Mutat Res Fund Mol Mech Mutagen 806 (2017) 27–30

Contents lists available at ScienceDirect

Mutat Res Fund Mol Mech Mutagen

journal homepage: www.elsevier.com/locate/mut

Mutational signatures efficiently identify different mutational processes underlying cancers with similar somatic mutation spectra

Nan Zhou, Yuan Yuan, Xin Long, Chuanfang Wu, Jinku Bao*

Key Laboratory of the State Ministry of Education for Bio-Resources and Ecologic Environment, College of Life Sciences, Sichuan University, Chengdu 610064, China

ARTICLE INFO

Keywords:
Mutational signature
Virus
Bladder
Cervix

ABSTRACT

Compared to analyzing mutations with conventional spectra, deciphering mutational signatures provides much greater resolution of biological processes that generate somatic mutations during cancer development. Previous studies of bladder urothelial cancer (BLCA) and cervical squamous cell carcinoma (CESC) mutational signatures failed to uncover different mutational processes underlying the two cancers, which diminishes the capability of mutational signature to differentiate between the two cancers. In this study, we deciphered and compared the mutational signatures of BLCA and CESC. Four BLCA mutational signatures were deciphered from 37,098 somatic mutations of 130 exomes. Five CESC mutational signatures were deciphered from 44,206 somatic mutations of 194 exomes. Three BLCA mutational signatures were very similar to the three CESC signatures. These mutational signatures exhibited common endogenous mutational processes during BLCA and CESC development. The respective BLCA and CESC mutational signature 4 revealed the role of viral infection in both cancers. Noticeably, CESC mutational signature 4 is a novel one that has not been described in other studies. In summary, we have demonstrated the similarities and differences between BLCA and CESC by deciphering mutational signatures. This study will shed light on the use of mutational signatures to clarify the mechanisms of endogenous and exogenous carcinogens that cause somatic mutations in human cancers.

1. Introduction

Cancer is the ultimate outcome of accumulation of somatic mutations in the genome [1,2]. During cancer development, somatic mutations result from mutational processes operative that affect DNA damage and repair mechanisms in addition to affecting the response to endogenous and exogenous carcinogens [3]. Each mutational process leaves a mutational signature on the cancer genome [4]. At present, signatures of mutational processes can be deciphered using several bioinformatics tools. Compared to conventional mutational spectra, which provide only the final mixture of strong exposures to dominant mutagenic processes, mutational signatures allow much greater resolution of insights into the diversity and complexity of somatic mutational processes underlying oncogenesis [4,5].

Bladder urothelial cancer (BLCA) and cervical squamous cell carcinoma (CESC) carry more somatic mutations than most other cancers. In our previous study, the spectra of single base substitutions in BLCA and CESC were found to be very similar [6]. BLCA and CESC mutational signatures have been studied together with other cancers [7]. However, those results were crude and did not explain the differences between BLCA and CESC. The association of mutational signatures with viral infection that plays a key role in CESC was not interpreted with a convincing argument. Therefore, we focused on BLCA and CESC in this study. The different mutational processes between BLCA and CESC have been described in detail. The different roles of viral infection in BLCA and CESC have also been explained.

2. Materials and methods

2.1. Data collection and preparation

Somatic mutations of BLCA and CESC were downloaded from version 23 of data release of the ICGC (International Cancer Genome Consortium) data portal [8]. The data originated from two studies in the TCGA (The Cancer Genome Atlas) studies concerning invasive urothelial bladder cancer and cervical squamous cell carcinoma [9,10]. The somatic mutations were called from exome sequences of these two cancer types. Duplicates were removed according to the ICGC mutation identifier. Somatic mutations other than single base substitutions were also removed. Finally, the somatic mutation data used in this study included 81,304 single base substitutions (37,098 for BLCA and 44,206 for CESC) and 324 exomes (130 for BLCA and 194 for CESC).

* Corresponding author.
E-mail address: baojinku@scu.edu.cn (J. Bao).

http://dx.doi.org/10.1016/j.mrfmmm.2017.07.004
Received 3 May 2017; Received in revised form 3 June 2017; Accepted 14 July 2017
Available online 19 July 2017
0027-5107/ © 2017 Published by Elsevier B.V.

2.2. Mutation motifs

Single base substitutions were classified into six categories: 1.) C > A; 2.) C > G; 3.) C > T; 4.) T > A; 5.) T > C; and 6.) T > G. Substitutions of G and A were converted to C and T on the complementary strand. Each of a 5′ and 3′ flanking base immediate to the mutated base was integrated, generating 16 substitution trinucleotides for each of the six categories. In total, 96 mutation motifs were generated to demonstrate all types of single base substitutions.

2.3. Mutational spectrum

All substitutions within a cancer were represented by a mutational spectrum consisting of 96 mutation motifs. Given a cancer with M substitutions in total, the absolute frequency of each of the 96 mutation motifs was counted. The empirical probability of each mutation motif was calculated by normalizing its absolute frequency by M. The mutational spectrum of the cancer was visualized as a heatmap.

