Journal of Genetics and Genomics 50 (2023) 330–340

Contents lists available at ScienceDirect

Journal of Genetics and Genomics

Journal homepage: www.journals.elsevier.com/journal-of-genetics-and-genomics/

Original research

Lack of evolutionary convergence in multiple primary lung cancer suggests insufficient specificity of personalized therapy

Hua Cheng [a, 1], Ziyan Guo [b, 1], Xiaoyu Zhang [b, 1], Xiao-Jin Wang [a, c, 1], Zizhang Li [b], Wen-Wen Huo [a, c], Hong-Cheng Zhong [a], Xiao-Jian Li [a], Xiang-Wen Wu [a], Wen-Hao Li [a], Zhuo-Wen Chen [a], Tian-Chi Wu [a], Xiang-Feng Gan [a], Bei-Long Zhong [a], Vassily A. Lyubetsky [d, e], Leonid Yu Rusin [d], Junnan Yang [b], Qiyi Zhao [f, g], Qing-Dong Cao [a, *], Jian-Rong Yang [b, c, h, *]

[a] Department of Thoracic Surgery, The Fifth Affiliated Hospital, Sun Yat-Sen University, Zhuhai, Guangdong 519000, China
[b] Department of Genetics and Biomedical Informatics, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, Guangdong 510080, China
[c] Program in Cancer Research, Zhongshan School of Medicine, The Fifth Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong 510080, China
[d] Kharkevich Institute for Information Transmission Problems Russian Academy Sciences, Moscow 127051, Russia
[e] Department of Mathematical Logic and Theory of Algorithms, Faculty of Mechanics and Mathematics, Lomonosov Moscow State University, Moscow 119991, Russia
[f] Department of Infectious Diseases, Third Affiliated Hospital, Sun Yat-Sen University, Guangzhou, Guangdong 510630, China
[g] Guangdong Provincial Key Laboratory of Liver Disease Research the Third Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong 510630, China
[h] Key Laboratory of Tropical Disease Control, Ministry of Education, Sun Yat-sen University, Guangzhou, Guangdong 510080, China

ARTICLE INFO

Article history:
Received 11 November 2022
Accepted 11 November 2022
Available online 19 November 2022

Keywords:
Multiple primary lung cancer
Cancer evolution
Convergent evolution

ABSTRACT

Multiple primary lung cancer (MPLC) is an increasingly prevalent subtype of lung cancer. According to recent genomic studies, the different lesions of a single MPLC patient exhibit functional similarities that may reflect evolutionary convergence. We perform whole-exome sequencing for a unique cohort of MPLC patients with multiple samples from each lesion found. Using our own and other relevant public data, evolutionary tree reconstruction reveals that cancer driver gene mutations occurred at the early trunk, indicating evolutionary contingency rather than adaptive convergence. Additionally, tumors from the same MPLC patient are as genetically diverse as those from different patients, while within-tumor genetic heterogeneity is significantly lower. Furthermore, the aberrant molecular functions enriched in mutated genes for a sample show a strong overlap with other samples from the same tumor, but not with samples from other tumors or other patients. Overall, there is no evidence of adaptive convergence during the evolution of MPLC. Most importantly, the similar between-tumor diversity and between-patient diversity suggest that personalized therapies may not adequately account for the genetic diversity among different tumors in an MPLC patient. To fully exploit the strategic value of precision medicine, targeted therapies should be designed and delivered on a per-lesion basis.

Introduction

Lung cancer is the second most common cancer globally and the leading cause of cancer deaths worldwide each year (Sung et al., 2021). A fraction of lung cancer patients bear multiple primary tumors at first diagnosis (multiple primary lung cancer, or MPLC). The incidence of MPLC has increased in recent years, probably owing to the application of high-resolution chest imaging and advanced lung

* Corresponding authors.
E-mail addresses: caoqd@mail.sysu.edu.cn (Q.-D. Cao), yangjianrong@mail.sysu.edu.cn (J.-R. Yang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jgg.2022.11.005

cancer screening (Chen et al., 2019). However, the genomic characteristics of MPLC, as well as the history of somatic evolution towards MPLC, remain poorly understood.

In lung cancer (Liu et al., 2016; Ma et al., 2017) and many other solid tumors (Ling et al., 2015; Tao et al., 2015; Wang et al., 2016; Zhai et al., 2017), genetic heterogeneity is a widely observed feature that is frequently associated with a poor prognosis (Zhou et al., 2011; Thress et al., 2015; Blakely et al., 2017; Rosenthal et al., 2019; Vitale et al., 2021). Understanding the origin and development of genetic heterogeneity could potentially aid in therapeutic target selection and drug development (Greaves and Maley 2012; McGranahan and Swanton 2015) and shed light on the somatic evolution history of cancer cells (Ling et al., 2015; Wu et al., 2016; Hu et al., 2019). For example, investigations of genetic heterogeneity in single or metastatic tumors have revealed the spatial organization of tumor cell subpopulations (Ling et al., 2015), discovered quantitative evidence of early metastatic seeding (Hu et al., 2019), and highlighted the genomic evolution that occurs from preneoplasia to tumor development and progression (Hu et al., 2019). For MPLC, the genetic heterogeneity among multiple lesions observed in the same patient has been used to confirm their nonmetastatic nature (Liu et al., 2016). More recently, the genetic heterogeneity of MPLC was found to be compatible with the convergent evolution of cancer cells, as only a limiting number of signaling pathways were mutated (Ma et al., 2017). This finding indicates that targeting these small number of convergent pathways might be a viable therapeutic strategy for MPLC (Ma et al., 2017). However, due to a lack of formal statistical or evolutionary model-based tests (Ma et al., 2017), such a theory of evolutionary/functional convergence of MPLC is still questionable. Meanwhile, multiple lines of evidence in favor of non-Darwinian divergent evolution have been reported for other types of cancer (Ling et al., 2015; Wu et al., 2016; Zhai et al., 2017). Clearly, a more detailed analysis of the genetic heterogeneity of MPLC is required to provide better diagnosis and treatment strategies for MPLC patients and to resolve the conflicting views of convergence versus divergence in cancer evolution.

Other than the biomedical significance mentioned above, investigation into the genetic heterogeneity of MPLC also bears important theoretical value. More than 30 years ago (Gould, 1990), the prominent evolutionary biologist Stephen Jay Gould conducted a gedanken experiment: if we rewind and replay the tape of evolution, should we expect the same outcome? The question, in essence, is whether evolution is dominated by deterministic evolutionary forces (and therefore convergent outcomes) or historical contingency (and therefore divergent outcomes). The answer to this now-famous question has been debated not only for the evolution of the natural population (Morris, 1998) but also for cancer cell evolution (Wu et al., 2016). Unfortunately, there is a general bias towards adaptationist storytelling that undervalues the role of nonadaptive forces such as drift or mutation (Gould and Lewontin, 1979), which is of particular importance in the genomic era (Lynch 2007; Koonin 2016). The multiple lesions in an MPLC patient are derived from an identical genetic background and subject to the same environmental exposure and thus represent repeated evolution. The systematic evaluation of the genetic heterogeneity of MPLC constitutes a unique opportunity to address Gould's question in the context of cancer cell evolution, preferably with proper assessment of the nonadaptive evolutionary null model.

In this study, we collected whole-exome sequencing data from 101 tumor/paratumor samples from 16 patients pathologically diagnosed with MPLC (including 78 newly sequenced samples). Comprehensive analysis of the genetic heterogeneity of the tumor samples confirmed their nonmetastatic nature and revealed a limited range of driver mutations during the early stage of cancer evolution. However, mutations occurring in later stages of cancer evolution suggested divergent rather than convergent evolution at the levels of both individual genes and functional pathways. Most importantly, our findings suggest that targeted therapies tailored for each patient might not be precise enough and need to be replaced by those tailored for each tumor, as different tumors found within the same patient were as genetically diverse as tumors from different patients.

