Bladder Cancer 9 (2023) 211–216
DOI 10.3233/BLC-230010
IOS Press

Research Report

Historical Pathogen-Driven Selection May Contribute to Contemporary Ethnic Difference in Bladder Cancer Susceptibility

Xiang-Yu Meng[a,b,1,*], Qiao-Li Wang[c,d,1], Ming-Jun Shi[e,1] and Hong-Yu Zhang[b,*]
[a]*Health Science Center, Hubei Minzu University, Enshi, China*
[b]*Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan, China*
[c]*Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden*
[d]*Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA*
[e]*Department of Urology, Beijing Friendship Hospital, Capital Medical University, Beijing, China*

Received 9 February 2023
Accepted 6 July 2023
Pre-press 1 August 2023
Published 25 September 2023

Abstract.
BACKGROUND: The rationale for ethnic differences in bladder cancer (BCa) susceptibility is an important open question. In this study, we raised the hypothesis that the APOBEC3-rs1014971 variant associated with BCa risk and APOBEC-mutagenesis probably contribute to ethnic differences.
METHODS: We calculated the ethnicity-stratified 5-year age-adjusted incidence rates of BCa using the US SEER database. We performed somatic mutational-signature analyses and compared the APOBEC-related mutational contribution across BCa tumors in patients of different ethnicities. We analyzed the allele frequency distribution of APOBEC3-related rs1014971 in contemporary populations of different ethnicities and in ancient human genomes. We also analyzed the natural selection profiles and ages of the investigated SNPs.
RESULTS: We validated the ethnic difference in BCa risk using US SEER data, revealing Caucasians to be at >2-fold greater risk than Asians / Pacific islanders. In contemporary populations, we observed a coherent ethnic distribution in terms not only of the allele frequency of APOBEC3-related rs1014971, but also the mutational contribution of APOBEC-mediated mutagenesis in BCa tumors. Population genetics and ancient genome analyses further suggested that the diverse ethnic distribution of rs1014971 could be rooted in human evolution.
CONCLUSIONS: It is possible that APOBEC3-related rs1014971 is involved in the different BCa incidence across ethnic groups, and this difference is potentially derived from human evolution. Our findings suggested an evolutionary link between contemporary population-level variations in malignancy susceptibility and pathogen-driven selection in the past, not unlike previously reported cases of certain autoimmune and metabolic disorders.

[1]These authors contributed equally.

*Correspondence to: Xiang-Yu MENG, Health Science Center, Medical School, Hubei Minzu University, Enshi 445000, China. E-mail: mengxy_whu@163.com. and Hong-Yu ZHANG, Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan 430070, China. E-mail: zhy630@mail.hzau.edu.cn.

INTRODUCTION

Bladder cancer (BCa) is the most common malignancy of the human urinary system, and the 6th most common cancer in all sites [1]. Established risk factors for BCa include age, male sex, tobacco smoking, and exposure to certain occupational hazards. It has also been noted that different ethnic populations have different susceptibilities to BCa; a factor which may have contributed to the geographic variability of BCa incidence [2]. The mechanism for this ethnic difference, however, remains unknown.

APOBEC-mediated mutagenesis contributes to the majority (~70%) of the mutation burden in BCa tumors [3–5]. APOBEC3 mutagens (particularly, *APOBEC3A* and *APOBEC3B*) played a crucial role in BCa carcinogenesis through the introduction of hotspot driver mutations in major BCa oncogenes and tumor suppressor genes [4]. An APOBEC3-related common genetic variant, rs1014971, a C>T substitution at position 39332623 of chr22 (hg19), was identified as being associated not only with BCa risk, but also with APOBEC3 gene expression and the APOBEC-mutagenesis burden in BCa tumors [6]. APOBEC3 enzymes are essential in the body's defense against pathogens, particularly retroviruses (e.g., HIV) which have exerted a strong selective pressure in human evolution [7]. These proteins have undergone rapid evolution, driven by the pathogen-host arms race [7]. It is also recognized that a variety of the infectious agents historically encountered by populations across different regions have resulted in various local adaptations of the pathogen-defense system, and empirical evidence has suggested a link between ancient pathogen invasions and contemporary susceptibility to diseases, such as certain metabolic and autoimmune disorders [8–10]. Given these findings, we hypothesized that the present-day ethnic variation in BCa susceptibility could be because APOBEC3-variants were differentially selected among ethnic populations during human evolution.

