Multiple roles of apolipoprotein B mRNA editing enzyme catalytic subunit 3B (APOBEC3B) in human tumors: a pan-cancer analysis

Jiacheng Wu[1,2], Ni Li[2], Linwen Zhu[2], Dawei Zhen[2], Mengqi Li[3], Hang Chen[1,2], Mengmeng Ye[1,2], Yiqin Wei[1,2] and Guofeng Shao[2*]

*Correspondence:
sgf1958@sina.com

[1] Medical College, Ningbo University, Ningbo, Zhejiang, People's Republic of China
[2] Department of Cardiothoracic Surgery, Lihuili Hospital Affiliated to Ningbo University, Ningbo 315000, Zhejiang, People's Republic of China
[3] Department of Geriatrics, Shaoxing People's Hospital (Shaoxing Hospital, Zhejiang University School of Medicine), Shaoxing, Zhejiang, People's Republic of China

Abstract

Although there have been some recent cell and animal experiments indicating that expression of the gene encoding apolipoprotein B mRNA editing enzyme catalytic subunit 3B (*APOBEC3B*) is closely related to cancer, it still lacks pan-cancer analysis. Here we analyzed the potential carcinogenic role of *APOBEC3B* in 33 tumors based on The Cancer Genome Atlas (TCGA). *APOBEC3B* was highly expressed in most tumors and weakly expressed in a few. Differences in expression level were significantly correlated with the pathological tumor stage and prognosis of affected patients. The high-frequency *APOBEC3B* changes were principally mutations and amplifications in some tumors, such as uterine corpus endometrial carcinomas or cutaneous melanomas. In testicular germ cell tumors and invasive breast carcinomas, *APOBEC3B* expression and CD8[+] T lymphocyte counts were correlated. In other cancers, such as human papilloma virus (HPV)-related head and neck squamous cell carcinomas or esophageal adenocarcinomas, there was also cancer-associated fibroblast infiltration. The *APOBEC3B* enzyme acts in the mitochondrial respiratory electron transport chain and in oxidative phosphorylation. This first pan-cancer study provides a comprehensive understanding of the multiple roles of *APOBEC3B* in different tumor types.

Keywords: *APOBEC3B*, Oncogenic role, Pan-cancer analysis

Introduction

Over the past decades, the incidence of cancer has increased rapidly. Cancers are among the leading causes of death worldwide, so the therapeutic success rate needs to be further improved [1, 2]. In this regard, the mechanisms of tumorigenesis and progression are complicated and deserve further investigation. Pan-cancer analysis has emerged as a widely used method to investigate the common features and heterogeneities of various tumor types. It allows for a thorough and profound understanding of human cancers [3]. The publicly funded The Cancer Genome Atlas (TCGA) project contain functional genomics datasets for different tumors, [4, 5] which enable pan-cancer analysis.

Wu *et al. BMC Bioinformatics*   (2022) 23:312

Page 2 of 16

The apolipoprotein B mRNA editing enzyme catalytic subunit 3B (*APOBEC3B*) protein, also known as A3B or ARP4, is a member of the cytidine deaminase gene family [6]. It is one of seven related genes or pseudogenes found in a cluster on human chromosome 22, thought to result from gene duplication. Members of the cluster encode proteins that are structurally and functionally related to the C-to-U RNA-editing cytidine deaminase APOBEC [7–9]. *APOBEC3B* is involved in the development and progression of breast cancers [10], hepatocellular carcinomas [11], ovarian cancers, [12] nasopharyngeal carcinomas, [13] and chondrosarcomas [14]. Comprehensive analysis of the expression patterns, prognostic significance and relationship with tumor microenvironment of *APOBEC3B* at pan-cancer level will help us deepen our understanding of this enzyme.

Here, we conducted a pan-cancer analysis of *APOBEC3B* using the TCGA project. We also analyzed the pathogenesis or potential molecular mechanisms of *APOBEC3B* in different cancers by gene expression, patient survival status, protein expression, gene mutation, immune infiltration, protein interactions and cellular pathways. These findings could strengthen our understanding of the biological functions of *APOBEC3B* in different cancer types.

Materials and methods

Gene expression analysis

We entered *APOBEC3B* into the 'GENE_DE' module of 'Exploration' on the Tumor Immune Estimation Resource (TIMER2, v. 2) web page (http://timer.cistrome.org/) [15–17] and observed the differential expression of *AOPBEC3B* in different tumors and adjacent normal tissues in the TCGA project. For certain tumors that lacked normal adjacent tissue samples (e.g., TCGA-ACC for adrenocortical carcinomas; TCGA-OV for ovarian serous cystadenocarcinomas), we used the UCSC XENA database (https://xenabrowser.net/datapages/) [18] downloaded the TCGA and the Genotype-Tissue Expression (GTEx) of transcripts per million reads (TPM) RNAseq data in the format processed uniformly by the Toil process [19]. Then we performed $\log_2$ transformation of these data, compared the expression between samples of tumor tissues and corresponding normal tissues, and have presented the results in box plots. Gene expression differences were visualized with the 'ggplot2' R package. The R language software (R-3.6.3, 64 bit; http://www.r-project.org/) was used in this analysis. In addition, we used the 'Expression DIY-Pathological Stage Plot' of the Gene Expression Profiling Interactive Analysis, web services (GEPIA2 v. 2; http://gepia2.cancer-pku.cn/analysis) [20] to obtain violin plots of *APOBEC3B* expression levels in different pathological stages (stages I through IV) of all TCGA tumors. All the above box plots or violin plots were based on $\log_2(\text{TPM}+1)$ transformed expression data.

We performed protein expression analysis on the Clinical Proteomic Tumor Analysis Consortium (CPTAC) dataset using the UALCAN web resource (http://ualcan.path.uab.edu/index.html), [21] a comprehensive, interactive web for analyzing cancer omics data. We entered '*APOBEC3B*' to analyze differences in *APOBEC3B* total protein levels between tumors and adjacent normal tissues. The available data sets for eight tumor types were selected: breast cancers, ovarian cancers, renal clear cell carcinomas (RCC), uterine corpus endometrial carcinomas (UCEC), lung adenocarcinomas (LUAD), head

Wu *et al. BMC Bioinformatics*   (2022) 23:312

Page 3 of 16

and neck squamous carcinomas (HNSC), pancreatic adenocarcinomas (PAAD), and liver hepatocellular carcinomas (LIHC).

Survival analysis

We used the 'Survival Analysis-Survival Map' module of GEPIA2 to obtain the survival heat map of overall survival (OS) and disease-free survival (DFS) of *APOBEC3B* for patients with all TCGA tumors, estimated using the Mantel–Cox test. We used months as the survival time unit, high cutoff (50%) and low cutoff (50%) values as the segmentation criteria of high expression and low expression cohorts and $P < 0.05$ as the measure of significance level. Subsequently, we analyzed tumors with significant differences in patient survival in heat maps through the 'survival analysis' module of GEPIA2 and obtained the corresponding survival plots.

