Article

APOBEC3-Mediated RNA Editing in Breast Cancer is Associated with Heightened Immune Activity and Improved Survival

Mariko Asaoka [1,2], Takashi Ishikawa [2], Kazuaki Takabe [1,2,3,4,5,*] and Santosh K. Patnaik [3,6,*]

1   Department of Breast Surgery, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263, USA;
    590mariko@gmail.com
2   Department of Breast Surgery and Oncology, Tokyo Medical University, Tokyo 160-8402, Japan;
    tishik55@gmail.com
3   Department of Surgery, Jacobs School of Medicine and Biomedical Sciences, State University of New York,
    Buffalo, NY 14263, USA
4   Department of Surgery, Niigata University Graduate School of Medical and Dental Sciences, Niigata
    951-8510, Japan
5   Department of Surgery, Yokohama City University, Yokohama 236-0004, Japan
6   Department of Thoracic Surgery, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263, USA
*   Correspondence: kazuaki.takabe@roswellpark.org (K.T.); santosh.patnaik@roswellpark.org (S.K.P.);
    Tel.: +1-716-8455540 (K.T.); +1-716-8458359 (S.K.P.); Fax: (+1)-716-8451668 (K.T.); (+1)-716-8458922 (S.K.P.)

check for
updates

Abstract: APOBEC3 enzymes contribute significantly to DNA mutagenesis in cancer. These enzymes are also capable of converting C bases at specific positions of RNAs to U. However, the prevalence and significance of this C-to-U RNA editing in any cancer is currently unknown. We developed a bioinformatics workflow to determine RNA editing levels at known APOBEC3-mediated RNA editing sites using exome and mRNA sequencing data of 1040 breast cancer tumors. Although reliable editing determinations were limited due to sequencing depth, editing was observed in both tumor and adjacent normal tissues. For 440 sites (411 genes), editing was determinable for ≥5 tumors, with editing occurring in 0.6%–100% of tumors (mean 20%, SD 14%) at an average level of 0.6%–20% (mean 7%, SD 4%). Compared to tumors with low RNA editing, editing-high tumors had enriched expression of immune-related gene sets, and higher T cell and M1 macrophage infiltration, B and T cell receptor diversity, and immune cytolytic activity. Concordant with this, patients with increased RNA editing in tumors had better disease- and progression-free survivals (hazard ratio = 1.67–1.75, $p < 0.05$). Our study identifies that APOBEC3-mediated RNA editing occurs in breast cancer tumors and is positively associated with elevated immune activity and improved survival.

Keywords: APOBEC; breast cancer; RNA editing; sequencing; tumor immune microenvironment

1. Introduction

Base modifications by deamination of adenine to inosine (A-to-I) and of cytidine to uracil (C-to-U) are the major types of RNA editing in higher eukaryotes. Inosine and uracil in RNA are read as guanosine and thymine, respectively, during RNA translation. Thus, RNA editing has the capability of altering protein sequences encoded by DNA. The activation-induced deaminase (AID), APOBEC, and cytidine deaminase (CDA) proteins of mammals harbor the cytidine deaminase motif for hydrolytic deamination of C to U [1]. CDA is involved in the pyrimidine salvaging pathway. While AID causes C-to-U mutation of DNA, multiple studies have failed to identify any RNA editing activity of the protein [2]. Humans have ten (*APOBEC1, 2, 3A–D, 3F–H,* and *4*) and mice have four (*APOBEC1–4*)

APOBEC genes. Expression of *APOBEC1*, *APOBEC2*, and *APOBEC4* genes is largely restricted to intestine, muscle, and gonad tissues, respectively, whereas *APOBEC3* genes are expressed the most in cells of the immune system [3]. RNA sequencing studies have identified wide-spread C-to-U RNA editing in mammals (e.g., [4–6]).

APOBEC3 proteins cause C-to-U deamination of single-stranded DNA, and this DNA deamination activity underlies the proteins' well-established role in restricting viruses and retrotransposon elements [7,8]. Lately, the DNA deamination activity of APOBEC3 proteins has also been identified as a major contributor to mutagenesis in multiple human cancers, including that of breast [9–12]. Furthermore, individuals with particular structural changes in *APOBEC3A* and *APOBEC3B* genes have been found to be at an increased risk for development of breast, ovarian, and liver cancer [13,14].

While all human APOBEC3 proteins have been known for a while to bind RNA [15], it is only recently that their C-to-U RNA editing activity has come to light. In both independent and collaborative studies, our group has demonstrated such hitherto unknown C-to-U RNA editing activities for human APOBEC3A, APOBEC3B, and APOBEC3G enzymes, and shown physiologically-affected occurrence of such APOBEC3-mediated RNA editing in immune cells [16–19]. Although the prevalence and significance of APOBEC3-mediated DNA mutagenesis in multiple types of cancer are now known, such knowledge is lacking for APOBEC3-mediated RNA editing for any cancer. Because APOBEC3s are well-studied for breast cancer (e.g., [12,20–23]), we sought to examine this cancer to survey the prevalence of APOBEC3-mediated RNA editing and investigate its biological and clinical relevancies.

2. Results

2.1. Compilation of a Set of APOBEC3-Mediated RNA Editing Sites

In order to estimate APOBEC3-mediated C-to-U RNA editing in tissues of The Cancer Genome Atlas (TCGA)-breast carcinoma (BRCA) cohort, we collected a list of genome positions (editing sites) for which C-to-U RNA editing by APOBEC3 enzymes is known to occur in human cells. Monocytes and macrophages express multiple *APOBEC3* but not *APOBEC1*, *APOBEC2*, or *APOBEC4* genes [3]. Except for 96 sites identified by us in HEK-293T/APOBEC3B transfectants (manuscript in preparation), the sites were collected from published data. In all these studies, RNA editing at the sites was established using RNA sequencing-based approaches, although allele-specific polymerase chain reaction (PCR) or Sanger sequencing of cDNA amplicons was also utilized for a few dozen sites.

A total of 5208 sites, which were all the APOBEC3-mediated editing sites that were known when we performed this investigation, were thus collected (Table S1). A majority of the sites (79%) were identified with HEK-293T/APOBEC3A transfectants (Figure 1a). A small fraction (5.5%) had been discovered in multiple studies. Most (93.9%) of the sites lie within coding (62.7%) or untranslated (31.2%) regions of mRNAs. C-to-U editing of the mRNAs is predicted to affect their encoded protein sequences for a significant number of sites (29.7%), through either non-synonymous codon change (1195 sites) or gain of a stop codon (353 sites; Figure 1a).

2.2. Estimation of RNA Editing Levels in TCGA-BRCA Tissues

We identified 1040 breast cancer tumor and 93 adjacent normal tissues in the TCGA-BRCA project for which RNA and DNA (whole exome) sequencings had been performed with the same tissue portion. TCGA sequencing data that had been aligned with the reference human genome was obtained from Genomic Data Commons (GDC) portal and analyzed to estimate C-to-U RNA editing levels at each of the 5208 sites. A bioinformatics pipeline was developed for this study, the schematics of which is shown in Figure S1.