MATERIALS AND METHODS

Estimation of bladder cancer incidence

Data were derived from the Surveillance, Epidemiology, and End Results (SEER) Program, SEER*Explorer Database, of 22 areas, including San Francisco, Connecticut, Hawaii, Iowa, New Mexico, Seattle, Utah, Atlanta, San Jose-Monterey, Los Angeles, Alaska Native Registry, Rural Georgia, California (excluding SF/SJM/LA), Kentucky, Louisiana, New Jersey, Georgia (excluding ATL/RG), Idaho, New York, Massachusetts, Illinois, and Texas. The study population represented 48% of the entire US population for the period of 2015 to 2019.

Five-year age-adjusted incidence rates for 2015–2019 are presented as per 100,000, and calculated by adjustment to the 2000 US Standardized Population (19 age groups - Census P25-1130). The 95% confidence intervals were estimated based on gamma distribution [11].

Somatic mutation analysis

The whole-exome, sequencing-derived, somatic-mutation data of 602 bladder cancer tumors were compiled as described in our previous work [4]. *De novo* mutational-signature extraction and mutational-etiology origin analysis was performed using the Palimpsest package as previously described [12].

Natural selection analysis

Non-neutral signatures of skews in the site-frequency spectrum were analyzed using D_{FL} and F_{FL} metrics, and the threshold for statistical significance ($\alpha = 0.05$) was also estimated as previously described [13].

Allele frequency analysis in contemporary and ancient genomes

Allele frequency in contemporary populations of various ancestry (European, American, African, South Asian, East Asian) was calculated using 1000Genomes project (TGP) data (Sample size, $n = 5{,}008$) [14]. Allele frequency in ancient genomes collected from various regions (West Eurasia, Americas, Africa, South Asia, East Asia) was estimated using Allen Ancient DNA Resource (AADR) project data (1240K v54.1; https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/). Joint analysis of present ($n = 2{,}504$) and ancient ($n = 4{,}092$) genomes was performed to examine the present vs. ancient difference in rs1014971 allele frequency. Odds ratios were calculated, along with 95% confidence intervals and *P*-values. For the ancient genomes from West Eurasia ($n = 3{,}209$), we calculated the number of samples and rs1014971 allele frequency corresponding to each 1k-year unit (time to present),

and aggregated neighboring units into epochs by similarity in rs1014971 allele frequency, also conducting Fisher's tests to examine the statistical difference in rs1014971 allele frequency across these epochs.

For rs12628403, which was not included in the 1240K data, the data in the PGG.SNV database were used for the analysis [15].

Variant age estimation and eQTL analysis

Variant age was estimated as described above, using the joint clock of the mutation and recombination clocks [16]. The eQTL relationships between rs1014971/rs12628403 and *APOBEC3A*/*APOBEC3B* and the associated summary statistics (-log10 adjusted *P*-value) considered in this study are those calculated from pan-tissue meta-analysis by Metasoft as provided in the GTEx database (v8) [17].

