SPANDIDOS PUBLICATIONS

ONCOLOGY REPORTS 26: 1539-1546, 2011

Remarkable difference of somatic mutation patterns between oncogenes and tumor suppressor genes

HAOXUAN LIU, YUHANG XING, SIHAI YANG and DACHENG TIAN

State Key Laboratory of Pharmaceutical Biotechnology, Department of Biology, Nanjing University, Nanjing, P.R. China

Received July 15, 2011; Accepted August 19, 2011

DOI: 10.3892/or.2011.1443

Abstract. Cancers arise owing to mutations that confer selective growth advantages on the cells in a subset of tumor suppressor and/or oncogenes. To understand oncogenesis and diagnose cancers, it is crucial to discriminate these two groups of genes by using the difference in their mutation patterns. Here, we investigated >120,000 mutation samples in 66 well-known tumor suppressor genes and oncogenes of the COSMIC database, and found a set of significant differences in mutation patterns (e.g., non-3n-indel, non-sense SNP and mutation hotspot) between them. By screening the best measurement, we developed indices to readily distinguish one from another and predict clearly the unknown oncogenesis genes as tumor suppressors (e.g., ASXL1, HNF1A and KDM6A) or oncogenes (e.g., FOXL2, MYD88 and TSHR). Based on our results, a third gene group can be classified, which has a mutational pattern between tumor suppressors and oncogenes. The concept of the third gene group could help to understand gene function in different cancers or individual patients and to know the exact function of genes in oncogenesis. In conclusion, our study provides further insights into cancer-related genes and identifies several potential therapeutic targets.

Introduction

Cancer is responsible for one in eighth deaths all over the world (1), and it is well accepted that cancer is a genetic disease caused by a sequential mutation of oncogenes and tumor suppressor genes (2). Oncogenes are mutated in ways that render the gene constitutively active or active under conditions in which the wild-type gene is not. Taking oncogene BRAF for example, the activated BRAF kinase was able to phosphorylate downstream targets such as extracellular signal-regulated kinase leading to uncontrolled growth (3). Tumor suppressor genes, which suppress tumorigenesis are mutated to reduce the activity of the gene product (4).

Correspondence to: Dr Dacheng Tian, Department of Biology, Nanjing University, Nanjing 210093, P.R. China
E-mail: dtian@nju.edu.cn

Key words: cancer, tumor suppressor gene, oncogene, mutation pattern

Nowadays, the central aim of cancer research has been to identify the mutated genes that are causally implicated in oncogenesis (5). As the sequencing method becoming cheaper and easier, many large-scale studies have been published identifying mutations both in coding regions and in whole genome of human tumors (6-12). Cancer research emphasis is more and more on large-scale sequence of cancer genome, in 2010, the international cancer genome consortium was launched to investigate genome sequences of 25,000 tumors (13). With the databases flooded with massive information, Bert Vogelstein (the Ludwig Center for Cancer Genetics and Therapeutics at Johns Hopkins), pointed out the obstacle of cancer research: 'The difficulty is going to be figuring out how to use the information to help people rather than to just catalogue lots and lots of mutations.'

With the massive mutational data, a great progress has been made to understand the somatic mutation pattern of cancer-related genes. The different patterns of mutations were noted between oncogene and tumor suppressor gene. In particular, tumor suppressor genes are characterized by diverse mutation types, ranging from SNPs and small indels to whole gene deletion, which have the common result of abolishing of the function of the gene product, oncogenes are mutated more conserved, both with respect to the type of mutation and its location in the gene, the mutations usually recurrent and are nearly always missense (14,15). It has also been observed that the distribution of 3n and non-3n indels in oncogenes and tumor suppressor genes is non-random and different, in which tumor suppressor genes have much more proportion of non-3n indels than oncogenes (16).

The distinct mutation patterns of the functionally-different genes in oncogenesis could be very helpful to detect oncogenic mutations at early stage and to discriminate the roles of a gene in this process. To reach these goals, it is essential to find appropriate measurements to characterize the detail mutation patterns of individual genes. Here, we analyzed a large number of both cancer-related and non-related genes as controls, and searched various parameters to define mutational patterns for each of these genes. Our analyses covered most of the well-known tumor suppressor genes and oncogenes with >120,000 mutational samples from the COSMIC database. In the total of 37 tumor suppressor genes and 29 oncogenes, we found a remarkable difference in the mutational patterns between them. In addition, our analysis confirmed a consistent mutation pattern for a gene in different tissues. Based on the highly

consistent results, a role played by a gene could be predicted. Indeed, some of oncogenesis-unknown genes can be identified as tumor suppressor (e.g., ASXL1, HNF1A and KDM6A) or as oncogenes (e.g., FOXL2, MYD88 and TSHR). These indices, developed by our study, could be very useful in the functional prediction of genes in oncogenesis and cancer diagnosis.