Results

MPLC samples and whole-exome sequencing

We sampled a cohort of 11 patients, each diagnosed with MPLC according to the comprehensive histologic assessment criteria used in our previous work (Cheng et al., 2017) (Fig. 1A; Materials and methods). In each patient, two to three primary tumors were detected, the majority of which were detected by computed tomography (Fig. 1B–1L). We collected tumor samples from at least 3 foci within each tumor, as well as corresponding paratumor/normal samples, yielding a total of 78 samples (Table S1). These samples were subjected to whole-exome sequencing on BGISEQ-500 to an average coverage of 140.14, with approximately 90% of the targeted regions covered to a depth of ~42× or more (Fig. 2A; Table S2; Materials and methods). Furthermore, we downloaded raw data of previously sequenced MPLC samples (Ma et al., 2017), and discarded those from patients with only one tumor sample sequenced. As a result, we obtained an additional dataset consisting of 18 tumor samples and 5 paired normal samples. Taken together, our full dataset contains whole-exome sequencing data for a total of 101 samples from 16 patients (Tables S1 and S2). As inferred by segmented copy number and allelic fraction values for somatic point mutation profiles, the tumor samples had an average tumor purity of 0.24 (Fig. 2B). Following the best practice pipeline suggested by Genome Analysis Toolkit (GATK) (Materials and methods), we identified 9 to 1418 somatic alterations per tumor sample, including ~10,288 single nucleotide variations (SNVs) (Table S3), which correspond to a rate of 0.23–35.6 mutations per megabase (Fig. 2C).

A list of putative lung cancer driver genes was compiled by requiring that the genes appear in at least two of the five commonly used driver gene datasets, including Catalogue of Somatic Mutations in Cancer (COSMIC) (Tate et al., 2019), DriverDBv3 (Liu et al., 2020), Integrative Onco Genomics (IntOGen) (Martinez-Jimenez et al., 2020), OncoVar (Wang et al., 2021), and 12 general tumorigenesis drivers inferred by Trigos and colleagues (Trigos et al., 2017) (Fig. 2D). One striking pattern was the high prevalence of missense mutations in *Epidermal Growth Factor Receptor* (*EGFR*), which is indeed one of the genes reported to show recurrent alterations in lung cancer (Cancer Genome Atlas Research Network 2014). Mutations that appeared in other driver genes appeared moderately consistent within the same tumor and displayed substantial variation between different tumors from the same patient. For example, only 57.8% and 9.0% of the driver gene mutations were shared within the same tumor and between different tumors from the same patient, respectively (See Materials and methods).

Reconstructed evolutionary history suggested early onset of driver gene mutations at the trunk of the evolutionary tree

From a theoretical point of view, recurrence of a driver gene mutation in samples from different tumors of the same MPLC patient could be explained by two mutually exclusive mechanisms (Fig. 3). On the one hand, driver gene mutations could occur at the early trunk in the evolutionary tree of cancer cells, making them shared by all descendant cancer cells. On the other hand, driver

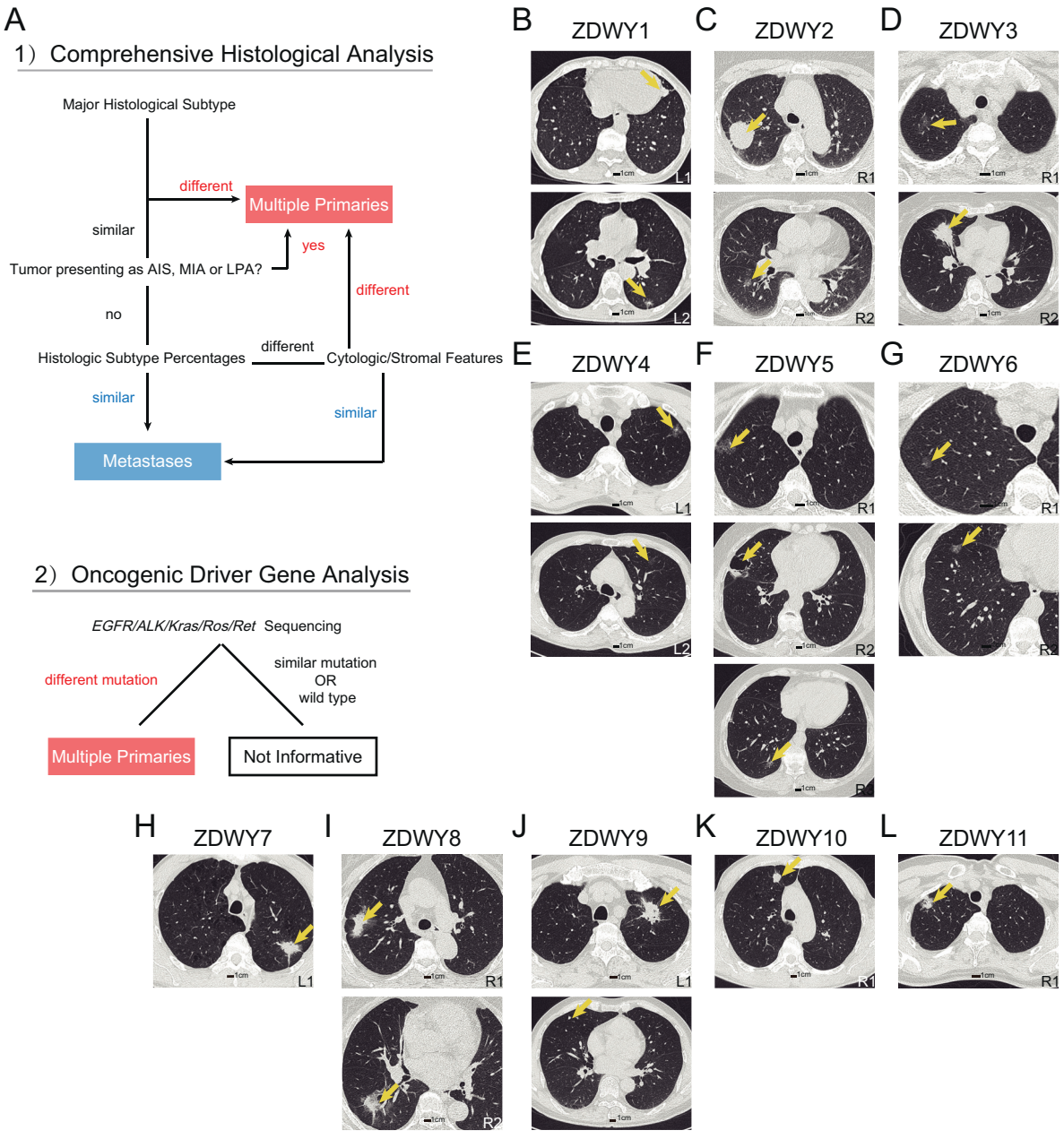

Fig. 1. MPLC samples. **A**: Comprehensive histologic assessment criteria for MPLC diagnosis. AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; LPA, lepidic predominant invasive adenocarcinoma; *EGFR*, epidermal growth factor receptor; *ALK*, ALK receptor tyrosine kinase; *Kras*, KRAS proto-oncogene; *Ros*, ROS proto-oncogene 1, receptor tyrosine kinase; *Ret*, ret proto-oncogene. **B**–**L**: Computed tomography (CT) diagnosis of each MPLC patient collected in this study, with yellow arrows pointing at each individual tumor. The scale bar on the right has tick marks with 1 cm increment. For (**A**–**L**), the patient names are listed on top of each panel, which are all named with prefix "ZDWY," followed by the number of the patient (1–11). The name of each tumor is indicated at the bottom right corner of the image, where "L" or "R" indicating that the tumor was found in the left or right lung. Note that not all tumors were captured on CT. The list of all samples used in this study can be found in Table S1.

gene mutations (identical mutations or mutations from functionally interchangeable genes within a pathway) could occur later but were subsequently selected due to the growth advantage they conferred to the cells, ultimately making the cancer cell population evolutionarily converge (as previously argued). We will hereinafter refer to these two models as the "early-trunk mutation" and "adaptive convergence" models. Note that the exclusivity of these two models lies in the fact that in the early-trunk mutation model, all cancer cells share driver gene mutations; therefore, these mutations are considered to confer no growth advantage relative to other cancer cells (not relative to normal cells). In other words, these mutations have no adaptive value in the early-trunk mutation model, making adaptive convergence irrelevant for the recurrence of driver gene mutations.

