Original Article

Clinical relevance of mutant-allele tumor heterogeneity and lung adenocarcinoma

Hengyu Mao

Department of Thoracic Surgery, Fudan University Shanghai Cancer Center, Shanghai 200032, China
Correspondence to: Hengyu Mao, MD. Department of Thoracic Surgery, Fudan University Shanghai Cancer Center, 270 Dong-An Road, Shanghai 200032, China. Email: hengyu.mao@gmail.com.

Background: Intra-tumor heterogeneity (ITH) plays an important role in the progressing of lung adenocarcinoma (LUAD). We used a mutant-allele tumor heterogeneity (MATH) algorithm to measure ITH and the correlation between ITH and clinical parameters and its prognosis.
Methods: We assessed 230 LUAD patients from The Cancer Genome Atlas (TCGA). We calculated the MATH values from whole-exome sequencing data and further investigated their correlation with clinical characteristics.
Results: The patients were divided into low and high MATH groups. People with high MATH were more likely to be female, smoking and EGFR mutation. And a high MATH may predict a poor prognosis.
Conclusions: MATH is a new method for describing the internal heterogeneity of tumors in lung cancer. We used TCGA data to demonstrate that groups with higher MATH values are more likely to be female, smoker and EGFR mutations, and may have a poor prognosis.

Keywords: Prognosis; non-small cell lung cancer (NSCLC); mutant-allele tumor heterogeneity (MATH)

Introduction

The incidence and mortality of lung cancer have increased steadily worldwide in the past several decades (1). Lung adenocarcinoma (LUAD) is the most common histological subtype of non-small cell lung cancer (NSCLC) (2). In the previously published literature, we confirmed a number of factors associated with LUAD (3). However, for LUAD, we still have a lot of unknown. For this reason, we need better tools to describe the characteristics of LUAD and predict the prognosis.

Several studies have shown that tumors are considered to be derived from the conversion of multiple genes in normal cells (4). This mutation exists in the entire process of tumor formation while these cells will continue to mutate in the tumor (4). ITH denotes the coexistence of subpopulations of cancer cells that differ in their genetic, phenotypic or behavioral characteristics within a given primary tumor, and between a given primary tumor and its metastasis (5).

There are many theories to explain the formation and function of ITH. A reasonable explanation is that tumor development is a Darwinian evolutionary process (6), involving the interplay between cancer sub-clones and the local immune microenvironment. Mutual mutations and evolution between tumor cells cause changes in ITH. Thereby, the tumor develops in a direction favorable to its survival and enhances its drug resistance. The more heterogeneous a tumor has the greater potential for its evolution, the stronger its ability to survive. The category of ITH includes linear, branching, neutral, punctuated and mixed (7).

Mutated cells produce clones inside the tumor and counting these clones can detect heterogeneity. In general, greater genetic heterogeneity may cause worse prognosis. This diversity can be attributed to genetic and epigenetic factors and non-hereditary mechanisms such as adaptive responses or fluctuation in signaling pathways. Larger genetic heterogeneity can contain many subtypes. These

are more likely to include some genotypes that are more capable of transferring more resistant and more. This means that highly heterogeneous tumors may be detrimental to drug resistance and easier to metastasize (8-10). In other words, it is easier to make the tumor to develop a worse direction. Making tumors more prone to adverse mutations. Unfortunately, there is currently not a particularly good standard for assessing heterogeneity, so new approaches to assessing heterogeneity are imminent.

Next-generation sequencing (NGS) of tumor DNA has been a widely used method in scientific and clinical research (11). Unlike other single-cell heterogeneity evaluations, which are directed at the detection of the overall heterogeneity of a tumor (12). At the same time, its requirements for the organization are simple. The distribution of mutant-allele fractions among loci thus provides a straightforward measure of one type of ITH, called mutant-allele tumor heterogeneity (MATH) (13). We noticed that intra-tumor heterogeneity (ITH) noted in a range of tumor types. A measure of this ITH might provide clinically significant information.

Genetically distinct subpopulations of cells in a tumor lead to differences among mutated loci in terms of the fraction of sequence reads that show a mutant allele and the role of MATH in many recent cancers has been revealed. In head and neck squamous cell carcinoma, MATH is mainly associated with disruptive TP53 mutations, human papillomavirus (HPV) status and increasing exposure to cigarette smoke. Higher MATH was strongly associated with shorter overall survival (OS), too (13,14). In breast cancer, we know that T classification and both clinical and molecular subtypes PAM50 were related to MATH (15). This article will explore the relationship between MATH and clinical status and prognosis in LUAD.

