, 2/6) C-SCLC patients, respectively. In our cohort, we found that all EGFR mutations co-occurred with TP53 inactivation and RB1 inactivation (mutation and/or loss) (Supplementary Data 6). The TP53/RB1/EGFR mutations were independent of clinical (tumor...
Fig. 2 Copy number alterations in our cohort. a Arm level CNVs identified by GISTIC2.0 in SCLC (n = 40). False discovery rate (FDR) corrected p value represents significant changes from Benjamini-Hochberg testing. b The genome chromosome plots depict significant cytobands identified by GISTIC2.0. c The significant somatic focal CNVs of pure SCLC and combined lung cancer are shown for each region of the individual patient. Cytobands with genes involved in cosmic drivers and those that occurred in at least 50% of patients are shown. d Counts in the trunk and branch of CNVs for each patient; Percentage of branch CNVs for each patient (n = 40). SCLC small cell lung cancer, Amp amplification, Del deletion, CNVs copy number variations.
Fig. 3 Phylogenetic trees and evolution in SCLC. a Phylogenetic trees for each patient (n = 40) stratified according to stages. b The evolution mode in all patients (n = 40). P-SCLC pure small cell lung cancer, C-SCLC combined small cell lung cancer, pre-GD pre-genome doubling.

stage and tumor size), and genomic features (TMB, ITH, and WGD) in SCLC (Supplementary Fig. 6a). Intriguingly, EGFR/RB1/TP53-mutant patients exhibited higher ploidy than those with wild-type (p = 0.017). And WGD occurred in all of the EGFR/RB1/TP53 mutant patients (Supplementary Fig. 6a). Besides, these mutations were not associated with disease-free survival (DFS) or OS in the absence or presence of treatment after surgery (Supplementary Fig. 6b, c).

Supplementary Fig. 3 and Fig. 5a show the basic clinicopathological information in this cohort. Patients with P-SCLC/C-SCLC, smoker/non-smoker, EGFR mutant/wild type had similar levels of ITH, TMB, and mutation clusters, and they exhibited no discrepancy in their gene signature and mutation landscape (Fig. 4, Supplementary Fig. 4b, c). Remarkably, a higher TMB/cluster correlated with better DFS using univariate analysis, while the SNV ITH was correlated to OS (Fig. 5b, c). However, no significant correlation was observed among DFS or OS and TMB, mutation cluster, or tumor stage (Fig. 5b, c, Supplementary Fig. 6d, e). In a multivariate analysis adjusted for age, tumor size, tumor stage, and smoking status, only TMB/cluster were
Fig. 4 The ITH and clinicopathological characteristics of SCLC. The comparison of a SNV ITH, b CNV ITH, c TMB, d average TMB per cluster between pure SCLC (n = 34) and combined lung cancer (n = 6), EGFR mutant (n = 7), and wild type (n = 33), as well as smoking (n = 31) and nonsmoking (n = 8) subgroups. p Value from two-sided Mann-Whitney U test. Boxplots are represented by a centerline, median; box limits, the 25th and 75th percentiles; whiskers extend represent the lower and upper values within 1.5 * inter-quartile range. e, f The proportion of driver genes, passenger genes, and other genes in the trunk and branch. p Value from two-sided Fisher’s exact test. SNV single-nucleotide variant, CNV copy number variation, ITH intratumoral heterogeneity, P-SCLC pure small cell lung cancer, C-SCLC combined small cell lung cancer, TMB tumor mutation burden, TSG tumor suppressor gene.
associated with better DFS, and SNV ITH is also linked to worse OS of SCLC (Fig. 5d, e).