Gene alteration analysis

We selected 'TCGA pancancer atlas studies' in the 'quick select' option on the home page of the cBioPortal website (https://www.cbioportal.org/), [22] and then entered '*APOBEC3B*' to query its gene alterations. In the 'Cancer Type Summary' module, the alteration frequency, mutation type and copy number alteration (CNA) results of all TCGA tumors were observed. In the 'Mutations' module, the mutation site information of *APOBEC3B* is displayed in the protein structure track or three-dimensional (3D) structure diagrams. After that, we analyzed the overall, disease-specific, disease-free and progression-free data of TCGA cancer cases with *APOBEC3B* gene alterations through the 'Comparison/Survival' module, and obtained the corresponding Kaplan–Meier plots with log-rank test P-values.

Immune infiltration analysis

We used the 'Immune gene' module of the TIMER2 website to evaluate the correlation between *APOBEC3B* and immune infiltration expression in all TCGA tumors, entered '*APOBEC3B*' in gene expression and selected 'T-cells CD8+' and 'cancer-associated fibroblasts' in the immune information menu. Then, the immune infiltration data were analyzed by TIMER, EPIC, MCPCOUNTER, XELL, TIDE, CIBERSORT, CIBERSORT-ABS algorithms, and then the P-values and partial correlation values were obtained using a purity-adjusted Spearman's rank correlation test. Finally, heatmaps and scatter diagrams were obtained after the data were visualized.

APOBEC3B-related gene enrichment analysis

After logging into the STRING (https://cn.string-db.org/) website, [23] we first entered the protein name '*APOBEC3B*' and selected the organism as '*Homo sapiens*', Then we set the parameters as follows: meaning of network edges ('Confidence'), active interaction sources ('Textmining', 'Experiments', 'Databases', 'Co-expression', 'Gene Fusion', and 'Co-occurrence'), minimum required interaction score ('low confidence 0.150') and maximum number of interactors to show 'no more than 50 interactors'. Ultimately, we obtained the top 50 *APOBEC3B*-related evidence-based proteins and corresponding network diagrams.

Wu *et al. BMC Bioinformatics*   (2022) 23:312

Page 4 of 16

To search for genes with an expression pattern similar to *APOBEC3B* in different cancer types, we entered the gene '*APOBEC3B*' and 'top 100' in the 'Expression Analysis-Similar Gene Detection' module of the GEPIA2 website, and selected all TCGA tumor data sets for analysis. Subsequently, we selected the top five genes from the list, entered the gene'*APOBEC3B*' and 'top5 genes' in the 'expression analysis correlation analysis' module of GEPIA2, selected the Pearson correlation coefficient, and performed a visual correlation analysis on the entire TCGA tumor dataset. The dot plots obtained were scaled by $\log_2$ TPM and contained the P-values and the correlation coefficients (R values). In addition, we used the 'Exploration Gene_Corr' module of TIMER2 to obtain the heatmap data related to the top five genes and *APOBEC3B* in various cancer types, which gave us the purity-adjusted partial Spearman's rho value as the degree of their correlation.

We used the 'Draw Venn Diagram' website (http://bioinformatics.psb.ugent.be/webtools/Venn/) to cross-analyze the previously acquired 'Top50 gene' of STRING and 'Top100 gene' data from GEPIA2. Subsequently, we performed Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis for the above genes. We entered the gene list on the David (https://david.ncifcrf.gov/) [24] website and selected 'OFFICAL_GENE_SYMBOL' and '*Homo sapiens*' to finally obtain functional annotation data. In addition, we used the 'clusterprofiler' (v. 3.14.3) for enrichment analysis and the 'org. Hs.eg.db' (v. 3.10.0 for ID transformation) R packages to perform gene ontology (GO) enrichment analysis. The resulting KEGG and GO analysis including biological process (BP), cellular component (CC), and molecular function (MF) data described above were visualized as bubble plots and network diagrams via the 'ggplot2' and 'clusterprofiler' R packages (v. 3.3.3). All R languages in this article are based on the R software (R-3.6.3, 64bit; http://www.r-project.org/).

Results
Expression analysis of APOBEC3B
In this study, we principally explored the potential role that *APOBEC3B* (NM_004900.4 for the mRNA or NP_004891.4 for the protein; Fig. S1a) might play during cancer initiation and progression. The gene encoding *APOBEC3B* is conserved in the chimpanzee, dog, cow, mouse, and rat, and 20 organisms have orthologs with human *APOBEC3B*. [25] We investigated the expression profile of the *APOBEC3B* gene in cells from different normal tissues, as shown in Additional file 2: Figure S1b, the mRNA expression of *APOBEC3B* showed the highest level in bone marrow, followed by the appendix and colon, based on the Human Protein Atlas dataset.

We analyzed *APOBEC3B* expression levels in different cancer types from TCGA using the TIMER2 website. As shown in Fig. 1a, the expression level of *APOBEC3B* was higher in the tumor tissues of bladder urothelial carcinomas (BLCA), breast invasive carcinomas (BRCA), cholangiocarcinomas (CHOL), esophageal carcinomas (ESCA), glioblastoma multiforme (GBM), HNSC, kidney renal clear cell carcinomas (KIRC), kidney renal papillary cell carcinomas (KIRP), LIHC, LUAD, lung squamous cell carcinomas (LUSC), stomach adenocarcinomas (STAD), UCEC ($P < 0.001$), cervical squamous cell carcinomas and endocervical adenocarcinomas (CESC), kidney chromophobes (KICH), prostate adenocarcinomas (PRAD) ($P < 0.01$) than in the corresponding normal tissues.
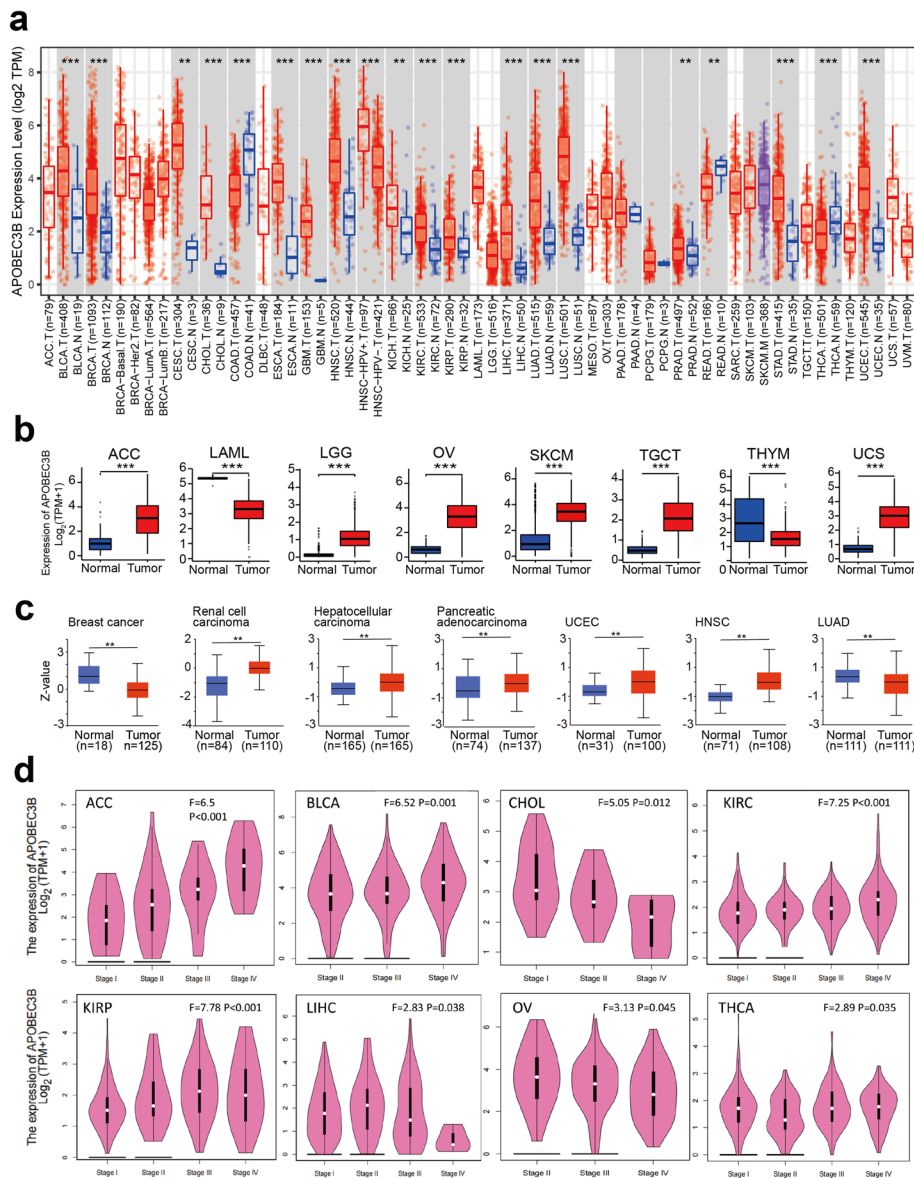
Wu *et al. BMC Bioinformatics*   (2022) 23:312

