



OPEN The role of distinct *APOBEC/ADAR* mRNA levels in mutational signatures linked to aging and ultraviolet radiation

Ahmadreza Niavarani

The *APOBEC/AID* family is known for its mutator activity, and recent evidence also supports the potential impact of *ADARs*. Furthermore, the mutator impacts of *APOBEC/ADAR* mutations have not yet been investigated. Assessment of pancancer TCGA exomes identified enriched somatic variants among exomes with nonsynonymous *APOBEC1*, *APOBEC3B*, *APOBEC3C*, *ADAR*, and *ADARB1* mutations, compared to exomes with synonymous ones. Principal component (PC) analysis reduced the number of potential players to eight in cancer exomes/genomes, and to five in cancer types. Multivariate regression analysis was used to assess the impact of the PCs on each COSMIC mutational signature among pancancer exomes/genomes and particular cancers, identifying several novel links, including SBS17b, SBS18, and ID7 mainly determined by *APOBEC1* mRNA levels; SBS40, ID1, and ID2 by age; SBS3 and SBS16 by *APOBEC3A/APOBEC3B* mRNA levels; ID5 and DBS9 by DNA repair/replication (DRR) defects; and SBS7a-d, SBS38, ID4, ID8, ID13, and DBS1 by ultraviolet (UV) radiation/*ADARB1* mRNA levels. *APOBEC/ADAR* mutations appeared to potentiate the impact of DRR defects on several mutational signatures, and some factors seemed to inversely affect certain signatures. These findings potentially implicate certain *APOBEC/ADAR* mutations/mRNA levels in distinct mutational signatures, particularly *APOBEC1* mRNA levels in aging-related signatures and *ADARB1* mRNA levels in UV radiation-related signatures.

Keywords *APOBEC/ADAR* mRNA levels, *APOBEC/ADAR* mutation, Cancer mutation, Mutational signature

Abbreviations

<i>A1</i>	<i>APOBEC1</i>
<i>A2</i>	<i>APOBEC2</i>
<i>A3A</i>	<i>APOBEC3A</i>
<i>A3B</i>	<i>APOBEC3B</i>
<i>A3C</i>	<i>APOBEC3C</i>
<i>A3D</i>	<i>APOBEC3D</i>
<i>A3F</i>	<i>APOBEC3F</i>
<i>A3G</i>	<i>APOBEC3G</i>
<i>A3H</i>	<i>APOBEC3H</i>
<i>A4</i>	<i>APOBEC4</i>
ACC	Adrenocortical carcinoma
ADAR	Adenosine deaminase acting on RNA
AID	Activation-induced deaminase
<i>APOBEC</i>	Apolipoprotein B mRNA-editing enzyme catalytic polypeptide
BER	Base excision repair
BLCA	Bladder carcinoma
BRCA	Breast carcinoma
CESC	Cervical/endocervical cancer
CHOL	Cholangiocarcinoma

Digestive Oncology Research Center, Digestive Disease Research Institute, Tehran University of Medical Sciences, Tehran, Iran. email: arniavarani@gmail.com

COADREAD	Colorectal adenocarcinoma
DBS	Double-base substitution
DLBC	Diffuse large B-cell lymphoma
DNA-Rep	DNA replication pathway
DRR	DNA repair/replication
eID	Exome indel
eSBS	Exome single-base substitution
ESCA	Esophageal carcinoma
FATLS	Fanconi anemia-translesion synthesis
gDBS	Genome double-base substitution
gID	Genome indel
gSBS	Genome single-base substitution
GBM	Glioblastoma multiforme
HNSC	Head and neck squamous cell carcinoma
HR	Homologous recombination
ID	Insertion/deletion; indel
KICH	Kidney chromophobe
KIRC	Kidney clear cell carcinoma
KIRP	Kidney papillary cell carcinoma
LAML	Acute myeloid leukemia
LGG	Low-grade glioma
LIHC	Hepatocellular carcinoma
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
MESO	Mesothelioma
MMR	Mismatch repair
NER	Nucleotide excision repair
NHEJ	Nonhomologous end-joining
OV	Ovarian serous cystadenocarcinoma
PAAD	Pancreatic adenocarcinoma
PCPG	Pheochromocytoma/paraganglioma
PRAD	Prostate adenocarcinoma
SARC	Sarcoma
SBS	Single-base substitution
SKCM	Skin melanoma
SNV	Single-nucleotide variant
STAD	Stomach adenocarcinoma
TGCT	Testicular germ cell tumor
THCA	Thyroid carcinoma
THYM	Thymoma
UCEC	Uterine corpus endometrial carcinoma
UCS	Uterine carcinosarcoma
UVM	Uveal melanoma
UV	Ultraviolet

The APOBEC/AID (apolipoprotein B mRNA editing enzyme, catalytic polypeptide/activation-induced cytidine) family of cytidine deaminases, which are involved in a wide range of physiological and developmental activities, whether in DNA (reviewed in¹) or RNA^{2,3}, have also been implicated in mutagenesis across various cancers⁴⁻⁷. Two single base substitution mutational signatures with predominant C>T (SBS2) and C>G (SBS13) variants have been attributed to the APOBEC/AID family^{5,7}, in addition to potentially one double base substitution signature with predominant CC>NN (DBS11) variants⁸. While both APOBEC3B (A3B)^{6,9} and APOBEC3A (A3A)^{9,10} are strongly linked to APOBEC mutational signatures, the implication of other family members with nuclear distribution, including APOBEC1 (A1)¹¹, APOBEC3C (A3C)¹², APOBEC3F (A3F)¹³, APOBEC3H (A3H)¹⁴, and AID¹⁵, cannot be excluded. AID has long been implicated in B-cell malignancies^{16,17}. Although much attention has been given to the high mRNA levels of *APOBEC/AICDA* genes^{6,18} and their copy number⁹ or single nucleotide variations¹⁹, these might not fully explain all APOBEC mutagenic impacts observed in cancer genomes. For instance, Kanke et al. reported some tumors of the breast, ovary, and uterus with a predominant SBS2, which did not show high *APOBEC* mRNA levels²⁰. It seems that the implication of other factors, including *APOBEC* somatic gene mutations, has been simply overlooked. On the other hand, there is evidence showing that known adenosine deaminases acting on RNA (ADARs) may also act on DNA. These include the DNA mutator activity of ADAR1 in *MYCC* and a class switch recombination region (Ig- μ)²¹, deamination of adenosine at dA-C mismatches of the DNA–RNA hybrids by both ADAR1 and ADAR2²²; and some indirect evidence^{23,24}. However, the role of *APOBEC/ADAR* mutations has not yet been clarified in cancer. Our pancancer analysis shows that these genes are themselves subject to somatic mutations, although at a low frequency.