As a first examination of the early-trunk mutation and adaptive convergence models, we built phylogenetic trees using all SNVs in the tissue samples (tumor or paratumor/normal) from each individual patient (Fig. S1). By assigning the paratumor samples as the

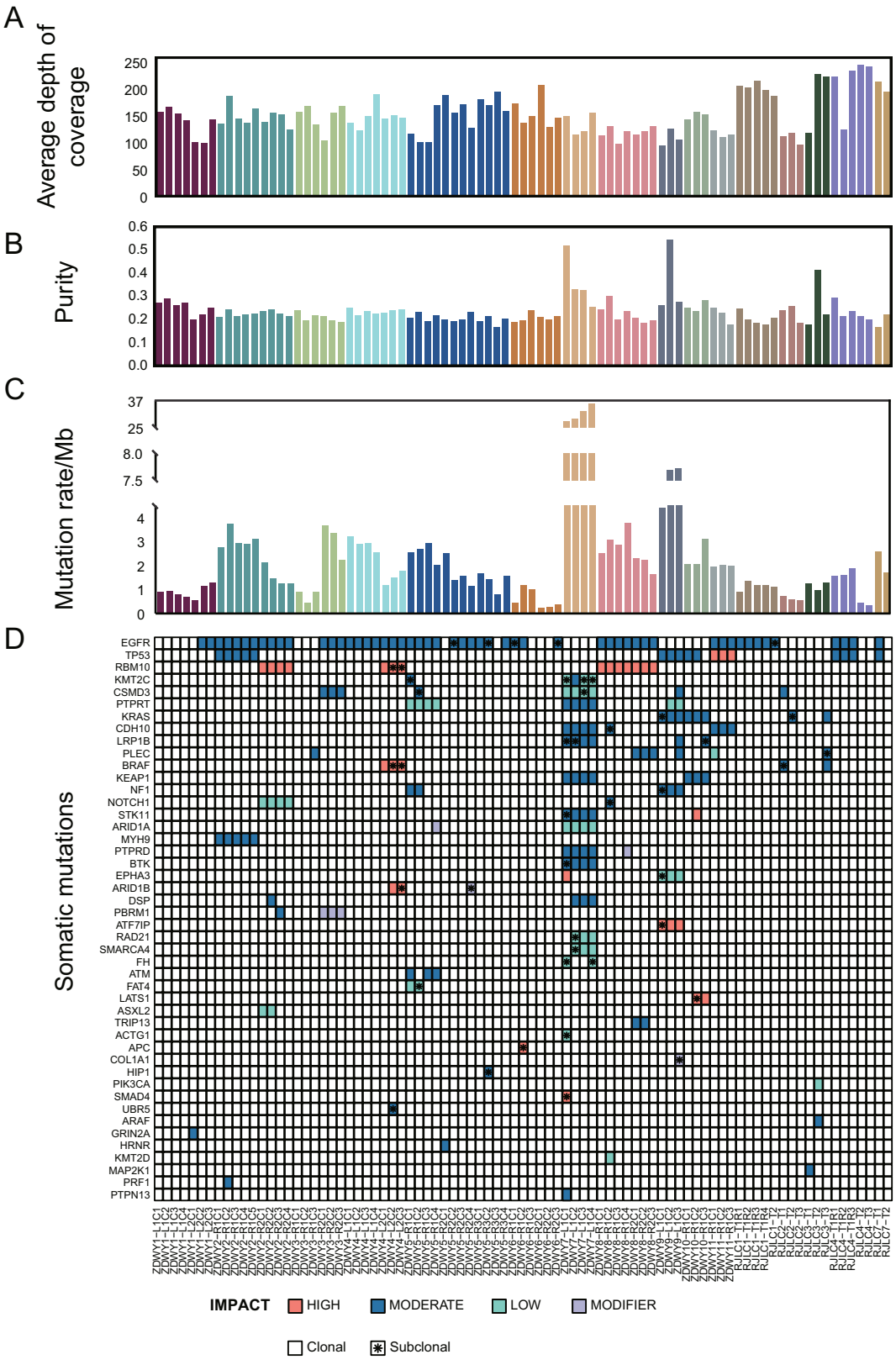

Fig. 2. Sequencing overview. The sequencing depth on the exome (**A**), estimated tumor purity (**B**), mutation rate (**C**) and identified single nucleotide somatic mutations on driver genes of lung cancer (**D**) are listed for each sample (*x* axis). Subclonal mutations, as indicated by asterisks, were called by ABSOLUTE as a cancer cell fraction < 80%. The functional impact of each mutation as predicted by EnsEMBL VEP is shown as the color of each tile as indicated by the legend on the bottom. The sample of each column is listed at the bottom. The samples are all named by the patient name (e.g., "ZDWY1"), a hyphen, the tumor name (e.g., "L1"), and the number of the sample within the tumor as C1 to C3 (or C4). See also Tables S2 and S3.

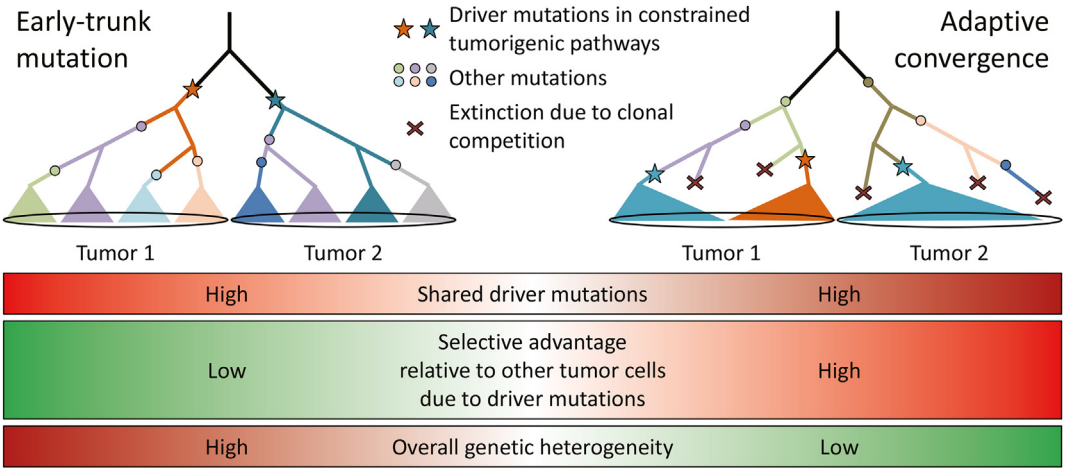

Fig. 3. Schematic diagram for the early-trunk model and the adaptive convergence model. The recurrence of the same "driver" mutations in different tumor samples might be explained by two different models. That is, the mutations could happen at the "early-trunk" of cancer evolution, or they could have been selected by adaptive convergence. The different phenotypic consequences are also listed below.