The exact criteria used for the RNA editing determination are noted in the Materials and Methods section. These criteria were developed such that the degree of their stringency permitted reliably confident RNA editing determination considering the type of data that was available to us. Nevertheless, RNA editing could not be calculated for a large fraction (97%) of determinations due to inadequate

sequencing depth in the data. This can be seen in the information provided in Table S2, which lists criteria and their fulfillment in the editing determination workflow.
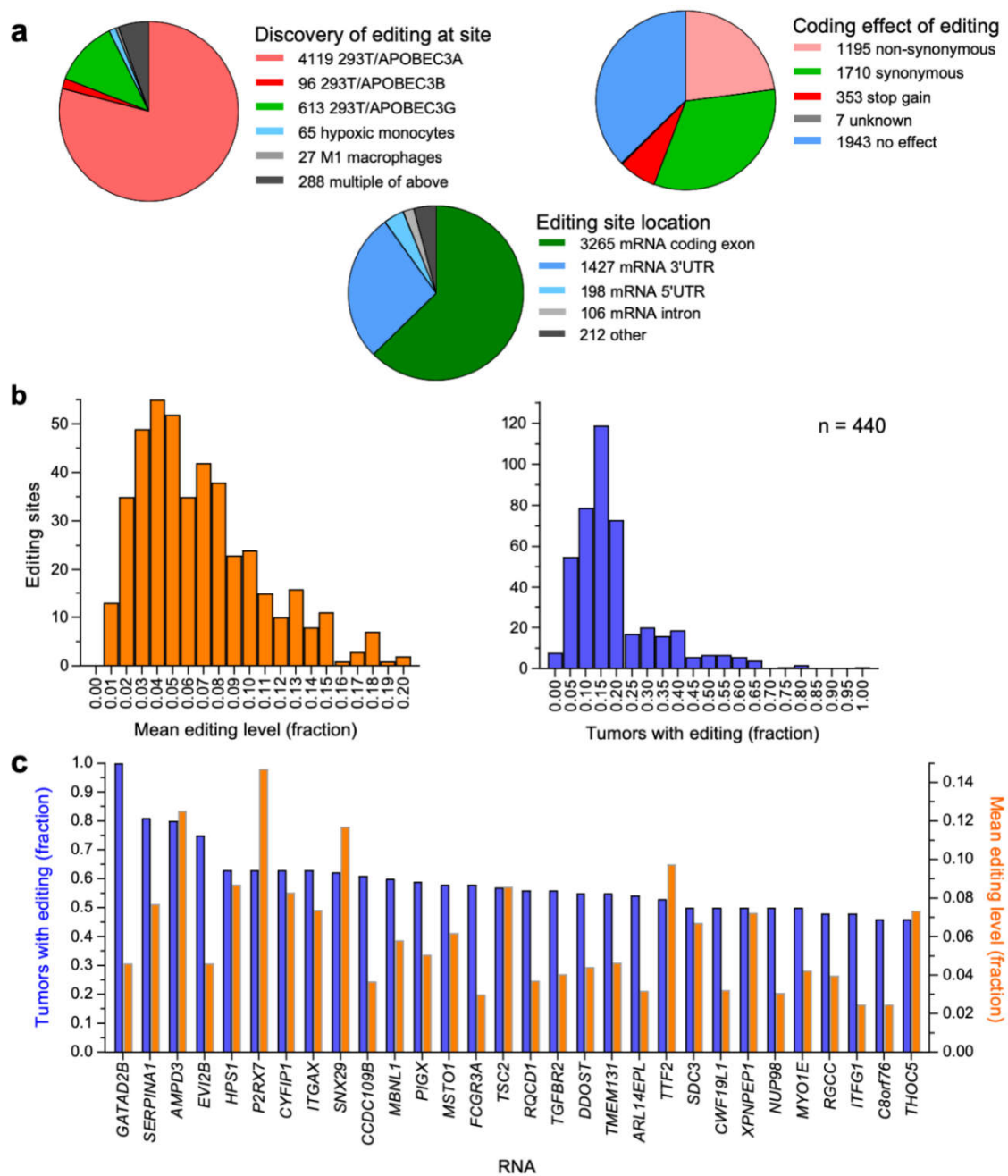

Figure 1. APOBEC3-mediated RNA editing sites. (**a**) Characteristics of the sites that are examined in this study are illustrated with pie-charts. UTR, untranslated region. (**b**) Frequency histograms are shown for the fraction of The Cancer Genome Atlas (TCGA)-BRCA tumors with detectable RNA editing and the mean level of such editing for 440 sites for which editing was detected in the tumor cohort. (**c**) Fraction of tumors with editing and their mean editing level are shown for the 30 most commonly edited RNAs of the cohort.

C-to-U editing determinations could not be made for any sample for 664 of the 5208 known APOBEC3-mediated editing sites. For the remaining 4544 sites, editing was determinable for 1–371 samples (mean = 11, standard deviation (SD) = 7). Among tumors, editing determination was possible for ≥5 tumors for only 2082 sites. A total of 1625 C-to-U editing events (editing level >0)

were identified with the workflow. With a set of 198 negative control sites that are transcribed as C but are not known to undergo APOBEC3-mediated C-to-U RNA editing (Table S1), 21 editing events were observed. To further affirm the strength of the editing determination workflow, the 5208 editing sites were also evaluated for the hypothetical C-to-A type of RNA editing and only 152 such events were identifiable.

Editing determinations that were programmatically made with the bioinformatics pipeline were confirmed for multiple determinations through visual assessment of aligned RNA and DNA sequencing data. A few examples of such visualization of editing events are depicted in Figure S2. Because of the unavailability of TCGA tissues or their nucleic acid preparations, validation of editing through experimental methods such as allele-specific PCR or Sanger sequencing could not be performed. To account for the possibility of misestimation of editing due to artifacts arising from read sequence alignment, we reassessed C-to-U RNA editing levels for 24 randomly selected editing events after aligning RNA sequencing data with a different mapping software (Subread subjunc [24], instead of STAR, which was used for the GDC data). As shown in Figure S3, editing levels with the originally aligned and Subread-aligned sequencing data had good correlation (Pearson $r = 0.85$), with a mean deviation value of 13.6%.

2.3. APOBEC3-Mediated C-to-U RNA Editing is Prevalent in Breast Cancer Tumors

C-to-U RNA editing events at the 5208 known APOBEC3-mediated editing sites were identified among the TCGA-BRCA tissues. For 440 sites, in a total of 411 gene transcripts, editing was determinable for ≥5 tumors, with editing occurring at the sites in 0.6%–100% of the tumors (mean = 20%, SD = 14%) at an average level of 0.6%–20% (mean = 7%, SD = 4%). Distributions of these values are plotted in Figure 1b. Among these 440 sites, editing was most prevalent for *GATAD2B* (100% of tumors with a determinable editing level), *SERPINA1* (81%), and *AMPD3* (80%) RNAs. Mean editing levels were highest for *SH3PXD2A* (20%), *FAM208A* (20%), and *DSCR3* (19%) RNAs. The number of tumors with RNA editing was largest for *SERPINA1* (160), followed by *DDOST* (60) and *EVI2B* (43). Mean editing levels and editing frequency among tumors for the 30 RNAs for which editing was most prevalent among tumors are plotted in Figure 1c. Table S3 has genome coordinates and other annotations for the 50 most commonly edited sites among tumors. Of the 411 genes representing the 440 sites, 80 are in the gene ontology (GO) biological process set of genes for "immune system process," which is a 1.5-fold overrepresentation with false discovery rate (FDR) of 0.03.

No editing event was observed for 409 tumors and 32 normal tissues that had determinable editing for ≥5 sites. Among the 448 tumors and 44 normal tissues that had editing at ≥1 site and ≥5 editing-determinable sites, an average of 6.6% of determinable sites had editing (SD = 4.2%), with an average editing level of 6.6% (SD = 3.4%). To summarize editing levels across multiple editing sites in a sample, we used the ratio of editing-positive sites and determinable sites as an editing score. This editing score was calculated only for samples with ≥5 editing-determinable sites. Samples with editing scores of 0 and >0 were considered as editing-low and editing-high samples, respectively, for further analyses. There were thus 409 editing-low and 448 editing-high tumors in the TCGA-BRCA cohort.