RESULTS AND DISCUSSION

First of all, we validated the differences in BCa incidence across ethnic groups, based on the estimated age-adjusted incidence rates using the US Surveillance, Epidemiology, and End Results data. A substantial difference in incidence was observed in ethnic groups overall, and also between groups stratified by age and sex, with Caucasians at a >2-folds risk as compared to Asians / Pacific islanders (Fig. 1A–C). We next calculated the ancestry-specific prevalence of the alternative allele of APOBEC3-rs1014971 in contemporary populations, using 1000Genomes (TGP) data. Interestingly, the T allele, associated with both increased BCa risk and APOBEC mutagenesis burden in BCa tumors, presented at a much higher frequency in Caucasians (65%) as opposed to Asians (35% in South Asian and 28% in East Asian) (Fig. 1D) [6]. We then analyzed the ethnic distribution of the APOBEC mutagenesis burden, using the whole-exome-seq somatic mutation BCa cohort ($n = 602$) compiled in our previous work [4]. Coherently, a much higher APOBEC-mutagenesis burden was observed in Caucasians than in other ethnic groups for both non-muscle-invasive and muscle-invasive BCa tumors (Fig. 1E-F). We also considered whether such differences in global APOBEC-mutagenesis would further lead to differential mutation profiles in BCa-driver genes across ethnic groups. In our previous work, *FGFR3* S249C has been shown to be induced by APOBEC among all recurrent mutations in the BCa driver gene *FGFR3* [3, 18]. The current study notably found that the proportion of *FGFR3* S249C in BCa tumors bearing recurrent *FGFR3* mutations was significantly higher in Caucasian than in East Asian BCa patients (Chi-squared test, $P = 5.0 \times 10^{-4}$; Fig. 1G). This was further verified in additional targeted-sequencing samples and meta-analyses of WES and targeted-sequencing cohorts (targeted-sequencing, S249C prevalence in 257 Caucasians vs. 10 Asians with *FGFR3* recurrent mutations, OR = 1.99, $P = 0.32$; meta-analysis of WES and targeted-sequence samples, OR = 3.34, $P = 0.002$; Table S1 and Fig. S1). In contrast, RAS mutations (*HRAS* and *KRAS* genes) that were basically associated with non-APOBEC mutagenic processes were found in a much lower proportion in Caucasian than in Asian BCa patients (Chi-squared test, $P = 4.7 \times 10^{-4}$; Fig. 1H and Fig. S2). These findings suggest that the APOBEC3-related rs1014971 was not only involved in the different BCa risk profiles and APOBEC-mutagenesis burden, but also associated with the divergent mutational profiles in mechanistically pivotal BCa-driver genes across ethnic groups.

In view of the fact that the rs1014971 is a common variant with strong ethnic heterogeneity in contemporary human beings, and that APOBEC3 enzymes play crucial roles in immunity and pathogen restriction, we further explored whether the distribution of its alleles was potentially shaped by the history of human evolution. Non-neutral signatures of skews in the site-frequency spectrum were identified using the D_{FL} and F_{FL} metrics. These two metrics are population-genetic parameters commonly used to test the likelihood that a given genetic variant is under natural selection, and a negative value surpassing the threshold of statistical significance suggests influence by natural selection (Fig. 1I, upper for D_{FL} and lower for F_{FL}, respectively). Moreover, the joint analysis of contemporary (TGP data, $n = 2{,}504$) and ancient human genomes (Allen Ancient DNA Resource, AADR 1240K data; $n = 4{,}072$) showed a significantly higher frequency of the alternative allele [T] in contemporary populations than in ancient populations in the Americas, whereas a lower frequency was found in contemporary Asian populations than in their ancient counterparts (Americas, OR = 2.77, 95% CI = [2.20, 3.50], $P = 5.9 \times 10^{-18}$; South Asia, OR = 0.42, 95% CI = [0.30, 0.57], $P = 2.4 \times 10^{-8}$; East Asia, OR = 0.56, 95% CI = [0.46, 0.69], $P = 1.4 \times 10^{-11}$; Fig. 1J, upper part; Table S2).

