IJC
International Journal of Cancer

*Short Report*

# Multiregion sequencing reveals the intratumor heterogeneity of driver mutations in TP53-driven non-small cell lung cancer

Le-Le Zhang[1], Mengyuan Kan[1], Man-Man Zhang[1], Sha-Sha Yu[2], Hui-Jun Xie[2], Zhao-Hui Gu[2], Hai-Ning Wang[2], Shuang-Xia Zhao[1], Guang-Biao Zhou[3*], Huai-Dong Song[1,2*] and Cui-Xia Zheng[1*]

[1] Department of Respiration, Center Lab in Research Center for Clinical Medicine, Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China
[2] State Key Laboratory of Medical Genomics, Shanghai Jiaotong University School of Medicine, Shanghai, China
[3] State Key Laboratory of Biomembrane and Membrane Biotechnology, Institute of Zoology, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, China

**Intratumor heterogeneity (ITH) in non-small cell lung cancer (NSCLC) may account for resistance after a period of targeted therapies because drugs destroy only a portion of tumor cells. The recognition of ITH helps identify high-risk patients to make effective treatment decisions. However, ITH studies are confounded by interpatient heterogeneity in NSCLC and a large amount of passenger mutations. To address these issues, we recruited NSCLC patients carrying *TP53* mutations and selected driver mutations within recurrently mutated genes in NSCLC. A total of 12-paired normal-tumor tissues were subjected to whole-genome/whole-exome sequencing. From these, 367 non-silent mutations were selected as driver mutations and deeply sequenced in 61 intratumoral microdissections. We identified a universal prevalence of heterogeneity in all 12 tumors, indicating branched evolution. Although *TP53* mutations were observed in single biopsy of all 12 tumors, most tumors consist of both *TP53* mutated and non-mutated cells in separate regions within the same tumor. This suggests the late molecular timing of the acquisition of *TP53* mutations; therefore, the detection of *TP53* mutations in a single biopsy may simply not reflect the early malignant potential. In addition, we identified regions of loss of heterozygosity surrounding *TP53* and *CDKN2A* mutations in tumor 711, which also exhibited heterogeneity in different regional samples. Because the ITH of driver mutations likely has clinical consequences, further efforts are needed to limit the impact of ITH and to improve therapeutic efficiency, which will benefit NSCLC patients receiving targeted treatments.**



**Key words:** intratumor heterogeneity, non-small cell lung cancer, TP53 driver mutation, multiregion sequencing

**Abbreviations:** ITH: intratumor heterogeneity: LOH: loss of heterozygosity: NSCLC: non-small cell lung cancer

Additional Supporting Information may be found in the online version of this article.

L.L.Z., M.K and M.M.Z. contributed equally to this work

**Grant sponsor:** National Natural Science Foundation of China; **Grant numbers:** 81472177, 31571296, 81500602, 81430019, 81370888; **Grant sponsor:** National Basic Research Program of China (973); **Grant number:** 2012CB517604

**DOI:** 10.1002/ijc.30437

**History:** Received 21 Aug 2016; Accepted 31 Aug 2016; Online 20 Sep 2016

***Correspondence to**: Cui-Xia Zheng, Shanghai Ninth People's Hospital, 639 Zhizaoju Road, Shanghai 200011, China, E-mail: zcx9566@163.com; Tel: +86 21 23271699 extend to 5816; Huai-Dong Song, Shanghai Ninth People's Hospital, Building 12, 833 Zhizaoju Road, Shanghai 200011, China, E-mail: huaidong_s1966@163.com; Tel: +86 021 23271699 extend to 6525; Guang-Biao Zhou, Institute of Zoology, 1 West Beichen Road, Beijing 100101, China, E-mail: gbzhou@ioz.ac.cn; Tel: +86 010 64807951

## Introduction

Lung cancer is one of the leading causes of cancer-related mortality worldwide.[1] Non-small cell lung cancer (NSCLC) accounts for >80% of lung cancer cases. The majority of NSCLC patients suffer from metastases, relapse or drug resistance, even after a few years of treatment. Evidence shows that during targeted treatment, intratumor heterogeneity (ITH) may foster tumor adaptation and drug resistance through the selection of subclonal populations.[2] For example, the anti-cancer drug gefitinib, a type of epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor, is effective in only cancers with *EGFR* mutations. This led to a reduced response to gefitinib because the drug destroys only a portion of the tumor cells, thereby decreasing overall survival.[3,4] Therefore, it is of great clinical importance to design effective personalized therapy for NSCLC based on knowledge regarding ITH.