A preliminary analysis showed that pancancer exomes with any of the fourteen *APOBEC/ADAR* genes mutated showed significantly more single nucleotide variants (SNVs) and insertions/deletions (indels). Since this could suggest either a “driver” or a “passenger” role for *APOBEC/ADAR* mutations in cancer hypermutation, the burden of genomic variants was assessed in exomes with nonsynonymous versus synonymous *APOBEC/*

ADAR mutations to check their true driver impact, showing that mutations in *A1*, *A3B*, *A3C*, and *ADAR* are correlated with indel variants, while *ADARB1* mutations are correlated with certain SNVs. Follow-up analyses estimated the roles of *APOBEC/ADAR* mutations and mRNA levels in mutational signatures among various cancer types, as adjusted for defects in DNA repair/replication (DRR) pathways.

Results

Assessment of 10,295 tumor exomes identified 3,600,963 unique somatic variants, of which 3,427,680 (95.2%) were found to be SNVs and 173,109 (4.8%) were of the indel type. Approximately 2,710,172 (75.3%) of the somatic variants were found to be nonsynonymous variants occurring within coding regions, and the remaining 24.7% (890,791) were either intronic or synonymous exonic, herein called synonymous mutations. Collectively, 874 exomes (8.5%) showed at least one nonsynonymous somatic mutation in one or more *APOBEC/ADAR* genes, including *APOBEC1*, *APOBEC2*, *APOBEC3A*, *APOBEC3B*, *APOBEC3C*, *APOBEC3D*, *APOBEC3F*, *APOBEC3G*, *APOBEC3H*, *APOBEC4*, *AICDA*, *ADAR*, *ADARB1*, and *ADARB2*, with 188 exomes (21.5%) showing more than one mutated gene. The mean number of total *APOBEC/ADAR* mutations varied between 0 per exome in PCPG, TGCT, THYM, and UVM to 0.36 and 0.62 in SKCM and UCEC, respectively (Supplementary Table S1). The most common nonsynonymous mutations were seen in *ADARB2* (198; 1.92% per exome), whereas *A3A* (40; 0.39% per exome) was found to be the least mutated. In general, the *ADAR* family tended to have a higher mean number of mutations (178; 1.73% per exome) than the *APOBEC* family (68; 0.66% per exome). None of the nonsynonymous *APOBEC/ADAR* mutations reported by the TCGA-MC3 catalog were annotated as having a *low Impact*, 68 (5.3%) were classified as having a *modifier Impact*, and 1,204 (94.7%) were classified as having a *moderate/high Impact*. Similarly, none of the nonsynonymous DRR mutations were annotated as having a *low Impact*, 562 (2.8%) were classified as having a *modifier Impact*, and 19,292 (97.2%) were classified as having a *moderate/high Impact*.

Essentially, all *APOBEC/ADAR*-mutated exomes showed a higher number of somatic variants

The mean SNV burden was found to be enriched in those exomes with any *APOBEC/ADAR* mutations compared to those with wild-type ones, varying between 13.9-fold for *ADARB2*-mutated exomes to 25.5-fold for *A2*-mutated ones (Supplementary Table S2). Likewise, the mean indel burden was enriched in essentially all *APOBEC/ADAR*-mutated exomes compared to their wild-type counterparts, varying between 5.9-fold for *AICDA*-mutated exomes and 11.9-fold for *A1*-mutated ones.

T > N SNVs were correlated with *ADARB1* mutations, as were certain indels with *A1*, *A3B*, *A3C*, and *ADAR* mutations

In a more stringent assessment, the mean numbers of genomic T > N, C > N, Ind-T, and Ind-C variants were compared in exomes with nonsynonymous versus synonymous *APOBEC/ADAR* mutations. Assessment of the SNVs showed that the mean number of T > N variants was enriched 2.8-fold in nonsynonymous *ADARB1* mutant exomes (Supplementary Table S3). On the other hand, the mean number of Ind-T variants was enriched 6.0-fold in nonsynonymous *A1* mutant exomes, 5.0-fold in nonsynonymous *A3B* mutants, 22.2 times higher in nonsynonymous *A3C* mutants, and 2.3-fold in nonsynonymous *ADAR* mutants, and the mean number of Ind-C variants was 17.2-fold in nonsynonymous *A3C* mutant exomes and 2.1-fold in nonsynonymous *ADAR* mutant exomes (Supplementary Table S3).

Assessing the impacts of *APOBEC/ADAR* mutations and mRNA levels on cancer mutational signatures

Regression followed by clustering analysis was performed using 10,126 samples with WES- and/or WGS-identified mutational signatures, identifying that many of the potential endogenous factors closely co-clustered with each other, including all DRR defects and *APOBEC/ADAR* mutations; *A3A* and *A3B* mRNA levels co-clustered with each other; and *A3C*, *A3D*, *A3F*, *A3G*, and *A3H* mRNA levels co-clustered with each other (Fig. 1). Hence, it was next attempted to reduce the number of potentially implicated covariates using principal component (PC) analysis. Assessment of 37 potential covariates, including age and UV factor, along with *APOBEC/ADAR* factors and DRR defects, reduced the number of potential covariates to eight PCs (explaining 49.5% of the variability) in pancancer exome (9,493 samples)/genome (773 samples) analyses (Supplementary Table S4). These included (1) seven DRR defects and mutations in *ADARB1*, *ADAR*, and *A3F*; (2) *A3C*, *A3D*, *A3F*, *A3G*, and *A3H* (*A3C-H*) mRNA levels; (3) UV factor and *ADARB1/A3C/ADAR* mRNA levels; (4) *A3A* and *A3B* mRNA levels; (5) age and *A1* mRNA levels; (6) *A3D* and *A3G* mutations; (7) *A3B* mutation; and (8) *A4* and *AICDA* mRNA levels. A similar analysis of the mean exome/genome covariates among 33 cancer types resulted in five PCs (explaining 77.6% of the variability; Supplementary Table S5), including (1) seven DRR defects, all *APOBEC/ADAR* mutations, and *A4* mRNA levels; (2) *A3C-H* and *AICDA* mRNA levels; (3) UV factor and *ADARB1/A3C/ADARB2* mRNA levels; (4) *A3A* and *A3B* mRNA levels; and (5) age and *A1/ADAR* mRNA levels. Multivariate correlations of the pancancer exome and genome signatures with the aforementioned PCs are shown in Figs. 2 and 3, respectively, and clustering analyses of the positive correlations are illustrated in Supplementary Figure S1. Furthermore, the correlations of the exome and genome signatures with PCs across 33 particular cancers are depicted in Figs. 4 and 5, respectively, as well as Supplementary Figures S2 and S3, respectively. Collectively, these findings can be summarized below, as illustrated in Fig. 6.