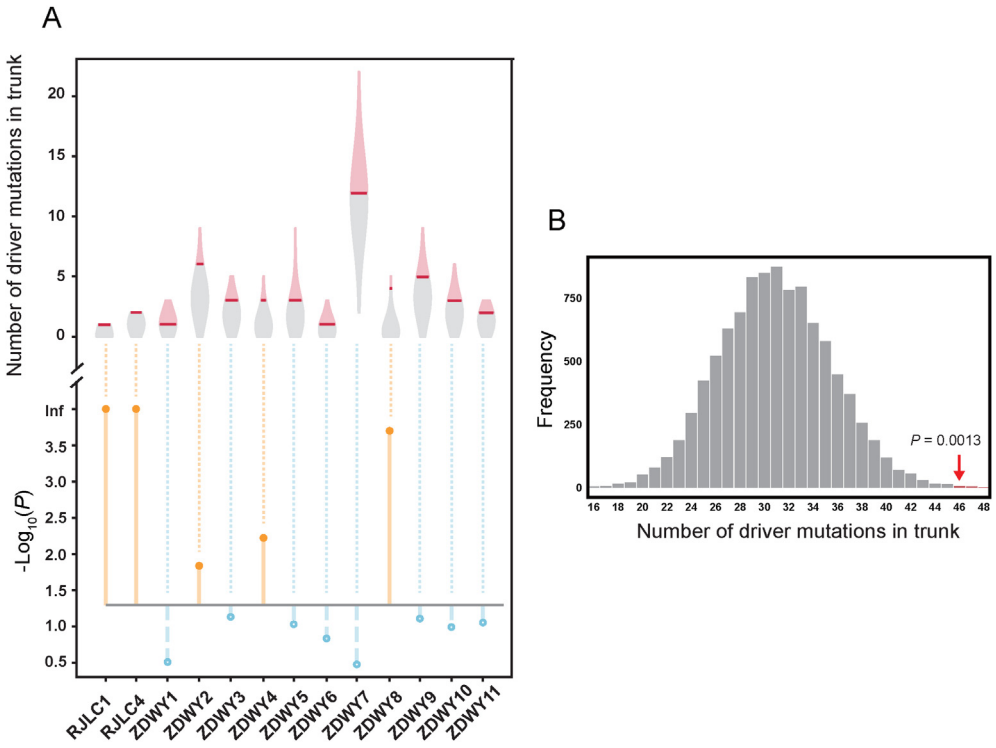

Fig. 4. Driver mutations at the early trunk. The observed number of early-trunk mutations on driver genes for each sample (**A**) or the pooled sample (**B**) are indicated as red lines or arrows, whereas their respective random expectations are indicated as accompanying violin plots or histograms.

outgroups in every phylogenetic tree, we were able to reconstruct the ancestral genotypes and locate the evolutionary branches in which the driver gene mutations occurred. We found that many driver gene mutations occurred at the early trunk in the phylogenetic tree (dark or light red branches in Fig. S1). Indeed, significant enrichment of driver gene mutations in early trunk can be identified in five patients, but none of the patients displayed significant results in the opposite direction (lower half of Fig. 4A, permutation tests; see Materials and methods). Moreover, when we pooled and analyzed data from all patients together, we again found that driver gene mutations are significantly enriched at the evolutionary trunk of all MPLCs (Fig. 4B, permutation test; see Materials and methods). Therefore, the observed mutation profile of MPLC appeared to be better explained by the early-trunk mutation model.

Genetic heterogeneity suggested a lack of evolutionary convergence

The multiple lesions found in an MPLC patient are derived from an identical genetic background and subject to the same environmental exposure. They are therefore repeated evolutions of cancer cells. As such, we can test evolutionary convergence by quantifying the genetic heterogeneity within a tumor ("within-tumor heterogeneity") and between different tumors of the same patient ("between-tumor

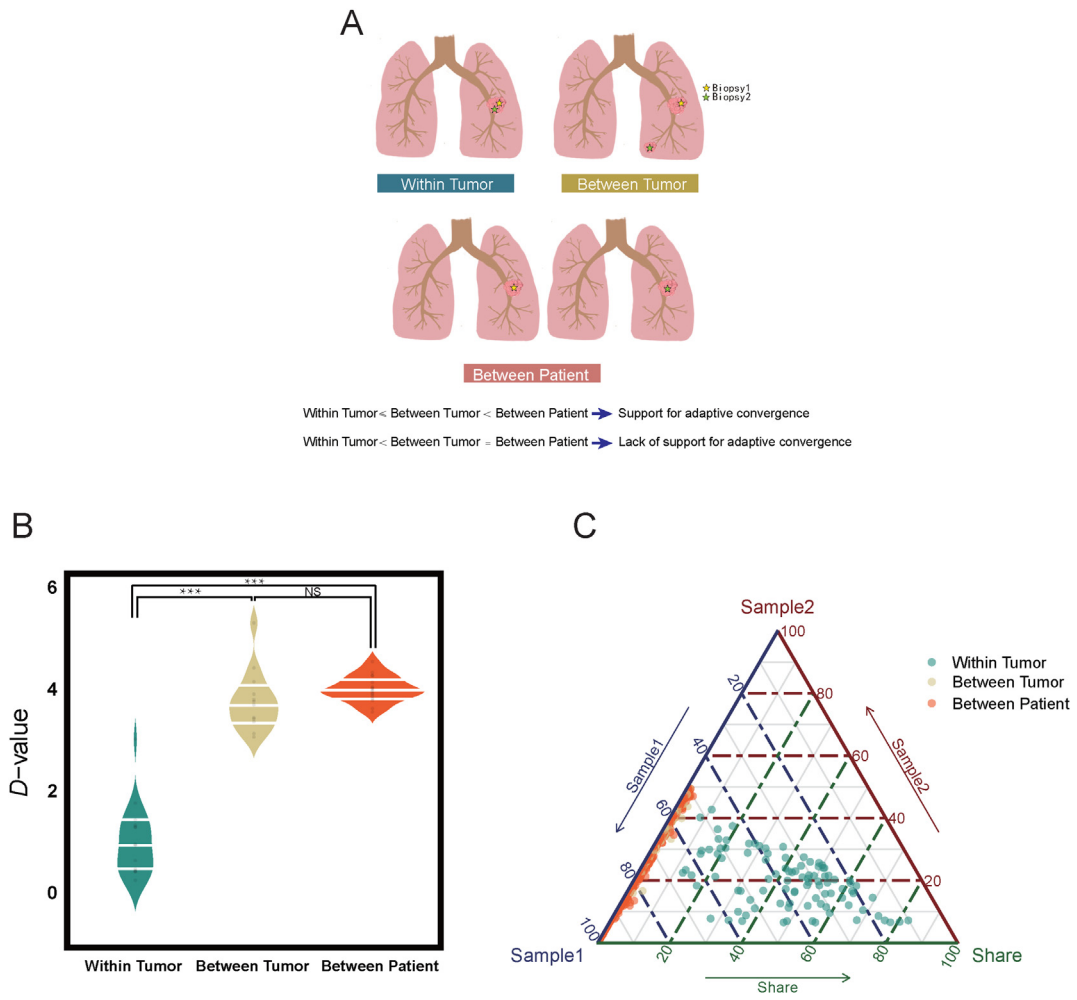

Fig. 5. Genetic heterogeneity and the general lack of evolutionary convergence. **A**: Schematic diagram for three types of comparisons among tumor samples. Possible outcomes and their corresponding conclusions are listed at the bottom. **B**: Averaged Nei's genetic distances (*D*-value) of each indicated group for each patient is shown as individual dots and their distribution is shown as violin plots. Nominal *P* values from the Wilcoxon signed-rank tests are indicated as ***, $P < 0.001$; NS, not significant. **C**: Statistics in the modulus space of evolutionary trees. Each dot represents a pair of tumor samples, in which one sample with more mutations is named "Sample 1" and the other "Sample 2." The coordinates correspond to the probability that Sample 1 (blue axis) or Sample 2 (red axis) is more distant or that they are similarly distant (green axis) from their common ancestor.

heterogeneity"). It is expected that the between-tumor heterogeneity should be higher than the within-tumor heterogeneity (Liu et al., 2016; Ma et al., 2017). But if there is strong evolutionary convergence, different tumors within the same patient should genetically converge, and as a result, the between-tumor heterogeneity should be comparable to the within-tumor heterogeneity. It should be noted that even if the between-tumor heterogeneity is higher than the within-tumor heterogeneity, there may still be intermediate levels of evolutionary convergence, which will reduce the between-tumor heterogeneity relative to its convergence-free expectation. As a further test, we used the genetic heterogeneity of tumors between different patients ("between-patient heterogeneity"), which have different genetic backgrounds and evolutionary histories in different environments, to approximate the expected heterogeneity without evolutionary convergence among tumors within the same patient. The between-patient heterogeneity should be higher than the between-tumor heterogeneity if there is intermediate level of evolutionary convergence between tumors from the same patient. In contrast, if the between-patient heterogeneity was not significantly higher than the between-tumor heterogeneity, we considered it a lack of evidence for evolutionary convergence. From a quantitative perspective, we can examine the value of between-tumor heterogeneity relative to the lower bound defined by within-tumor heterogeneity and the upper bound defined by between-patient heterogeneity to assess the strength of evidence for evolutionary convergence (Fig. 5A).