Page 5 of 16

Fig. 1 Expression level of the *APOBEC3B* gene in different tumors and pathological stages. **a** The mRNA expression of the *APOBEC3B* in different tumor subtypes and adjacent normal tissues. **P < 0.01; ***P < 0.001. **b** The mRNA expression of the *APOBEC3B* in ACC, LAML, LGG, OV, SKCM, TGCT, THYM, and UCS tumor types in the TCGA dataset, corresponding normal tissues of the GTEx database were included as controls. ***P < 0.001. **c** The protein levels of *APOBEC3B* between normal tissues and primary tissues of breast cancers, clear cell RCC, hepatocellular carcinomas, PAAD, UCEC, HNSC, and LUAD, based on the CPTAC dataset. **P < 0.01. **d** The relationship between *APOBEC3B* expression and clinical stages in ACC, BLCA, CHOL, KIRC, KIRP, LIHC, OV, and THCA tumor types

However, it was lower in the tumor tissues of colon adenocarcinomas (COAD), thyroid carcinomas (THCA) ($P < 0.001$) and PRAD ($P < 0.01$).

For those TCGA tumors with missing normal tissue data, we acquired the normal tissue data of the GTEX dataset and used them as a control group to further analyze the expression differences of *APOBEC3B* in normal tissues and tumor tissues of Adrenocortical carcinoma (ACC), acute myeloid leukemia (LAML), OV, testicular germ cell tumors

Wu *et al. BMC Bioinformatics*   (2022) 23:312

Page 6 of 16

(TGCT), uterine carcinosarcomas (UCS), brain lower grade gliomas (LGG), skin cutaneous melanomas (SKCM) and thymomas (THYM) (Fig. 1b, $P < 0.001$). Among them, LAML and THYM were weakly expressed in tumor tissues. Nevertheless, we did not find significant differences in the expression of *APOBEC3B* in lymphoid neoplasm diffuse large B-cell lymphomas (DBLC), mesotheliomas (MESO), pheochromocytomas and paragangliomas (PCPG), or uveal melanomas (UVM) tissues.

To further explore the proteomics data of *APOBEC3B* in patients with cancers, we conducted an analysis using the CPTAC dataset. The *APOBEC3B* protein expression level was higher in renal clear cell carcinomas, UCEC, HNSC, PAAD and LIHC and lower in breast cancers and LUAD (Fig. 1c, all $P < 0.01$) than in normal tissues. Interestingly, the proteomics data suggested lower expression of the *APOBEC3B* protein in breast cancer and LUAD, opposite of that observed at the mRNA level. These inconsistent results might be caused by post-transcriptional or translational regulation.

In addition, we used the 'Pathological Stage Plot' module of GEPIA2 to analyze the relationship between the expression of *APOBEC3B* and the pathological stages of cancers. As shown in Fig. 1d, we obtained the pathological stage plots of ACC, BLCA, CHOL, KIRC, KIRP, LIHV, OV and THCA (all $P < 0.05$). To be specific, in ACC, BLCA, KIRC, and KIRP tissues, *APOBEC3B* RNA expression levels were consistently associated with later clinical stages; in CHOL, LIHC, and OV mRNA expression level of *APOBEC3B* was significantly higher in the early stage of these diseases.

Thus, *APOBEC3B* has different expression levels in different tumor tissues; the expression levels were closely related to the pathological state of the tumor, but the correlation was not consistent across different cancer species. These findings indicated a diverse biological function of this enzyme in different tumor types.

Survival analysis data

Taking the median expression level of *APOBEC3B* in tumor cases as the standard, we divided the tumor cases in TCGA datasets into high- and low-expression groups, and explored the relationship between the expression of *APOBEC3B* and the prognoses for different patients. As shown in Fig. 2a, the tumors associated with high expression of *APOBEC3B* and poor OS were ACC ($P < 0.001$), LGG ($P < 0.001$), LIHC ($P = 0.03$) and UVM ($P = 0.032$) within the TCGA project In the DFS analysis (Fig. 2b). A high expression level of *APOBEC3B* was associated with poor prognosis for patients with ACC ($P < 0.001$), KIRP ($P = 0.015$), LGG ($P = 0.0016$), LIHC ($P = 0.037$), and THCA ($P = 0.0068$). Furthermore, low expression of *APOBEC3B* was related to poor OS and DFS prognoses for patients with CESC (Fig. 2a, $P = 0.0012$; Fig. 2b, $P = 0.029$). These inconsistent results in different cancer species suggested that *APOBEC3B* might play different roles in different tumor types.