According to the tumor clonal theory proposed by Nowell,[5] cancer is an evolutionary process that parallels Darwinian evolution. Subclonal populations with multiple driver mutations conferring selective growth advantage expand during tumor progression, which shows branched evolution

 

**What's new?**

**Intratumor heterogeneity (ITH) in non-small cell lung cancer (NSCLC) may account for resistance during targeted treatment. Identifying ITH could help make effective treatment decisions, but ITH studies have so far been hampered by confounding factors. To reduce the interference of interpatient heterogeneity and passenger mutations, the authors recruited NSCLC patients with TP53 mutations and selected potential driver mutations. They identified extensive ITH of driver mutations in all NSCLC tumors and demonstrated the late molecular acquisition of TP53 mutation during tumor progression in most patients. The data highlight the clinical potential of limiting the impact of ITH for NSCLC targeted treatment.**



leading to ITH. Recent studies demonstrated the universal prevalence of ITH in various tumors including NSCLC.[6–14] Two studies sequenced multiregional samples from individual NSCLC tumors using next generation sequencing technology and revealed clonal evolution by characterizing somatic mutations that were either ubiquitous in all regions (trunk mutations) or not in all regions (branch mutations).[8,14] Although most known cancer gene mutations in NSCLC tumors are trunk mutations, few cancer driver genes are shared among patients, which may be due to the complexity of NSCLC subtypes. The lack of driver genes in common increases the difficulty of controlling interpatient heterogeneity. To address this issue, *TP53* was realized. As a well-characterized tumor suppressor gene, *TP53* is frequently mutated in lung cancer, contributing to 37% of adenocarcinoma (AD) and 44% of squamous cell carcinoma (SCC) patients from the COSMIC database. Mutated *TP53* plays an important role in the tumorigenesis of lung epithelia cells. We identified NSCLC patients carrying *TP53* hotspot mutations to refine our study under the same molecular subtype, and we also focused on whether the acquisition of *TP53* mutations is an early event in all NSCLC tumors. Moreover, previous studies that included all somatic mutations or somatic non-silent mutations could introduce many passenger mutations, which have little effect on cancer process. To eliminate a large number of passenger mutations from analyses and to obtain results with greater clinical implication, we selected potential NSCLC driver mutations from a variety of somatic non-silent mutations. The objectives of our study were as follows: (1) determine whether a single tumor exhibits the ITH of driver mutations in NSCLC; and (2) determine whether the evolutionary behavior of *TP53* mutations is consistent in different *TP53*-driven NSCLC patients. The workflow is shown in Figure 1.

## Material and Methods

### Patients

Tumor samples from 12 NSCLC patients (Supporting Information Table S1) were surgically removed and confirmed with >80% tumor content by experienced lung cancer pathologists using an hematoxylin and eosin (H&E)-stained slide. The adjacent normal tissues were free of tumor cells. All tumor samples were analyzed for *TP53* hotspots using Sanger sequencing (Supporting Information Table S2).

### Whole-exome/whole-genome sequencing (WGS/WES) and somatic non-silent mutation

A total of 12-paired tumor and adjacent normal tissues underwent either WGS or WES with an equal number (Supporting Information Table S1). WGS for six-paired samples was previously reported, and the data were obtained with an average depth of 65X for tumor samples and 42X for normal samples.[15] For the remaining six-paired samples, exome enrichment and paired-end sequencing were performed (Supporting Information Note S1), with an average depth of 70X for tumor samples and 67X for normal samples (Supporting Information Table S3). Raw data were processed for alignments, duplication removal and recalibration steps. Single nucleotide variants (SNVs) were called and annotated. Somatic non-silent mutations were selected based on their predicted functional consequence and pathogenic potentiality (Supporting Information Note S2). Finally, we identified 2,342 somatic functional mutations within 1,975 genes in 12 NSCLC samples (Supporting Information Table S4).