Tumor and normal tissues had similar editing scores as suggested by *p*-value of 0.06 in paired analysis with standard *t*-test. The incidence of editing-high tumors was similar among groups of tumors of the American Joint Committee on Cancer pathological stages I–IV (Figure 2). Incidence rates were also similar between lymph node metastasis-negative (TNM N0) and -positive (TNM N1–3) tumors. Incidence rates were significantly different among TNM T1–T4 tumor groups, and among different breast cancer subtypes, with editing-high tumors about 40% more prevalent for HER2 compared to luminal subtype. Both estrogen and progesterone receptor negativity were associated with increased editing-high incidence, whereas the fraction of editing-high tumors was less among HER2 receptor-negative compared to -positive tumors. High mitosis and nuclear pleomorphism scores were associated with increased editing-high incidence. The relation of editing with cell proliferation that was suggested by this finding was also noted with the observation of increased editing-high

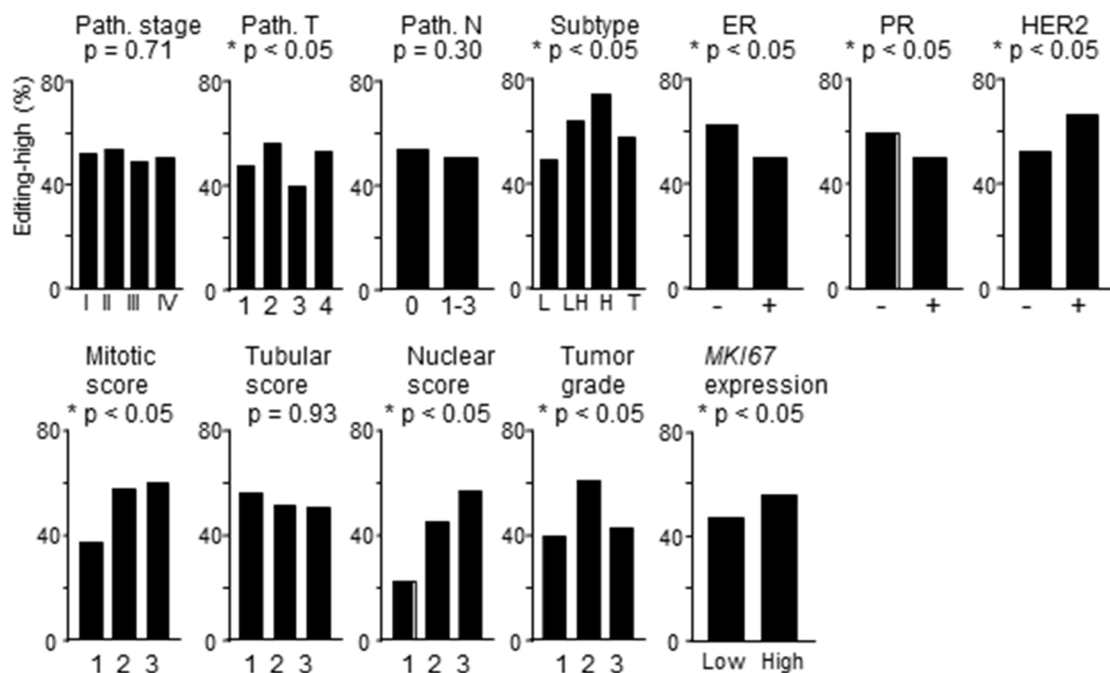incidence among tumors with high gene expression of *MKI67*, a well-known proliferation marker (Figure 2).

Figure 2. RNA editing and clinico-pathological features. Incidence of RNA editing-high tumors are plotted for different features of breast cancer. Pathological (Path.) stage and T and N TNM values are as per the American Joint Committee on Cancer staging system. Indicators of breast cancer subtypes are as follows: L, luminal type; LH, luminal-HER2; H, HER2; T, triple negative. For *MKI67* gene expression, the median gene expression value was used to identify low and high groups. *p*-values were determined with Fisher's exact test. ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2.

2.4. APOBEC3 Gene Expression Correlates with RNA Editing Better Than APOBEC-Mediated Mutations

As expected, editing scores were associated with tumor gene expression of *APOBEC3* genes. Expressions among editing-high tumors were significantly raised for all seven *APOBEC3*s, with the largest expression difference (about 3×) seen for *APOBEC3A* (Figure 3a). In line with this, editing scores had the best correlation with *APOBEC3A* gene expression (Pearson $r = 0.31$; Figure 3b). Association of *APOBEC3* gene expression with editing at specific sites was also examined. In general, tumors with editing at specific sites had higher expression of multiple *APOBEC3* genes compared to tumors without editing. Representative examples are shown in Figure S4. While positive correlations of *APOBEC3* gene expressions and RNA editing were modest, except for *APOBEC3B* ($r < 0$), they were decidedly higher than those observed for APOBEC-mediated DNA mutation burden of tumors (Figure 3b). Consistent with this, editing-high and -low tumors had similar single nucleotide mutation burden, APOBEC-mediated as well as all, and predicted neoantigen load (Figure 3c). However, editing-high tumors had more genome copy number alterations, aneuploidy, intra-tumoral clonal heterogeneity, and homologous recombination deficiency than editing-low tumors.
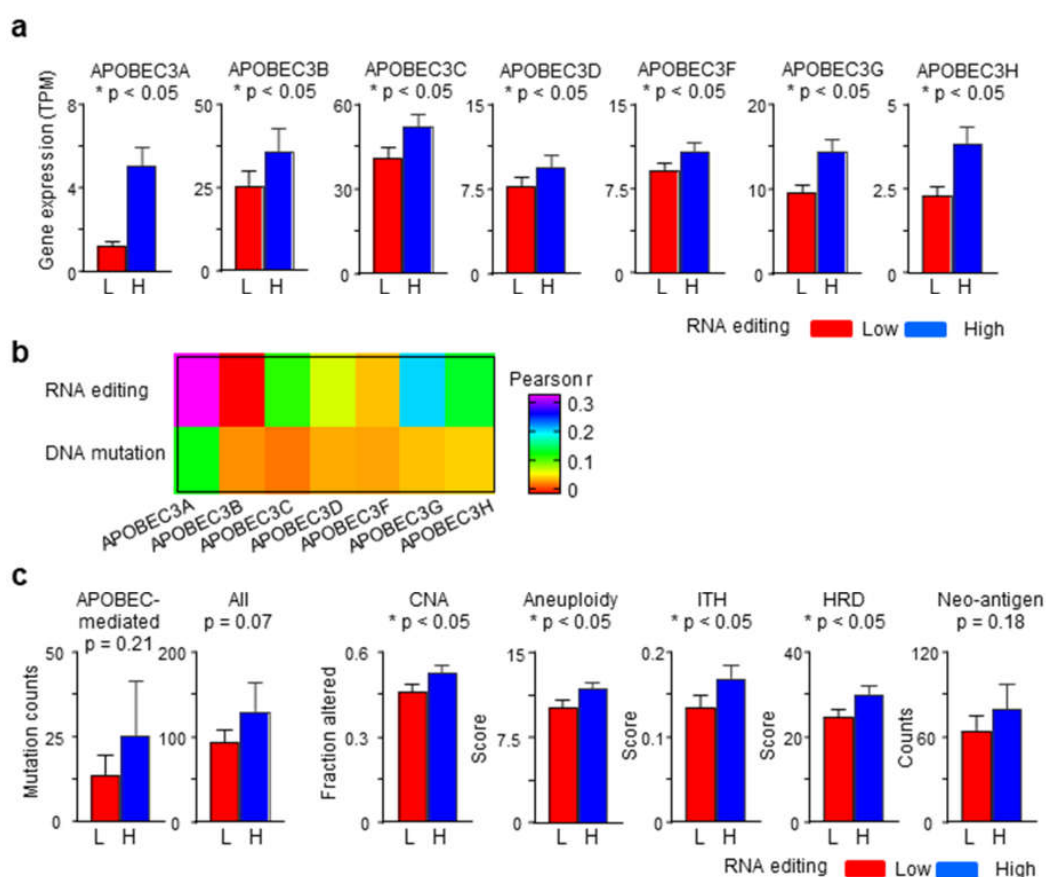
Figure 3. RNA editing and APOBEC3 gene expression and genome characteristics of tumors. (**a**) Tukey boxplots show *APOBEC3* gene expression of RNA editing-high (H) and -low (L) tumors. TPM, transcripts per million. (**b**) Heatmap shows Pearson's correlation coefficients for association of *APOBEC3* gene expression levels with APOBEC3-mediated C-to-U RNA editing score and APOBEC signature DNA mutation burden of tumors. (**c**) Barplots are shown for comparison of tumor genome characteristics of the RNA editing-high and -low tumors. *p*-values in two-group comparisons were calculated with Fisher's exact or standard *t*-tests. CNA, copy number alteration; HRD, homologous recombination deficiency; ITH, intra-tumor heterogeneity; SNV, single nucleotide variation.