Materials and methods

Data source. All mutation data are obtained from the COSMIC database (the Catalogue of Somatic Mutations in Cancer; http://www.sanger.ac.uk/cosmic). This large-scale database, founded by the Welcome Trust Sanger Institute, is designed mainly to store and catalog somatic mutation information with regard to human cancers. Data in COSMIC are gathered from publications in scientific literature and the output of the genome-wide screens from the Cancer Genome Project (CGP) at the Sanger Institute (17,18). The frequently mutated genes usually are oncogenes and tumor suppressors that are involved in the generic processes including cell cycle control, signal transduction and stress responses (19). COSMIC was initiated in 2004 and by now is providing over 160,000 mutations in almost 19000 genes for investigation (20). The data can also be queried by tissue, which allows us to analyze different mutations occurring within different tissues.

Analysis of mutation pattern in cancer-related genes. Sixty-six genes with >20 mutated samples in COSMIC database were selected to analyze their mutation patterns, 37 of the 66 genes are suggested to be tumor suppressor genes while 29 of them are recommended to be oncogenes by previous studies. We defined mutation pattern of a gene by five statistic standards: the portion of indel mutation number to all indels and SNPs combined (indel/indel + SNP), the portion of non-3n-indel number to total indels (non-3n-indel/indel) (genes with one indel only were removed), the portion of non-sense mutation number to total SNPs (non-sense/SNP), the portion of synonymous SNP to total SNPs (synonymous/SNP) and the portion of missense SNP to total SNPs (missense/SNP).

Definition of mutation hotspots in cancer-related genes. Within the 66 genes mentioned above, 60 genes with >20 single-base substitution samples, 21 genes with >20 insertion samples and 31 genes with >20 deletion samples were selected for analysis of their mutation hotspots. We define one amino acid site (3 bp) with most mutations of each gene as the first unit, and denote the second to the fifth unit as the second to the fifth abundant mutation unit (3-bp region) in this gene. Based on these definitions, we calculate the portion of mutations in the first unit to the total number of mutations (first/total). Then we continued to calculate the following parameters, (first + second)/total, (first + second + third)/total, (first + second + third + fourth)/total and (first + second + third + fourth + fifth)/total. The higher the proportion, the more centralized the mutation distribution is within the gene.

Mutational analysis of the data from the same cancer gene in different tissues. Within these 66 genes, 16 tumor suppressor genes and 15 oncogenes, which mutated in more than one tissue and had no <20 mutational samples within each tissue, have been selected for the mutational analysis. We calculated the three good parameters (non-3n-indel/indel, non-sense/SNP and missense/SNP) for each of these genes.

Selection of genes as controls. The 1000 Genomes Project aims to characterize human genome sequence variation (21). The sequences in this project contain three parts of data: low-coverage sequencing of 179 individuals from four populations; high-coverage sequencing of two mother-father-child trios; and exon-targeted sequencing of 697 individuals from seven populations. The exon-targeted sequencing targeted capture of 8,140 exons from 906 randomly selected genes (total of 1.4 Mb) and found 5,708 synonymous SNPs, 7,063 non-synonymous SNPs, 59 small in-frame indels and 37 small frameshift indels. In total, the three parts of data identified 60,157 synonymous SNPs, 68,300 non-synonymous SNPs, 714 small in-frame indels and 954 small frameshift indels. We used the exon-targeted part of the project as control 1, and the whole project (three parts combined) as control 2. All five statistic parameters used in this study were calculated for these data.

Results

Difference of mutational types between tumor suppressor genes and oncogenes. To characterize mutational patterns for individual genes, we selected cancer-related genes as many as possible. In total, 37 tumor suppressor genes and 29 oncogenes satisfied our criterion: >20 mutational samples in the COSMIC database. We used mutational parameters to measure different types of mutations as many as possible for the characterization of cancer-related genes. These standards are the ratio of indel/(indel + SNP), the ratio of non-3n-indel/indel, the ratio of non-sense/SNP, the ratio of synonymous/SNP and the ratio of missense/SNP. In general, there are significant differences (t-test; $P<0.01$) between tumor suppressor genes and oncogenes in each of the above parameters (Fig. 1), suggesting that different mutational patterns exist between these two functional gene groups.

To evaluate effectiveness among parameters for the discrimination of two functional-known cancer-related genes, we used the average ratio and standard variation for each parameter. There is a large difference in the average ratios of all the five standards between tumor suppressor genes and oncogenes. In particular, the average ratios of indel/(indel + SNP) are 0.102 and 0.450 for oncogene and tumor suppressor gene, while the standard variations are almost the same, 0.202 and 0.193, respectively. The greatest difference is observed in the average ratios of non-3n-indel/indel, 0.167 and 0.891 with the standard variation 0.278 and 0.121, respectively, for the two groups of genes. The greatest ratio difference and the relatively small standard variation suggest that this parameter is the best one to distinguish these two functionally-different groups of cancer genes. On the other hand, the average values (± standard variations) of non-sense/SNP are 0.006±0.0129 (for oncogenes) and 0.404±0.226, where the tumor suppressor genes are nearly 70 times higher than oncogenes. The extremely low average ratio and standard variation in oncogenes indicate that the oncogenes separate in a much smaller range in this value and the non-sense/SNP may be a uniquely good parameter to characterize oncogenes. The average ratios of