UV-*ADARB1* mRNA cluster

Seven signatures are known to be UV radiation-related, which were also found to be correlated with PC3 representing UV factor and *ADARB1/A3C/ADAR* mRNA levels (Fig. 6; Supplementary Figure S4). These include

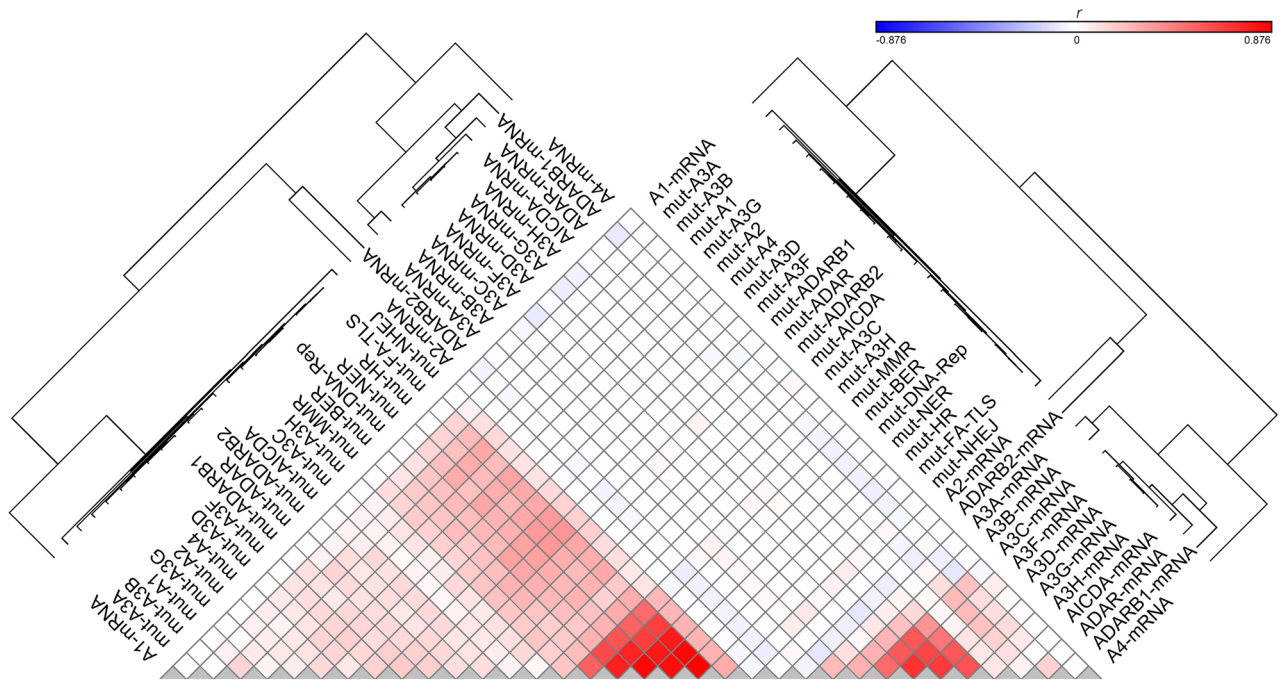


Figure 1. Clustering analysis of the correlation (r) between potential covariates implicated in cancer mutagenesis, including *APOBEC/ADAR* mutations and mRNA levels as well as DNA repair/replication (DRR) defects. Pearson regression analysis was used in order to estimate the correlations. DRR defects were very closely clustered with each other, followed by *APOBEC/ADAR* mutations. On the other hand, *A3A* and *A3B* mRNA levels were closely clustered to each other, and *A3C*, *A3D*, *A3E*, *A3F*, *A3G*, and *A3H* mRNA levels were also closely co-clustered.

SBS7a,b,c,d, SBS38, ID13, and DBS1. Furthermore, the unknown genome ID4 (gID4) and the so-called artifact exome signatures SBS45 (eSBS45) and SBS49 (eSBS49) were mainly determined by this component. Not surprisingly, these signatures were primarily associated with significant DRR defects as well. More detailed analysis showed that SBS7a-d, SBS38, ID13, and DBS1 were merely or mainly determined by UV factor, while ID4 was determined by only the *ADARB1* mRNA level, and ID8 was determined by both *ADARB1* and *ADAR* mRNA.

A3A/B mRNA cluster

Pancancer analysis across both samples and cancer types identified both exome and genome signatures SBS2 and SBS13 highly determined by *A3A/A3B* (*A3A/B*) mRNA levels (Fig. 6; Supplementary Figure S5). These findings were consistent with analyses of the exome mutational signatures across particular cancers, showing that eSBS2 was determined by *A3A/B* mRNA (THYM, OV, ESCA, THCA, STAD, HNSC) as well as *A3C-H* mRNA (THYM, ACC, HNSC, THCA, BLCA, BRCA), age-*A1* mRNA (KIRP, THCA, HNSC, BRCA, LUAD), and DRR defects in various cancers (Fig. 4). eSBS13 was also determined by *A3A/B* mRNA (THYM, OV, DLBC, ESCA, THCA, STAD, HNSC, UCEC, SKCM, LGG, LUAD), *A3C-H* mRNA (THYM, HNSC, THCA, BLCA, PAAD, ACC, OV), age-*A1* mRNA (OV, THCA, HNSC, BRCA, LUAD), and DRR defects in various cancers. However, the genome signatures SBS2 (gSBS2) and SBS13 (gSBS13) were found to be determined solely by DRR defects in particular cancers (Fig. 5). As expected, gDBS11 was also determined by the *A3A/B* mRNA level. Furthermore, HR-related eSBS3 and alcohol-related gSBS16 were found to be highly determined by *A3A/B* mRNA levels.

Age-APOBEC1 (A1) mRNA cluster

The so-called 5FU-related eSBS17b and the ROS-related gSBS18 were found to be determined by age-*A1* mRNA level, as did the unknown SBS40 (Fig. 6; Supplementary Figure S6). Assessment of individual cancers (CESC, GBM, KIRC, LIHC, PRAD) across exomes; KIRP, KIRC, COADREAD across genomes) corroborated the correlation of SBS40 with age-*A1* mRNA level. A combined impact of the DRR-*ADARB1/ADAR/A3F* mutations and age-*A1* mRNA levels (BLCA, BRCA, CESC, COADREAD, ESCA, KICH, LGG, OV, STAD, THCA, UCEC across exomes; BRCA, CESC, GBM, KICH, SARC, THCA across genomes) was seen in aging-related SBS1. Likewise, the combined impact of the DRR-*ADARB1/ADAR/A3F* mutations and age-*A1* mRNA levels was also seen for indel signatures ID1, ID2, and ID7 in pancancer analyses as well as in certain cancers. A more detailed analysis of PC1 covariates identified SBS40, ID1, and ID2 to be determined mainly by age, while SBS17b, SBS18, and ID7 were determined mainly by *A1* mRNA level.