All the cases with recurrence received systemic chemotherapy in our cohort. No ITH discrepancies were observed in patients according to the recurrence status and systemic chemotherapy (Supplementary Fig. 6f). ITH and TMB/cluster were not associated with survival outcomes in the recurrent cases ($p > 0.05$, $n = 11$, Supplementary Fig. 6g). Cases that received systemic chemotherapy had a superior overall outcome (Supplementary Fig. 6g), suggesting the favorable role of chemotherapy after surgery in the treatment of SCLC.
Discussion
Many SCLC patients are sensitive to initial treatment, but all patients inevitably face the dilemma of chemoresistance. It has been speculated that ITH is common in treatment-naïve SCLC, with many drug-resistant subclones. Yet, because of the lack of available tumor samples, this gap is still vacant in the field of SCLC research. Moreover, research in the field has mainly utilized traditional genomic sequencing of a single site which is unable to capture the full genomic landscape. Whereas MRS is superior in traditional genomic sequencing of a single site which is unable to including SNV ITH, CNV ITH, mutation cluster, and TMB per cluster. The forest plot showing multiple covariate Cox regression analysis of disease-free survival and overall survival by subgroups including age, smoking, tumor size, stage, and ITH in SCLC. A two-sided, unpaired, Wilcoxon rank test was performed for the statistical comparison among subgroups. WGD, whole-genome duplication; GIL genome instability index, MSI microsatellite instability, SNV single-nucleotide variant, CNV copy number variation, ITH intratumoral heterogeneity, P-SCLC pure small cell lung cancer, C-SCLC combined small cell lung cancer, TMB tumor mutation burden, HR hazard ratio, CI confidence interval.

EGFR mutations are a rare occurrence in either de novo SCLC or in cases of transformed EGFR-mutant (EGFR-mt) adenocarcinoma. In our study, the frequency of classic EGFR mutations in P-SCLC was 5.9%. These data were comparable with previous reports of 2.6% in Taiwanese and 2.0% in a Chinese cohort. Our EGFR-mutant SCLC patients did not receive EGFR-TKI therapy, and EGFR mutation status is not associated with recurrence after surgery (Supplementary Fig. 3d). An EGFR mutation is considered an early clonal event in our analysis. However, a lower driver dominant EGFR score did not support its role as a driver gene in SCLC, which is distinct from common NSCLC (Supplementary Fig. 4a). In other words, an EGFR mutation was not a predominant driver gene in SCLC. Currently, there is no targeted therapy in EGFR-mutant SCLC. The majority of de novo EGFR-mt SCLC are resistant to EGFR-TKI therapy, compared with EGFR-mt NSCLC, which may be due to focusing much more on the driver gene “EGFR” and neglecting of passenger mutations’ effect. EGFR passenger mutations may also collaborate synergistically with driver mutations to trigger tumorigenesis in SCLC. Previous researchers have shown that EGFR/RB1/TP53 are key events that transform NSCLC to SCLC after EGFR-TKI treatment. In our treatment-naïve SCLC cohort, we also found that all EGFR mutations co-occurred with TP53 and RB1 mutations. EGFR/ RB1/TP53 mutant patients had WGD events and exhibited higher ploidy than those with wild-type (Supplementary Fig. 6a). Yet, the TP53/RB1/EGFR mutations were independent of clinicopathologic features and not associated with prognosis. Based on the tumor evolutionary algorithm model proposed by Swanton et al., we conferred that TP53 and EGFR mutations were early events in the evolution of SCLC, while the RB1 mutation and loss occurred later, indirectly suggesting a key role of RB1 inactivation in SCLC evolution. However, this hypothesis needs validation in further studies.

We sought to explore the relationship between ITH scores and clinicopathological features. We were particularly interested in the six patients with C-SCLC in this study cohort. Comprehensive research showed that this group of patients behaved much in the same way as P-SCLC patients, both in terms of mutation distribution, ITH, TMB, mutation clusters, and gene signatures. This condition is also present in patients with EGFR mutations and those with a history of smoking. Among diagnosed SCLC patients, most patients have a history of smoking. We paid special attention to the evolutionary tree of non-smoking SCLC patients and found there was no obvious difference compared with smoker patients (Supplementary Fig. 5). To some extent, the intratumoral heterogeneity of the SCLC genome is independent of common clinicopathological features, such as pathological types, smoking history, and driver gene mutation status, but there is still a relatively uniform moderate level of intratumoral heterogeneity. A previous study reported widespread ITH in chemotherapy-treated SCLC and found that it may lead to poor treatment response and prognosis. We observed the same performance of SNG ITH in treatment-naïve LD SCLC patients. Multivariable COX analysis supported the independent prognostic role of SNG ITH for OS. We turned our perspective to another tumor heterogeneity assessment algorithm, TMB per cluster, which seems to be another potential prognosis biomarker. We found that more TMB per cluster is linked to early disease recurrence and progression. It indicated complex mutations inside the tumor may lead to the failure of anti-cancer treatment. Further research on its relationship with treatment sensitivity and resistance is needed.