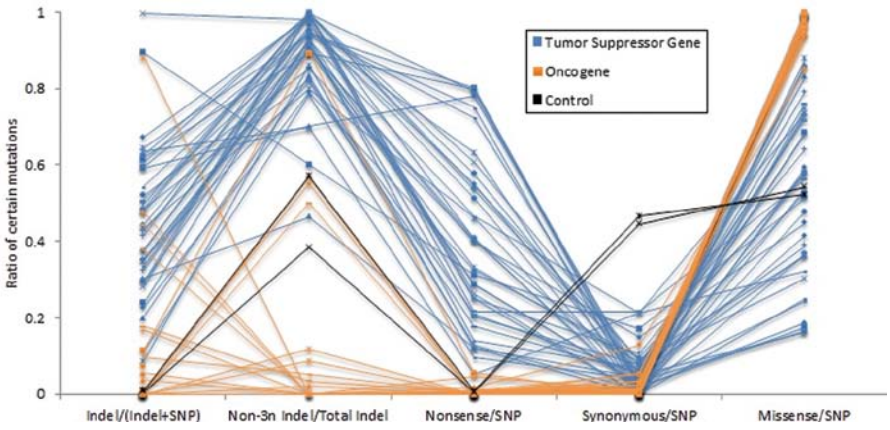

Figure 1. Mutation patterns of 66 tumor suppressor genes and oncogenes. Tumor suppressor genes are shown in blue lines while oncogenes are shown in yellow lines. Two control groups of genes are shown in black lines.

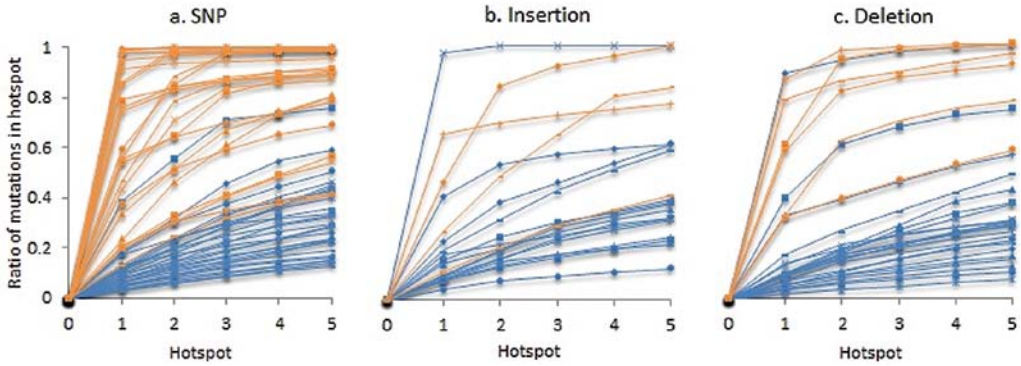

Figure 2. Mutation distribution of tumor suppressor genes and oncogenes. (a), (b) and (c) show the distribution of SNP, insertion and deletion, respectively. The horizontal axis (hotspot) stands for the first amino acid site (3-bp nucleotides) with the largest number of mutations, the first and the second sites (total 6-bp), ... , and all the five sites (first + second + third + fourth + fifth; total 15-bp). The definition of each amino acid site is described in Materials and methods. The vertical axis stands for to the proportion of mutations in corresponding number of hotspots to total mutations in a tumor suppressor or an oncogene. For example, if all mutations occur in the first site of amino acid, the proportion is equal to 1 at the first site. Of course, the proportion is still equal to 1, although no mutation occurs in the other four amino acid sites. Tumor suppressor genes are shown in blue lines while oncogenes are shown in yellow lines.

missense/SNP are 0.977±0.031 and 0.535±0.219, respectively for the groups. The smallest difference is found to be the average ratios of synonymous/SNP (0.014±0.027 and 0.065±0.064) in oncogenes and tumor suppressor genes, respectively, indicating that this is not a good parameter to separate two gene groups.

In addition, we could visually check the exotic gene numbers within a gene group. For example at indel/(indel + SNP) position in Fig. 1, there are 7 oncogenes within the ratio range of tumor suppressor genes. For the other parameters (orderly in Fig. 1), there are 3 oncogenes crossing over the border of tumor suppressor genes, no crossover between two gene groups, all the oncogenes located within the range of tumor suppressor genes, and only 1 exotic oncogene within the range of tumor suppressor genes, respectively. In fact, the visual inspection is highly consistent with the average ratio and standard variation for a parameter. Therefore, we conclude that non-3n-indel/indel, non-sense/SNP and missense/SNP are three good parameters to distinguish oncogenes and tumor suppressor genes, synonymous/SNP is not good, but indel/(indel + SNP) could be used although it is not as good as the top three parameters.