Defective DRR cluster

This constituted the largest cluster by far, contributing to many mutational signatures. In addition to SBS21 and SBS26, which are currently attributed to some MMR defects, both unknown ID5 and DBS9 were also determined

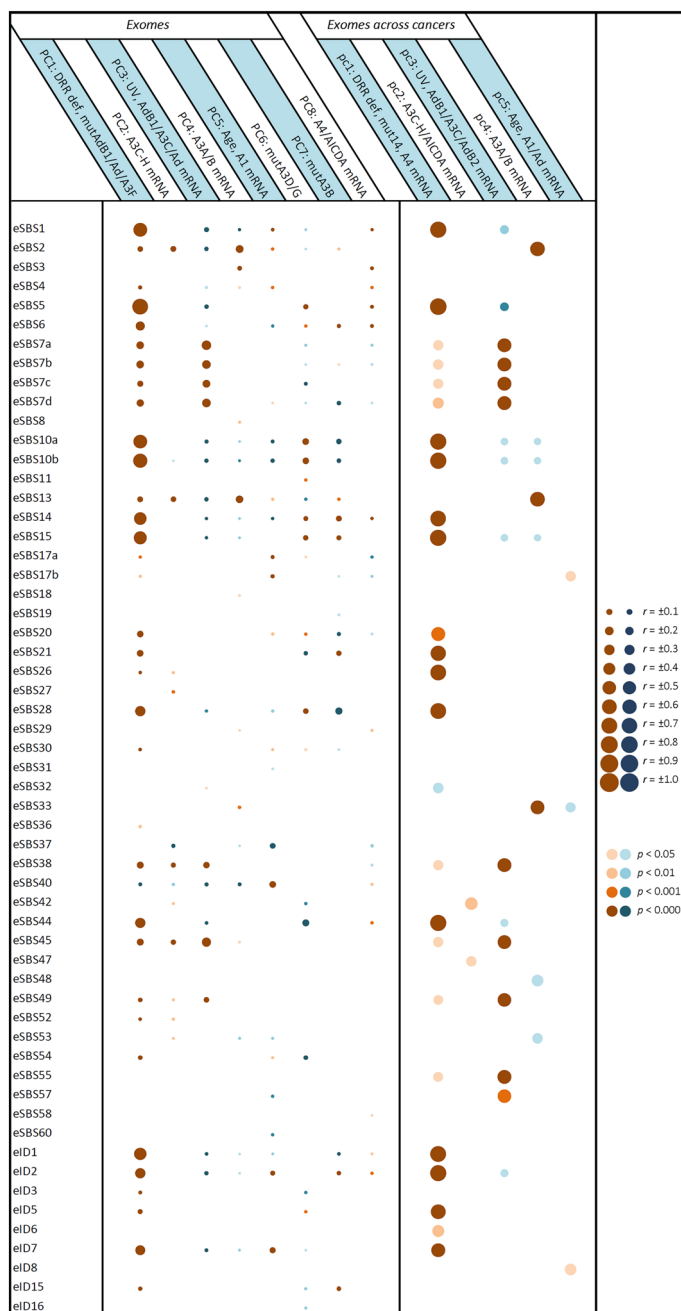


Figure 2. Correlation of pancancer exome mutational signatures with principal components potentially implicated in cancer mutagenesis. Multivariate forward linear regression analysis was performed using eight/five varying PCs. Left; correlation across exome samples regardless of the cancer types. Right; correlation of mean exome values across 33 cancers. Brown and blue bubbles indicate positive and negative impacts, respectively, with darker shadows indicative of more significant correlations. The main determinant PCs were at least partially concordant among exomes and cancers in the following mutational signatures: eSBS1, eSBS2, eSBS5, eSBS7a-d, eSBS10a,b, eSBS13, eSBS14, eSBS15, eSBS17b, eSBS20, eSBS21, eSBS26, eSBS28, eSBS33, eSBS38, eSBS44, eSBS45, eSBS49, eID1, eID2, eID5, and eID7. A1, *APOBEC1*; A2, *APOBEC2*; A3, *APOBEC3*; A4, *APOBEC4*; Ad, *ADAR*; AdB1, *ADARB1*; AdB2, *ADARB2*; mut14, mutation of fourteen *APOBEC/ADAR* genes. The “e” prefix denotes the exome signatures.

by DRR defects across exomes and/or genomes (Fig. 6). One subcluster consisted of signatures that were also determined by tobacco exposure, including the known signatures SBS4, ID3, and DBS2, as well as the so-called aging-related SBS5. Another subcluster consisted of SBS6, SBS10a,b, SBS14, SBS15, SBS20, SBS28, SBS44, ID1, ID2, and DBS10, which are currently known or predicted to be related to DRR defects, but they were also found