### NSCLC driver gene selection

In brief, potential NSCLC driver genes were selected as follows: (1) genes that are significantly mutated in NSCLC samples (collected from the COSMIC database, v68) compared to the gene mutated frequency in 87 healthy controls (Supplementary Information Note S3); or (2) genes with a mutated frequency ≥7% in the COSMIC database. We then compared these selected genes with 1,975 mutated genes identified from the WGS/WES. A total of 130 mutated genes from the WGS/WES data overlapped with the potential driver genes selected from step 1, and an additional 93 genes overlapped with the potential driver genes selected from step 2 (Supporting Information Table S5). In total, 367 non-silent somatic mutations within these 223 genes were considered driver mutations to be sequenced in multiple intratumoral regions.

### Multiregion-targeted sequencing

We obtained 3–8 separate tumor regions from each single tumor using laser capture microdissection (LCM) (Supporting Information Table S1, Note S4). Seventy-three intratumoral regions were collected and further reviewed and confirmed by experienced lung cancer pathologists (Supporting Information Fig. S1, Table S1). A pool of multiplex PCR primers for the 367 selected driver mutations was designed for targeted enrichment. 12 of 73 dissections failed to amplify

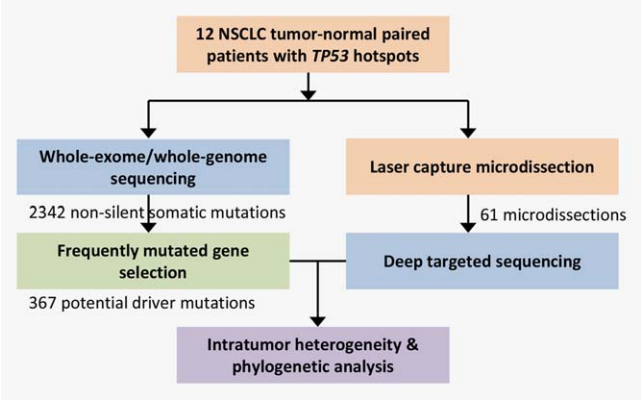


**Figure 1.** Workflow for multiregion sequencing. Twelve-paired normal-tumor tissues were collected from *TP53*-driven NSCLC patients and subject to WGS/WES. 2,342 non-silent somatic mutations were identified from WGS/WES. 367 of them within selected frequently mutated genes were considered driver mutations. 61 regional samples were obtained using LCM and subjected to deep sequencing which targeted those 367 driver mutations. Intratumor heterogeneity of the driver mutations was investigated and phylogenetic analysis was performed

and were not used for deep-targeted sequencing. Paired-end sequencing was applied to all dissected samples to an average depth of 7,582X per mutation site (Supporting Information Note S5). A total of 295 variants were identified after the variant calling (Supporting Information Table S6) and filtration steps described in Supporting Information Note S5.

### Phylogenetic analysis

All 295 driver mutations validated from the multiregion-targeted sequencing were used in the phylogenetic analysis to designate trunk (early) and branch (late) mutations. Mutation sites were coded as 1 for being mutated and 0 otherwise. We excluded mutation sites without mutated alleles in all intratumoral regions or with missing values in $\geq$30% of genotype calls in the regional samples from the analysis. The phylogenetic relationship was implemented using PHYLIP (version 3.695) with the Wanger parsimony method. The mutation status of adjacent normal tissues was used as an outgroup. Trees were redrawn with the length of trunks and branches proportional to the number of mutations.

### Hierarchical clustering

Variant allele frequency (VAF) was computed using the read depth of the mutated allele *vs.* the total read depth for a mutation site. Both intratumoral regions and driver mutations were clustered by the Euclidean distance according to their VAFs implemented with R. Mutation sites without mutated alleles in all intratumoral regions or with missing values in $\geq$30% of genotype calls in regional samples were excluded from the analysis.