We also analyzed the clinical prognostic significance of DNA methylation of *APOBEC3B* in the website tool (https://biit.cs.ut.ee/methsurv/). We have listed the results in Additional file 1: Table S1. Methylation at cg06837067 had the most prognostic significance, predicting a poor overall survival in ACC (HR: 2.179, $P = 0.04$), BLCA (HR: 1.924, $P < 0.01$), CESC (HR: 2.034, $P < 0.01$), HNSC (HR: 1.556, $P = 0.01$), KIRC (HR: 2.353, $P < 0.01$), KIRP (HR: 2.258, $P = 0.01$), SKCM (HR: 1.735, $P < 0.01$), and STAD (HR: 1.526, $P < 0.01$), however, in LGG (HR: 0.343, $P < 0.01$), MESO (HR: 0.564, $P = 0.045$), and
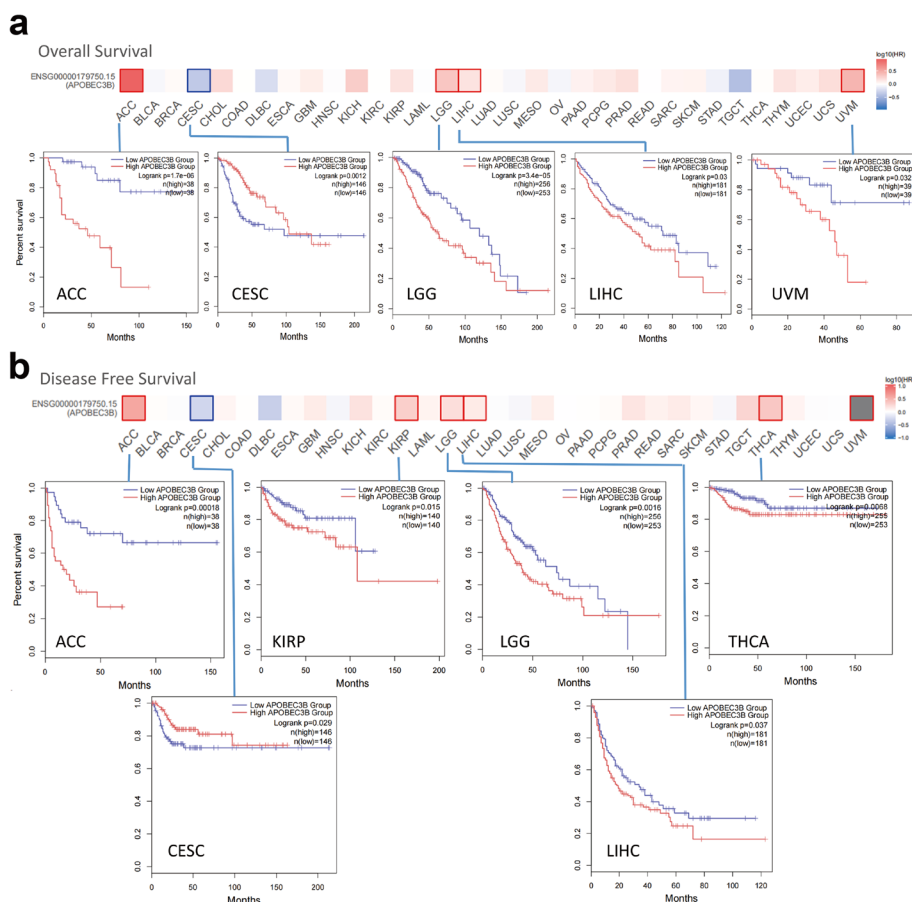
Wu *et al. BMC Bioinformatics* (2022) 23:312

Page 7 of 16

Fig. 2 Correlation between *APOBEC3B* gene expression and survival prognosis for patients with cancers in TCGA. We performed overall survival (OS) (**a**) and disease-free survival (DFS) (**b**) analyses of different tumors in TCGA according to *APOBEC3B* gene expression. The survival maps and Kaplan–Meier curves with positive results were provided

UVM (HR: 0.229, *P* = 0.046), the methylation level at cg06837067 predicted greater overall survival (Additional file 1: Table S1). This further suggested that *APOBEC3B* might play different roles in different cancer types.

Gene alteration analysis

We analyzed the gene alteration status of *APOBEC3B* in all tumor samples in the TCGA dataset through cBioPortal. As shown in Fig. 3a, the tumor type with the highest alteration frequency of *APOBEC3B* was UCEC (> 4%), mainly of the 'mutation' type. The second most frequent alteration was in SKCM (~ 3.7% frequency). In cases with ovarian serous cystadenocarcinomas, the main type was 'amplification', and the alteration frequency was ~ 2%. It is worth mentioning that all types of genetic alteration (~ 2% frequency) in cases of thymomas were of the 'deep deletion' type. In addition, the types, loci and corresponding cases of *APOBEC3B* gene alteration are shown in Fig. 3c. Missense mutations were found to be the main type of *APOBEC3B* genetic alterations: an R114H/S alteration in the APOBEC-like N-terminal domain, which was detected in one case of GMB, one case of PRAD and one case of STAD (Fig. 3b), and a P355S alteration
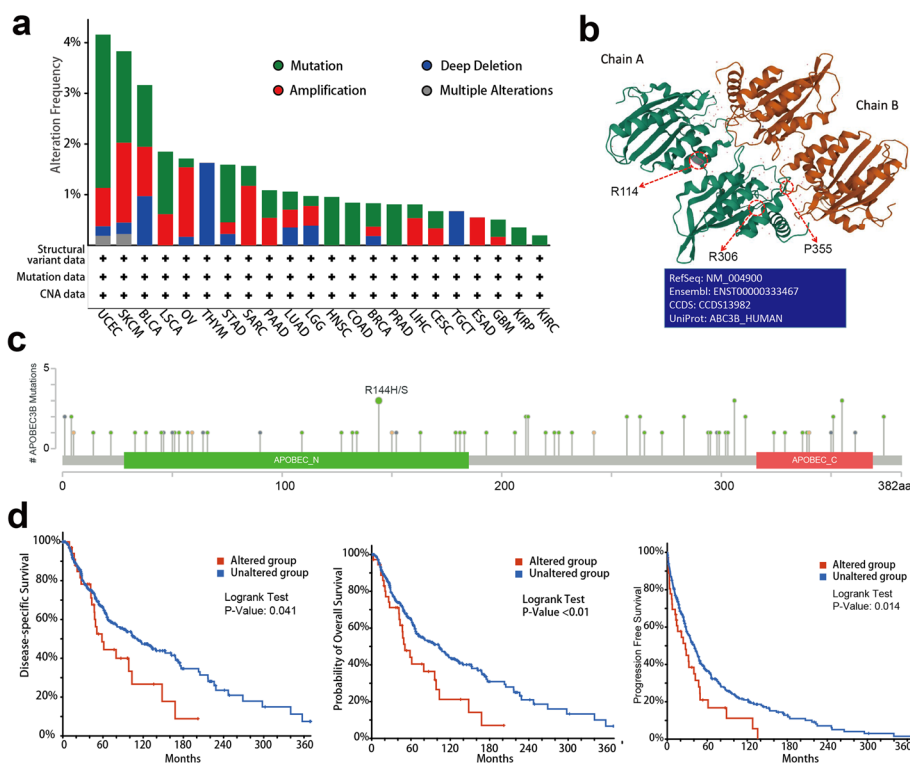
Fig. 3 Mutation feature of *APOBEC3B* in different tumors listed in TCGA. **a, b** The mutation features of *APOBEC3B* for the TCGA-listed tumors using the cBioPortal tool. The alteration frequencies with mutation type (**a**) and mutation site (**b**) were displayed. **c** The highest alteration frequency of mutation sites (R114, R306, R355) of *APOBEC3B* were shown in the 3D structure. **d** The potential correlation between mutation status and disease-specific, overall and progression-free survival in patients with SKCM

in APOBEC-like C-terminal domain, which was detected in one case of LUSD and two cases of SKCM. These alterations are able to induce a frame shift mutation of the *APOBEC3B* gene, from R (Arginine) to H (Histidine) or S (Serine) at site 114 and P (Proline) to S at site 355 of the *APOBEC3B* protein, and subsequent *APOBEC3B* protein truncation. We also noted the sites of R144, R306 and P335 in the 3D structure of *APOBEC3B* protein (Fig. 3c). In addition, we also analyzed the association between *APOBEC3B* gene alteration and clinical survival and prognosis with different types of cancers. As shown in Fig. 3d, the *APOBEC3B* altered group in patients with SKCM showed better prognosis in DSS ($P = 0.0411$), OS ($P < 0.01$) and progression-free ($P = 0.0137$) survival, but in UCEC cases overall, disease-specific, disease-free and progression-free patient survival rates were not statistically significant.