Difference of mutational distribution between tumor suppressor genes and oncogenes. Our analysis on mutational types revealed significant differences in the pattern of mutations between tumor suppressor genes and oncogenes. The results indicate that different distribution pattern along a gene may exist between these two groups of cancer-related genes. To detect the uneven distribution of mutations, we defined the first to the fifth 3-bp unit which has the most to the least abundant mutations (Materials and methods for details). Then a graph was drawn to show the distribution of these mutations (Fig. 2). Generally, mutation hotspots exit in both types of genes. However, for most of the genes, the mutation hotspots dominate in oncogenes compared with those in tumor suppressor genes. We analyzed mutation hotspots of single base substitution, insertion, deletion respectively in tumor suppressor genes and in oncogenes. The results are roughly the same, mutations in oncogenes usually bind to the same site or several amino acid sites while mutations in tumor suppressor genes are more separately located. We detected whether these hotspot correlates with GC% variance of gene sequence, and no clear relationship was found between the

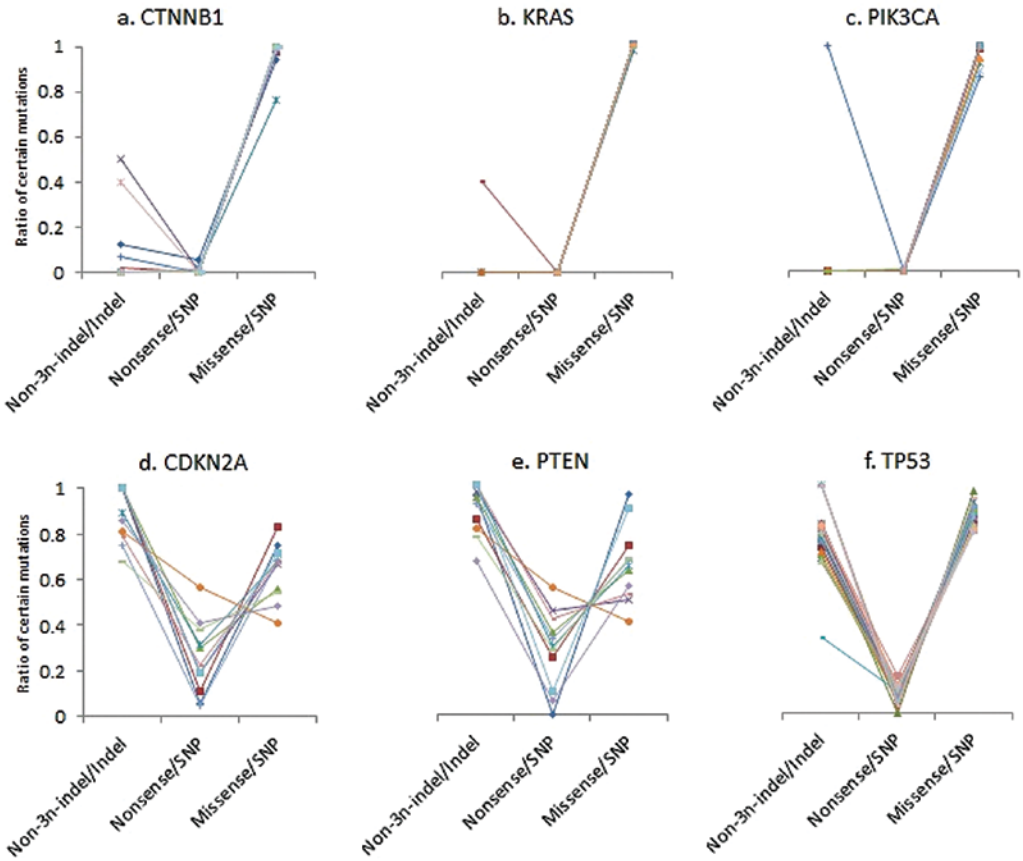

Figure 3. Mutation patterns of oncogenes and tumor suppressor genes in different tissues. Each color stands for one tissue for a gene. (a), (b) and (c) are oncogenes which have the most mutated tissues; (d), (e) and (f) are tumor suppressor genes with the most mutated tissues.

mutation hotspot location and GC% variance. Thus the hotspots are assumed to bind to gene function. In this scenario, for example in oncogenes, only the mutations in specific amino acid sites enable the gene to have the function in uncontrolled cell proliferation, resulting in such apparent mutational hot spots observed.

Mutation patterns of cancer genes in different tissues. In the COSMIC database, the mutation samples were from different cancer tissues, e.g., from kidney, lung, ovary, bone, brain and others (17). The different mutation patterns between oncogene and tumor suppressor gene could be only present in some tissues. To examine whether there are consistently different patterns between these two gene groups, we calculated the average ratios of three good parameters (non-3n-indel/indel, non-sense/SNP and missense/SNP) of the genes with enough samples in each tissue (Materials and methods for details).

First, we inspected whether there is a significant difference among tissues in the ratio of each parameter for each gene by Chi-square test (some examples in Fig. 3). Although certain variations exist in some ratios, especially in the parameter non-3n-indel/indel in oncogenes, no significant difference was found except for the gene PIK3CA in the liver tissue, in which the ratio of non-3n-indel/indel is significantly higher than that from the other tissues (P=0.05). This result indicates that there is a consistent mutation pattern among different tissues for most genes. However, due to the obvious variation in some ratios of certain genes, our result cannot exclude this possibility: that the function (oncogenesis or tumor suppressor) of certain genes may be different in some tissues (e.g., the great ratio variation of 3 oncogenes and 3 tumor suppressor genes in different tissues are shown Fig. 3).