## Results

### ITH prevalence in NSCLC tumors

There was clear evidence of the spatial heterogeneity of driver mutations in all 12 NSCLC tumor tissues, with a mean of 88.6% heterogeneous mutations (range 33.3–100%; Supporting Information Fig. S2, Table S7). Trunk and branch mutations in phylogenetic trees demonstrated the various molecular timing of acquisition of driver mutations (Fig. 2, Supporting Information Figs. S3 and S4), supporting a model of branched evolution during tumor progression, resulting in ITH, consistent with other NSCLC studies.[8,14] Only a small proportion of driver mutations were mapped to the trunk (mean: 11.4%, range 0–66.7%; Supporting Information Table S7) compared to proportions in previous studies (mean: 70%[8] and 76%[14]). We did not observe significant differences in the proportion of late mutations in smokers and non-smokers ($p > 0.05$, Supporting Information Fig. S2).

### Early or late molecular timing of the acquisition of *TP53*

Our study also revealed the spatial heterogeneity of *TP53* mutations in the same tumor. As a selected marker in common, *TP53* mutations were identified in all 12 single tumor tissues by WGS/WES with sufficient VAF (Fig. 3a), which included different mutation types, i.e. missense, nonsense and splicing sites (Fig. 3b, Supporting Information Table S2). In only three tumors (711, 823 and BB22) *TP53* mutations were ubiquitous in all intratumoral regions, indicating an early event during tumor formation and progression (Figs. 2a, 3c and 3d). In the remaining nine tumors, *TP53* mutations were present in some of the intratumoral regions as branch mutations, reflecting the relatively late timing of acquisition (Figs. 2b, 3c and 3d, Supporting Information Fig. S3). Interestingly, in tumor 781, *TP53* mutation was entirely absent in all intratumoral regions (Fig. 3d, Supporting Information Fig. S3*b*), which we also designated as branch mutations (Fig. 3c) because the mutation may occur in other regions in the same tumor that were not dissected. The heterogeneity of *TP53* mutations suggests that mutations in canonical cancer genes such as *TP53* may not be simply considered a marker of early malignant potential.

### Molecular timing in known cancer genes

We further characterized other known cancer genes in the phylogenetic trees (Fig. 2, Supporting Information Figs. S3, S4 and Table S8). Most tumors displayed heterogeneous populations of these known cancer gene mutations, indicating a late mutation event, such as *TP53* mutation (Fig. 2, Supporting Information Fig. S3). For example, we found *CDKN2A* mutation in tumor 711 present in 7 regional samples while absent in one region. Another example is *TTN* gene which was frequently mutated in our sample set (6/12 tumors). Only in two tumors (803 and 823), we found *TTN* mutations present in all intratumoral regions, whereas in the other four tumors (711, 759, 829 and RJ11), *TTN* mutations only occurred in some regional samples as branch mutations. Previous studies of NSCLC reported that most known cancer genes mutations mapped to the trunks; therefore, single-region sampling may be sufficient to identify the majority of