Immune infiltration analysis data

Tumor-infiltrating immune cells play an important role in the occurrence, progression and metastasis of tumors [26]. We used the TIMER, EPIC, MCPCOUNTER, XELL, TIDE, CIBERSORT, and CIBERSORT-ABS algorithms to explore correlations between the infiltration level of various immune cells and the expression level of *APOBEC3B* gene in different TCGA tumors. Cancer-associated fibroblasts are the main component of the stroma and secrete growth factors, inflammatory ligands and

Wu *et al. BMC Bioinformatics*   (2022) 23:312

Page 9 of 16

extracellular matrix proteins to promote tumor proliferation, therapeutic resistance and immune rejection. After a series of analyses, we observed that the infiltration value of cancer-associated fibroblasts in ESCA, GBM, HNSC-HPV⁻, LGG, PCPG, rectum adenocarcinoma (READ), TGCT, and THCA type tumors were positively correlated with the expression level of *APOBEC3B*, but negatively correlated with BRCA and HNSC-HPV⁺ tumor types (Fig. 4a). Moreover, we observed a statistically significant positive correlation between the immune infiltration of CD8⁺ T-cells and the expression level of *APOBEC3B* in HNSC0-HPV⁺ and UVM tumors based on most algorithms, while the ESCA tumor type was negatively correlated (Fig. 4b). Finally, as shown in Fig. 5c, we selected an algorithm to obtain a scatterplot of this tumor. For example, based on XCELL algorithm, the expression level of *APOBEC3B* in BRCA was negatively correlated with the infiltration level of cancer-associated fibroblasts (Fig. 5c, $R = -0.316$, $P = 1.85^{e-24}$).
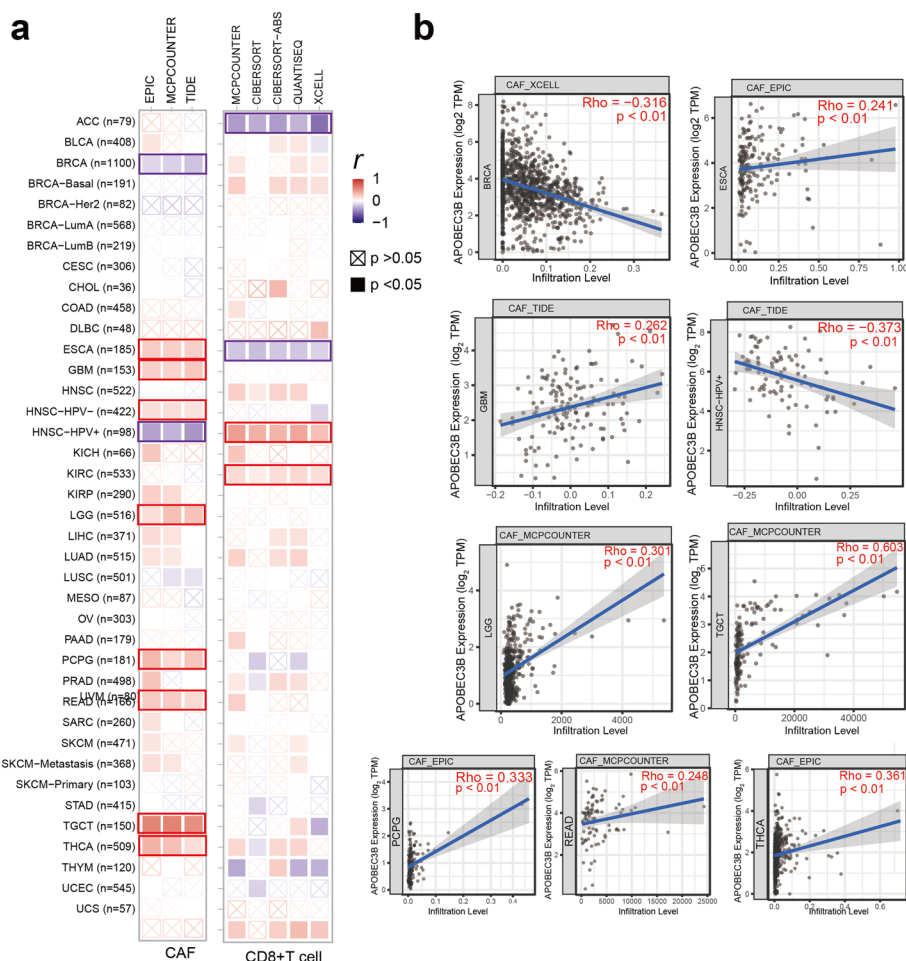

Fig. 4 Correlation analysis between *APOBEC3B* expression and immune infiltration of cancer-associated fibroblasts and CD8⁺ T-cells. **a** Heatmap showing the correlation between the expression levels of the *APOBEC3B* and the infiltration level of cancer-associated fibroblasts/CD8⁺ T-cells obtained by different algorithms. **b** The correlation between *APOBEC3B* expression and immune infiltration cancer-associated fibroblasts in some TCGA-listed tumors
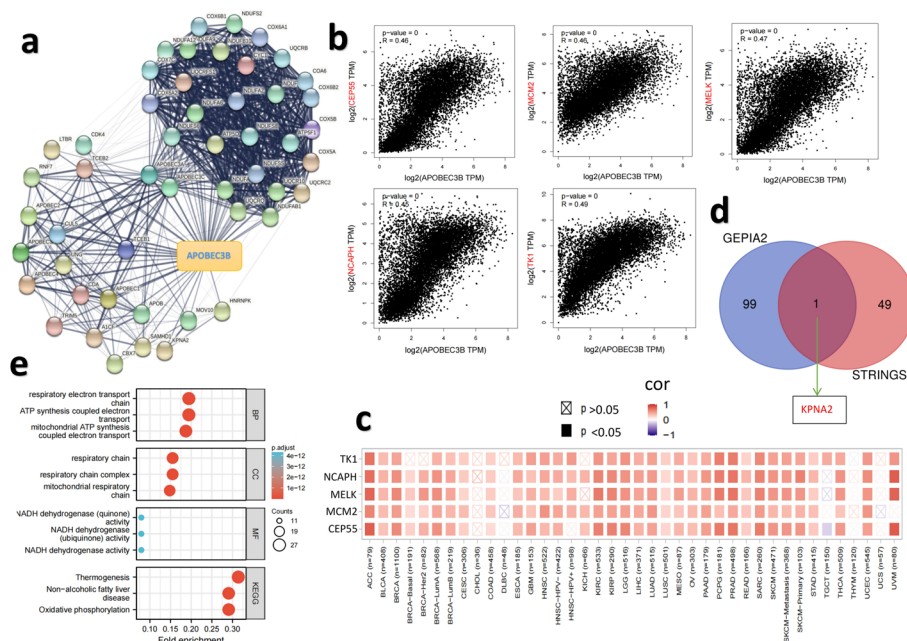
Fig. 5 *APOBEC3B*-related gene enrichment analysis. **a** The *APOBEC3B*-binding proteins identified using the STRING tool. **b** The correlation between the expression of *APOBEC3B* and top 5 genes co-expression with *APOBEC3B* (TK1, MELK, CEP55, MCM2, and NCAPH). **c** Heatmap showing the correlation between the expression of *APOBEC3B* and top 5 genes co-expression with *APOBEC3B* (TK1, MELK, CEP55, MCM2, and NCAPH) in the detailed cancer types. **d** Venn diagram showing the intersection analysis of the *APOBEC3B*-binding and correlated genes. **e** Bubble chart of KEGG pathway analysis and GO enrichment analysis based on the *APOBEC3B*-binding and interacting genes