2.5. RNA Editing is Associated with Enriched Expression of Immune- but not Cancer Development-Related Gene Sets

To obtain an insight into the biological basis for the variability in RNA editing among the TCGA-BRCA tumors, a differential gene expression analysis was performed to compare editing-high and -low tumor groups. Expression of 66 and 20 genes was significantly up- and downregulated with a ≥2-fold change, respectively, in the editing-high compared to -low group (adjusted *p* < 0.05 in limma's Bayes-moderated *t*-test). *APOBEC3A* was the most upregulated gene (Table S4); other upregulated genes included those encoding chemokine ligands, such as *CXCL10* and *CXCL13*, and interferons, such as *IFNG* and *IFNL1*. Of the 66 upregulated genes, 44 are in the PANTHER GO) biological process dataset of genes for immune system process, a 5.1-fold overrepresentation with FDR < 2e-20. Consistent with this, gene set enrichment analysis using the gene set variation analysis (GSVA) method [25] showed that, compared to the editing-low group, the editing-high group had significantly enriched expression for six of the seven immune-related gene sets of the Molecular Signatures Database (MSigDB) Hallmark collection (FDR < 0.05; Figure 4a). In contrast, enrichment was not observed for gene sets that are typically associated with cancer development, such as those for angiogenesis and epithelial mesenchymal transition (Figure 4b). Overall, enrichment was observed for 22 of the 50 hallmark gene sets.
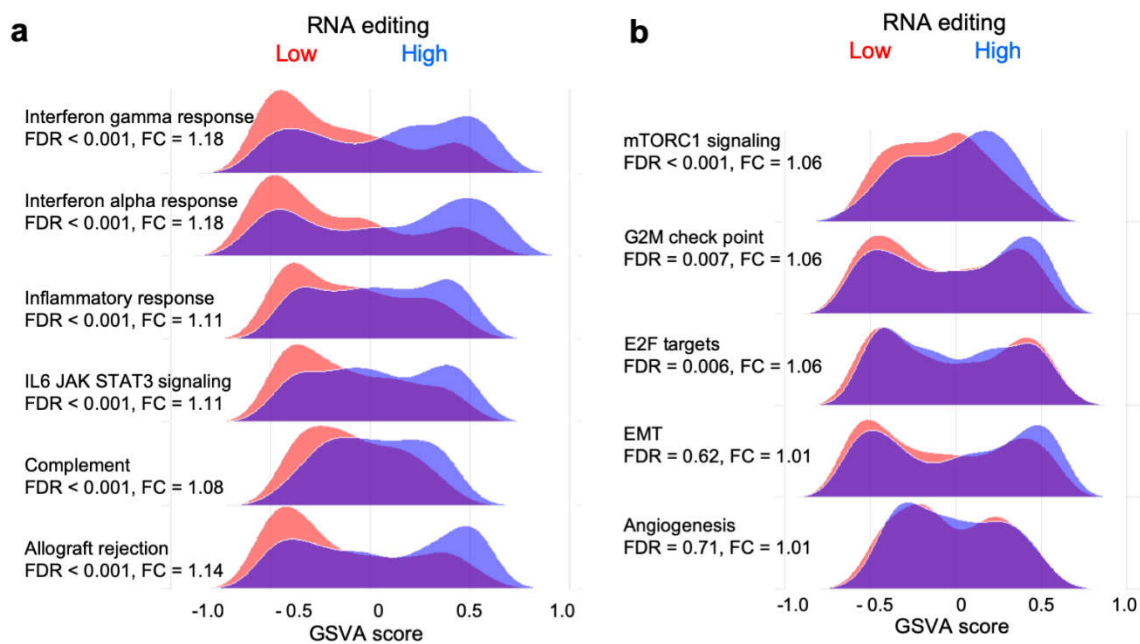
Figure 4. Immune-related gene sets have enriched expression in RNA editing-high tumors. Shown are examples of ridgeline plots of gene set variation analysis (GSVA) scores of RNA editing-high and -low tumors for (**a**) six immune-related gene sets and (**b**) five other gene sets of Molecular Signatures Database (MSigDB) Hallmark collection. Fold change (FC) values comparing scores of editing-high and -low tumors and false discovery rate (FDR) values in significance analysis of the comparisons were calculated with a Bayes-moderated *t*-test. EMT, epithelial mesenchymal transition.

2.6. Immune Activity is Elevated in Tumors with High RNA Editing

The immune response-related results of gene expression analyses that are described above, and the overrepresentation of immune response genes among the genes undergoing RNA editing in tumors, suggested a relevance of RNA editing in the tumor microenvironment. To test this, we compared editing-high and -low tumor groups for various features of the immune microenvironment. Tumor-infiltrating immune cell subsets were quantified from tumor gene expression data using the CIBERSORT algorithm [26]. Although both editing-high and -low tumor groups were similar for overall lymphocytic infiltration, the editing-high tumors had significantly more T cells as well as various T cell subtypes (CD8$^+$, CD4$^+$, T-reg, and gamma delta T; fold changes of 1.08–1.67). On the other hand, the fraction of B cells was reduced compared to editing-low tumors (fold change of 0.88; Figure 5a). Similarly, while the two groups were similar for macrophages, the M1 subtype of macrophages was more prevalent in editing-high group (fold change of 1.21), whereas the M2 subtype was less prevalent (fold change of 0.88) compared to the editing-low group (Figure 5a).

As may be expected from the positive association of RNA editing with T cell and M1 macrophage infiltration, RNA editing-high tumors showed markers of heightened anticancer immune response. Both B cell receptor (BCR) and T cell receptor (TCR) diversity (Shannon scores) and immune cytolytic activity (CYT score) were significantly higher (fold changes of 1.14–1.89) in the editing-high group of tumors (Figure 5b).
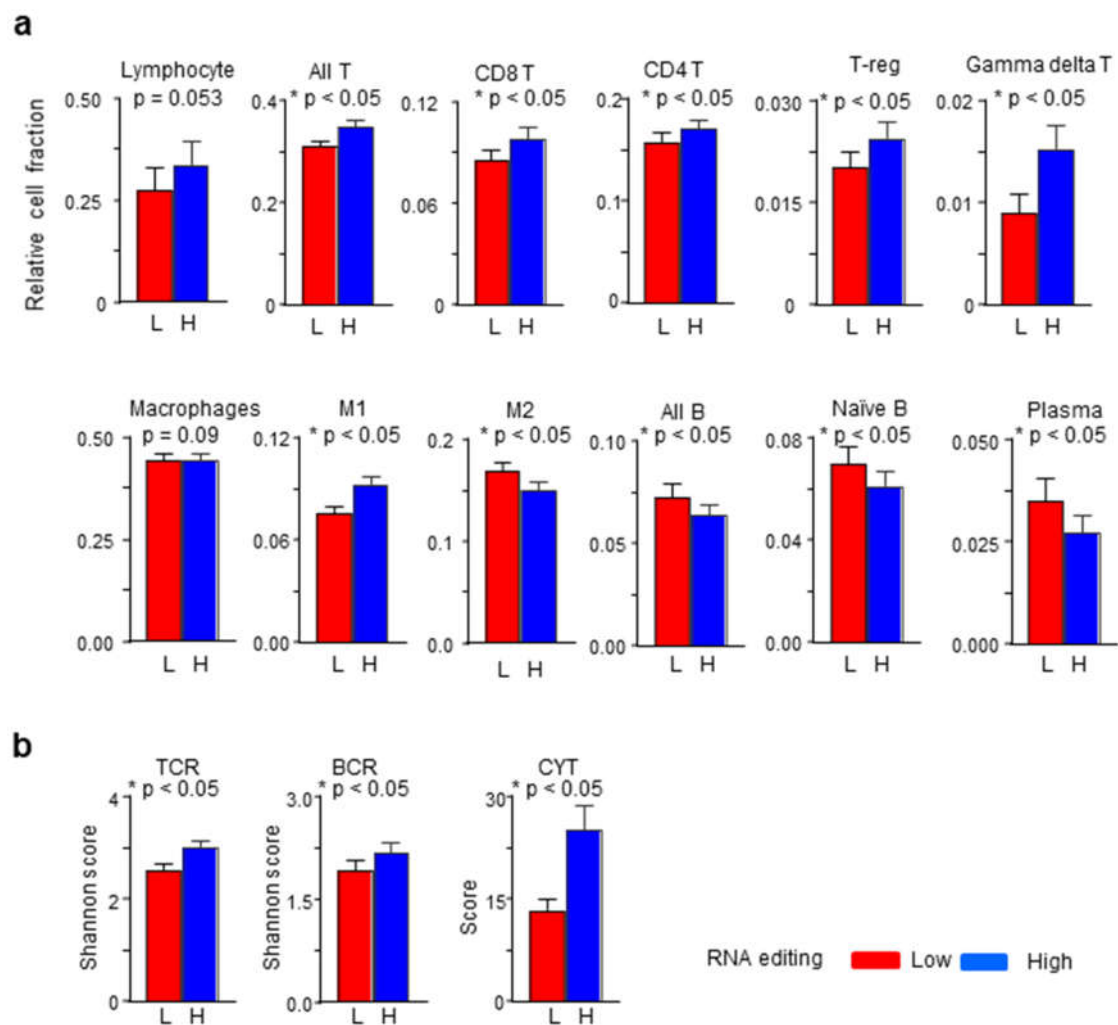
Figure 5. RNA editing and immune microenvironment features. Barplots of values for (**a**) relative fractions of tumor-infiltrating immune cells of different types and (**b**) T cell receptor (TCR) and B cell receptor (BCR) diversity (Shannon scores) and immune cytolytic activity (CYT score) are shown for the groups of RNA editing-high (H) and -low (L) tumors. The *p*-values in two-group comparisons were calculated with standard *t*-test.