As a pairwise metric for genetic heterogeneity, we calculated Nei's genetic distance (*D*-Value) between all tumor samples (Materials and methods). We found that the within-tumor heterogeneity was significantly lower than the between-tumor heterogeneity (Fig. 5B), which was consistent with previous reports of MPLC (Liu et al., 2016; Ma et al., 2017). Unexpectedly, the between-tumor heterogeneity and between-patient heterogeneity were not significantly different from each other (Fig. 5B). More importantly, as reasoned above, the significantly higher between-tumor heterogeneity than within-tumor heterogeneity, along with the similar between-tumor and between-patient heterogeneity, suggested a lack of convergence in MPLC evolution. To further corroborate our findings, we computed statistics on the evolutionary modulus spaces (Fig. 5C; Materials and methods). We found that pairwise comparisons between samples within the same tumor revealed intermediate levels of shared mutations (Fig. 5C, green points), whereas comparisons between samples of different tumors from the same patients or between samples from different patients all yielded results indicating a similar depletion of shared mutations for these two groups of comparisons (Fig. 5C, yellow and red points). Collectively,

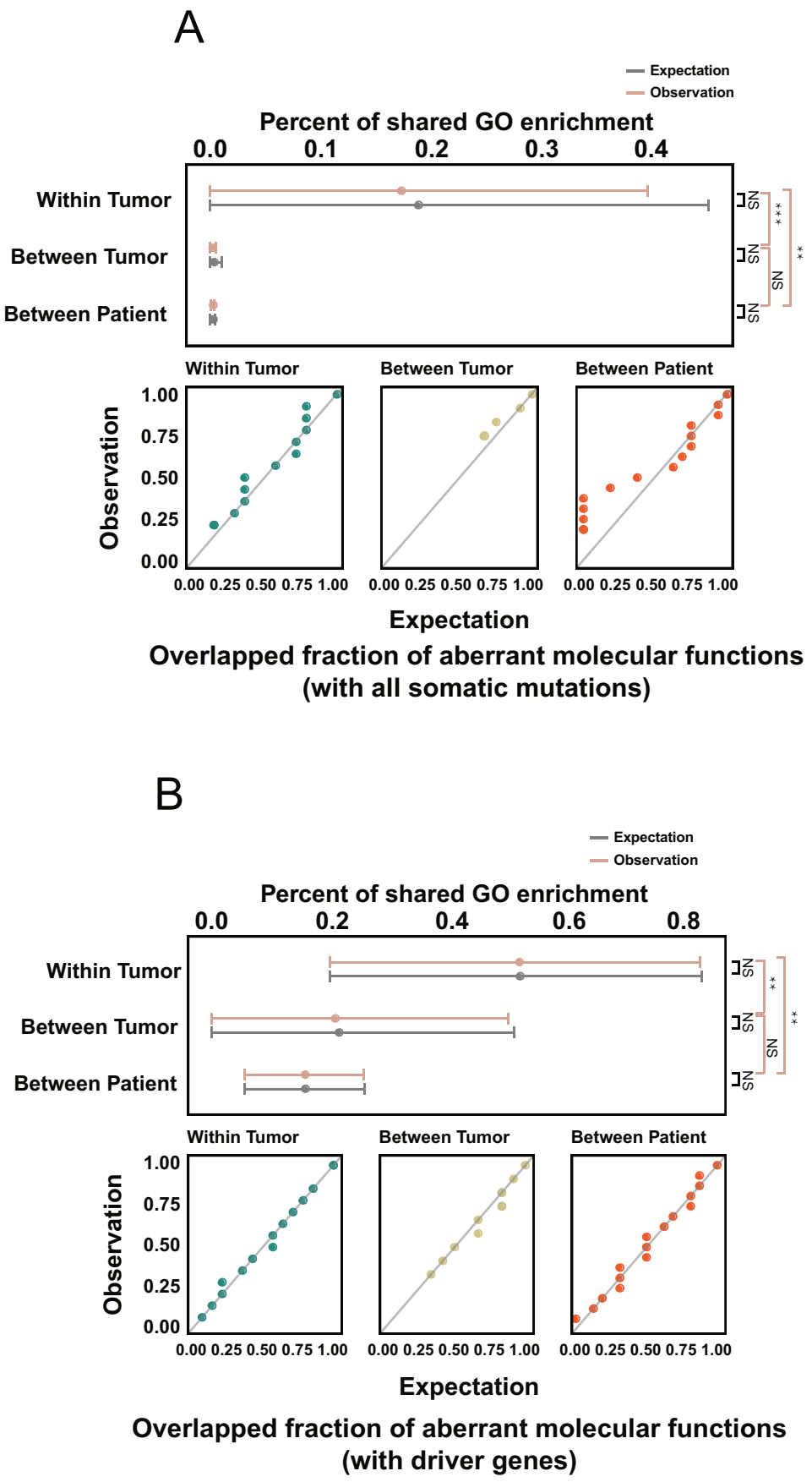

Fig. 6. Evolutionary convergence is absent at the level of molecular function. **A**: The overlapping fraction of aberrant molecular functions between tumor samples of each indicated group is shown along with their corresponding null expectation (see Materials and methods). **B**: Same as (**A**) except that only GO terms containing driver genes of lung cancer are considered. In both panels, nominal *P* values from the Wilcoxon rank-sum test are indicated as NS, not significant; **, $P < 0.01$; ***, $P < 0.001$.

our analyses of the genetic heterogeneity of MPLC suggested a general lack of evolutionary convergence at the level of individual mutations. More importantly, this result indicates that personalized targeted therapies might be of little help for patients with MPLC relative to general therapies, as tumors within the same patient were found to be as genetically diverse as tumors from different patients. Instead, MPLC needs targeted therapy with higher precision, such that treatments should be tailored for each tumor. This notion is supported by a clinical case report (Ye et al., 2016) of MPLC with heterogeneous *EGFR* and *KRAS* mutations in different lesions of the same patient, which shows that treatments tailored per lesion are effective.

Lack of convergence at the functional pathway level

Although we have shown the lack of convergence at the mutation level, whether convergence is also absent at the level of functional pathways remained unanswered. On the one hand, a previous report implied that mutations in different tumors of the same patient converged to a narrow list of functional pathways, thereby leading to constrained phenotypic and biomedical outcomes (Ma et al., 2017). On the other hand, studies on other types of cancers have suggested that evolutionary divergence caused by either contingent mutations or adaptive divergence is more prevalent (Ling et al., 2015; Wu et al., 2016).

To resolve this anomaly in MPLC with a formal statistical test, we categorized the molecular function of mutated genes by Gene Ontology terms and then inferred the aberrant molecular function by identifying the functional categories enriched with the mutated genes within each sample (Materials and methods). The overlap in aberrant molecular functions for within-tumor comparisons was found to be higher than that in between-tumor comparisons, which was similar to that observed in the between-patient comparisons (Fig. 6A). In other words, the functional convergence between different tumors from the same patient is not stronger than that between patients, and this observation directly contradicts previous claim that the evolutionary convergence at the functional level among tumors from the same patient (Ma et al., 2017).

To further assess the null expectation of the above analysis, we randomly shuffled the mutations found in each sample such that each sample retained the same total number of mutations in all genes and each gene retained the same total number of mutations in all samples. The resulting shuffled list of mutations suggested that the expected level of overlapped aberrant molecular functions was not significantly lower than the observed level between different samples from the same tumor, the same patient, or different patients (Fig. 6A). This result again suggested a general absence of evolutionary convergence at the functional level between tumors from the same patient.