APOBEC3B-related gene enrichment analysis data

To further study how expression of the *APOBEC3B* gene affects tumorigenesis, we screened *APOBEC3B*-binding proteins and *APOBEC3B* expression-related genes, and carried out a series of pathway and function enrichment analyses. Through the STRING tool, we screened the top 50 *APOBEC3B* binding proteins based on text mining, experiments, databases, co-expression, Gene Fusion, and co-occurrence. Figure 5a shows the interaction network of these proteins. We then obtained the top 100 genes related to *APOBEC3B* expression based on all tumor expression data of TCGA using the GEPIA2 tool. As shown in Fig. 5b, the expression level of *APOBEC3B* was positively correlated with those for the genes encoding thymidine kinase 1 (TK1; $R = 0.49$), maternal embryonic leucine zipper kinase (MELK; $R = 0.47$), centrosomal protein 55 (CEP55; $R = 0.46$), minichromosome maintenance complex component 2 (MCM2; $R = 0.46$) and non-SMC condensin I complex subunit H (NCAPH; $R = 0.45$) (all $P < 0.001$). The corresponding heatmap data also showed that the expression level of *APOBEC3B* was positively correlated with the above five genes in most tumor types (Fig. 5c). The top 50 *APOBEC3B*-binding proteins and the top 100 *APOBEC3B* expression-related-genes were cross analyzed to obtain a common gene, namely that encoding karyopherin subunit alpha 2 (KPNA2; Fig. 5d).

To further clarify the function of *APOBEC3B* in tumors, we combined the top 50 *APOBEC3B*-binding proteins and the top 100 co-expression genes above for KEGG (www.kegg.jpkegg/kegg1.html) and GO enrichment analysis. KEGG data show that the

terms 'thermogenesis', 'non-alcoholic fatty liver disease' and 'oxidative phosphorylation' might be involved in the mechanism of *APOBEC3B* affecting tumorigenesis. The GO enrichment analysis data showed that most of these genes are involved in or regulate the mitochondrial oxidative respiratory chain and its related pathways, such as the respiratory electron transport chain, ATP synthesis-coupled electron transport, mitochondrial ATP synthesis-coupled electron transport, and the mitochondrial respiratory chain among others. These data of KEGG and GO enrichment analysis have been visualized as a bubble plot (Fig. 5e).

Discussion

Gene mutations are closely related to the occurrence of tumors, and the sources that mediate mutation are divided into exogenous and endogenous types [27, 28]. Endogenous mutation sources involve enzymes acting in the DNA repair system, and there are reports that the mutation mode of APOBEC cytidine deaminases might play a role in carcinogenic somatic mutations and eventually lead to genomic instability [29, 30]. The human APOBEC protein family has seven members and are important members of the innate immune system [31]. Thus, *APOBEC3B* had the highest expression in bone marrow among normal tissues in our study. It not only participates in innate immunity, but also participates in a variety of biological processes in different cells. In particular, the *APOBEC3B* protein with cytosine deaminase activity can lead to base substitutions in tumor-causing genes [32, 33]. Many publications have reported that *APOBEC3B* is closely related to the occurrence and development of a variety of tumors [10, 13, 14, 34]. However, there is still a lack of in-depth research on whether *APOBEC3B* can play a role in different tumors through molecular mechanisms. We searched the literature of major databases and failed to retrieve any pan-cancer analysis publications on *APOBEC3B* from the perspective of tumors as a whole. Therefore, based on the TCGA and CPTAC databases, we comprehensively analyzed the gene expression, gene changes and immune infiltration of the *APOBEC3B* gene in 33 different tumors.

APOBEC3B proved to be highly expressed in most tumor tissues, but weakly in a few types, such as COAD, LAML, READ, THCA, and THYM. To minimize errors caused by individual differences in tumor data, we downloaded the RNAseq pan-cancer data in level 3 HTseq-FPKM format from the TCGA project, selected paired sample data, and then analyzed the differential expression of *APOBEC3B* through the 'ggplot2' R package. The results obtained are basically consistent with Fig. 1a (see also Additional file 2: Figure. S1c). We also found that in some tumors, the expression of *APOBEC3B* was mainly positively correlated with the tumor's pathological stage, but in CHOL and OV tumors, the expression of *APOBEC3B* was negatively correlated with pathological stage. These results indicate that the expression trend of *APOBEC3B* is different in different tumors, and it may play a different role in their progression.

In the patient survival analysis, we found that the relationship between *APOBEC3B* expression and survival obtained by GEPIA2 was basically the same for OS and DFS. In most tumors, a high expression level of *APOBEC3B* indicated a poor prognosis, such as ACC, KIRP, LGG, LIHC, THCA, and UVM (but not in CESC). To further confirm the reliability of survival analysis data, we used Kaplan–Meier plotter tool for progression-free survival (PFS) analysis of the above cancers with different prognoses. The results

(Additional file 2: Figure. S2) showed that high expression levels of *APOBEC3B* in ACC, LGG, LIHC, THCA and UVM were associated with poor PFS prognosis. It was associated with the low expression of *APOBEC3B* in CESC, but it was not statistically significant in KIRP. Thus, the expression level of *APOBEC3B* is closely linked to the prognosis and survival of patients in some tumor cases, but in different ways. Therefore, a larger sample size is needed to test the role of *APOBEC3B* in the survival and prognosis of different types of cancers.

Gene alterations play important roles in tumorigenesis. We found that an *APOBEC3B* gene alteration was associated with patient prognosis and survival in SKCM, and the prognosis for the gene-altered group was significantly worse than that of the unaltered group. However, in other tumors with an *APOBEC3B* gene alteration, we did not find an impact of gene alterations on patient prognosis and survival. Because of the small number of altered group cases for most tumors, we think it is necessary to further expand the data sample for verification in the future. In addition, we also analyzed the relationship between the expression levels of *APOBEC3B* and tumor mutational burden (TMB) and microsatellite instability (MSI) in all TCGA tumors. MSI often reflects a defect in the DNA mismatch repair system and the TMB reflects cancer mutation quantity. [35, 36] Additional file 2: Figure S3 shows that the expression level of *APOBEC3B* was positively correlated with the TMB of ACC, BLCA, BRCA, CESC, COAD, LIHC, LUAD, OV, PRAD, PAAD, SKCM, THCA, and THYM. The expression level of *APOBEC3B* was also negatively correlated with the MSI of BRCA, HNSC, KIRC, LIHC, STAD, and TGCT. These findings deserve further analysis.