2.4. K-S test

The Kolmogorov-Smirnov (K-S) test is a nonparametric test of the equality of probability distributions. It can be used to compare two samples by quantifying the largest distance between their cumulative distribution functions [11]. The two-sample K-S test is one of the most useful and general nonparametric methods for comparing two samples as it compares overall distribution shapes rather than specific central tendencies, dispersions, or other parameters. The K-S test was used to test if the mutational spectra of BLCA and CESC were drawn from the same distribution. The two-sample K-S test was implemented by the ks.test function in R [12]. The significance level was set to 0.05. A p-value of $> = 0.05$ indicates that there is insufficient evidence to reject the null hypothesis that BLCA and CESC have the same mutational spectra.

2.5. Deciphering mutational signatures

Mutational signatures of BLCA and CESC were deciphered by the Wellcome Trust Sanger Institute framework (WTSI) [13]. It is a computational framework that has been widely used to decipher signatures of mutational processes operative in human cancer [14,15]. Identifying the number (S) of mutational processes operative in a cancer is required prior to deciphering mutational signatures [13]. WTSI was set to evaluate different values of S ranging from 1 to 10. The value selected as the number of mutational signatures was the largest value of S for which the stability was > 0.9 with low error. WTSI then was used to decipher the mutational signatures of the cancer with this value.

2.6. Similarity between mutational signatures

When a mutational signature is presented as a numerical vector, the similarity between two mutational signatures corresponds to the correlation between two vectors. This is usually quantified as the cosine of the angle between the vectors [16]. In our study, cosine similarity was calculated to evaluate the similarity between two mutational signatures. A cosine similarity of 1 indicated two mutational signatures are identical while 0 indicated distinct mutational signatures. Given two mutational signatures, Sa and Sb, the cosine similarity between them was calculated using the equation:

$$Similarity = \cos(\theta) = \frac{Sa \cdot Sb}{|Sa||Sb|}$$

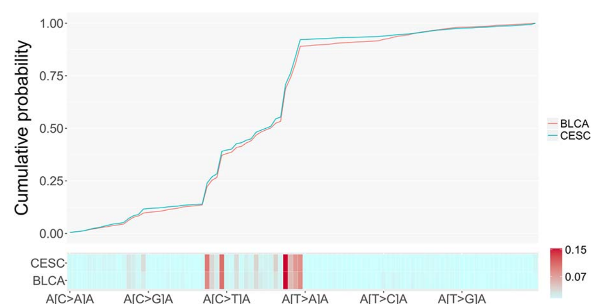

Fig. 1. Comparison of BLCA and CESC mutational spectra. Top panel, K-S test of the equality of the mutational spectra of BLCA and CESC. Bottom panel, heatmap showing the spectra of 96 mutation motifs of BLCA and CESC. The 96 motifs of substitutions are ordered as A[C > A]A/C/G/T, C[C > A]A/C/G/T, …, T[T > G]A/C/G/T.

3. Results

3.1. BLCA and CESC mutational spectra

As shown in Fig. 1, BLCA and CESC had very similar mutational spectra. C > G transversions and C > T transitions were the dominate substitutions in BLCA and CESC, while other types of substitutions only made up a very small proportion. There was a high prevalence of C > G transversions at TpCpA and TpCpT and high prevalence of C > T transitions at TpCpN (N can be any base). The preference of C > T transitions at NpCpG was also obvious in BLCA and CESC.

The K-S test indicated that the mutational spectra of BLCA and CESC were drawn from the same distribution, with p-value = 0.19. It can be clearly seen from Fig. 1 that BLCA and CESC were very similar in shape and position of the cumulative distributions across 96 mutation motifs.

3.2. BLCA and CESC mutational signatures

After evaluating different values of S for the mutational signatures, WTSI deciphered four mutational signatures from BLCA and five mutational signatures from CESC. As can be seen in Table 1, the stability and error for deciphering four BLCA mutational signatures were 0.96 and 191.49, respectively. The same evaluations of deciphering five CESC mutational signatures were 0.96 and 245.27, respectively. The high stability and low error indicated that BLCA and CESC mutational signatures were correctly deciphered by WTSI.

Three mutational signatures of BLCA and CESC were quite similar. They included the numbered mutational signatures 1, 2, and 3 and are shown in Fig. 2. Mutational signature 1 had a cosine similarity of 0.99 between BLCA and CESC. The cosine similarities of mutational signatures 2 and 3 between BLCA and CESC were 0.98 and 0.96, respectively. Mutational signature 1 was dominated by C > T transitions at

Table 1
Evaluation of number of BLCA and CESC mutational signatures.

Number of signatures	BLCA		CESC	
	Stability [a]	Error [b]	Stability	Error
1	1	503.94	1	1094.5
2	0.99	343.46	0.99	657.73
3	0.78	248.87	0.99	354.06
4	0.96	191.49	0.58	563.97
5	0.84	186.06	0.96	245.27
6	0.57	187.38	0.71	244.06
7	0.33	194.45	0.52	276.56
8	0.33	192.02	0.45	270.65
9	0.17	223.39	0.42	246.92
10	0.15	254.87	0.37	237.86

[a] Indicates the stability of deciphered mutational signatures.
[b] Indicates to what extent the mutational signatures were deciphered correctly.

Download English Version:

https://daneshyari.com/en/article/5528650

Download Persian Version:

https://daneshyari.com/article/5528650

Daneshyari.com