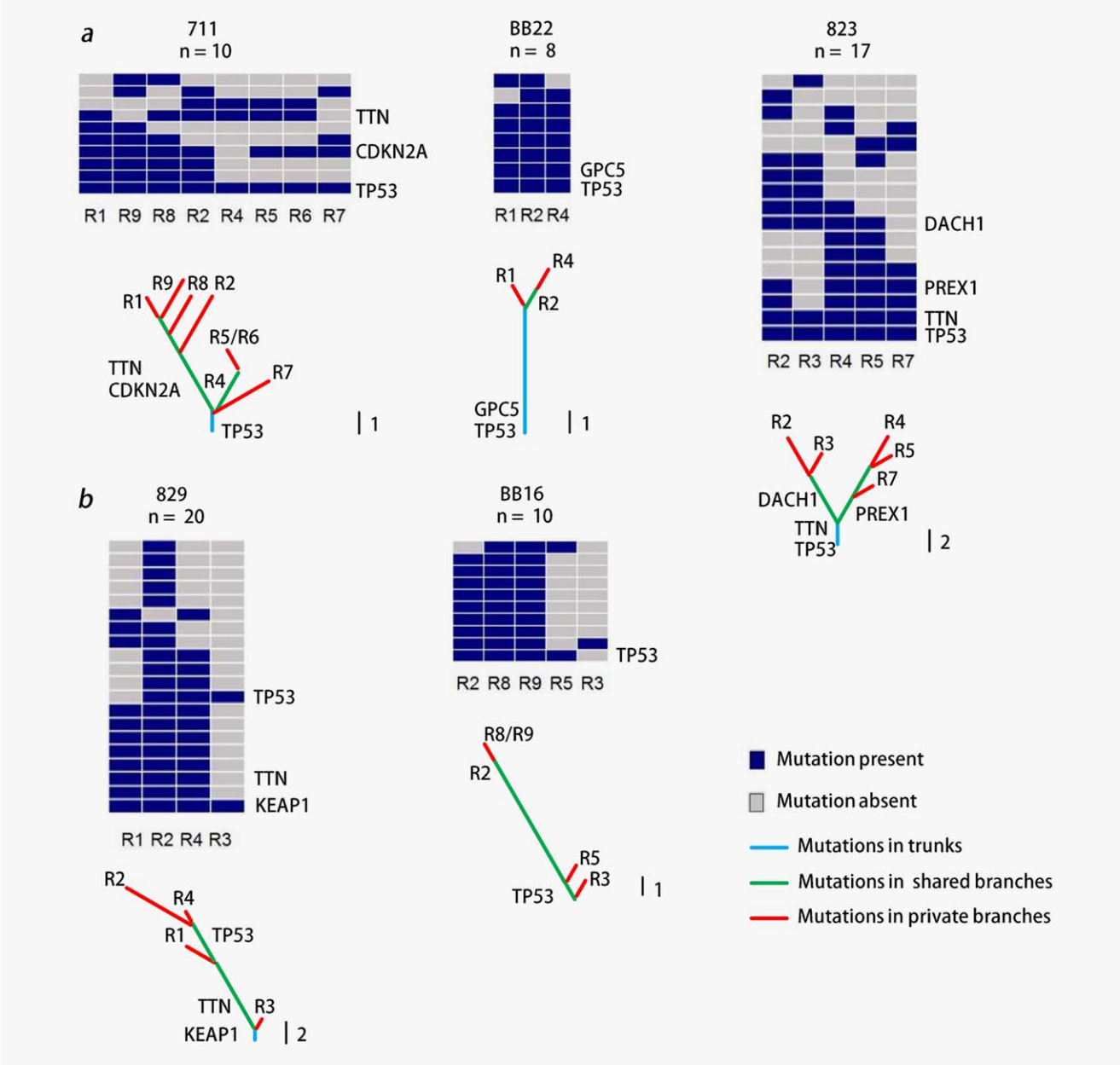


**Figure 2.** Intratumor heterogeneity of driver mutations in five NSCLC tumors with *TP53* mutations mapped to (*a*) trunk and (*b*) branch. Heatmaps in the upper panel show the regional distribution of driver mutations; present (blue), absence (gray). The total number of driver mutations (*n*) is provided for each tumor. Phylogenetic trees in the bottom were re-drawn based on the evolutionary relationship between regional samples. Trunk (blue), branch (green) and private branch (red) lengths are proportional to the number of driver mutations. The scale bar indicates the number of driver mutations for each sample. Known cancer gene mutations are annotated in the corresponding positions; those that were absent in all the regional samples are presented in gray in the phylogenetic trees.

these genes,[8,14] although our results were not consistent with this.