Cancer-associated fibroblasts are considered to be a form of mutant cells with negative epithelial, endothelial and leukocyte markers, with slender morphology not found in cancer cells [37]. They have a variety of functions, including matrix deposition and remodeling, extensive signal interaction with cancer cells, and cross-talking with invasive leukocytes [38]. Therefore, they are potential factors affecting cancer treatment strategies. CD8$^+$ cytotoxic T lymphocytes are immune cells that can target cancer, and tumor-associated fibroblasts, macrophage type 2 cells and regulatory T cells can form an immune barrier to the anti-tumor immune response mediated by such cells [39]. Our study is the first to suggest an association between the expression of *APOBEC3B* and the level of immune cell infiltration of tumor-associated fibroblasts and CD8$^+$ T cells in some tumors. In addition, we integrated gene information related to *APOBEC3B* binding or affecting *APOBEC3B* expression in all tumors via the STRING and GEPIA2 databases, and conducted a variety of interaction, correlation and enrichment analyses. In addition, we integrated gene information related to *APOBEC3B* binding or affecting *APOBEC3B* expression in all tumors through a variety of databases, and conducted a variety of interaction, correlation and enrichment analyses (Table S2). Finally, we found that the GO terms 'ATP synthesis-coupled electron transport', 'thermogenesis' and 'oxidative phosphorylation' might be involved in the pathogenesis of cancer. Moreover, we screened for the KPNA2 gene based on two databases, and found that its expression was highly positively correlated with the expression level of *APOBEC3B*. KPNA2 is a member of the nuclear transport protein family. In many studies, KPNA2 has been linked to tumorigenesis (such as non-small cell lung cancer, breast cancers, and epithelial ovarian cancer), [40–42] and the gene is highly expressed in many cancerous tissues, which

is consistent with the expression of the *APOBEC3B* gene in this study. It has a cancer promoting effect in in vivo and in vitro experiments, and is often related to a poor prognosis for patients [43]. We hypothesize that there might be some mechanism connecting *APOBEC3B* and KPNA2 that promotes the occurrence and development of tumors, which needs to be further tested by in vivo and in vitro experiments in the future.

Conclusions

Here we conducted pan-cancer analysis of *APOBEC3B,* explored the relationship between *APOBEC3B* expression levels and clinical prognosis, immune cell infiltration, TMB or MSI, as well as the mutation sites, interactive genes and involved pathways of the *APOBEC3B* gene. These findings indicate that APOBEC3B might have different biological functions in different cancer species. Their main significance lies in the need to take such differences into account when designing drugs targeted for individual genes (Additional file 3: Table S2).

Abbreviations

ACC	Adrenocortical carcinoma
APOBEC3B	Apolipoprotein B mRNA editing enzyme catalytic subunit 3B
BLCA	Bladder urothelial carcinomas
BP	Biological process
CC	Cellular component
CEP55	Centrosomal protein
CESE	Cervical squamous cell carcinomas and endocervical adenocarcinomas
CHOL	Cholangiocarcinomas
CNA	Copy number alteration
COAD	Colon adenocarcinomas
CPTAC	Clinical proteomic tumor analysis consortium
DBLC	Lymphoid neoplasm diffuse large B-cell lymphomas
DFS	Disease-free survival
ESCA	Esophageal carcinomas
GBM	Glioblastoma multiforme
GEPIA	Gene expression profiling interactive analysis
GO	Gene ontology
GTE	Genotype-tissue expression
HNSC	Head and neck squamous carcinomas
HPV	Human papilloma virus
KEGG	Kyoto encyclopedia of genes and genomes
KICH	Kidney chromophobes
KIRC	Kidney renal clear cell carcinomas
KIRP	Kidney renal papillary cell carcinomas
KPNA2	Karyopherin subunit alpha 2
LAML	Acute myeloid leukemia
LGG	Brain lower grade gliomas
LIHC	Liver hepatocellular carcinomas
LUAD	Lung adenocarcinomas
LUSC	Lung squamous cell carcinomas
MCM2	Minichromosome maintenance complex component 2
MELK	Maternal embryonic leucine zipper kinase
MESO	Mesotheliomas
MF	Molecular function
MSI	Microsatellite instability
NCAPH	Non-SMC condensin I complex subunit H
OV	Ovarian cancers
PAAD	Pancreatic adenocarcinomas
PCPG	Pheochromocytomas and paragangliomas
PFS	Progression-free survival
PRAD	Prostate adenocarcinomas
READ	Rectum adenocarcinoma
SKCM	Skin cutaneous melanomas
STAD	Stomach adenocarcinomas
TCGA	The cancer genome atlas
TGCT	Testicular germ cell tumors

THCA	Thyroid carcinomas
THYM	Thymomas
TIMER	Tumor immune estimation resource
TK1	Thymidine kinase 1
TMB	Tumor mutational burden. TPM: transcripts per million reads
UCEC	Uterine corpus endometrial carcinomas
UCS	Uterine carcinosarcomas
UVM	Uveal melanomas

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-022-04862-0.

Additional file 1. Clinical prognostic significance of DNA methylation of APOBEC3B.

Additional file 2. An overview of apobec3b and the relationship between its expression and PFS or TMB.

Additional file 3. GO and KEGG enrichment analysis of APOBEC3B related genes.

Availability of data and materials
The datasets generated and analyzed during the current study are available in the TCGA repository, https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors report no competing interests in this work.

Published online: 02 August 2022

References
1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics 2021. CA Cancer J Clin. 2021;71(1):7–33. https://doi.org/10.3322/caac.21654.
2. Rahib L, Wehner MR, Matrisian LM, Nead KT. Estimated projection of US Cancer incidence and death to 2040. JAMA Netw Open. 2021;4(4): e214708. https://doi.org/10.1001/jamanetworkopen.2021.4708.
3. Ganini C, Amelio I, Bertolo R, et al. Global mapping of cancers: the cancer genome atlas and beyond. Mol Oncol. 2021;15(11):2823–40. https://doi.org/10.1002/1878-0261.13056.
4. Tomczak K, Czerwinska P, Wiznerowicz M. The cancer genome atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol. 2015;19(1A):A68–77. https://doi.org/10.5114/wo.2014.47136.
5. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 2013. https://doi.org/10.1093/nar/gks1193.
6. Yang B, Li X, Lei L, Chen J. APOBEC: from mutator to editor. J Genet Genomics. 2017;44(9):423–37. https://doi.org/10.1016/j.jgg.2017.04.009.
7. Dunham I, Shimizu N, Roe BA, et al. The DNA sequence of human chromosome 22. Nature. 1999;402(6761):489–95. https://doi.org/10.1038/990031.
8. Jarmuz A, Chester A, Bayliss J, et al. An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. Genomics. 2002;79(3):285–96. https://doi.org/10.1006/geno.2002.6718.
9. Wedekind JE, Dance GS, Sowden MP, Smith HC. Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. Trends Genet. 2003;19(4):207–16. https://doi.org/10.1016/S0168-9525(03)00054-4.

Wu *et al. BMC Bioinformatics*   (2022) 23:312

Page 15 of 16

10. Burns MB, Lackey L, Carpenter MA, et al. *APOBEC3B* is an enzymatic source of mutation in breast cancer. Nature. 2013;494(7437):366–70. https://doi.org/10.1038/nature11881.