Although our study presents several findings, there are several limitations. First, our results would have been more reliable with more patients from other centers. Related to our limited sample, we did not perform dynamic genome monitoring for each patient. We also did not provide a better understanding of the tumor microenvironment of SCLC. In addition, we should notice that the presence of technical noise in sequencing data is common, and genuine intratumor genetic heterogeneity is hard to distinguish from these sequencing artifacts. It may lead to the overestimation of ITH. Therefore, we used two mutation calling algorithms and strict criteria to filtering out these private artifacts, and to minimize the impact of artifacts. Due to the unavailability of the samples, we could not validate our results in the same sample. Nevertheless, further studies with high depth sequencing are required to accurately quantifying ITH.

Fig. 5 The relationship between heterogeneity and clinical characterization in SCLC. a A heatmap displaying the clinical information and genomic features for each patient (n = 40). The Kaplan–Meier plot depicts the estimation of disease-free survival (b) and overall survival (c) with parameters including SNV ITH, CNV ITH, mutation cluster, and TMB per cluster. The p value and hazard ratio were determined using the two-sided log-rank test. The forest plot showing multiple covariate Cox regression analysis of disease-free survival (d) and overall survival (e) by subgroups including age, smoking, tumor size, stage, and ITH in SCLC. A two-sided, unpaired, Wilcoxon rank test was performed for the statistical comparison among subgroups. WGD, whole-genome duplication; GIL genome instability index, MSI microsatellite instability, SNV single-nucleotide variant, CNV copy number variation, ITH intratumoral heterogeneity, P-SCLC pure small cell lung cancer, C-SCLC combined small cell lung cancer, TMB tumor mutation burden, HR hazard ratio, CI confidence interval.
We demonstrated the ITH landscape of surgically resected SCLC. Despite a moderate mutation burden, SCLC showed a medium intratumoral heterogeneity with high SNV and CNV ITH at the early stage, which may explain the difficult treatment dilemma faced by SCLC patients.

Methods

Patients and samples. Forty enrolled SCLC patients underwent thoracic surgery at Sun Yat-Sen Cancer Center between September 2009 and September 2018. The diagnosis of SCLC was confirmed by two pathologists via immunohis- tochemistry. None of the patients received any previous systemic anti-cancer therapy. We collected 120 surgically resected FFPE tumor tissues from 40 patients (3 tumor regions in different quadrants for each patient). A paired peripheral blood sample was obtained during the surgery. The study protocol was approved by the institutional review board of Sun Yat-Sen University Cancer Center. We have complied with all relevant ethical regulations for work with human participants, and that written informed consent was obtained.

Multi-region whole-exome sequencing. For each region of the patient, DNA was extracted from the FFPE kit (Promega) according to the manufacturer’s instructions. We constructed the sequencing libraries from native DNA using the zGen® Exome DNA library preparation kit (Exome Technologies, IA, USA) and the NEB Next Ultra DNA Library Prep Kit (Lot: NEB-0311611, NEB, UK) with a KAPA polymerase (KapaBiosystems, Wilmington, MA, USA). Whole-exome sequencing was performed using GeneSeq-2000 (GeneTech-Suzhou, Suzhou, China) and 100-bp paired-end sequencing. The data preprocessing and variant filtering were performed using the institutional review board of Sun Yat-Sen University Cancer Center. We have compiled with all relevant ethical regulations for work with human participants, and that written informed consent was obtained.