The previous claim of functional convergence between different tumors of an MPLC patient was based on an additional constraint that only driver genes/pathways were considered (but not all genes/pathways) (Ma et al., 2017). Was this additional requirement responsible for the detected functional convergence? We repeated the analysis performed for Fig. 6A but only considered the aforementioned list of putative driver genes. Similar to the above result, we found that the overlap of aberrant molecular functions between tumors was less than that of within-tumor comparisons and similar to that of between-patient comparisons. Note that the levels of convergence for the within-tumor, between-tumor and between-patient comparisons increased from 0.19, 0.0027, and 0.0020 (red points in upper Fig. 6A) to 0.51, 0.21, and 0.15 (red points in upper Fig. 6B), respectively. These numbers appeared to indicate that functional convergence is strong when only cancer driver genes/pathways are considered. However, we found that their corresponding null expectations also increased similarly (gray points, Fig. 6B compared with Fig. 6A), such that the observed levels of functional convergence remained indistinguishable from their null expectations (Fig. 6B). This result therefore suggested that constraining the analysis within cancer driver genes/pathways might artifactually increase the signal of functional convergence. However, the observed signal of functional convergence was actually comparable to its null expectation, lending no support for evolutionary convergence among tumors in an MPLC patient and therefore suggesting that the genetic heterogeneity of MPLC is better described by evolutionary contingency, but not evolutionary convergence.

Discussion

The genetic heterogeneity of tumors not only has major implications for targeted cancer therapies but also informs us of the evolutionary history of cancer cells. In the current study, we analyzed the whole exome of 101 tumor or paratumor samples from 16 MPLC patients and performed a comprehensive analysis of their genetic heterogeneity. We identified recurrent mutations in putative cancer driver genes, which were inferred as events at the early trunk of the evolutionary history of cancer cells rather than evolutionary convergence of independent mutational events. Taking advantage of multiple primary lesions in the same MPLC patients as repeated evolutionary processes within the same genetic/environmental background, we further tested the conjecture of functional convergence of cancer cells. As a result, the level of genetic heterogeneity between tumors from the same patient was found to be significantly increased relative to that within the same tumor to a level comparable to that between different patients. This observation, which was confirmed at both the levels of genes and functional pathways, directly contradicted the evolutionary convergence and thereby cast doubt on the potential application of personalized MPLC therapies relying on the evolutionary convergence of different tumors in the same patient.

There are a few potential caveats to our study worth discussing. First, it is possible that sample patients with metastasis were mistakenly diagnosed with MPLC. In this case, the genetic similarities between tumors from the patient should be overestimated relative to actual MPLC, thereby further diminishing the signal for adaptive convergence between tumors of the same patient. Second, although our results are consistent with the early-trunk model and evolutionary contingency underlying cancer evolution in MPLC, we emphasize that the precise nature of this result is a lack of evidence for adaptive convergence; in other words, we cannot reject the null hypothesis of historical contingency driven by largely neutral evolution (i.e., compared to other tumor cells, no mutation appears to be advantageous). It is also possible that the tumors found are in their early stage and therefore have not been subjected to strong enough natural selection. Regardless, our results have critical implications for the clinical treatment of MPLC. That is, tumors from the same MPLC patient are already as genetically diverse as tumors from different patients, which in theory would render personalized targeted therapies useless relative to generalized therapies. Third, we have shown that the assessment of convergence at the level of functional pathways could have been skewed by limiting the analysis within cancer driver genes/pathways. Although we have used well-accepted driver gene lists (Martinez-Jimenez et al., 2020), it is still possible that other key driver genes/pathways exist for lung cancer or MPLC. If these additional drivers were to be discovered, we would repeat our analyses again to test for functional convergence. However, a total reversion of the vast difference between the observed and expected overlap of aberrant molecular functions (Fig. 6B) is highly unlikely. We therefore believe that our conclusion of a lack of functional convergence will be robust to changes in the list of cancer driver genes/pathways.

The evolutionary convergence of cancer cells has long been presumed without being formally tested (Wu et al., 2016). On the one hand, this conjecture is not totally unfounded because, after all, cancers phenotypically converge, as we have defined cancer by, for example, phenotypic hallmarks (Hanahan and Weinberg 2011). On the other hand, a high level of genetic heterogeneity also appears to be the norm (Liu et al., 2016). How can we reconcile such an apparent inconsistency between genotypic divergence and phenotypic convergence? In the case of MPLC, it was purported that convergence was achieved by constraints at the level of functional pathways, which, according to our analyses, was an artifact with no actual statistical support (Fig. 6). More importantly, the whole premise of precision medicine is founded on the idea that despite their phenotypic convergence, a type of cancer such as lung cancer could have a vast amount of clinically relevant genetic heterogeneity. In the case of MPLC, testing whether the between-tumor (within-patient) heterogeneity is indeed limited due to convergent evolution, or is just as high as the between-patient heterogeneity would indicate whether the level of precision required for MPLC should be per-patient or per-tumor. There is therefore a clear clinical relevance to our results beyond their theoretical importance.

Our results are consistent with a previously proposed model of carcinogenesis, where a few founder cancer cells with a limited number of driver mutations undergo further near-neutral evolution to give rise to a large population of cancer cells with a high level of heterogeneity. Stem cells or other specific functional groups of cells are possible sources of these founder cells (Chen et al., 2021). A contrasting, unsupported model was that driver mutations occurring much later in the evolutionary history of cancer cells were positively selected and fixed in the cancer cell population, which should lead to relatively low genetic heterogeneity due to the effect of the selective sweep caused by positive selection. Indeed, pervasive non-Darwinian evolution has been observed by intratumor heterogeneity analyses in hepatocellular carcinoma (Ling et al., 2015; Zhai et al., 2017), among others (Williams et al., 2016). To the best of our knowledge, the results presented herein represent the first evidence to support non-Darwinian evolution in MPLC.

In the context of evolution theory, our findings that between-tumor heterogeneity was significantly higher than within-tumor heterogeneity strongly suggested that evolution starting from the same genotype and within the same environment cannot give rise to convergent outcomes. More surprisingly, the genotypic outcomes arising from such repeated evolutions were no more similar than those arising from different genetic backgrounds and different environments (different patients). Such observations are compatible with the notion that the evolution of cancer cells is mostly driven by historical contingency but not adaptive convergence.

As for the mechanism underlying the lack of adaptive convergence, we suspected that local dispersal among tumors of the same patient or even among regions within the same tumor might have limited the competition among cancer cells and therefore the efficacy of natural selection (Kerr et al., 2002). Additionally, non-cell-autonomous drivers for tumor growth and clonal interference (Marusyk et al., 2014) might have also contributed to the genetic heterogeneity of MPLC. Resolving detailed mechanism for the observed non-Darwinian evolution would be an important future direction for MPLC and other type of tumor with significant heterogeneity.

Materials and methods

Tumor samples

Patients who had more than one synchronous lung tumor and who underwent surgical resection with curative intent were eligible for inclusion. The tumor size was at least one centimeter. We used a comprehensive histologic assessment (see below) to distinguish these tumors as multiple primary lung cancers or intrapulmonary metastases. This study was reviewed and approved by the Medical Ethical Committee of The Fifth Affiliated Hospital of Sun Yat-sen University, China. All participants provided written informed consent. Note that ZDWY7 and ZDWY10 were tumor samples taken from two different surgeries of the same patient, both after neoadjuvant chemotherapy. Therefore, these two samples were never paired as between-patient samples.

Whole-exome sequencing data from a previous report (Ma et al., 2017) of MPLC were also downloaded from NCBI SRA under the accession number SRP095985. Each downloaded sample is named according to this previous report through our study. It should be noted that this dataset contains a total of seven patients (RJLC1-7), but two of them (RJLC5 and RJLC6) have only one sample sequenced, which is therefore useless for our analysis and discarded.

Comprehensive histologic assessment

The diagnostic criteria of multiple primary lung cancers were based on comprehensive histologic assessment, as used in our previous work (Cheng et al., 2017). Briefly, we semiquantitatively evaluated the relative percentage of each histological subtype, including lepidic, acinar, papillary, micropapillary, and solid components, in 10% increments. Additional histological features such as grade, cytological features, and collagen and inflammatory stromal characteristics were also considered when comparing tumors. Paired tumors exhibiting similar histologic features were considered lung metastases, and those showing different histologic features were considered multiple primaries. Tumors presenting as adenocarcinoma in situ (AIS), minimally invasive adenocarcinoma (MIA), or lepidic predominant invasive adenocarcinoma (LPA) were considered multiple primaries. We also incorporated a targeted gene (*EGFR*/*ALK*/*Kras*/*Ros*/*Ret*) panel test to confirm the multiple primaries. After MPLC was confirmed, we performed whole-exome sequencing (WES) on fresh frozen samples from multiple subsections of each tumor and paired germline DNA samples.