2.7. Patients with Tumors with High RNA Editing in Tumors have Better Survival

The finding of a heightened immune activity in editing-high tumors indicated that patients who have high RNA editing in tumors may have a better clinical outcome. To investigate this, we performed survival analyses to compare the editing-low and -high groups of tumors. Recently released, uniformly curated, and filtered TCGA data for survival endpoints was used for this analysis [27]. Both disease-free and progression-free survivals were better for patients with editing-high tumors compared to those with editing-low tumors (Figure 6). Log-rank *p*-values in this analysis were <0.05, and hazard ratio (HR) values in Cox regression models were 1.75 and 1.67, respectively. For disease-specific and overall survival as well, the trend of improved survival for the editing-high group was observed, although *p*-values were >0.05 (both 0.08).
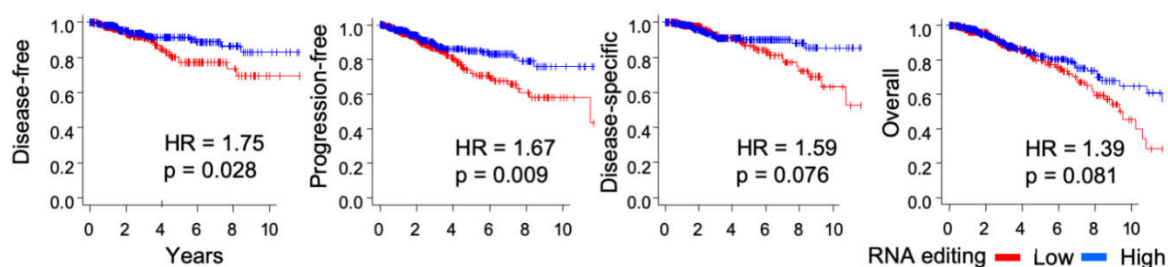
Figure 6. RNA editing is associated with improved survival. Kaplan–Meier survival plots along with log-rank test *p*-value and Cox proportional hazard ratio (HR) value are shown for association of RNA editing in tumors with disease-free, progression-free, disease-specific, and overall survival of patients. Groups of RNA editing-high and -low cases are compared.

3. Discussion

Base conversions in RNAs through editing by enzymes like APOBEC3s can have a consequence at the protein level because of an effect of editing on not only codon sequences but also on intron splicing, RNA stability, microRNA targeting, etc. Resulting changes in phenotypes of cancerous cells as well as cells of the tumor microenvironment therefore have the potential to affect cancer biology (e.g., [28,29]). These changes may include the generation of neoantigens on cancer cells and alterations in anticancer response pathways in immune cells [30]. The A-to-I type of RNA editing has been examined in tumors of multiple cancers, including breast cancer [31–33]. For C-to-U RNA editing, our knowledge of its prevalence in cancers is limited. APOBEC1-mediated RNA editing has been studied for one cancer [34], and cancers have not been surveyed for the prevalence of APOBEC3-mediated C-to-U RNA editing. We therefore investigated the prevalence and significance of such editing in tumors of breast cancer, for which APOBEC3-mediated DNA mutagenesis has been substantially explored. APOBEC3s are significant contributors of DNA mutagenesis in cancers, including breast cancer [35]. Human APOBEC3A, APOBEC3B, and APOBEC3G possess RNA editing capability [16–19], whereas it is unknown if the other APOBEC3 proteins too have such ability.

Our study, the first examination of APOBEC3-mediated C-to-U RNA editing in cancer, identifies that this phenomenon occurs in breast cancer tumors. Transcriptome and exome sequencing data were examined for this. Stringent criteria were used in the study to identify such editing in order to reduce false positivity because of sequence variations arising from inherited DNA polymorphisms, DNA mutations, sequencing errors, and read sequence alignment artifacts. Strictness of the criteria was limited by the depth of coverage in the aligned sequencing data that was available to us. With the criteria that were used, approximately 97% of editing estimations were indeterminable because of shallow coverage (Table S2). Considering only base calls of high quality (phred33 ≥20), the median coverage at the examined genome positions was 72 and 31, respectively, for RNA and DNA sequencing data (mean = 195 and 56, respectively). Nevertheless, C-to-U RNA editing events could be identified in the tumor tissues. Editing was observed for 448 tumors with editing determinable at ≥5 sites, and editing occurring at average 20% of the determinable sites at an average level of 7%. Reliability of the editing estimations is also likely to be high because the identified C-to-U editing events were at sites known to undergo APOBEC3-mediated RNA editing. Reliability was also attested through assessments for the hypothetical C-to-A type of RNA editing at the known C-to-U editing sites, and of C-to-U editing at a negative control set of genome positions that are not known to undergo APOBEC3-mediated RNA editing. While editing determinations could be somewhat validated through the examination of RNA sequencing data after realigning it with a different alignment software (Figure S4), validation with a different experimental technique was not possible because of unavailability of specimens. We were also unable to find an independent breast cancer tumor cohort of sufficient size with both RNA and exome or genome sequencing data of adequate depth coverage for validation purposes.

We devised an editing score to simplify further examination of the multi-dimensional RNA editing determinations. RNA editing across multiple sites of a sample was summarized into the editing score for the sample. The score was simply the fraction of sites for which editing was observed in the sample. We chose not to use the mean of editing levels among the sample's sites because the precision of editing level calculations was not high due to limitations of depth of sequencing coverage. The approach of targeted sequencing with a focus on known editing sites, which permits much deeper coverage, should be considered in future studies for accurate determination of RNA editing levels.

As expected, tumors with high editing scores (value >0) had higher expression of multiple *APOBEC3* genes compared to tumors with a score of 0 (Figure 3a). This association of editing with higher *APOBEC3* gene expression was also seen in the examination of editing at specific sites (Figure S4). Interestingly, the correlations of tumor *APOBEC3* gene expression with burden of mutations with APOBEC signature [12] were poorer compared to the RNA editing score (Figure 3b). This may be expected because, unlike DNA mutation (which gets fixed in the genome), edited RNAs have a limited life and their detection is contingent upon concurrent expression of RNA editing enzymes.

Because of the constrained site-specific editing determinability, determination being possible for an average of only 11 sites among tumors, we did not examine the biological or clinical relevance of editing of individual RNAs. In examinations using the editing score as an overall marker of APOBEC3-mediated RNA editing, numerous insights were obtained. Tumors with increased RNA editing had more cancer-related changes such as aneuploidy and copy number aberrations in their genomes (Figure 3c). Whether the increase in editing is in response to such changes remains unknown.

Not only were immune-related genes overrepresented among genes for which editing was detected, but differential gene expression and gene set enrichment analyses also revealed that expression of numerous genes associated with immune response was higher among tumors with high scores in contrast to tumors with low scores. Furthermore, tumors with high editing had more T compared to B cells. Infiltration levels of multiple subtypes of T cells, including $CD4^+$ and $CD8^+$, were higher among these tumors (Figure 5a). Not only was T cell infiltration high, their activity was also likely high as indicated by the increased T cell receptor diversity and cytolytic activity that were observed in the editing-high tumors (Figure 5b). These tumors also had more M1 and less M2 macrophages. This is consistent with the known observation that both *APOBEC3* gene expression and C-to-U RNA editing are markedly elevated in M1 compared to M2 macrophages from peripheral blood [18]. Concordant with these indications of a heightened anticancer immune response in the tumor microenvironment, patient survival was significantly better for those whose tumors had high RNA editing (Figure 6).