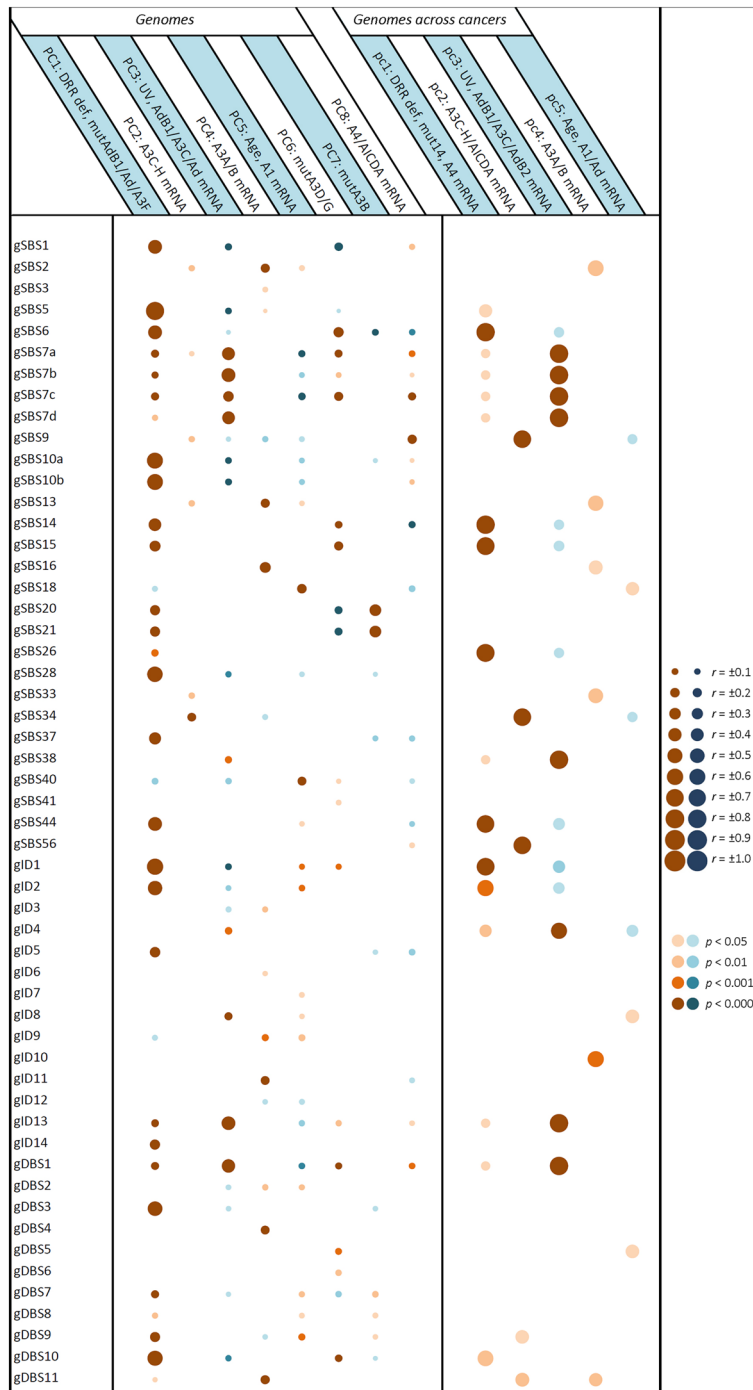


Figure 3. Correlation of pancancer genome mutational signatures with principal components potentially implicated in cancer mutagenesis. Multivariate forward linear regression analysis was performed using eight/five varying PCs. Left; correlation across genome samples regardless of the cancer types. Right; correlation of mean genome values across 33 cancers. Brown and blue bubbles indicate positive and negative impacts, respectively, with darker shadows indicative of more significant correlations. The main determinant PCs were at least partially concordant among genomes and cancers in the following mutational signatures: gSBS2, gSBS5, gSBS6, gSBS7a-d, gSBS9, gSBS13, gSBS14, gSBS15, gSBS16, gSBS18, gSBS26, gSBS34, gSBS38, gSBS44, gSBS56, gID1, gID2, gID4, gID8, gID13, gDBS1, gDBS10, and gDBS11. A1, *APOBEC1*; A2, *APOBEC2*; A3, *APOBEC3*; A4, *APOBEC4*; Ad, *ADAR*; AdB1, *ADARB1*; AdB2, *ADARB2*; mut14, mutation of fourteen *APOBEC/ADAR* genes. The “g” prefix denotes the genome signatures.

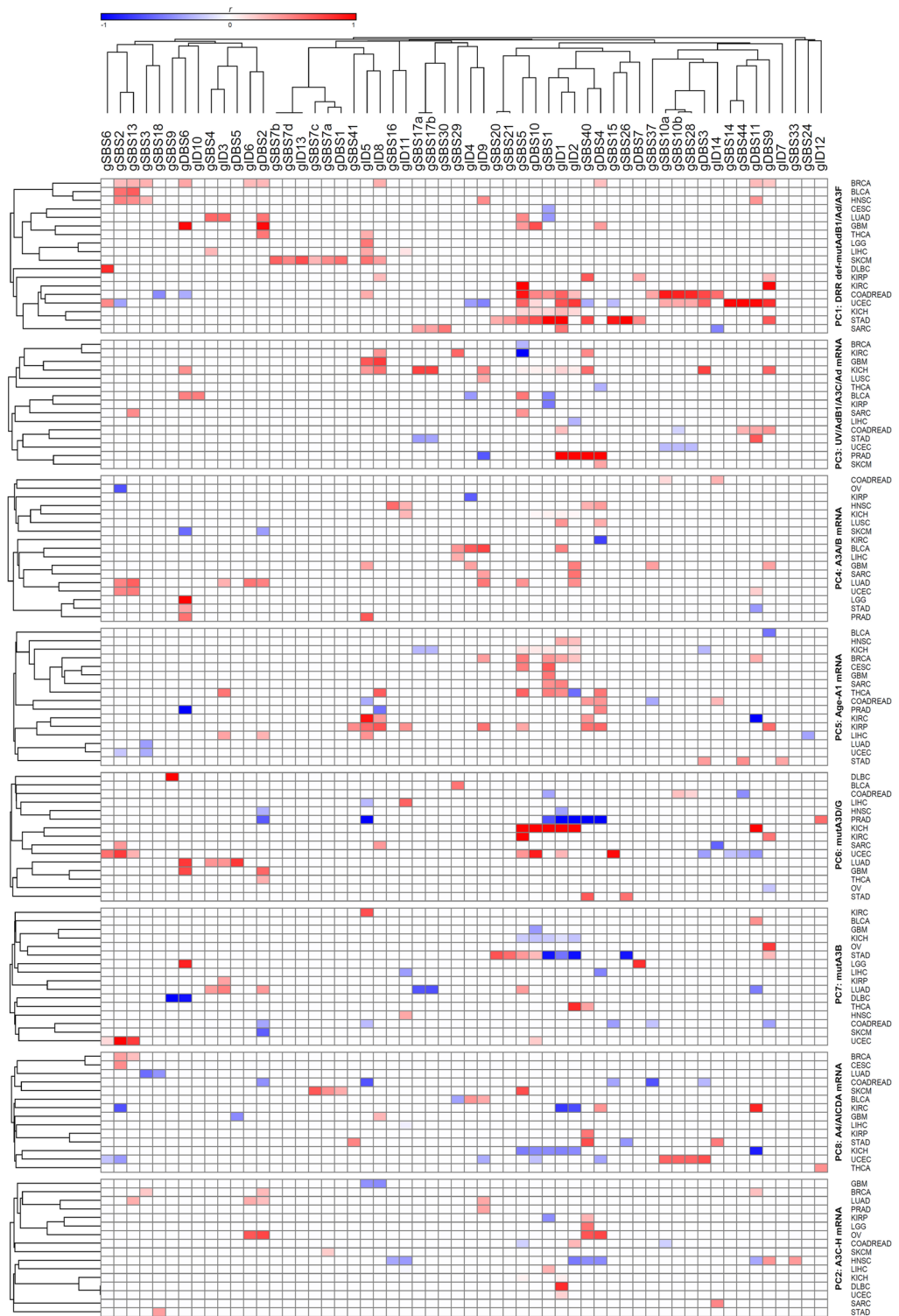


Figure 5. Detailed correlation of the genome signatures with principal components (PCs) potentially implicated in cancer mutagenesis in particular cancers, as stratified by PCs and clustered using Pearson regression r coefficient. The PCs that positively determine the mutational signatures in at least net 10% of cancers (three or more) included PC1 (DRR defects-mutAdB1/Ad/A3F) correlated with gSBS1, gSBS2, gSBS5, gSBS13, gID1, gID2, gID5, gID8, gDBS2, gDBS9, gDBS10, and gDBS11; PC2 (A3C-H mRNA level) correlated with gID1 and gDBS2; PC3 (UV-AdB1/A3C/Ad mRNA level) correlated with gSBS40 and gID8; PC4 (A3A/B mRNA level) correlated with gID1, gID2, gID9, and gDBS6; PC5 (age-A1 mRNA level) correlated with gSBS1, gSBS5, gSBS40, gID1, gID2, gID5, gID8, and gDBS4; PC6 (mutA3D/G) correlated with gSBS5; and PC7 (mutA3B) correlated with gSBS2. A1, *APOBEC1*; A3, *APOBEC3*; Ad, *ADAR*; AdB1, *ADARB1*.