### Mutation spectra of trunk and branch mutations

We examined the timing of mutational processes during NSCLC evolution and observed a predominant proportion of C > A changes (52.7%) compared to other types of nucleotide changes (Supporting Information Fig. S5*a*). Although C > A changes are attributed to the mutagenic effects of tobacco smoke,[16] we did not observe significant changes in the proportions of total C > A transversions between smokers and non-smokers ($p > 0.5$, one-way ANOVA and thereafter). Interestingly, 7 of 12 NSCLC tumors had C > A transversions in the *TP53* gene (Supporting Information Table S2), consistent with a previous study showing that C > A transversions in the *TP53* gene are the most common change.[16] When comparing trunk (early) *vs.* branch (late) mutations, we observed a strong enrichment of C > T changes in branch

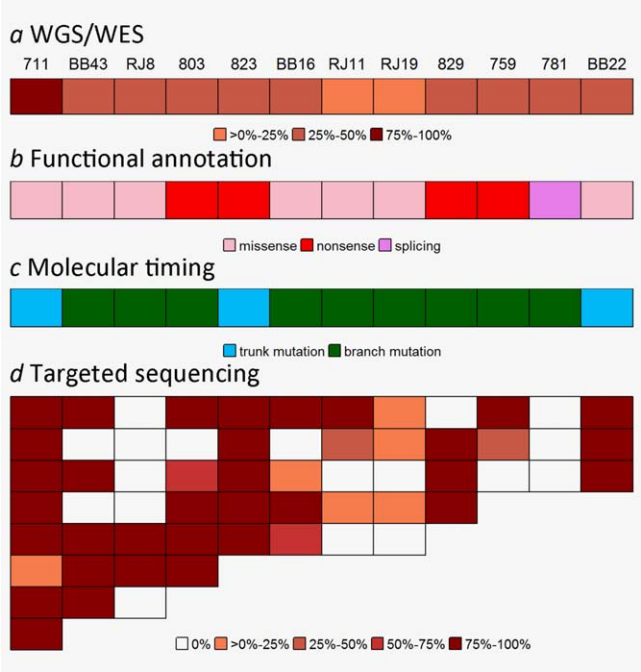


**Figure 3.** *TP53* mutational profiles in 12 NSCLC tumors. (*a*) VAF for *TP53* mutations identified from WGS/WES, (*b*) functional annotations, (*c*) molecular timing characterized by trunk (early) or branch (late) mutations and (*d*) VAF for *TP53* mutations identified from multiregion targeted sequencing in each intratumoral region. Tumor samples are ordered by the number of dissected regions obtained from a single tumor

($p = 0.005$) compared to trunk mutations (Supporting Information Fig. S5*b*). We also identified a decrease in the proportion of C > A changes in non-smokers, whereas there was an increase in smokers; however, the changes between early and late mutations were not significantly different ($p > 0.5$, Supporting Information Fig. S5*b*).

### Clonal similarity in intratumoral regions

VAF profiles of driver mutations showed both spatial heterogeneity and clonal similarity (Supporting Information Fig. S6). In some tumors, driver mutations in known cancer genes are likely to be clustered in the same branch, e.g., mutations in *TP53*, *CDKN2A* and *TTN* in tumor 711 and mutations in *TP53*, *TTN* and *DACH1* in tumor 823. Regional samples in some tumors may be clustered (e.g., BB16-2, BB16-8 and BB16-9, RJ8-5 and RJ8-6). Moreover, these clustered regions are not necessarily from neighboring dissected regions (Supporting Information Fig. S1*a*).

### Regions of loss of heterozygosity for *TP53* and *CDKN2A* in tumor 711

We observed the most driver mutations from multiregion sequencing with a higher VAF (≥75%) in the intratumoral regions (Supporting Information Fig. S6). Mutations with a high VAF were usually identified as homozygous mutations using Sanger sequencing because the proportion of non-mutated alleles was too small to be detected. We validated 23 mutations with a VAF ≥75% by Sanger sequencing, which are likely homozygous mutations, and obtained 100% agreement.