Tumor collection and processing

All surgically resected tumors were cut fresh, and the largest section of tumor of 3 mm thick was sent for pathologic diagnosis without compromising the bronchial or pleural resection margins. The rest of the specimens were dissected into multiple subsections for DNA extraction. We cut each tumor with new, sterile scalpel blades on a clean surface to avoid contamination with other DNA. At least two subsections from each tumor, separated by at least 3 mm, were collected for research purposes. Areas that were obviously necrotic, fibrotic or hemorrhagic were avoided to maximize the number of viable tumor cells. Spatially separated tumor regions were collected and snap frozen in liquid nitrogen and stored at $-80^{\circ}C$ for subsequent DNA extraction.

DNA extraction

Approximately $3\times3\times3$ mm of tumor tissue from each subsection was used for genomic DNA extraction using a modification of the DNA/RNA AllPrep Kit (Qiagen). Peripheral blood or normal lung tissue was also collected at the time of surgery. DNA was quantified by Qubit (Invitrogen), and DNA integrity was assessed by TapeStation (Agilent Technologies).

Whole-exome sequencing

Whole-exome sequencing was conducted by BGI Shenzhen on a BGISEQ-500 platform using the DNA extracted from each sample.

Sequence quality control and variant calling

The whole pipeline of sequence quality control and variant calling generally followed the best practices workflows proposed by the GATK team (McKenna et al., 2010). Specifically, raw short reads obtained from BGI Shenzhen were first processed by Trimmomatic (Bolger et al., 2014) to remove sequencing adaptors and short reads with poor sequencing quality, after which all reads were mapped to the human genome GRCh38 (Church et al., 2011) by BWA (Li and Durbin, 2009). The mapping statistics can be found in Table S2. We then used GATK to mark and remove duplicated reads and recalibrate the base quality score before calling the variants in each sample by MuTect2. The somatic mutations in each tumor sample were called by comparing them with the corresponding normal sample via MuTect2 parameters. We considered the full list of somatic mutations, along with their residing genes (intergenic genomic regions), and the annotation by EnsEMBL Variant Effect Predictor (VEP) (McLaren et al., 2016) can be found in Table S3. Finally, the tumor purity of each tumor sample and cancer cell fraction (CCF) of each mutation were estimated by ABSOLUTE (Carter et al., 2012). The fraction of (driver gene) mutations shared by a pair of samples was calculated as the number of mutations shared by the two samples, divided by the number of unique mutations found in the two samples.

Evolutionary analyses

For each patient, we compiled a list of segregating sites by the union of all the genomic coordinates that appeared somatically mutated in at least one tumor sample from the patient. Either the mutated allele or the germline allele was then extracted from each sample according to the GATK MuTect2 result and concatenated with those from other segregating sites to form a genotype sequence containing only the segregating sites. All genotype sequences from the different samples of the same patient were used to construct a maximum likelihood-based phylogenetic tree in MEGA (Kumar et al., 2018), where the evolutionary relationships of the samples as well as the ancestral genotypes were also inferred. Each individual somatic mutation was then assigned to one of the branches on the phylogenetic tree. These genotype sequences were also used to calculate Nei's genetic distance between samples and compute statistics on the modulus spaces of phylogenetic trees (Zairis et al., 2014). Nei's genetic distance (D) was calculated as

$$D = -\log \frac{\Sigma(x_i y_i + (1-x_i)(1-y_i))}{\sqrt{\left(\Sigma\left(x_i^2 + (1-x_i)^2\right)\right)\left(\Sigma\left(y_i^2 + (1-y_i)^2\right)\right)}}$$

where, x is all CCFs of sample 1 and y is all CCFs of sample 2. Since the numbers of sample pairs in the three groups (within-tumor, between-tumor, and between-patient) are vastly different, we averaged, for each patient, the D values of all relevant sample pairs within a group, such that all three groups have the same number of data points and allows paired comparison among groups.

To assess the null expectations of the number of driver mutations, for each patient, we randomly picked a group of mutations that was the same number of early-trunk mutations and checked the number of driver mutations with the group. This process was repeated 10000 times to infer the null distribution of the number of driver mutations occurring at the early trunk during cancer evolution. All mutations from all patients were combined and similarly analyzed as the "pooled" sample (Fig. 4B).

Gene Ontology analyses

The list of mutated genes in each tumor sample was subjected to Gene Ontology (GO) analysis to infer aberrant molecular functions, which were defined as GO terms significantly (by hypergeometric test) enriched among the mutated genes. GO analyses were performed using the R packages "clusterProfiler" (Yu et al., 2012) and "org.Hs.eg.db" (Carlson 2019).

CRediT authorship contribution statement

Hua Cheng: Methodology, Data curation, Funding acquisition, Supervision, Writing - Review & Editing. **Ziyan Guo**: Methodology, Formal analysis, Writing - Original draft, Writing - Review & Editing. **Xiaoyu Zhang**: Methodology, Formal analysis, Funding acquisition, Writing - Review & Editing. **Xiao-Jin Wang**: Data curation, Writing - Review & Editing. **Zizhang Li**: Methodology, Formal analysis. **Wen-Wen Huo**: Data curation. **Hong-Cheng Zhong**: Data curation. **Xiao-Jian Li**: Data curation. **Xiang-Wen Wu**: Data curation. **Wen-Hao Li**: Data curation. **Zhuo-Wen Chen**: Data curation. **Tian-Chi Wu**: Data curation. **Xiang-Feng Gan**: Data curation. **Bei-Long Zhong**: Data curation. **Vassily A. Lyubetsky**: Methodology, Formal analysis. **Leonid Yu. Rusin**: Methodology, Formal analysis. **Junnan Yang**: Formal analysis. **Qiyi Zhao**: Data curation, Formal analysis. **Qing-Dong Cao**: Data curation, Funding acquisition, Writing - Review & Editing, Supervision. **Jian-Rong Yang**: Conceptualization, Methodology, Formal analysis, Funding acquisition, Writing - Original draft, Writing - Review & Editing, Supervision.

Data availability

All the raw reads from BGISEQ have been uploaded to the Genome Sequence Archive (https://bigd.big.ac.cn/gsa/) under the accession number HRA002361.

Custom codes were used in the data analysis, which are available on GitHub (https://github.com/laveny/MPLC_WES).

Conflict of interest

The authors declare that they have no conflicts of interest to disclose.

Acknowledgments

This work was supported by the National Key Research and Development Program of China to J.-R. Y. (2021YFF1200904 and 2021YFA1302500), the National Natural Science Foundation of China to J.-R. Y. (31871320 and 81830103), and by Science and Technology Planning Project of ZhuHai, China to H. C., and by Science and Technology Planning Project of Guangdong Province, China to X. Z. (2014A030304053).

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jgg.2022.11.005.