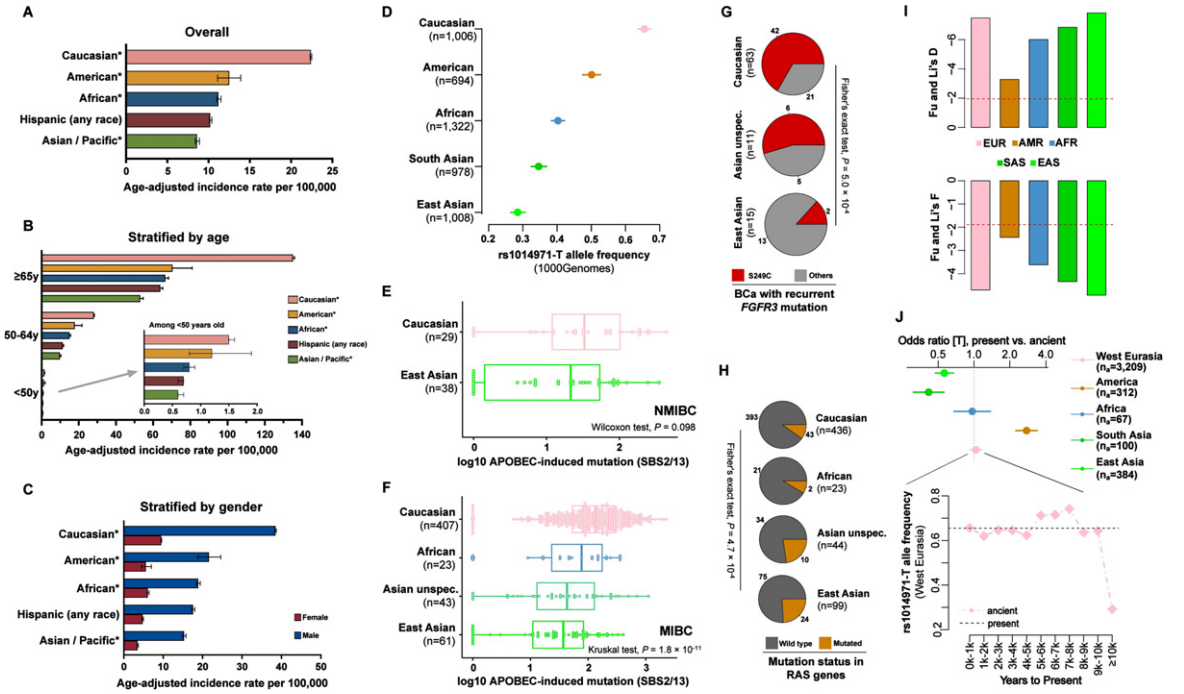

Fig. 1. Ethnic profiles of BCa susceptibility, BCa-related APOBEC3-variants rs1014971, APOBEC-induced mutation burden, *FGFR3* and RAS gene mutation status, and the natural selection and ancient genome dynamics of APOBEC3-rs1014971. (A–C) Ethnic differences in BCa-incidence rate and 95% confidence intervals revealed in US Surveillance, Epidemiology, and End Results (SEER) data. * indicates non-Hispanic ethnicity. (D) Ethnic difference in rs1014971-T allele frequency in 1000Genomes populations (Sample size, $n = 5{,}008$). (E-F) Ethnic difference in APOBEC-mutagenesis burden in whole-exome-seq BCa cohort ($n = 602$). NMIBC, non-muscle-invasive bladder cancer; MIBC, muscle-invasive bladder cancer. (G) Proportion of *FGFR3* S249C vs. other recurrent *FGFR3* mutations in BCa tumors bearing recurrent *FGFR3* mutations across ethnic groups. (H) Mutation status in RAS genes (*HRAS* and *KRAS*) in BCa tumors across ethnic groups. (I) Fu & Li's *D* and *F* statistics for test of neutrality in 1000Genomes populations, shown respectively in the upper and lower panel. A value <0 indicates potential natural selection. The dashed lines represent thresholds of statistical significance ($\alpha = 0.05$). EUR, European; AMR, American; AFR, African; SAS, South Asian; EAS, East Asian. (J) Joint analysis of current (TGP, $n = 2{,}504$) and ancient (AADR 1240K, $n = 4{,}092$) genomes regarding rs1014971 allele distribution. Upper part, odds ratios of rs1014971-T allele frequency between current vs. ancient samples. Lower part, chronological dynamics of rs1014971-T allele frequency in ancient human genomes of West Eurasian origin, by 1k-year unit. n_a, number of ancient genomes analyzed.

Although the rs1014971-T allele frequency was globally comparable in the contemporary and ancient genomes of populations of West Eurasian origin, apparent chronological dynamics were observed, which were characterized by a dramatically lower [T] frequency in samples from ≥10,000 years before present (YBP), as well as a higher frequency in samples between 5,000 and 8,000 YBP (Fig. 1J, lower part; Table S3-S4). Selection and adaptation in response to past pathogen invasions, potentially associated with changes in climate and environment, could be a possible explanation for these dynamics, which have contributed to the differentiated distribution of alleles in different modern ethnic groups [8]. We also evaluated another APOBEC3-variant, the rs12628403, which has important implications in cancer mutagenesis and is similar to rs1014971. It is a proxy for a 30-kb deletion between *APOBEC3A* and *APOBEC3B* in non-African populations, and, correlated with the APOBEC-mutagenesis pattern in pan-cancer, represents the strongest signal [19]. Interestingly, the alternative-allele frequency of this variant is also highly heterogeneous across ethnic groups in contemporary populations (TGP data, Fig. S3A). Although it was detected in a few ancient genomes ($n = 3$ out of 48 samples in the PGG.SNV database, mainly West Eurasia origin; all heterozygous, with the ~45ka Ust-Ishim individual being the most ancient), due to the lack of data in East Asian samples, its chronological dynamics are yet to be established (Fig. S3B). Additionally, both variants were found to be strong expression-quantitative trait loci (eQTL) of *APOBEC3A* and *APOBEC3B*, with the rs12628403 being the most significant for both genes, while the estimated age of the two variants was quite ancient (>~20,000 generations equal to roughly