Second, we checked the ratio differences among tissues between oncogene and tumor suppressor gene. Fig. 3 visually shows that the differences in all these ratios are consistently present between the two groups of genes, except for PIK3CA. PIK3CA is a known oncogene and contrarily high ratio of non-3n-indel/indel is the indicator of tumor suppressor gene, so in liver this gene may act differently. Overall in each of three parameters calculated, there is a significant difference between these two groups of genes by t-test (P<0.01). These results strongly suggest that in general, the oncogene and tumor suppressor gene have different mutational patterns in all tissues.

Finally, the mutation rate of oncogenes and tumor suppressor genes was studied. It is well known that mutated genes in cancer vary from one individual or one tissue to another. Some cancer genes often mutate in different tissues while the others tend to mutate in specific tissues (obviously shown in the COSMIC database). It is still unknown fully whether the function of a cancer gene changes from one tissue to another. Some cancer genes, like TP53 and CDKN2A, found to be mutated in diverse tissues but the mutation rates vary greatly. For example, TP53 has a mutation rate of 42% (4509/10626) in large intestine tumors but a mutation rate of 6% (66/1199) in cervix tumors. Although mutation rates of TP53 vary, its mutation patterns tend to remain unchanged in different tissues (also apply to most of

Table I. Candidate indices used to identify oncogenes and tumor suppressor genes.

Candidate indices	Ratio difference	Crossover genes
Non-3n-indel/indel	0.7241	16
Non-sense/SNP	0.3974	2
1-missense/SNP	0.4416	4
(Non-3n-indel + non-sense)/(indel + SNP)	0.6010	0
(SNP-missense + non-3n-indel)/(indel + SNP)	0.6204	0
(Non-sense + SNP-missense)/(indel + SNP)	0.4266	26
2/3*(non-sense + SNP-missense + non-3n-indel)/(indel + SNP)	0.5493	0
1/2*non-sense/SNP + 1/2*non-3n-indel/indel	0.5952	5
1/2*(1-missense)/SNP + 1/2*non-3n-indel/indel	0.6173	3
1/2*non-sense/SNP + 1/2*(1-missense)/SNP	0.4195	0
1/3*non-sense/SNP + 1/3*(1-missense)/SNP + 1/3*non-3n-indel/indel	0.5440	2

The ratio difference is the average difference between oncogene and tumor suppressor gene for each index. The crossover genes stand for the total number of gene overlapping between these two gene groups. (1-missense/SNP) is used here instead of missense/SNP, so generally the values of all the three single indices (non-3n-indel/indel, non-sense/SNP and 1-missense/SNP) of tumor suppressor genes are higher than oncogenes.

the other genes). This result indicates that the variable mutation rates for a gene in different tissues may not affect its function and that its mutation pattern is a more important reflection of its function.

Indices for identification of tumor suppressor genes and oncogenes according to their mutation pattern. According to the strong relationship between mutation pattern and gene function, an oncogene and a tumor suppressor gene could be identified from the oncogenesis-unknown genes. In COSMIC database, there are 11 oncogenesis-unknown genes with sufficient mutational samples (>20). To discriminate the function of these genes, it is necessary to know the best parameter or index of several parameters. In principle, there are two ways to develop such parameters or indices: one is the use of a single parameter and another of weighted parameters. To assess which way is the best, we still used the difference of average ratio and the number of crossover genes between two gene groups by using the 66 functionally known genes.

Table I shows the difference average ratio and the number of crossover genes. Clearly, the larger the difference or the smaller the number, the better it is for the parameter or index. Based on the values in Table I, each single parameter is not good for functional discrimination. For example for the parameter non-3n-indel/indel, the difference is the largest (0.724) but the number is also the largest (16). Therefore, we tried to use different approaches to calculate the weighted index. Technically, we could use two or three parameters and weigh them by the total number of mutations or either of indels and SNPs. For example, the parameter non-3n-indel/indel can be weighted by indel/total mutations (= indel + SNP), and the non-sense/SNP by SNP/total mutations. Then sum ratio of these weighted two parameters can be used as an index to measure the tendency of function in oncogenesis-unknown genes. In fact, this sum ratio is equal to (non-3n-indel + non-sense)/total mutations. By searching many approaches of calculations (orderly in Table I), three best ones were found (the third, fourth and sixth in Table I). In each of these indices, oncogenes (yellow ones) and tumor suppressor genes (blue ones) were well separated without overlapping (Fig. 4).