References

Blakely, C.M., Watkins, T.B.K., Wu, W., Gini, B., Chabon, J.J., McCoach, C.E., McGranahan, N., Wilson, G.A., Birkbak, N.J., Olivas, V.R., et al., 2017. Evolution and clinical impact of co-occurring genetic alterations in advanced-stage *EGFR*-mutant lung cancers. Nat. Genet. 49, 1693–1704.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.
Cancer Genome Atlas Research Network, 2014. Comprehensive molecular profiling of lung adenocarcinoma. Nature 511, 543–550.
Carlson, M., 2019. Genome Wide Annotation for Human. https://doi.org/10.18129/B9.bioc.org.Hs.eg.db.
Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al., 2012. Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. 30, 413–421.
Chen, C., Huang, X., Peng, M., Liu, W., Yu, F., Wang, X., 2019. Multiple primary lung cancer: a rising challenge. J. Thorac. Dis. 11, S523–S536.
Chen, Y., Xue, Y., Jin, Y., Ji, H., 2021. Lung stem cells in regeneration and tumorigenesis. J. Genet. Genomics 48, 268–276.
Cheng, H., Lei, B.F., Peng, P.J., Lin, Y.J., Wang, X.J., 2017. Histologic lung cancer subtype differentiates synchronous multiple primary lung adenocarcinomas from intrapulmonary metastases. J. Surg. Res. 211, 215–222.
Church, D.M., Schneider, V.A., Graves, T., Auger, K., Cunningham, F., Bouk, N., Chen, H.C., Agarwala, R., McLaren, W.M., Ritchie, G.R., et al., 2011. Modernizing reference genome assemblies. PLoS Biol. 9, e1001091.
Gould, S.J., 1990. Wonderful Life: The Burgess Shale and the Nature of History. W. W. Norton & Company.
Gould, S.J., Lewontin, R.C., 1979. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc. R. Soc. Lond. B Biol. Sci. 205, 581–598.
Greaves, M., Maley, C.C., 2012. Clonal evolution in cancer. Nature 481, 306–313.
Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. Cell 144, 646–674.
Hu, X., Fujimoto, J., Ying, L., Fukuoka, J., Ashizawa, K., Sun, W., Reuben, A., Chow, C.W., McGranahan, N., Chen, R., et al., 2019a. Multi-region exome sequencing reveals genomic evolution from preneoplasia to lung adenocarcinoma. Nat. Commun. 10, 2978.
Hu, Z., Ding, J., Ma, Z., Sun, R., Seoane, J.A., Scott Shaffer, J., Suarez, C.J., Berghoff, A.S., Cremolini, C., Falcone, A., et al., 2019b. Quantitative evidence for early metastatic seeding in colorectal cancer. Nat. Genet. 51, 1113–1122.
Kerr, B., Riley, M.A., Feldman, M.W., Bohannan, B.J., 2002. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. Nature 418, 171–174.
Koonin, E.V., 2016. Splendor and misery of adaptation, or the importance of neutral null for understanding evolution. BMC Biol. 14, 114.
Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. Mega X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.
Ling, S., Hu, Z., Yang, Z., Yang, F., Li, Y., Lin, P., Chen, K., Dong, L., Cao, L., Tao, Y., et al., 2015. Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. Proc. Natl. Acad. Sci. U. S. A. 112, E6496–E6505.
Liu, S.H., Shen, P.C., Chen, C.Y., Hsu, A.N., Cho, Y.C., Lai, Y.L., Chen, F.H., Li, C.Y., Wang, S.C., Chen, M., et al., 2020. DriverDBv3: a multi-omics database for cancer driver gene research. Nucleic Acids Res. 48, D863–D870.
Liu, Y., Zhang, J., Li, L., Yin, G., Zhang, J., Zheng, S., Cheung, H., Wu, N., Lu, N., Mao, X., et al., 2016. Genomic heterogeneity of multiple synchronous lung cancer. Nat. Commun. 7, 13200.
Lynch, M., 2007. The evolution of genetic networks by non-adaptive processes. Nat. Rev. Genet. 8, 803–813.
Ma, P., Fu, Y., Cai, M.C., Yan, Y., Jing, Y., Zhang, S., Chen, M., Wu, J., Shen, Y., Zhu, L., et al., 2017. Simultaneous evolutionary expansion and constraint of genomic heterogeneity in multifocal lung cancer. Nat. Commun. 8, 823.
Martinez-Jimenez, F., Muinos, F., Sentis, I., Deu-Pons, J., Reyes-Salazar, I., Arnedo-Pac, C., Mularoni, L., Pich, O., Bonet, J., Kranas, H., et al., 2020. A compendium of mutational cancer driver genes. Nat. Rev. Cancer 20, 555–572.
Marusyk, A., Tabassum, D.P., Altrock, P.M., Almendro, V., Michor, F., Polyak, K., 2014. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. Nature 514, 54–58.
McGranahan, N., Swanton, C., 2015. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. Cancer Cell 27, 15–26.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.
McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., Cunningham, F., 2016. The Ensembl Variant Effect Predictor. Genome Biol. 17, 122.
Morris, S.C., 1998. The Crucible of Creation: the Burgess Shale and the Rise of Animals. Oxford University Press.
Rosenthal, R., Cadieux, E.L., Salgado, R., Bakir, M.A., Moore, D.A., Hiley, C.T., Lund, T., Tanic, M., Reading, J.L., Joshi, K., et al., 2019. Neoantigen-directed immune escape in lung cancer evolution. Nature 567, 479–485.
Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F., 2021. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 71, 209–249.
Tao, Y., Hu, Z., Ling, S., Yeh, S.-H., Zhai, W., Chen, K., Li, C., Wang, Y., Wang, K., Wang, H.-Y., et al., 2015. Further genetic diversification in multiple tumors and an evolutionary perspective on therapeutics. bioRxiv. https://doi.org/10.1101/025429.
Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al., 2019. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 47, D941–D947.
Thress, K.S., Paweletz, C.P., Felip, E., Cho, B.C., Stetson, D., Dougherty, B., Lai, Z., Markovets, A., Vivancos, A., Kuang, Y., et al., 2015. Acquired *EGFR* C797S mutation mediates resistance to AZD9291 in non–small cell lung cancer harboring *EGFR* T790M. Nat. Med. 21, 560–562.
Trigos, A.S., Pearson, R.B., Papenfuss, A.T., Goode, D.L., 2017. Altered interactions between unicellular and multicellular genes drive hallmarks of transformation in a diverse range of solid tumors. Proc. Natl. Acad. Sci. U. S. A. 114, 6406–6411.
Vitale, I., Shema, E., Loi, S., Galluzzi, L., 2021. Intratumoral heterogeneity in cancer progression and response to immunotherapy. Nat. Med. 27, 212–224.
Wang, J., Cazzato, E., Ladewig, E., Frattini, V., Rosenbloom, D.I., Zairis, S., Abate, F., Liu, Z., Elliott, O., Shin, Y.J., et al., 2016. Clonal evolution of glioblastoma under therapy. Nat. Genet. 48, 768–776.
Wang, T., Ruan, S., Zhao, X., Shi, X., Teng, H., Zhong, J., You, M., Xia, K., Sun, Z., Mao, F., 2021. OncoVar: an integrated database and analysis platform for oncogenic driver variants in cancers. Nucleic Acids Res. 49, D1289–D1301.
Williams, M.J., Werner, B., Barnes, C.P., Graham, T.A., Sottoriva, A., 2016. Identification of neutral tumor evolution across cancer types. Nat. Genet. 48, 238–244.
Wu, C.I., Wang, H.Y., Ling, S., Lu, X., 2016. The ecology and evolution of cancer: the ultra-microevolutionary process. Annu. Rev. Genet. 50, 347–369.
Ye, C., Wang, J., Li, W., Chai, Y., 2016. Novel strategy for synchronous multiple primary lung cancer displaying unique molecular profiles. Ann. Thorac. Surg. 101, e45–e47.
Yu, G., Wang, L.G., Han, Y., He, Q.Y., 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16, 284–287.
Zairis, S., Khiabanian, H., Blumberg, A.J., Rabadan, R., 2014. International Conference on Brain Informatics and Health. IEEE Computer Society.
Zhai, W., Lim, T.K., Zhang, T., Phang, S.T., Tiang, Z., Guan, P., Ng, M.H., Lim, J.Q., Yao, F., Li, Z., et al., 2017. The spatial organization of intra-tumour heterogeneity and evolutionary trajectories of metastases in hepatocellular carcinoma. Nat. Commun. 8, 4565.
Zhou, Q., Zhang, X.C., Chen, Z.H., Yin, X.L., Yang, J.J., Xu, C.R., Yan, H.H., Chen, H.J., Su, J., Zhong, W.Z., et al., 2011. Relative abundance of *EGFR* mutations predicts benefit from gefitinib treatment for advanced non-small-cell lung cancer. J. Clin. Oncol. 29, 3316–3321.