>~500,000 YBP) (Fig. S3C-D, upper part). These characteristics of the two variants suggested the possible scenario that ancient functional mutations of APOBEC3 enzymes were differentially selected among ethnic groups during diverse local adaptation following varying pathogen exposures, which in turn shaped the contemporary ethnic differences in cancer susceptibility through the modulations of abnormal editing effects on the genomes of somatic cells (similar to the case of the *HFE* mutations which were involved in *Yersinia pestis*-driven selection in the past, and which modify iron-metabolism phenotypes in current populations) [9]. The ethnic distribution of the analyzed APOBEC3 variants could be linked with geographic variations in the prevalence of local viral species and the frequency of a certain infection. Nevertheless, although human APOBEC3 genes have been shown to undergo positive natural selection, the impact of population history / genetic drift may also have played a role in shaping the ethnic differentiation of the rs1014971 and APOBEC3A_3B-deletion variants associated with APOBEC-induced somatic mutagenesis and the risk of cancer [20]. Further spatiotemporally-resolved evolutionary analysis of the APOBEC-related genetic variants may lead to a better understanding of their functional roles in both physiological conditions and the carcinogenesis of multiple cancer types. *APOBEC3A* would be a target of special interest for further investigation, because it plays a major role in somatic deamination among all APOBEC3 enzymes and there is an unexplained significance of the over-representation of ancient mutations among its eQTLs (Fig. S3C, lower part) [21]. Our findings highlighted the significance of evolution-related genetic effects in ethnic differences in the risk of bladder cancer, a malignancy for which genetic heritability is estimated to account for 30% of the variation in susceptibility [22], however, other factors may also play a role in explaining ancestry-specific, age-adjusted incidence rates, such as the differences in smoking habits between different ethnic groups in the US.

CONCLUSIONS

In summary, the present study suggests a possible involvement of APOBEC3-related rs1014971 in the differing BCa incidence across ethnic groups, and such a difference could potentially be the result of human evolution. This adds to the collection of empirical evidence indicating that contemporary, population-level variations in disease susceptibility could be an inherited 'side-effect' of pathogen-driven selection in the past. Integrated analysis of cancer genomics and evolutionary-population genetics in large samples might prove to be a powerful, novel approach to further decoding the underlying mechanisms of human carcinogenesis.

ACKNOWLEDGMENTS

We thank the authors of the tools and publicly available data that supported the analysis of this study.

FUNDING

HYZ was supported by the Fundamental Research Funds for the Central Universities (2662021JC008). MJS was supported by the National Natural Science Foundation of China (82002672) and by the Beijing Hospitals' Authority Youth Program (QML20230114). The sponsors played no direct role in the study.

AUTHOR CONTRIBUTIONS

Study concept and design: XYM, HYZ.

Acquisition of data: XYM, QLW, MJS.

Analysis and interpretation of data: XYM, QLW, MJS.

Drafting of the manuscript: XYM, QLW, MJS.

Funding acquisition: HYZ.

Critical revision of the manuscript for important intellectual content: HYZ.

Statistical analysis: XYM, QLW.

Administrative, technical, or material support: HYZ.

Supervision: HYZ.

CONFLICT OF INTEREST

XYM, QLW, MJS and HYZ report no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in the US Surveillance, Epidemiology, and End Results database at https://seer.cancer.gov/data-software/, in the NCBI dbSNP database at https://ncbi.nlm.nih.gov/snp/, in the Allen Ancient DNA Resource database at

https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/ (V54.1), in the PGG.SNV database at https://www.pggsnv.org/, in the VannoPortal database at http://www.mulinlab.org/vportal, in the Human Genome Dating database at https://human.genome.dating/, and in the Genotype-Tissue Expression database at https://commonfund.nih.gov/GTEx. Other data that support the findings of this study are available from the corresponding author upon reasonable request.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/BLC-230010.