Methods

Information of 230 patient is downloaded from The Cancer Genome Atlas (TCGA) at 2017.8.1. All patients have complete clinical information and follow-up information. The median absolute deviation (MAD) is a robust measure of the sample bias for univariate numerical data. It is also a population parameter about the sample. For the univariate datasets X1, X2, ..., XN, MAD is defined as the median of the absolute deviation of the data points to the median:

$$MAD = median[|Xi - median(X)|]$$

The calculation of MAD is followed the default in R,

with values scaled by a constant factor (1.4826) so that the expected MAD of a sample from a normal distribution equals the standard deviation (13).

Each tumor's MATH value was calculated from the (MAD) and the median of its mutant-allele fractions at tumor-specific mutated loci:

MATH $=100\times$ MAD/median[(t_alt_count)/(t_alt_count + t_ref_count) $\times100]$

Results

MATH primarily indicates the width of the center ratio of mutant allele fractions in the tumor-specific mutation loci. Each patient used in this article has a clear relevant data and follow-up information. In order to exclude interference from non-tumor-related factors in the study, we excluded the follow-up data eight years after the operation. *Figure 1* shows some MATH related information. The chart shows that the distribution of MATH generally tends to be a normal distribution. It's 5% percentile is 30.9 and its 95% percentile is 90.4. The 50% percentile is 54.6. The highest peak (45–55%) of the MATH value appears mainly between 53.4–56.22. However, the distribution may not be a very strict normal distribution because of the small sample size (P=0.031). At the same time, we intuitively demonstrated a relationship between MATH and age, gender, race and stage. In the *Figure 1C,D,E*, Groups 1 to 6 refer to the change in the clinical characteristics of the MATH as the MATH increases, from the small to the large, and the size is divided into six groups by size. We can also clearly see that there is a clear relationship between MATH and smoking (P=0.025). Smokers apparently have higher MATH than non-smokers (*Table 1*).

In the meantime, in order to clarify the clinical information related to MATH and better guide clinical application, we divided the study into two groups according to the median MATH value (54.6). After comparison, we found some clinical information related to MATH. People with high MATH are more likely to be women and more likely to smoke. More EGFR mutation was found in a group of people with high MATH (*Table 2*).

More importantly, MATH is directly related to the patient's prognosis. The results showed that patients with higher MATH may have a poorer OS (*Figure 2*). This is in line with our clinical expectations and has great guiding significance for the prognosis.
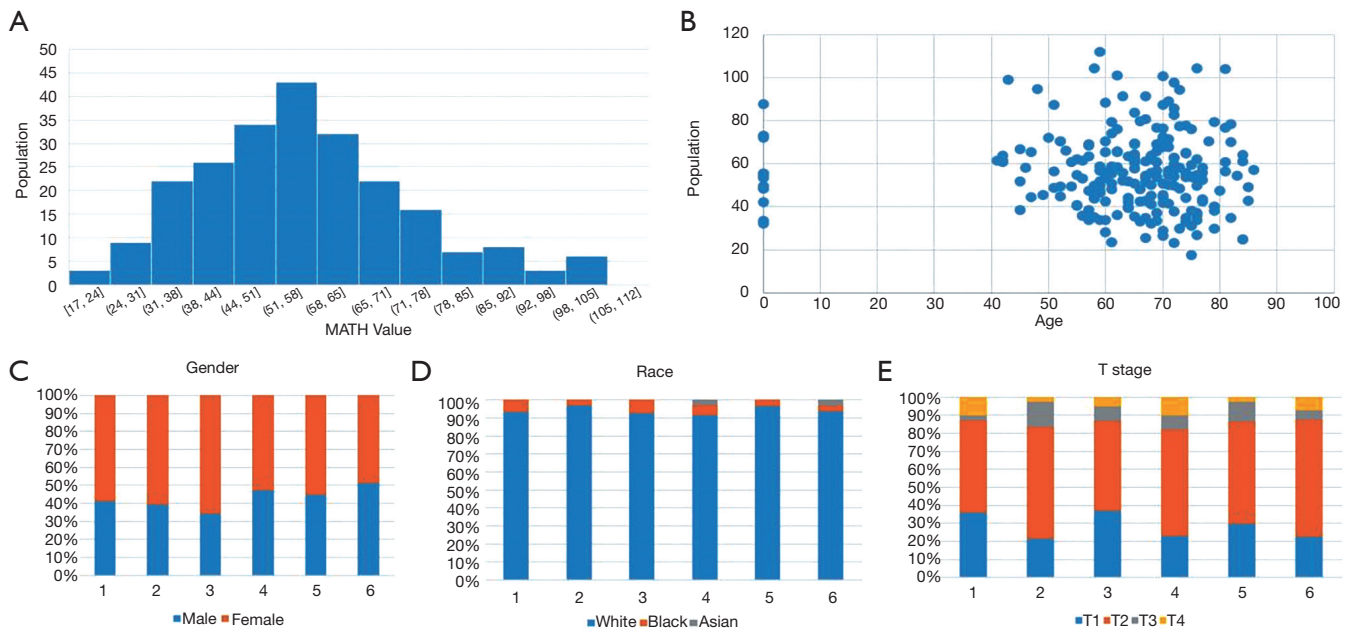
Figure 1 Distribution of the mutant-allele tumor heterogeneity (MATH) value and its relationship with clinical factors. (A) Histogram of MATH values among patients. (B) Relationship between MATH and age. (C,D,E) Relationship between gender, race, and T stage.