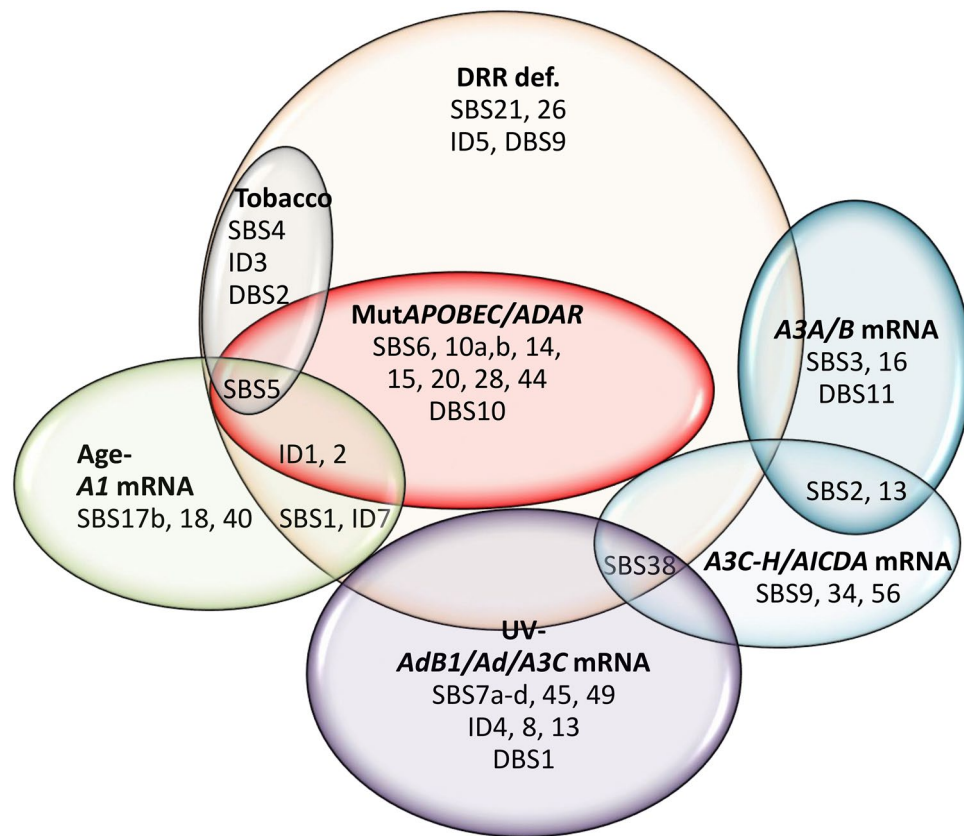


Figure 6. Collective clustering of mutational signatures linked to different factors across various cancers. Those mutational signatures that were found to be correlated with a particular factor in at least two analyses (Figs. 2, 3, 4, 5) were considered to belong to that cluster.

to be determined by *APOBEC/ADAR* mutations (Fig. 6; Supplementary Figure S7). As discussed earlier, the mutational signatures SBS1, ID1, ID2, and ID7 were also determined by age-*A1* mRNA levels. Intriguingly, mutA3B (PC7) was also seen to inversely impact several mutational signatures across various cancers, particularly SBS1, SBS5, ID1, ID2, and ID7 (Fig. 4). Likewise, an apparently inverse impact was seen for other covariates in certain cancers, particularly those of mutA3D/G (most commonly affecting SBS1, SBS44, and ID2) and UV-*ADAR1/A3C/ADAR* mRNA levels (most commonly affecting SBS10a,b and SBS28).

A3C-A3H/AICDA mRNA cluster

Although the variations in *A3C*, *A3D*, *A3F*, *A3G*, and *A3H* (*A3C-H*) and *AICDA* mRNA levels tended to be separate across particular exomes/genomes, they were very close across different cancer types. In light of this, three mutational signatures, including *PolH*-related SBS9, unknown SBS34, and the so-called artifact SBS56, were found to be determined by *A3C-H/AICDA* mRNA levels across both genomes and cancer types (Fig. 6). As mentioned earlier, both SBS2 and SBS13 were also determined by *A3C-H* mRNA levels in certain cancers.

Discussion

In this study, DRR defects were found to be a major determinant of various mutational signatures, sometimes even for those primarily known to be related to other factors such as aging, *A3A/B* mRNA levels, or UV exposure. Since the high mRNA levels of *A1*, *A3A/B*, and *ADAR1* were all found to determine distinct mutational signatures in both wild-type and mutated DRR subgroups, it is reasonable to believe that these factors can act independently of the DRR defects as well. This is not surprising, considering the independent regulation of the expression levels of the aforementioned genes. However, this was not the case for the *APOBEC/ADAR* mutations, which almost always occurred in the presence of a DRR defect. Therefore, it can be concluded that while the *APOBEC/ADAR* mutations have arisen from a DRR defect, they potentiate the impact of the latter on specific mutational signatures. Several mutational signatures were found to be solely determined by DRR defects, including SBS21, SBS26, ID5, and DBS9. Boichard et al. showed that mutations in the *A1*, *A4*, *AICDA*, and *A3* subfamily predicted the mutational burden regardless of MMR or *POLD1/POLE* defects, with no assessment of individual *APOBEC* mutations or other DRR defects²⁵. Our preliminary analysis showed that rather infrequent *APOBEC/ADAR* mutations occur more commonly in hypermutated exomes. Since synonymous mutations are supposed to be of no functional impact and occur rather constantly over time²⁶, in this study the burden of

genomic variants was assessed among those exomes with nonsynonymous *APOBEC/ADAR* mutations compared to those with synonymous ones, showing that T > N SNVs were enriched in *ADARB1*-mutant exomes, while all indels were found to be enriched in *A3C* and *ADAR*-mutant exomes and Ind-T variants were enriched in *A1* and *A3B*-mutant exomes.

The high number of potentially implicated covariates made a PC analysis inevitable, although complicating the attribution of distinct mutational signatures to individual *APOBEC/ADAR* aberrations, particularly infrequent *APOBEC/ADAR* mutations. PC analysis reduced the number of potentially implicated factors, and multivariate models using the measured principal components proposed novel correlations for the so-called artifact/unknown mutational signatures, including SBS34 and SBS56 (determined by *A3C-H/AICDA* mRNA level), SBS40 (determined by age), ID5 and DBS9 (determined by DRR defects), and SBS45, SBS49, and ID4 (determined by *UV-ADARB1/A3C/ADAR* mRNA level). SBS34 is commonly seen among DLBC, STAD, and PAAD and shows asymmetry toward the intergenic regions, as does SBS9²⁷. SBS40 has been proposed to be related to environmental factors because of its accumulation with age in some cancer types²⁸. SBS45 and SBS49 are claimed to be possible sequencing artifacts, the former due to 8-oxo-guanines introduced during sequencing²⁷. SBS56 has been reported to be a marker of AKT inhibitor sensitivity in some cancer cell lines²⁹, further undermining an artifact nature.