In this study, we used a well-established method, CIBERSORT, to estimate tumor-infiltrating immune cells using RNA sequencing data of whole tumor, which, besides cancer cells, contains unknown cells as well as closely related types of cells such as immune cell subsets. Accurate estimation of tumor immune cells with this algorithm has limitations because of inadequacies in cell-type signatures that the algorithm relies on, phenotypic malleability, and disease-induced alterations in cellular differentiation or transcription. Therefore, abundances of some specific tumor-infiltrating immune cell types could have been over- or under-estimated in our study, and the associations of abundances with RNA editing that we observed may be different if infiltrating immune cells had been quantified by flow cytometric or histologic methods.

The findings of our study imply that APOBEC3 enzymes are relevant in breast cancer because of not only their DNA mutagenic but also their RNA editing activity, and they highlight the pertinence of finer dissection of such editing in further studies. While we could detect C-to-U RNA editing events in the tumors, the cellular origin of such events remains unclear. Overrepresentation of edited sites among immune-related genes, enriched expression of such genes in editing-high tumors, and the fact that *APOBEC3* gene expression is much higher among immune compared to epithelial cells [7] suggests that the editing occurs in the tumor microenvironment. Further investigations implementing methods such as single-cell sequencing and isolation of sub-populations of cells from tumors are needed to definitively know if the editing occurs in cancerous epithelial or immune cells of breast

tumors. The biological consequences of the editing events on cancer development, progression, and immune response also remain unknown.

4. Materials and Methods

4.1. Collation of APOBEC3-Mediated C-to-U RNA Editing Sites

A total of 5208 sites of 3630 gene transcripts known to undergo APOBEC3-mediated C-to-U RNA editing were identified from our published and ongoing studies [16–19]. Specifically, genomic coordinates for the sites were collated from supplementary data of the publications. C-to-U RNA editing at these sites was observed in human 293T fibroblasts that were transiently transfected with expression constructs for human APOBEC3A, APOBEC3B, or APOBEC3G, and/or in human monocytes and macrophages that, respectively, were subjected to hypoxia and M1 polarization. The LiftOver tool at https://genome.ucsc.edu/cgi-bin/hgLiftOver was used to convert the published GRCh37 human genome reference-based coordinates to GRCh38 in the case of sites identified in macrophages or monocytes [18]. The list of 5208 sites is provided in Table S1. The ANNOVAR tool (Feb. 2016 version) and its RefSeq-based database (20151211 build) were used to annotate the sites with information on gene name, RNA region, effect of editing on protein sequence, etc. For negative control examination, 198 genome positions that are transcribed as C and are adjacent to one of the 5208 editing sites but are not known to undergo APOBEC3-mediated RNA editing were randomly selected. These control sites are listed in Table S1.

4.2. Access to Controlled Sequence Alignment Data of TCGA Project

Approval for access to aligned DNA (whole exome) and RNA sequencing data was obtained from National Institutes of Health, Bethesda, MD, USA, under study accession phs000178 of Database of Genotypes and Phenotypes (dbGaP). The data, in BAM format, was downloaded in early 2018 from the Genomic Data Commons (GDC) portal of National Cancer Institute, Bethesda, MD, USA, using the gdc-client GDC data transfer tool (version 1.3). Data in GDC were processed for all TCGA samples in a harmonized and uniform manner, using the GRCh38.p0 human genome reference [36]. Briefly, alignment of RNA sequencing data was done with STAR software (version 2.4) in a 2-pass workflow [37]. For whole exome DNA sequencing reads, BWA aligner software (version 0.7.12) was used in a workflow that used Picard MarkDuplicates (version 1.1.38) and GATK IndelRealigner (version 3.5) for duplicate-marking and co-cleaning [38].

4.3. Estimation of RNA Editing Levels in TCGA Tissues

Aligned DNA and RNA sequencing data were examined and compared to determine editing levels in 1040 tumor and 93 adjacent normal tissues. DNA and RNA data were from the same portion of TCGA tissue samples as determined from their biospecimen data identifiers. The data were analyzed with samtools (version 1.3) to generate pileups of base calls of phred33 quality $\geq$20 for each examined genome position. Pileups were then parsed with custom scripts in awk and R languages to count base calls (A/C/G/T) and to determine editing level from the count data, as outlined in Figure S1. In the examination for C-to-U RNA editing, reference and variant bases were C and T for genes that are transcribed from the + chromosomal strand, and G and A otherwise. First, RNA data were examined. Editing was deemed not determinable if coverage (sum of all base calls, N) was poor (<10), or if the reference base count (R) suggested the presence of DNA polymorphism, either homozygous ($R < 6$ for $N \leq 20$, or $< 0.4\times N$ for $N > 20$) or heterozygous ($R < 0.8\times N$, otherwise), or if variant base count (V) was 1. For $V = 0$, editing level was estimated as 0 if R was $> 3\times$ mean of R/V ratios of all samples with $V > 0$; otherwise, editing was considered indeterminable. For $V \geq 2$, editing was indeterminable if DNA sequencing coverage was <10, or if the p-value in the comparison of R and V values of RNA and DNA data with one-sided Barnard's test (Boschloo method) was >0.05. Otherwise, editing level was estimated as the $V/(R + V)$ fraction if V was 0 for DNA data or $V/(R + V)$ of RNA data was $\geq$10$\times$ that

of DNA data, and, if *V* was > 2, there was no strong (>8×) or significant (Fisher exact test $p \leq 0.01$) RNA sequencing strand bias of variant compared to reference bases. Strand bias was not tested for *V* = 2. Editing was considered indeterminable for all other cases. Table S2 lists these various criteria along with the number of RNA editing determinations made for the 1133 tissues and 5208 sites that fulfilled the criteria.

4.4. Validation of RNA Editing Level Estimates Using a Different Sequencing Read Alignment

For 24 C-to-U RNA editing events that were identified in four tumor samples, the RNA editing levels were redetermined after realigning RNA sequencing data. For realignment, samtools fastq utility was used to extract read data from the aligned data that had been generated by GDC. Only paired read sequences with concordant mapping were extracted. The sequences were then realigned using Subread [24] subjunc aligner software (version 1.5.0-p1) against version 81 of the Ensembl GRCh38 reference genome. With the realigned data, C-to-U RNA editing level was estimated as the ratio of variant to sum of variant and reference base calls of phred33 quality ≥ 20.

4.5. Collection of Clinical and Pathologic Information of TCGA Subjects

Tumor histologic scores for mitotic counts, nuclear pleomorphism, and tubular formation as per the Nottingham system were collected through manual examination of pathology reports with the aid of data and the search engine in the Text Information Extraction System (TIES) Cancer Research Network [39], a federated network of four cancer centers in the USA. All three scores could be recorded for 591 of the total 1093 TCGA-BRCA patients; at least one or at least two types of scores, respectively, could be recorded for 669 and 625 patients. TCGA Pan-Cancer Clinical Data Resource [27], the standardized and curated dataset for survival endpoints for the TCGA project, was the source for survival data for the TCGA-BRCA patients. Information for other clinical and pathologic variables was obtained from TCGA through the cBio Cancer Genomics Portal [40] in late 2017.

4.6. Quantification of TCGA Tumor Genome Features

Scores for intratumoral clonal heterogeneity, as well as tumor aneuploidy, copy number changes, homologous recombination deficiency, and neoantigen load were collated from a TCGA Pan-Cancer Atlas study [41]. To count all single nucleotide substitution mutations and obtain a minimum estimate of counts of mutations attributable to DNA deaminase activity of APOBEC enzymes as per the P-MACD method [12], publicly available mutation annotation format (MAF) files from TCGA were analyzed with the maftools [42] Bioconductor package (version 1.8.0).