REFERENCES

[1] Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022;72(1):7-33.
[2] Anton-Culver H, Lee-Feldstein A, Taylor TH. The association of bladder cancer risk with ethnicity, gender, and smoking. Ann Epidemiol. 1993;3(4):429-33.
[3] Shi MJ, Meng XY, Lamy P, Banday AR, Yang J, Moreno-Vega A, et al. APOBEC-mediated mutagenesis as a likely cause of FGFR3 S249C mutation over-representation in bladder cancer. Eur Urol. 2019;76(1):9-13.
[4] Shi MJ, Meng XY, Fontugne J, Chen CL, Radvanyi F, Bernard-Pierrot I. Identification of new driver and passenger mutations within APOBEC-induced hotspot mutations in bladder cancer. Genome Med. 2020;12(1):85.
[5] Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, et al. Comprehensive molecular characterization of muscle-invasive bladder cancer. Cell. 2017;171(3):540-56 e25.
[6] Middlebrooks CD, Banday AR, Matsuda K, Udquim KI, Onabajo OO, Paquin A, et al. Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors. Nat Genet. 2016;48(11):1330-8.
[7] Ito J, Gifford RJ, Sato K. Retroviruses drive the rapid evolution of mammalian APOBEC3 genes. Proc Natl Acad Sci U S A. 2020;117(1):610-8.
[8] Klunk J, Vilgalys TP, Demeure CE, Cheng X, Shiratori M, Madej J, et al. Evolution of immune genes is associated with the Black Death. Nature. 2022;611(7935):312-9.
[9] Ye K, Cao C, Lin X, O'Brien KO, Gu Z. Natural selection on HFE in Asian populations contributes to enhanced non-heme iron absorption. BMC Genet. 2015;16:61.
[10] Dominguez-Andres J, Netea MG. Impact of historic migrations and evolutionary processes on human immunity. Trends Immunol. 2019;40(12):1105-19.
[11] Anderson RN, Rosenberg HM. Age standardization of death rates: Implementation of the year 2000 standard. Natl Vital Stat Rep. 1998;47(3):1-16, 20.
[12] Letouze E, Shinde J, Renault V, Couchy G, Blanc JF, Tubacher E, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. Nat Commun. 2017;8(1):1315.
[13] Fu YX, Li WH. Statistical tests of neutrality of mutations. Genetics. 1993;133(3):693-709.
[14] Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.
[15] Zhang C, Gao Y, Ning Z, Lu Y, Zhang X, Liu J, et al. PGG.SNV: Understanding the evolutionary and medical implications of human single nucleotide variations in diverse populations. Genome Biol. 2019;20(1):215.
[16] Albers PK, McVean G. Dating genomic variants and shared ancestry in population-scale sequencing data. PLoS Biol. 2020;18(1):e3000586.
[17] Consortium GT. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369(6509):1318-30.
[18] Shi MJ, Meng XY, Chen CL, Dyrskjot L, Radvanyi F, Prokunina-Olsson L, et al. Reply to Alexander Yang, Vincent L. Cannataro, Jeffrey P. Townsend's Letter to the Editor, re: Ming-Jun Shi, Xiang-Yu Meng, Philippe Lamy, et al. APOBEC-mediated Mutagenesis as, a Likely Cause of FGFR3 S249C Mutation Over-representation in Bladder Cancer. Eur Urol. 2019;76:9-13. Eur Urol. 2020;77(1):e26-e7.
[19] Consortium ITP-CAoWG. Pan-cancer analysis of whole genomes. Nature. 2020;578(7793):82-93.
[20] Wang Y, Schmitt K, Guo K, Santiago ML, Stephens EB. Role of the single deaminase domain APOBEC3A in virus restriction, retrotransposition, DNA damage and cancer. J Gen Virol. 2016;97(1):1-17.
[21] Petljak M, Dananberg A, Chu K, Bergstrom EN, Striepen J, von Morgen P, et al. Mechanisms of APOBEC3 mutagenesis in human cancer cells. Nature. 2022;607(7920):799-807.
[22] Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000;343(2):78-85.