These indices provide better measurements to predict the function of oncogenesis-unknown genes in oncogenesis (the red ones in Fig. 4). Based on their positions, 3 out of 11 genes (ASXL1, HNF1A and KDM6A) locate within the range of tumor suppressor genes in all three indices as typical tumor suppressor genes. Three genes (FOXL2, MYD88 and TSHR) lie in the range of typical oncogenes in these indices. For the other 5 genes, they locate around the border between two groups of genes. The function of these genes cannot be determined here.

Discussion

Differences of mutation patterns between oncogenes and tumor suppressor genes. Cancers arise from somatic mutations that confer selective growth advantages on the cells (14). Therefore, extensive studies have recently focused on the detection of somatic mutation patterns in tumorigenesis (7,22). High mutation rates have been observed in cancer-related genes for a long time (5), and mutation hotspots have repeatedly been revealed in certain locations in these genes. Much higher proportion and greater diversity of indels are also commonly present in somatic mutation of cancer-related genes (16). Surprisingly, a much higher proportion of 3n-indels in oncogenes than that in tumor suppressor genes is found to be a general pattern, indicating that this distinctive characteristic could be used for cancer diagnosis (16). Recently, an example of the mutational difference in these two groups of genes was reported in the ovarian clear cell carcinoma (15). Based on this criterion, the authors obtained the prediction of PPP2R1A as an oncogene and ARID1A as a tumor-suppressor gene.

These studies display a great potential to use the mutational difference in two gene groups for the prediction of gene

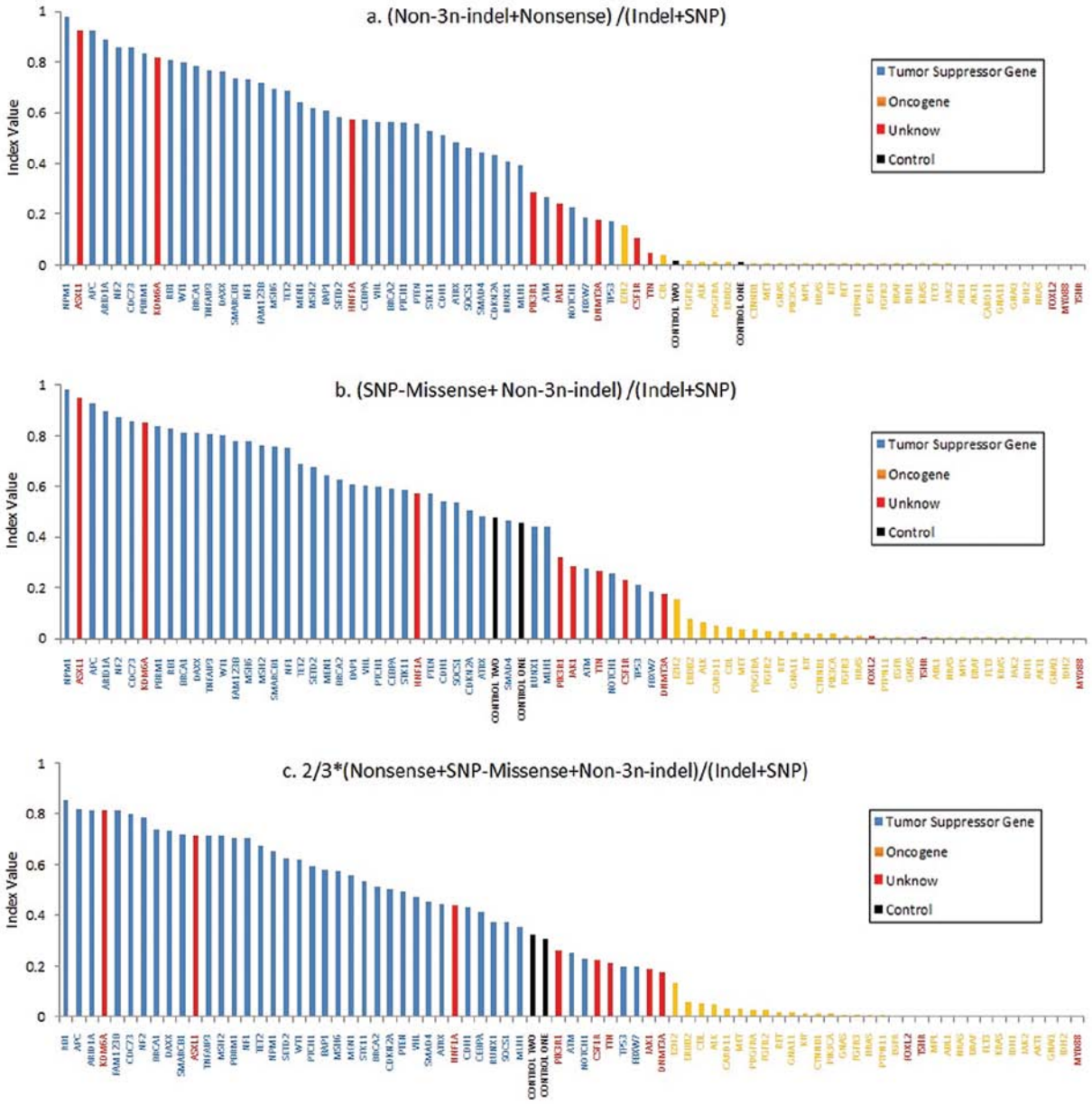

Figure 4. Index values of tumor suppressor genes, oncogenes, oncogenesis-unknown genes and control groups. (a), (b) and (c) described the weighted parameter (non-3n-indel + non-sense)/(indel + SNP), (SNP-missense + non-3n-indel)/(indel + SNP) and 2/3*(non-sense + SNP-missense + non-3n-indel)/(indel + SNP) for all these genes, respectively.