Somatic variant detection. Single nucleotide variants (SNVs) were called by the Somatic Genomics Pipeline (version somatic-genomics-201808)37 with parameters as follows (sentience driver -t 16 -r hs37d5.fa -algo VarCal -v SNP.vcf -resource_param 1000 G_phase1.snps.high_coverage.b37.vcf -resource_param 1000 G_omni2.5.b37.vcf -resource_param amni.known = false,training -true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 15.0 -resource dbsnp_138.b37.del100.vcf.gz -resource_param dbsnp,known = true,training -r hs37d5.fa -algo VarCal -v SNP.vcf -resource 1000 G_phase1.snp.high_confidence.b37.vcf -resource 1000 G_omni2.5.b37.vcf -resource_param amni.known = false,training -true,truth = false,prior = 2.0 -resource dbnp_138.b37.del100.vcf.gz -resource_param dbnp,known = true,training = false,prior = 0.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0 -resource hapmap_3.3.b37.pop.stratified.gz -resource_param hapmap known = false,training = true,truth = false,prior = 12.0

Tumor neoantigen detection. Tumor neoantigen was identified via netMHCPan (v4.0)35. Missense and nonsense mutations were correlated with the TNB counts using Spearman’s coefficient.

Mutational signature analysis. The mutational signatures were analyzed using deconstructSigs (v1.8.0) and MutationalPatterns (v2.0.0)36. The mutational signature contribution for each patient was compared with COSMIC SBS signature V2 (https://cancer.sanger.ac.uk/cosmic/signatures_v2.v26)

Classification of driver genes, oncogene, and tumor suppressor genes. Genes in the COSMIC cancer gene census (https://cancer.sanger.ac.uk/cosmic) were defined as driver genes. The oncogene and tumor suppressor genes (TSG) were classified based on the driver gene list.

Phylogenetic tree construction. All nonsilent somatic mutations excluding those co-localized within the LOH were used to construct phylogenetic trees via tools “ape” (v5.4-1), “phantom” (v2.5.3), and “ggtree” (v2.2.4)37. Phylogenetic trees were built on the basis of the binary presence/absence matrices obtained from the regional distribution of variants within the tumor. Trunk mutations occurred in all regions of the tumor. The length of each tree’s branch was calculated according to the number of mutations on each branch.

Cluster and timing of genomic alterations. All nonsilent somatic mutations were clustered by PyClone VI (https://github.com/Roth-Lab/pyclone-vi)38 and corrected by copy number and purity. The number of clusters identified by PyClone was defined as mutation clusters. The average TMB in each mutation cluster identified by PyClone VI was calculated as TMB cluster.

The timing of SNVs was determined by EstimateClonality (v1.0)10. Briefly, we estimated the cellular prevalence of somatic mutations based on tumor purity and CNV and mutation copy number. Early mutations were defined as a mutation copy number of >1, whereas, late ones were classified as a mutation copy number of <1. The mutations in neutral copy number clusters were clustered by sciClone (v1.1.0)39, then the results were used for evolution estimation through ClonEvol (v0.99.11)40 and plotted by fishplot41.

CVN gain was timed by the average mutation copy number of at least five mutations within each segment. The CVN gain was defined as “early” if the average mutation copy number was >1, and “late” if it was <1. Regarding CNV loss, clonal CNV loss coupled with genome doubling was classified as “early”, whereas, CNV loss unrelated to genome doubling was classified as “late”.

ITH evaluation. Clonal SNV/indels were defined as mutations in the PyClone VI cluster with a maximum cellular prevalence, while other SNV/indels in each tumor were defined as subclonal ones. SNV ITH was calculated by the number of subclonal mutations to all mutations.

CVN ITH was evaluated for each patient based on the presence of each CNV in different tumor regions with more than one variation and presented as the mean Jaccard distance among variation sets of each three regions42. ITH ranged from 0 to 1 (all branch events to all trunk events).

Comparison with published multi-regional whole-exome sequencing data. To compare the genomic heterogeneity between SCLC and NSCLC, the multi-regional WES data for NSCLC of the TRACERx study was downloaded10, and the SNV ITH was recalculated for each sample using the same algorithm.