We hypothesized that these mutations may be located in the regions of loss of heterozygosity (LOH). Because tumor 711 has driver mutations in two tumor suppressor genes, *TP53* and *CDKN2A*, both of which are heterogeneous in regional samples and with a VAF ≥75% (Supporting Information Fig. S7*a*), we then investigated these potential regions of LOH. First, *TP53* and *CDKN2A* mutations in all intratumoral regions in 711 were validated using Sanger sequencing (Supporting Information Fig. S7*b*). Second, we selected markers for regions of LOH that are heterozygous single nucleotide polymorphisms (SNPs) in normal tissues and are located within 20 kb distance to the *TP53* and *CDKN2A* mutations. These SNPs were genotyped by Sanger sequencing in all intratumoral regions, and those with a genotype missing rate >30% were excluded (Supporting Information Fig. S7*c* and Table S9). Surprisingly, all heterozygous markers appear to be homozygous in the intratumoral regions (Supporting Information Fig. S7*d*); therefore, the studied gene regions may be haploid, and the deletion occurs in only tumor cells and not in normal tissue cells. Moreover, genotypes of these homozygous SNPs are the same, regardless of whether the *TP53* or *CDKN2A* mutated allele is present in the intratumoral regions. It is likely that tumor cells in tumor 711 lost one copy of *TP53* or *CDKN2A* and then acquired a deleterious mutation in the remaining allele as the second hit after the deletion event (Supporting Information Figs. S7*c* and S7*d*). Furthermore, both of the regions of LOH also displayed heterogeneous cell populations with both *TP53*/*CDKN2A* mutated and non-mutated cells in the same tumor 711.

## Discussion

Tumor heterogeneity is a well-established phenomenon in different types of cancers.[6–10,13,14] We performed multiregion sequencing of 12 localized *TP53*-driven NSCLC tumors. Our study revealed the universal prevalence of ITH in NSCLC, demonstrating a model of branched evolution, consistent with previous NSCLC studies.[8,14] We observed that most tumors have a larger proportion of late mutations compared to early mutations, inconsistent with previous studies.[8,14] This may be because we selected potential driver mutations from non-silent mutations and only included these in our study. This indicates that the transformation from normal cells to tumor cells in different regions is likely triggered by various driver mutations in the same tumor.

In most studies documenting ITH in malignancies, driver mutations within well-known cancer genes are likely to be present in the trunks of phylogenetic trees,[8,13,14] except for one study of clear cell renal carcinoma.[10] This is consistent with the hypothesis that driver mutations are required for tumorigenesis and progression.[17] In our study, a mutated

driver gene *TP53* carried by all 12 NSCLC patients was investigated. Although 12 single biopsies were confirmed with *TP53* mutations prior to multiregion sequencing, most (9/12 tumors) showed heterogeneity with both *TP53* mutated and non-mutated cells in separate small regions from the same tumor, which indicates the late timing of the acquisition of *TP53* mutations during tumor progression. The detection of *TP53* mutations from a single tumor biopsy may not sufficiently reflect the early malignant potential of the tumor. Except for *TP53* mutations, other known cancer gene mutations were also found and mapped to the branch. Another example is *TTN* mutations detected in 6 tumors, 4 of which had a late molecular timing of acquisition of the *TTN* mutation. The ITH of driver mutations in known cancer genes indicated the complexity of clonal architecture in NSCLC tumors and the challenges to their treatment.

The C > A transversions were predominant in the detected driver mutations, which is similar to the mutation pattern observed in squamous cell carcinoma and *TP53*.[16,18] Although we did not observe significant shifts in the proportion of C > A transversions from early to late mutations, an opposite trend was observed between smokers and non-smokers, which indicated the impact of tobacco smoke on the mutational pattern during tumor progression.

LOH is the most common molecular genetic alteration observed in human cancers.[19–21] It occurs in a tumor suppressor genes in which one copy of the gene is lost, and the remaining allele contains a deleterious mutation that renders the gene inactive, recognized by Knudsen as the two-hit hypothesis.[22] Previous studies identified pronounced ITH in copy number aberrations in NSCLC,[8,14] whereas little evidence was shown for ITH in the regions of LOHs. Our results identified regions of LOH surrounding *TP53* and *CDKN2A* mutation sites in tumor 711, which provides additional evidence supporting the two-hit theory, and also demonstrated that these regions of LOH are heterogeneous with both mutated and non-mutated cells. Because we identified extensive ITH of other driver mutations, we propose that if these driver mutations with a higher VAF are likely located in the region of LOHs, LOHs are also likely to occur frequently and heterogeneously in NSCLC.