function, for cancer diagnosis and the mechanism of somatic mutation in cancer-related genes. Our study further shows that the essential difference lies in how deleterious the mutations are in these two gene groups. Clearly the non-3 sizes and abundant indel mutations are more destructive, and the non-sense and missense mutations are also harmful to gene function. By using a large number of genes, somatic mutations and various tissue samples, our results strongly demonstrate that the differences in destructive mutations exist as a general phenomenon between oncogene and tumor-suppressor gene. It is understandable that indels more easily cause a gene to lose function than single base substitution, among which non-3n indels are more destructive than 3n indels. Also the non-sense mutation will end up with pre-termination which is lethal to a gene. Accordingly, tumor suppressor genes have a higher portion of indel/(SNP + indel), non-3n-indel/indel and non-sense/SNP than oncogenes. The differences of mutation patterns between oncogenes and tumor suppressor genes are closely associated with their functional difference in oncogenesis, hence, the tumor suppressor genes are dysfunctional and oncogenes are activated (23).

Identification of tumor suppressor gene and oncogene. Personalized cancer therapy is based on targeting underlying genetic mutations involved in each patient, and presupposes that sustained inactivation of tumor suppressors and activation of oncogenes is essential in cancers (24). It has long been hoped that reactivation of tumor suppressor genes will finally cure cancer, and several methods have been used to search for tumor suppressor genes. Loss of heterozygosity is regarded as hallmark of chromosomal regions harboring tumor suppressor genes (25,26), and frequently promoter methylation in tumors

is another hallmark of a tumor suppressor gene that could be used for its identification. Recently, RNA interference is also used in identifying tumor suppressor genes (27). Now the difference of mutational patterns between tumor suppressor gene and oncogene could be a potential measurement for the prediction of gene function and cancer diagnosis.

Normally, the cancer-related genes are supposed to be either oncogenes or tumor suppressor genes (also known as dominant acting gene and recessive gene) (14,15). However, based on the distribution of these indices for the genes analyzed (Fig. 4), the function of the 11 genes could be categorized as three groups: typical oncogenes, typical tumor suppressor genes and functionally-uncertain genes (a transition state between two typical genes, five red bars in the middle in Fig. 4). This group consists of non-typical oncogenes or non-typical tumor suppressor genes. Maybe the function of this type of gene could vary. The concept of third gene group could help to understand gene function in different cancers or individual patients. With the adding of this group, the combinational use of those indices could provide more reliable results to predict the exact function of genes in oncogenesis. It is worth to note that all the indices developed are based on the genes with large mutation samples, it is difficult to distinguish the function of genes with few samples.

TP53, the best known tumor suppressor gene has been extensively studied for many years. A great progress has been made in cancer therapy by reactivation of TP53 in TP53-mediated tumors (24,28-30). On the other hand, several oncogenes, ABL, KIT and EGFR, have been used as the target for cancer therapy and shown significance clinical results (31-33). Thus, a method for the identification of tumor suppressor genes and oncogenes could be very useful for cancer therapy. The mutation-based identification developed in this study could be one of methods to effectively distinguish the function of cancer-related genes in oncogenesis.

Underlying mechanisms of specific mutation patterns in cancer-related genes. Oncogenesis is a development process analogous to Darwinian evolution (34). Historically, this process was considered to be a stepwise acquisition of new mutations, then the selection may eliminate less competitive cells or it may foster cells carrying mutations that confer competitive growth advantage, resulting in clonal expansion (35). Our study shows that mutations both in tumor suppressor genes and oncogenes share a hallmark of selection.

The ratio of non-synonymous/synonymous can be used to estimate the extent of selection overall on non-synonymous changes and often used for identifying driver mutations (7). The parameters and indices in this study may be better to find the driver forces for somatic mutations of cancer-related genes. For example, the non-3n-indel/3n-indel (small indels only) is also a strong indicator of selection. Non-3n-indel is more deleterious mutation analogous to non-synonymous SNP but 3n-indel is comparable to synonymous SNP. Compared with control groups, the values of non-3n-indel/3n-indel are two times more in 35 out of 37 tumor suppressor genes or six-times less in 28 out of 29 oncogenes. The significant difference must be a result of selection.

Though the underlying mechanism of the mutational patterns starts to be shown, the detail occurrence is still unknown. Maybe the mutation samples observed only reflect the final outcome which has experienced a long selection process. At the beginning of cancer occurrence, the mutations could take place randomly. Another possibility could be that the mutations are more or less induced at some fragile sites in cancer genes, and then these fragile sites mutate preferentially under certain circumstances such as exposure to carcinogens. If the fragile sites do exit, it will explain both the selection hallmark and the mutational hotspots. Further study is needed to explore the truth.