11. Wu PF, Chen YS, Kuo TY, Lin HH, Liu CW, Chang LC. *APOBEC3B*: a potential factor suppressing growth of human hepatocellular carcinoma cells. Anticancer Res. 2015;35(3):1521–7.

12. Du Y, Tao X, Wu J, Yu H, Yu Y, Zhao H. *APOBEC3B* up-regulation independently predicts ovarian cancer prognosis: a cohort study. Cancer Cell Int. 2018;18:78. https://doi.org/10.1186/s12935-018-0572-5.

13. Feng C, Zhang Y, Huang J, Zheng Q, Yang Y, Xu B. The prognostic significance of *APOBEC3B* and PD-L1/PD-1 in nasopharyngeal carcinoma. Appl Immunohistochem Mol Morphol. 2021;29(3):239–44. https://doi.org/10.1097/PAI.0000000000000852.

14. Jin Z, Han YX, Han XR. The role of *APOBEC3B* in chondrosarcoma. Oncol Rep. 2014;32(5):1867–72. https://doi.org/10.3892/or.2014.3437.

15. Li T, Fu J, Zeng Z, et al. TIMER20 for analysis of tumor-infiltrating immune cells. Nucleic Acids Res. 2020;48(W1):W509–14. https://doi.org/10.1093/nar/gkaa407.

16. Li T, Fan J, Wang B, et al. TIMER: A web server for comprehensive analysis of Tumor-Infiltrating immune cells. Cancer Res. 2017;77(21):e108–10. https://doi.org/10.1158/0008-5472.CAN-17-0307.

17. Li B, Severson E, Pignon JC, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol. 2016;17(1):174. https://doi.org/10.1186/s13059-016-1028-7.

18. Goldman MJ, Craft B, Hastie M, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020;38(6):675–8. https://doi.org/10.1038/s41587-020-0546-8.

19. Vivian J, Rao AA, Nothaft FA, et al. Toil enables reproducible, open source, big biomedical data analyses. Nat Biotechnol. 2017;35(4):314–6. https://doi.org/10.1038/nbt.3772.

20. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res. 2019;47(W1):W556–60. https://doi.org/10.1093/nar/gkz430.

21. Chandrashekar DS, Bashel B, Balasubramanya S, et al. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia. 2017;19(8):649–58. https://doi.org/10.1016/j.neo.2017.05.002.

22. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6(269): l1. https://doi.org/10.1126/scisignal.2004088.

23. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607–13. https://doi.org/10.1093/nar/gky1131.

24. Jiao X, Sherman BT, Huang DW, et al. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. Bioinformatics. 2012;28(13):1805–6. https://doi.org/10.1093/bioinformatics/bts251.

25. Burns MB, Temiz NA, Harris RS. Evidence for *APOBEC3B* mutagenesis in multiple human cancers. Nat Genet. 2013;45(9):977–83. https://doi.org/10.1038/ng.2701.

26. Sokratous G, Polyzoidis S, Ashkan K. Immune infiltration of tumor microenvironment following immunotherapy for glioblastoma multiforme. Hum Vaccin Immunother. 2017;13(11):2575–82. https://doi.org/10.1080/21645515.2017.1303582.

27. Shinohara M, Io K, Shindo K, et al. *APOBEC3B* can impair genomic stability by inducing base substitutions in genomic DNA in human cells. Sci Rep. 2012;2:806. https://doi.org/10.1038/srep00806.

28. Kuong KJ, Loeb LA. *APOBEC3B* mutagenesis in cancer. Nat Genet. 2013;45(9):964–5. https://doi.org/10.1038/ng.2736.

29. Barbieri CE, Baca SC, Lawrence MS, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet. 2012;44(6):685–9. https://doi.org/10.1038/ng.2279.

30. Stransky N, Egloff AM, Tward AD, et al. The mutational landscape of head and neck squamous cell carcinoma. Science. 2011;333(6046):1157–60. https://doi.org/10.1126/science.1208130.

31. Pak V, Heidecker G, Pathak VK, Derse D. The role of amino-terminal sequences in cellular localization and antiviral activity of *APOBEC3B*. J Virol. 2011;85(17):8538–47. https://doi.org/10.1128/JVI.02645-10.

32. Prohaska KM, Bennett RP, Salter JD, Smith HC. The multifaceted roles of RNA binding in APOBEC cytidine deaminase functions. Wiley Interdiscip Rev RNA. 2014;5(4):493–508. https://doi.org/10.1002/wrna.1226.

33. Bacolla A, Cooper DN, Vasquez KM. Mechanisms of base substitution mutagenesis in cancer genomes. Genes. 2014;5(1):108–46. https://doi.org/10.3390/genes5010108.

34. Vio CP, Martinez L, Gandolfi C, Leniz P. Effect of cyclosporine a on the distal nephron: the use of kallikrein as a specific morphological marker. Transplant Proc. 1990;22(4):1730–2.

35. Yang G, Zheng RY, Jin ZS. Correlations between microsatellite instability and the biological behaviour of tumours. J Cancer Res Clin Oncol. 2019;145(12):2891–9. https://doi.org/10.1007/s00432-019-03053-4.

36. Jardim DL, Goodman A, de Melo GD, Kurzrock R. The challenges of tumor mutational burden as an immunotherapy biomarker. Cancer Cell. 2021;39(2):154–73. https://doi.org/10.1016/j.ccell.2020.10.001.

37. Sahai E, Astsaturov I, Cukierman E, et al. A framework for advancing our understanding of cancer-associated fibroblasts. Nat Rev Cancer. 2020;20(3):174–86. https://doi.org/10.1038/s41568-019-0238-1.

38. Olumi AF, Grossfeld GD, Hayward SW, Carroll PR, Tlsty TD, Cunha GR. Carcinoma-associated fibroblasts direct tumor progression of initiated human prostatic epithelium. Cancer Res. 1999;59(19):5002–11. https://doi.org/10.1186/bcr138.

39. Farhood B, Najafi M, Mortezaee K. CD8(+) cytotoxic T lymphocytes in cancer immunotherapy: a review. J Cell Physiol. 2019;234(6):8509–21. https://doi.org/10.1002/jcp.27782.

40. Wang CI, Chien KY, Wang CL, et al. Quantitative proteomics reveals regulation of karyopherin subunit alpha-2 (KPNA2) and its potential novel cargo proteins in nonsmall cell lung cancer. Mol Cell Proteomics. 2012;11(11):1105–22. https://doi.org/10.1074/mcp.M111.016592.

41. Noetzel E, Rose M, Bornemann J, Gajewski M, Knuchel R, Dahl E. Nuclear transport receptor karyopherin-alpha2 promotes malignant breast cancer phenotypes in vitro. Oncogene. 2012;31(16):2101–14. https://doi.org/10.1038/onc.2011.403.

42. Zheng M, Tang L, Huang L, et al. Overexpression of karyopherin-2 in epithelial ovarian cancer and correlation with poor prognosis. Obstet Gynecol. 2010;116(4):884–91. https://doi.org/10.1097/AOG.0b013e3181f104ce.
43. Han Y, Wang X. The emerging roles of KPNA2 in cancer. Life Sci. 2020;241: 117140. https://doi.org/10.1016/j.lfs.2019.117140.

Publisher's Note