Table 1 Relationship between clinical characteristics and MATH values

Characteristic	Number	MATH ± SD	P value
Age in years			
≤67	100	56.96±17.26	0.4
>67	116	55.67±18.75	
Race			
Asian	2	74.85±28.35	0.302
Black or African American	9	50.24±21.32	
White	173	56.76±17.58	
T classification			0.31
1	65	54.58±17.03	
2	132	57.01±18.40	
3	18	57.1±15.10	
4	15	56.56±20.18	
N classification			
0	139	57.27±19.04	0.31
1	91	54.82±15.78	
Smoking			
Yes	72	60.20±18.41	0.025
No	158	54.53±17.32	
Gender			
Male	99	57.70±18.62	0.302
Female	131	55.25±17.20	

MATH, mutant-allele tumor heterogeneity.

Table 2 Relationship between characteristics and MATH value grouping

Characteristic	High, n (%)	Low, n (%)	P value
Gender			
Male	55 (47.8)	71 (61.7)	0.143
Female	60 (52.2)	44 (38.3)	
N			
0	69 (60.0)	70 (60.9)	0.893
1	46 (40.0)	45 (39.1)	
T			
1	29 (25.2)	36 (31.3)	0.779
2	69 (60.0)	63 (54.8)	
3	9 (7.8)	9 (7.8)	
4	8 (7.0)	7 (6.1)	
Smoking			
Yes	43 (37.4)	29 (25.2)	0.047
No	72 (62.6)	86 (74.8)	
Race			
Asian	2 (2.1)	0 (0.0)	0.357
Black or African American	4 (4.2)	5 (5.6)	
White	89 (93.7)	84 (94.4)	
EGFR			
No	34 (68.0)	38 (82.6)	0.099
Yes	16 (32.0)	8 (17.4)	
KRAS			
No	48 (78.7)	52 (86.7)	0.247
Yes	13 (21.3)	8 (13.3)	
ALK			
No	38 (90.5)	37 (92.5)	0.743
Yes	4 (9.5)	3 (7.5)	

MATH, mutant-allele tumor heterogeneity; EGFR, epidermal growth factor receptor; KRAS, Kirsten rat sarcoma viral oncogene; ALK, anaplastic lymphoma kinase.

Discussion

NGS is already a very mature technology, and it has also begun to be gradually applied clinically. Based on NGS, we also got a lot of information that we have not mentioned before (16). Unlike other methods, MATH does not depend on a particular gene. The relationship between MATH and the clinical features and prognosis of tumors has given MATH great potential as a clinical and scientific marker of tumors. The significance of MATH in head and neck cancer and breast cancer has been reported, and its use in lung cancer has great potential. Therefore, we used the TCGA database to study its role in adenocarcinoma. At first, we study the distribution of MATH. It tends to be normally distributed. This visually shows its distribution in the crowd. In addition, we found that their MATH
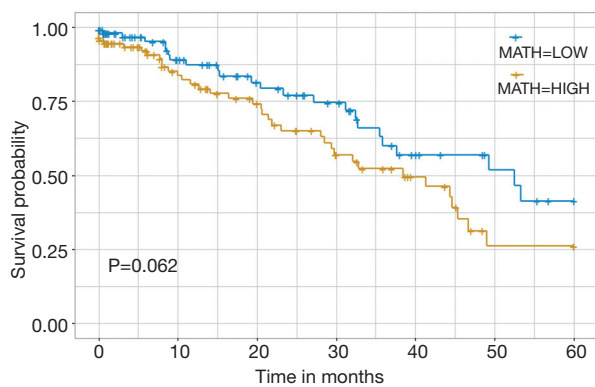
Figure 2 Kaplan-Meier survival curves for OS according to people with high MATH and low MATH. OS, overall survival; MATH, mutant-allele tumor heterogeneity.