Acknowledgements

This study was supported by funding from the National Natural Science Foundation of China (30930049, 31071062 and J1103512) to D.T.

References

1. Garcia M: Global cancer facts and figures 2007. ACS, 2007.
2. Hahn WC and Weinberg RA: Modelling the molecular circuitry of cancer. Nat Rev Cancer 2: 331-341, 2002.
3. Wan PT, Garnett MJ, Roe SM, *et al*: Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. Cell 116: 855-867, 2006.
4. Vogelstein B and Kinzler KW: Cancer genes and the pathways they control. Nat Med 10: 789-799, 2004.
5. Futreal PA, Coin L, Marshall M, *et al*: A census of human cancer genes. Nat Rev Cancer 4: 177-183, 2004.
6. Sjöblom T, Jones S, Wood LD, *et al*: The consensus coding sequences of human breast and colorectal cancers. Science 314: 268-274, 2006.
7. Greenman C, Stephens P, Smith R, *et al*: Patterns of somatic mutation in human cancer genomes. Nature 446: 153-158, 2007.
8. Wood LD, Parsons DW, Jones S, *et al*: The genomic landscapes of human breast and colorectal cancers. Science 318: 1108-1113, 2007.
9. The Cancer Genome Atlas Research Network: Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455: 1061-1068, 2008.
10. Ding L, Getz G, Wheeler DA, *et al*: Somatic mutations affect key pathways in lung adenocarcinoma. Nature 455: 1069-1075, 2008.
11. Jones S, Zhang X, Parsons DW, *et al*: Core signal pathways in human pancreatic cancers revealed by global genomic analyses. Science 321: 1801-1806, 2008.
12. Dalgliesh GL, Furge K, Greenman C, *et al*: Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature 463: 360-363, 2010.
13. The International Cancer Genome Consortium: International network of cancer genome projects. Nature 464: 993-998, 2010.
14. Stratton MR, Campbell PJ and Futreal PA: The cancer genome. Nature 458: 719-724, 2009.
15. Jones S, Wang T, Shih L, *et al*: Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. Science 330: 228-231, 2010.
16. Yang H, Zhong Y, Peng C, Chen J and Tian D: Important role of indels in somatic mutations of human cancer genes. BMC Med Genet 11: 128-138, 2010.
17. Forbes SA, Bhamra G, Bamford S, *et al*: The Catalogue of Somatic Mutations in Cancer (COSMIC). Curr Protoc Hum Genet, Chapter 10: Unit 10.11, 2008.
18. Forbes SA, Tang G, Bindal N, *et al*: COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. Nucleic Acids Res 38: 652-657, 2010.
19. Yeang C, McCormick F and Levine A: Combinatorial patterns of somatic gene mutations in cancer. FASEB J 22: 2605-2622, 2008.
20. Bamford S, Dawson E, Forbes S, *et al*: The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer 91: 355-358, 2004.
21. The 1000 Genomes Project Consortium: A map of human genome variation from population-scale sequencing. Nature 467: 1061-1073, 2010.

22. Kan Z, Jaiswal BS, Stinson J, *et al*: Diverse somatic mutation patterns and pathway alterations in human cancers. Nature 466: 869-873, 2010.
23. Bishop JM and Weinberg RA (eds): Molecular Oncology. Scientific American Inc., New York, 1996.
24. Feldser DM, Kostova KK, Winslow MM, *et al*: Stage-specific sensitivity to p53 restoration during lung cancer progression. Nature 468: 572-575, 2010.
25. Knudson AG Jr: Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci USA 68: 820-823, 1971.
26. Weinberg RA: Tumor suppressor genes. Science 254: 1138-1146, 1991.
27. Bric A, Miething C, Bialucha CU, *et al*: Functional identification of tumor-suppressor genes through an in vivo RNA interference screen in a mouse lymphoma model. Cancer Cell 16: 324-335, 2009.
28. Ventura A, Kirsch DG, McLaughlin ME, *et al*: Restoration of p53 function leads to tumour regression in vivo. Nature 445: 661-665, 2007.
29. Xue W, Zender L, Miething C, *et al*: Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. Nature 445: 656-660, 2007.
30. Junttila MR, Karnezis AN, Garcia D, *et al*: Selective activation of p53-mediated tumour suppression in high-grade tumours. Nature 468: 567-571, 2010.
31. Druker BJ, Talpaz M, Resta DJ, *et al*: Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med 344: 1031-1037, 2001.
32. Demetri GD, von Mehren M, Blanke CD, *et al*: Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. N Engl J Med 347: 472-480, 2002.
33. Lynch TJ, Bell DW, Sordella R, *et al*: Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 350: 29-39, 2004.
34. Hanahan D and Weinberg RA: The hallmarks of cancer. Cell 100: 57-70, 2000.
35. Nowell PC: The clonal evolution of tumor cell populations. Science 194: 23-28, 1976.