4.7. Generation of TCGA Tumor Gene Expression Dataset

Gene-level counts of mapped RNA sequencing reads were obtained from the GDC portal with the TCGAbiolinks (version 2.5.9) Bioconductor package [43] in mid-2018. The counts were generated with HTSeq software (version 0.6.1-p1) with Gencode 22 annotation and Ensembl gene identifiers. To generate gene expression values in transcripts per million (TPM) units, counts were divided by total exon length of the identifiers. Corresponding Human Genome Organization (HUGO) gene symbols for the gene identifiers were identified with biomaRt (version 2.38.0) Bioconductor package [44] and Ensembl human genome annotation (version 92). In the case of multiple identifiers with the same symbol, TPM values were aggregated by addition. Gene identifiers without an identifiable gene symbol were removed. For $\log_2$-transformation, TPM values were padded with 0.05.

4.8. Analyses of Gene Expression Data

$\log_2$-transformed TPM values were used. To identify genes that were differentially expressed between two groups, the empirical Bayes-moderated *t*-test in limma (version 3.38.2) Bioconductor package [45] was used with threshold values of 0.05 and 2, respectively, for *p*-value and absolute

fold change value. To determine the enrichment of expression of the 50 gene sets of the Molecular Signatures Database (MSigDB) Hallmark collection (version 6.2), the gene set variation analysis (GSVA) method implemented in the GSVA Bioconductor package (version 1.30.0) was used [25]. Two-group comparison of GSVA scores were with limma's *t*-test. To control false discovery in multi-testings of both differential gene expression and GSVA analyses, the Benjamini–Hochberg method was used to adjust raw *p*-values.

4.9. Measurements of TCGA Tumor Immune Microenvironment

Immune cytolytic activity in tumors was quantified from perforin 1 (*PRF1*) and granzyme A (*GZMA*) gene expression values as the CYT score [46]. Relative fractions of 22 types of tumor-infiltrating immune cells were inferred from tumor gene expression data (TPM values) with the CIBERSORT method. The online tool at https://cibersort.stanford.edu (jar version 1.06) was used with LM22 signatures and 100 permutations [26]. Values for certain types of cells were aggregated to generate relative fraction values for more comprehensive cellular classes such as lymphocytes [41]. The immune cell composition estimates for tumors for which the CIBERSORT *p*-value was ≥0.05 (about 10% of cases) were excluded from analysis. Other immune-related measurements of tumors such as B and T cell receptor diversities were from the study of Thorsson et al. [41].

4.10. Other

Unless noted otherwise, default option values were used for all software, group comparisons of continuous and categorical variables were with standard *t*-tests and Fisher's exact tests, respectively, and *p*-values below 0.05 were considered significant. Analyses and visualization of data were performed using Prism (version 7.0d; GraphPad Software®, San Diego, CA, USA) and R (version 3.5+). The survival package (version 2.44) for R was used for survival analyses with Cox regression and log-rank tests for the determination of HR and *p*-values. Statistical overrepresentation tests of gene lists with Fisher's exact test and false discovery calculation were conducted online at https://pantherdb.org with the complete GO biological process annotation dataset of PANTHER [47] database (version 14.1).

Supplementary Materials: The following are available online at http://www.mdpi.com/1422-0067/20/22/5621/s1, Figure S1. Estimation of RNA editing levels using RNA and DNA sequencing data. Figure S2. Example views of mapped sequencing read alignment data illustrating RNA editing events. Figure S3. Correlation of RNA editing levels determined with originally aligned and realigned RNA sequencing data. Figure S4. *APOBEC3* gene expression differences between tumors with editing absent and present for four editing sites. Table S1. Genome positions that were examined for RNA editing. Table S2. Characteristics of determinations made for the estimation of RNA editing levels. Table S3. Fifty most commonly edited sites among tumors. Table S4. Genes with significant differential expression between editing-high and -low tumors.

Author Contributions: Conceptualization, K.T. and S.K.P.; methodology, M.A. and S.K.P.; formal analysis, M.A. and S.K.P.; investigation, M.A. and S.K.P.; writing—original draft preparation, M.A. and S.K.P.; writing—review and editing, K.T. and T.I.; visualization, M.A. and S.K.P.; supervision, K.T. and S.K.P.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bransteitter, R.; Prochnow, C.; Chen, X.S. The current structural and functional understanding of APOBEC deaminases. *Cell Mol. Life Sci.* **2009**, *66*, 3137–3147. [CrossRef] [PubMed]
2. Fritz, E.L.; Rosenberg, B.R.; Lay, K.; Mihailovic, A.; Tuschl, T.; Papavasiliou, F.N. A comprehensive analysis of the effects of the deaminase AID on the transcriptome and methylome of activated B cells. *Nat. Immunol.* **2013**, *14*, 749–755. [CrossRef] [PubMed]
3. Smith, H.C.; Bennett, R.P.; Kizilyer, A.; McDougall, W.M.; Prohaska, K.M. Functions and regulation of the APOBEC family of proteins. *Semin Cell Dev. Biol.* **2012**, *23*, 258–268. [CrossRef] [PubMed]

4. Gu, T.; Buaas, F.W.; Simons, A.K.; Ackert-Bicknell, C.L.; Braun, R.E.; Hibbs, M.A. Canonical A-to-I and C-to-U RNA editing is enriched at 3'UTRs and microRNA target sites in multiple mouse tissues. *Plos ONE* **2012**, *7*, e33720. [CrossRef] [PubMed]

5. Peng, Z.; Cheng, Y.; Tan, B.C.; Kang, L.; Tian, Z.; Zhu, Y.; Zhang, W.; Liang, Y.; Hu, X.; Tan, X.; et al. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat. Biotechnol.* **2012**, *30*, 253–260. [CrossRef] [PubMed]

6. Bahn, J.H.; Lee, J.H.; Li, G.; Greer, C.; Peng, G.; Xiao, X. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res.* **2012**, *22*, 142–150. [CrossRef] [PubMed]

7. Harris, R.S.; Dudley, J.P. APOBECs and virus restriction. *Virology* **2015**, *479–480*, 131–145. [CrossRef]

8. Refsland, E.W.; Harris, R.S. The APOBEC3 family of retroelement restriction factors. *Curr Top. Microbiol. Immunol.* **2013**, *371*, 1–27.

9. Burns, M.B.; Temiz, N.A.; Harris, R.S. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Genet.* **2013**, *45*, 977–983. [CrossRef]

10. Alexandrov, L.B.; Nik-Zainal, S.; Wedge, D.C.; Aparicio, S.A.; Behjati, S.; Biankin, A.V.; Bignell, G.R.; Bolli, N.; Borg, A.; Borresen-Dale, A.L.; et al. Signatures of mutational processes in human cancer. *Nature* **2013**, *500*, 415–421. [CrossRef]

11. Leonard, B.; Hart, S.N.; Burns, M.B.; Carpenter, M.A.; Temiz, N.A.; Rathore, A.; Vogel, R.I.; Nikas, J.B.; Law, E.K.; Brown, W.L.; et al. APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. *Cancer Res.* **2013**, *73*, 7222–7231. [CrossRef] [PubMed]

12. Roberts, S.A.; Lawrence, M.S.; Klimczak, L.J.; Grimm, S.A.; Fargo, D.; Stojanov, P.; Kiezun, A.; Kryukov, G.V.; Carter, S.L.; Saksena, G.; et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* **2013**, *45*, 970–976. [CrossRef]

13. Nik-Zainal, S.; Wedge, D.C.; Alexandrov, L.B.; Petljak, M.; Butler, A.P.; Bolli, N.; Davies, H.R.; Knappskog, S.; Martin, S.; Papaemmanuil, E.; et al. Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat. Genet.* **2014**, *46*, 487–491. [CrossRef] [PubMed]

14. Caval, V.; Suspene, R.; Shapira, M.; Vartanian, J.P.; Wain-Hobson, S. A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3'UTR enhances chromosomal DNA damage. *Nat. Commun.* **2014**, *5*, 5129. [CrossRef] [PubMed]

15. Prohaska, K.M.; Bennett, R.P.; Salter, J.D.; Smith, H.C. The multifaceted roles of RNA binding in APOBEC cytidine deaminase functions. *Wiley Interdiscip Rev. RNA* **2014**. [CrossRef]

16. Sharma, S.; Patnaik, S.K.; Kemer, Z.; Baysal, B.E. Transient overexpression of exogenous APOBEC3A causes C-to-U RNA editing of thousands of genes. *RNA Biol.* **2017**, *14*, 603–610. [CrossRef]

17. Sharma, S.; Patnaik, S.K.; Taggart, R.T.; Baysal, B.E. The double-domain cytidine deaminase APOBEC3G is a cellular site-specific RNA editing enzyme. *Sci. Rep.* **2016**, *6*, 39100. [CrossRef]

18. Sharma, S.; Patnaik, S.K.; Taggart, R.T.; Kannisto, E.D.; Enriquez, S.M.; Gollnick, P.; Baysal, B.E. APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. *Nat. Commun.* **2015**, *6*, 6881. [CrossRef]

19. Patnaik, S.K.; Kannisto, E.D. APOBEC3B is a new RNA editing enzyme. In Proceedings of the RNA 2016, Annual Meeting of RNA Society, Kyoto, Japan, 28 June–2 July 2016.