values were significantly higher than those of head and neck squamous cell carcinoma and breast cancer. This indicates that lung cancer is more heterogeneous than head and neck squamous cell carcinoma and breast cancer, suggesting that lung cancer in the prognosis, metastasis, resistance and other aspects generally worse than the head and neck squamous cell carcinoma and breast cancer. This can still be mutually verified in our general conclusion. On the other hand, it explains the particularity of lung cancer. We also visually demonstrated a relationship between MATH and age, gender, race and staging. Clinical features make sense for our clinical reference. Especially when it comes to the characteristics associated with the patient's prognosis. We extracted the MATH and the patient's clinical features. Here, we found that people who smoked more MATH values, which may be interpreted as smoking as a recognized influencing factor, also had an impact on tumor internal heterogeneity. The existence of this effect in adenocarcinomas has important clinical implications. In the meantime, the association between high MATH and EGFR can also guide clinical development. Especially for the application of TKI drugs can have some guidance. In addition, regarding the relationship between TKI resistance and MATH, we think it is necessary to conduct future clinical trials to conduct research.

The most significant conclusion is that the MATH approach can serve as a clinical guideline as a way of predicting the patient's prognosis. One of the biggest purposes of clinical treatment is to improve the prognosis and to judge the prognosis of different patients especially in the increasingly popular NGS using today, the use

of NGS to help clinically is a very meaningful medical behavior. It is noteworthy that, in today's treatment system, whether it is chemotherapy or surgery, surgery, evaluation of the prognosis is often very important. In the current study, complete, long-term follow-up is relatively lacking. However, the importance of follow-up is beyond doubt, so we chose the TCGA database for research. At the same time, for different MATH worthy of patients we should also be flexible use of surgical radiotherapy and chemotherapy for treatment. In chemotherapy, the tumor can also be based on the level of heterogeneity to the most appropriate selection of the most appropriate drug combination in order to achieve the purpose of personalizing precision medical treatment, improve efficacy and reduce adverse reactions.

The shortcoming of this method lies in its complete dependence on NGS. Meanwhile, the relationship between specific MATH and medication also needs further clinical trials. At the same time, we can retrospectively conduct NGS based on the results of existing clinical trials to study the specific effects of various methods such as drugs, surgery and radiotherapy under various MATH values. This deserves our further consideration.

Acknowledgments

Footnote

References

1. Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. CA Cancer J Clin 2015;65:87-108.
2. Zhao Y, Wang R, Shen X, et al. Minor components of micropapillary and solid subtypes in lung adenocarcinoma are predictors of lymph node metastasis and poor prognosis. Ann Surg Oncol 2016;23:2099-105.
3. Xu-Welliver M, Carbone DP. Blood-based biomarkers

in lung cancer: prognosis and treatment decisions. Transl Lung Cancer Res 2017;6:708-712.

4. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646-74.

5. Nowell PC. The clonal evolution of tumor cell populations. Science 1976;194:23-8.

6. McGranahan N, Swanton C. Cancer Evolution Constrained by the Immune Microenvironment. Cell 2017;170:825-7.

7. Davis A, Gao R, Navin N. Tumor evolution: Linear, branching, neutral or punctuated? Biochim Biophys Acta Rev Cancer 2017;1867:151-61.

8. Greaves M, Maley CC. Clonal evolution in cancer. Nature 2012;481:306-13.

9. Carter SL, Cibulskis K, Helman E, et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol 2012;30:413-21.

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. Reis-Filho JS. Next-generation sequencing. Breast Cancer Research 2009;11:S12.

12. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248-9.

13. Mroz EA, Rocco JW. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. Oral Oncol 2013;49:211-5.

14. Mroz EA, Tward AD, Pickering CR, et al. High intratumor genetic heterogeneity is related to worse outcome in patients with head and neck squamous cell carcinoma. Cancer 2013;119:3034-42.

15. Ma D, Jiang YZ, Liu XY, et al. Clinical and molecular relevance of mutant-allele tumor heterogeneity in breast cancer. Breast Cancer Res Treat 2017;162:39-48.

16. Salk JJ, Fox EJ, Loeb LA. Mutational heterogeneity in human cancers: origin and consequences. Annu Rev Pathol 2010;5:51-75.