Moreover, some findings were apparently discrepant from the current literature, including SBS3, SBS16, SBS17b, SBS18, ID4, and ID8. SBS16 has been suggested to be alcohol-related in ESCC³⁰, while it was found here to be determined by *A3A/A3B* expression levels. Whether this discrepancy is due to the genomic hypomethylating impact of ethanol, which increases *A3C-H* mRNA levels³¹, or because of a shift in *A3A* function from so-called physiological 5hmC demethylation to potentially oncogenic C demethylation³² needs to be investigated in more detail. The so-called 5FU-related SBS17b³³ and ROS-related SBS18²⁷ were also found to be determined by *A1* mRNA levels. In addition, the clock-like signature SBS1³⁴ was found to be determined by DRR defects and *A1* mRNA levels, while the other clock-like signature SBS5³⁴, which has also been reported to be smoking-related³⁵, was determined by DRR defects, tobacco use, *APOBEC/ADAR* mutations, and age. Intriguingly, at least three *AID/APOBEC* proteins, including *AID*, *A3A*, and *A3B*, have been reported to efficiently deaminate dC neighboring DNA damage induced by oxidation or alkylation³⁶, a function that might be implicated in this case as well. *A1* mRNA levels have already been linked to cancer indels, in particular Ind-T ones¹⁸, and it was shown here that at least ID2 (Del-T) and ID7 (Del-C/T) signatures were determined by *A1* mRNA levels in addition to the known DRR defect impact^{37,38}, further characterizing the *A1* mRNA impact on cancer genomes. The impacts of age on ID1 (Ins-T) and ID8 (long Del) also seem to be mediated through *A1* mRNA. Unknown ID4, which has been suggested to be TOP1-related³⁹, was determined by *ADARB1* mRNA level, and the so-called radiation-related ID8 was determined by *ADARB1* and *ADAR* mRNA levels across genomes. Recent studies have shown that *ADAR1* (encoded by *ADAR*) can edit DNA:RNA hybrids that form during transcription and DNA replication^{40,41}. Specifically, *ADAR2* (encoded by *ADARB1*) has been reported to play a role in editing such hybrids needed for dsDNA break repair and genomic stability⁴⁰. This is the function which might be implicated in indel mutational signatures that are supposed to be UV-related, including ID4, ID8, and ID13. Since both *AID* and *A3A* have also been proposed to mediate skin cancer through chronic inflammation and mutational events, respectively^{42,43}, more detailed studies might be warranted to explore the true endogenous UV mediator. The supposedly HR-related SBS3 was determined by *A3A/B* mRNA levels. Whether this is a real link or *A3A/B* mRNA is a proxy for other factors (Fig. 1) needs to be determined by further investigation.

The expression of *APOBEC3* family members is regulated by different factors, including B-cell receptor signaling, noncanonical NF- κ B signaling, and SF3B1 inhibition^{44,45}. Additionally, the expression of *APOBEC3* genes can be regulated by epigenetic modifications, such as DNA methylation and histone acetylation⁴⁶. However, there is much evidence supporting some links between *APOBEC/ADAR* activation and environmental mutagens, including viruses, tobacco, and UV radiation. Not unexpectedly, HPV-positive HNSCs show higher mRNA levels of *A3A*, *A3B*, and *A3H* than HPV-negative HNSCs⁴⁷. It has been reported that increased expression of *A3B* in response to ionizing radiation could contribute to the acquisition of radiation resistance in cancer cells⁴⁸, and radiotherapy is followed by *APOBEC* mutagenesis⁴⁹. *A3G* has also been shown to be activated by UV radiation and rescue cells from its detrimental DNA effects^{50,51}, as well as enhancing double-strand break (DSB) repair in leukemia and lymphoma cells¹⁶, making them radioresistant. Likewise, assessment of tobacco-associated cancers suggested that the cellular machinery underlying *APOBEC* signatures was activated by tobacco smoke³⁵, and *APOBEC* rather than tobacco-associated mutagenesis predominated in two series of bladder cancers⁵². On the other hand, smoking has been shown to repress *ADAR* expression, enhancing intracellular oxidative stress^{53,54}.

One biologically plausible explanation for the inverse impacts of certain factors might be the competitive actions of homologous proteins, including those of *mutA3B* on *A1* mRNA-determined signatures SBS1, SBS5, ID1, ID2, and ID7. Certain *APOBEC/ADAR* proteins have been known to modulate each other through heterodimerization or coexpression. These include *A2* dimerizing with and inhibiting *A1*⁵⁵, *ADAR1* (encoded by *ADAR*) dimerizing with and sequestering *ADAR2* (encoded by *ADARB1*)⁵⁶, and *ADAR3* (encoded by *ADARB2*) downregulating *ADAR2*⁵⁷. Likewise, gain- or loss-of-function variants of *APOBEC/ADAR* proteins have been reported, including an *A3C* variant (S188I) with increased dimerization of the protein and hypermutation of target sequences⁵⁸ and an *ADAR* variant (P193A) destabilizing the protein-Z-DNA complex and affecting tumor cell proliferation⁵⁹. Although *ADAR3* has not been shown to have any catalytic activity thus far⁶⁰, some SNVs in *ADARB2* show a consistent link to longevity across populations⁶¹, suggesting a functional role.

One limitation of the present study is that PC2 does not differentiate between the distinct roles of the co-regulated *A3C-H* and *AICDA* genes. Given the cellular location of the gene products, however, it is reasonable to believe that *A3H* and *AID* are the potential players among *A3C-H/AID* proteins. I also acknowledge that some proposed novel links might seem unexpected by currently known mechanisms. Whether or not the reciprocal nucleic acid changes described elsewhere³ are implicated in these novel links would be an intriguing field of future studies, including the unexpected T > C changes in SBS16 found to be linked to *A3A/B* mRNA levels. The

fact that several highly significant associations were observed only in the much smaller genome rather than the exome subgroup (i.e., gSBS16, gSBS18, gSBS34, gSBS56, gID4, gID8, gID13, gDBS1, gSBS9, gDBS10, and gDBS11) might implicate some truly differential mutagenic mechanisms acting among intergenic versus genic regions.