In conclusion, this is the first investigation of ITH of driver mutations in *TP53*-driven NSCLC using multiregion sequencing. Our results revealed the universal spatial heterogeneity of driver mutations in NSCLC and demonstrated that the molecular timing of acquisition of driver mutations, such as *TP53* mutation, varies in different tumors during tumor formation and progression. Because the ITH of driver mutations is most likely to have clinical consequences, further endeavors to limit the impact of ITH and improve therapeutic efficiency are needed to benefit NSCLC patients undergoing targeted treatment.

## References


1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. *Cancer J Clin* 2015;65:5–29.
2. Bedard PL, Hansen AR, Ratain MJ, *et al.* Tumour heterogeneity in the clinic. *Nature* 2013;501:355–64.
3. Taniguchi K, Okami J, Kodama K, *et al.* Intratumor heterogeneity of epidermal growth factor receptor mutations in lung cancer and its correlation to the response to gefitinib. *Cancer Sci* 2008; 99:929–35.
4. Bai H, Wang Z, Wang Y, *et al.* Detection and clinical significance of intratumoral EGFR mutational heterogeneity in Chinese patients with advanced non-small cell lung cancer. *PLoS One* 2013;8:e54170
5. Nowell PC. The clonal evolution of tumor cell populations. *Science* 1976;194:23–8.
6. Boutros PC, Fraser M, Harding NJ, *et al.* Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat Genet* 2015;47:736–45.
7. Campbell PJ, Pleasance ED, Stephens PJ, *et al.* Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proc Natl Acad Sci U S A* 2008;105:13081–6.
8. de Bruin EC, McGranahan N, Mitter R, *et al.* Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* 2014;346:251–6.
9. Gerlinger M, Horswell S, Larkin J, *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet* 2014;46:225–33.
10. Gerlinger M, Rowan AJ, Horswell S, *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012;366:883–92.
11. Ling S, Hu Z, Yang Z, *et al.* Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. *Proc Natl Acad Sci U S A* 2015;112:E6496–E505.
12. Navin N, Krasnitz A, Rodgers L, *et al.* Inferring tumor progression from genomic heterogeneity. *Genome Res* 2010;20:68–80.
13. Sottoriva A, Spiteri I, Piccirillo SG, *et al.* Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A* 2013;110:4009–14.
14. Zhang J, Fujimoto J, Wedge DC, *et al.* Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 2014;346:256–9.
15. Yu XJ, Yang MJ, Zhou B, *et al.* Characterization of somatic mutations in air pollution-related lung cancer. *EBioMedicine* 2015; 2:583–90.
16. Pfeifer GP, Hainaut P. On the origin of G –> T transversions in lung cancer. *Mutat Res* 2003;526: 39–43.
17. Vogelstein B, Papadopoulos N, Velculescu VE, *et al.* Cancer genome landscapes. *Science* 2013; 339:1546–58.
18. Heist RS, Sequist LV, Engelman JA. Genetic changes in squamous cell lung cancer: a review. *J Thorac Oncol* 2012;7:924–33.
19. Cavenee WK, Dryja TP, Phillips RA, *et al.* Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* 1983;305:779–84.
20. Merajver SD, Frank TS, Xu J, *et al.* Germline BRCA1 mutations and loss of the wild-type allele in tumors from families with early onset breast and ovarian cancer. *Clin Cancer Res* 1995;1:539–44.
21. Thiagalingam S, Laken S, Willson JK, *et al.* Mechanisms underlying losses of heterozygosity in human colorectal cancers. *Proc Natl Acad Sci U S A* 2001;98:2698–702.
22. Knudson AG. Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 1971;68:820–3.