20. Cescon, D.W.; Haibe-Kains, B. DNA replication stress: A source of APOBEC3B expression in breast cancer. *Genome Biol.* **2016**, *17*, 202. [CrossRef]

21. Trevino, V. Integrative genomic analysis identifies associations of molecular alterations to APOBEC and BRCA1/2 mutational signatures in breast cancer. *Mol. Genet. Genom. Med.* **2019**, *7*, e810. [CrossRef] [PubMed]

22. Harris, R.S. Molecular mechanism and clinical impact of APOBEC3B-catalyzed mutagenesis in breast cancer. *Breast Cancer Res.* **2015**, *17*, 8. [CrossRef] [PubMed]

23. Sieuwerts, A.M.; Doebar, S.C.; de Weerd, V.; Verhoef, E.I.; Beauford, C.M.; Agahozo, M.C.; Martens, J.W.M.; van Deurzen, C.H.M. APOBEC3B Gene Expression in Ductal Carcinoma In Situ and Synchronous Invasive Breast Cancer. *Cancers (Basel)* **2019**, *11*, 1062. [CrossRef] [PubMed]

24. Liao, Y.; Smyth, G.K.; Shi, W. The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **2013**, *41*, e108. [CrossRef]

25. Hanzelmann, S.; Castelo, R.; Guinney, J. GSVA: Gene set variation analysis for microarray and RNA-seq data. *Bmc Bioinform.* **2013**, *14*, 7. [CrossRef]

26. Chen, B.; Khodadoust, M.S.; Liu, C.L.; Newman, A.M.; Alizadeh, A.A. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol. Biol.* **2018**, *1711*, 243–259.

27. Liu, J.; Lichtenberg, T.; Hoadley, K.A.; Poisson, L.M.; Lazar, A.J.; Cherniack, A.D.; Kovatich, A.J.; Benz, C.C.; Levine, D.A.; Lee, A.V.; et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **2018**, *173*, 400–416.e11. [CrossRef]

28. Hong, H.; An, O.; Chan, T.H.M.; Ng, V.H.E.; Kwok, H.S.; Lin, J.S.; Qi, L.; Han, J.; Tay, D.J.T.; Tang, S.J.; et al. Bidirectional regulation of adenosine-to-inosine (A-to-I) RNA editing by DEAH box helicase 9 (DHX9) in cancer. *Nucleic Acids Res.* **2018**, *46*, 7953–7969. [CrossRef]

29. Kung, C.P.; Maggi, L.B., Jr.; Weber, J.D. The Role of RNA Editing in Cancer Development and Metabolic Disorders. *Front. Endocrinol. (Lausanne)* **2018**, *9*, 762. [CrossRef]

30. Ruan, H.; Han, L. Regulatory RNA Editing Sites in Cancer: Prediction and Beyond. *EBioMedicine* **2018**, *27*, 7–8. [CrossRef]

31. Chigaev, M.; Yu, H.; Samuels, D.C.; Sheng, Q.; Oyebamiji, O.; Ness, S.; Yue, W.; Zhao, Y.Y.; Guo, Y. Genomic Positional Dissection of RNA Editomes in Tumor and Normal Samples. *Front. Genet.* **2019**, *10*, 211. [CrossRef] [PubMed]

32. Peng, X.; Xu, X.; Wang, Y.; Hawke, D.H.; Yu, S.; Han, L.; Zhou, Z.; Mojumdar, K.; Jeong, K.J.; Labrie, M.; et al. A-to-I RNA Editing Contributes to Proteomic Diversity in Cancer. *Cancer Cell* **2018**, *33*, 817–828.e7. [CrossRef] [PubMed]

33. Sagredo, E.A.; Blanco, A.; Sagredo, A.I.; Perez, P.; Sepulveda-Hermosilla, G.; Morales, F.; Muller, B.; Verdugo, R.; Marcelain, K.; Harismendy, O.; et al. ADAR1-mediated RNA-editing of 3'UTRs in breast cancer. *Biol. Res.* **2018**, *51*, 36. [CrossRef] [PubMed]

34. Saraconi, G.; Severi, F.; Sala, C.; Mattiuz, G.; Conticello, S.G. The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. *Genome Biol.* **2014**, *15*, 417. [CrossRef]

35. Swanton, C.; McGranahan, N.; Starrett, G.J.; Harris, R.S. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discov* **2015**, *5*, 704–712. [CrossRef]

36. Gao, G.F.; Parker, J.S.; Reynolds, S.M.; Silva, T.C.; Wang, L.B.; Zhou, W.; Akbani, R.; Bailey, M.; Balu, S.; Berman, B.P.; et al. Before and After: Comparison of Legacy and Harmonized TCGA Genomic Data Commons' Data. *Cell Syst.* **2019**, *9*, 24–34.e10. [CrossRef]

37. Wang, T.; Liu, J.; Shen, L.; Tonti-Filippini, J.; Zhu, Y.; Jia, H.; Lister, R.; Whitaker, J.W.; Ecker, J.R.; Millar, A.H.; et al. STAR: An integrated solution to management and visualization of sequencing data. *Bioinformatics* **2013**, *29*, 3204–3210. [CrossRef]

38. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef]

39. Jacobson, R.S.; Becich, M.J.; Bollag, R.J.; Chavan, G.; Corrigan, J.; Dhir, R.; Feldman, M.D.; Gaudioso, C.; Legowski, E.; Maihle, N.J.; et al. A Federated Network for Translational Cancer Research Using Clinical Data and Biospecimens. *Cancer Res.* **2015**, *75*, 5194–5201. [CrossRef]

40. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2012**, *2*, 401–404. [CrossRef]

41. Thorsson, V.; Gibbs, D.L.; Brown, S.D.; Wolf, D.; Bortone, D.S.; Ou Yang, T.H.; Porta-Pardo, E.; Gao, G.F.; Plaisier, C.L.; Eddy, J.A.; et al. The Immune Landscape of Cancer. *Immunity* **2018**, *48*, 812–830.e14. [CrossRef] [PubMed]

42. Mayakonda, A.; Lin, D.C.; Assenov, Y.; Plass, C.; Koeffler, H.P. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **2018**, *28*, 1747–1756. [CrossRef] [PubMed]

43. Colaprico, A.; Silva, T.C.; Olsen, C.; Garofano, L.; Cava, C.; Garolini, D.; Sabedot, T.S.; Malta, T.M.; Pagnotta, S.M.; Castiglioni, I.; et al. TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **2016**, *44*, e71. [CrossRef] [PubMed]

44. Durinck, S.; Moreau, Y.; Kasprzyk, A.; Davis, S.; De Moor, B.; Brazma, A.; Huber, W. BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* **2005**, *21*, 3439–3440. [CrossRef]

45. Smyth, G. Limma: Linear models for microarray data. In *Bioinformatics and Computational Biology Solutions using R and Bioconductor*; Gentleman, R., Carey, V.J., Huber, W., Dudoit, S., Irizarry, R.A., Eds.; Springer: New York, NY, USA, 2005; pp. 397–420.

46. Rooney, M.S.; Shukla, S.A.; Wu, C.J.; Getz, G.; Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **2015**, *160*, 48–61. [CrossRef]

47. Mi, H.; Thomas, P. PANTHER pathway: An ontology-based pathway database coupled with data analysis tools. *Methods Mol. Biol.* **2009**, *563*, 123–140.