Driver dominant score calculation. We calculated the driver dominant score, which measures the number of co-occurring drivers for each defined driver gene per tumor as Eq. (1)33. The ratio of patients carrying driver genes to the total number of patients was defined as an occurrence as Eq. (2). We downloaded the significant mutations for lung adenocarcinoma cancer (n = 10) and lung squamous cancer (n = 44)34,44. The driver genes were obtained from the mutation genes in lung adenocarcinoma and lung squamous cancers with q value < 0.1 by MutSig2CV results.

\[
\text{Dominant score} = \sum_{i=1}^{n} (\text{Frequency}) 
\times 1 \times \text{Frequency}
\]

\[
\text{Occurrence} = \frac{\text{Frequency}}{n}
\]

where n means the total number of patients of the cohort. The frequency represents the number of patients with the driver gene i. m means the number of driver genes.
Statistical analysis. The Mann–Whitney–Wilcoxon test was used to compare the continuous numbers in different groups. Fisher’s exact test was performed to analyze differences between proportional data. The Kaplan–Meier curve between clinical features and survival was performed using “survminer” (v0.4.7) and “survival” (v3.2-10) packages. The cutoff values for the two groups were determined by the best cutoff point for each parameter, excluding TMB. TMB was classified by an upper quantile value in all patients (\( n = 40 \)). The statistical significance was calculated using the Cox proportional hazards regression model and log-rank test for DFS and OS. All statistical analyses were performed with R v4.0.0 software. Statistical significance was defined as a two-sided \( p < 0.05 \).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The raw sequencing data generated in this study have been deposited in the GSA-Human (Genome Sequence Archive for Human in BIG Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences, http://gsa.big.ac.cn/gsa-human) under the accession code HRA000441. The data are available under controlled access. Access to the data may be requested by completing the application form via GSA-Human System and is granted by the corresponding Data Access Committee. The approximate response time for accession requests is about 10 working days. Additional guidance can be found at the Genomes Project [https://www.internationalgenome.org/data-portal/data-collection/phase-3], HapMap3, dbSNP, and ExAC. TCGA mutation data were downloaded from https://www.cbioportal.org/datasets. TRACERx data can be obtained from https://www.cbioportal.org/study/summary?id=tcga_cancer. Lung adenocarcinoma and lung squamous cancer can be obtained from https://www.cbioportal.org/study/summary?id=tcga_cancer. The continuous numbers in different groups. Fisher’s exact test was performed using the survminer, ggplot2, and gridR packages. The cutoff values for the two groups were determined by the best cutoff point for each parameter, excluding TMB. TMB was classified by an upper quantile value in all patients (\( n = 40 \)). The statistical significance was calculated using the Cox proportional hazards regression model and log-rank test for DFS and OS. All statistical analyses were performed with R v4.0.0 software. Statistical significance was defined as a two-sided \( p < 0.05 \).

Code availability

All custom code used in this work is available from https://github.com/LiyianJi-code/SLCCL_MRS.

Received: 2 September 2020; Accepted: 23 August 2021.
Published online: 14 September 2021

References

1. Gao, S. et al. DNA copy number and clonal heterogeneity in stage IIIA lung adenocarcinoma. *Cell* **164**, 883–904 (2016).
2. Zhang, M. et al. Concurrent RB1 and TP53 alterations define a subset of EGFR-mutant lung cancers at risk for histologic transformation and inferior clinical outcomes. *J. Thorac. Oncol.* **14**, 1784–1793 (2019).
3. Zhai, H., Moore, D. & Jamal-Hanjani, M. Inactivation of R1B and histological transformation in EGFR-mutant lung adenocarcinoma. *Ann. Oncol.* **31**, 169–170 (2020).
4. Shi, W. et al. Reliability of whole-exome sequencing for assessing intratumor genetic heterogeneity. *Cell Rep.* **25**, 1446–1457 (2018).
5. Callari, M. et al. Intersect-then-combine approach: improving the performance of somatic variant calling in whole exome sequencing data using multiple aligners and callers. *Genome Biol.* **9**, 35 (2017).
6. Cai, L., Yuan, W., Zhang, Z., He, L. & Chou, K. C. In-depth comparison of somatic point mutation callers on different tumor next-generation sequencing depth data. *Sci. Rep.* **6**, 36540 (2016).
7. Kendig, K. et al. The Sensitive DNAseq variant calling workflow demonstrates strong computational performance and accuracy. *Front. Genet.* **10**, 736 (2019).
8. McLaren, W. et al. The ensemble variant effect predictor. *Genome Biol.* **17**, 122 (2016).
9. Niu, B. et al. MSI-sensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* **30**, 1015–1016 (2014).
10. Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, e131 (2016).
11. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and concrete identification of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
12. McGaraghan, N. et al. Clonal status of actionable driver events and the timing of efficiencies and patterns of carcinoma evolution. *Sci. Transl. Med.* **7**, 283ra254 (2015).
13. Nahar, R. et al. Elucidating the genomic architecture of Asian EGFR-mutant lung adenocarcinoma through multi-region exome sequencing. *Nat. Commun.* **9**, 216 (2018).
14. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
15. Andreattu, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* **32**, 511–517 (2016).
16. Rosenthal, R., McGaraghan, N., Herrero, J., Taylor, B. S. & Swanton, C. DeconstructingNGS: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31 (2016).
17. Yu, G., Lam, T.-Y., Zhu, H. & Guan, Y. Two methods for mapping and visualizing associated data on phylogey using Gt recognizes. *Mol. Biol. Evol. 35*, 3041–3043 (2018).
18. Gillios, S. & Roth, A. PhyClone-VI: scalable inference of clonal population structures using whole genome data. *BMC Bioinform.* **21**, 571 (2020).
19. Miller, C. A. et al. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLoS Comput. Biol.* **10**, e1003665 (2014).
20. Dang, H. K. et al. CloneEv: clonal ordering and visualization in cancer sequencing. *Ann. Oncol.* **28**, 1076–1082 (2017).
21. Miller, C. A. et al. Visualizing tumor evolution with the fishplot package for R. *BMC Genomics* **17**, 880 (2016).
42. Zhang, Y. et al. Intratumor heterogeneity comparison among different subtypes of non-small-cell lung cancer through multi-region tissue and matched ctDNA sequencing. *Mol. Cancer* **18**, 7 (2019).

43. Cancer Genome Atlas Research, N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).

44. Cancer Genome Atlas Research, N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).

**Acknowledgements**

We thank all patients and researchers involved in this study. We are grateful to Mr. Christopher Lavender of Sun Yat-sen University Cancer Center for his editing assistance. This work was supported by the Natural Science Foundation of Guangdong Province (Grant no. 2020A151501129) and the Medical Scientific Research Foundation of Guangdong Province, China (Grant no. A2020153).

**Author contributions**

Study concept and design: Yaxiong Zhang, Ningning Zhou, and Huaqiang Zhou. Acquisition of data: Huaqiang Zhou, Liyan Ji, Hui Pan, Ting Zhou, Lanjun Zhang, Hao Long, Jianhua Fu, Zhesheng Wen, Siyu Wang, Xin Wang, Peng Lin, Haoxian Yang, and Junye Wang. Methods development: Liyan Ji, Mengmeng Song, Xin Yi, Ling Yang, Xuefang Xia, Yanfang Guan, and Pansong Li. Analysis of data: Huaqiang Zhou, Liyan Ji, Hui Pan, Yuanjuan Zhao, Yaxiong Zhang, and Ningning Zhou. Interpreting findings: Huaqiang Zhou, Yi Hu, Rongzhen Luo, Yuanjuan Zhao, Wenfeng Fang, Yunpeng Yang, Shaodong Hong, Yan Huang, Yaxiong Zhang, and Ningning Zhou. Drafting of the paper: Huaqiang Zhou, Yi Hu, Rongzhen Luo, Liyan Ji, and Yaxiong Zhang with the input of all authors. Critical revision of the paper for important intellectual content: All authors.

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-25787-x.

**Correspondence** and requests for materials should be addressed to Yaxiong Zhang or Ningning Zhou.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.