In conclusion, this pancancer approach links several exome/genome mutational signatures with novel factors among APOBEC and ADAR families, including SBS17b, SBS18, and ID7 (*A1* mRNA level); SBS3 and SBS16 (*A3A/B* mRNA level); ID4 and ID8 (*ADARB1* mRNA level); SBS40, ID1, and ID2 (age); and ID5 and DBS9 (DRR defects). While some *APOBEC/ADAR* mutations potentiate the impact of DRR defects on certain mutational signatures, the impact of high expression levels of *A1*, *A3A/B*, *A3C-H/AICDA*, and *ADARB1* on distinct signatures can be independent of other mutagenic factors, while still modulating them. It is proposed that the mutagenic impacts of aging are at least partly mediated through the *A1* mRNA level, while the UV impacts are mainly mediated through *ADARB1* mRNA levels. The precise mechanisms that are involved need to be investigated in detail.

Methods

Patients and samples

The findings of the current study are based on data generated by the TCGA Research Network. A catalog of all somatic variants identified by whole-exome sequencing of 10,295 paired tumor-normal samples was obtained from [sftp://tcgaftp.ncbi.nih.gov/tcgajamboree/mc3/pancan.merged.v0.2.7.PUBLIC.maf.gz](https://tcgaftp.ncbi.nih.gov/tcgajamboree/mc3/pancan.merged.v0.2.7.PUBLIC.maf.gz), comprising a total of 3,600,963 variants from 33 tumor types. In brief, the MC3 version of the TCGA variants had been consistently called using seven methods with proven performance including MuTect (<https://github.com/broadinstitute/mutect>), Pindel (<https://github.com/genome/pindel>), Radia (<https://github.com/aradenbaugh/radia>), VarScan2 (<http://dkoboldt.github.io/varscan/>), SomaticSniper (<https://github.com/genome/somatic-sniper>), MuSE (<https://github.com/danielfan/MuSE>), and Indelocator (<http://archive.broadinstitute.org/cancer/cga/indelocator>), as described before⁶². Non-synonymous *APOBEC/ADAR* variants used for analysis included all variants but the silent, intronic, and flanking ones. The same rule was applied for the mutations in DRR pathway genes. The list of key DRR pathway genes was obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG)⁶³, as follows: base excision repair (BER; hsa03410; 33 genes), nucleotide excision repair (NER) and transcription-coupled NER (hsa03420; 45 genes), mismatch repair (MMR; hsa03430; 22 genes), homologous recombination (HR; hsa03440; 41 genes), nonhomologous end-joining (NHEJ; hsa03450; 10 genes), and Fanconi anemia-translesion synthesis (FATLS; hsa03460; 54 genes), as well as DNA replication pathway (DNA-Rep; hsa03030; 35 genes), collectively 168 unique genes. All nonsynonymous exome variants were included for analysis of the potentially implicated pathways. A list of *APOBEC/AICDA/ADAR* gene mutations was extracted, in which mutations were classified into synonymous or nonsynonymous mutations, with the latter including missense, nonsense, splice site, and indel mutations.

The following open access (Level 3) databases were obtained from the Broad Institute GDAC Firehose (<http://gdac.broadinstitute.org/>): normalized *RNA-Seq* by Expectation–Maximization (RSEM) for genes (RSEM_genes_normalized_data, v.20160128) and clinical data, including vital status and days to death/last follow-up (Clinical_Pick_Tier1, v.20171224). The normalized RSEM data resulting from RNA sequencing were used as an estimate of TCGA gene expression profiles, and they were further normalized to *TATA-binding protein (TBP)* mRNA levels and then scaled to 1. Indels longer than one nucleotide were classified based on their 5'-end nucleotide. The absolute number of variants attributed to each mutational signature as detected by whole-exome or whole-genome sequencing (COSMIC release V89, May 2019) was obtained from Rosenthal's study⁶² and used for assessment of the correlation with *APOBEC/ADAR* aberrations as well as DRR defects. This version includes 93 mutational signatures: SBS1–SBS60, ID1–ID17, and DBS1–DBS11. Three signatures including SBS7, SBS10, and SBS17 were represented by four, two, and two sub-signatures, respectively.

Statistical analysis

Unpaired two-tailed Student's *t* test (IBM SPSS, v.22) was used to compare the mean number of indel variants and SNVs among wild-type versus mutant *APOBEC/ADAR* exomes, as well as nonsynonymous versus synonymous *APOBEC/ADAR* mutant exomes, and the *Benjamini–Hochberg* false discovery rate (FDR) was used to adjust for multiple testing, with an acceptable FDR of up to 0.05. Principal component (PC) analysis was performed using the Varimax method with Kaiser Normalization (IBM SPSS, v.22) across 10,126 samples with WES (9,493)/WGS (773)-identified mutational signatures (140 having both) to reduce the number of 37 covariates, including 14 *APOBEC/ADAR* mutations, 14 *APOBEC/ADAR* mRNA levels, seven DRR pathway defects, age, and UV factor (positive in SKCM and UVM). An acceptable eigenvalue cutoff of 1.0 was used to select the PCs and a cutoff of 0.4 for significant rotated components. A similar PC analysis was performed using the mean exome/genome values of the aforementioned 37 covariates across 33 cancer types, with an eigenvalue cutoff of 1.5 used to select the PCs and a cutoff of 0.4 for significant rotated components. Next, forward linear regression analysis (IBM SPSS, v.22) was used to estimate the correlation of the mutational signatures with PCs (eight PCs for pancancer exome/genome analysis, and five for mean exome/genome values across 33 cancers), with an acceptable *P* value of up to 0.05. For mutational signatures that were found to be correlated with PC3 (UV-*ADARB1/A3C/ADAR* mRNA level) and PC5 (age-*A1* mRNA level), repeat detailed analysis was performed to identify the exact factor correlated with each signature. For those mutational signatures whose univariate Pearson regression analysis (IBM SPSS, v.22) suggested a potential link to tobacco use, this factor was also included in multivariate regression analysis. The third kind of multivariate regression analysis involving mutational signatures was performed using varying PCs (aforementioned eight) in particular cancers, including adrenocortical carcinoma (ACC; 91 exomes/0 genomes), bladder carcinoma (BLCA; 389/22), breast carcinoma (BRCA; 930/86), cervical/endocervical cancer (CESC; 271/20), cholangiocarcinoma (CHOL; 36/0), colon and rectal adenocarcinoma (COAD-READ; 496/52), diffuse large B-cell lymphoma (DLBC; 31/6), esophageal carcinoma (ESCA; 182/0), glioblastoma

multiforme (GBM; 362/35), head and neck squamous cell carcinoma (HNSC; 464/43), kidney chromophobe (KICH; 50/45), kidney clear cell carcinoma (KIRC; 334/27), kidney papillary cell carcinoma (KIRP; 251/33), acute myeloid leukemia (LAML; 139/0), low-grade glioma (LGG; 512/18), hepatocellular carcinoma (LIHC; 311/52), lung adenocarcinoma (LUAD; 528/37), lung squamous cell carcinoma (LUSC; 427/48), mesothelioma (MESO; 82/0), ovarian serous cystadenocarcinoma (OV; 385/27), pancreatic adenocarcinoma (PAAD; 174/0), pheochromocytoma/paraganglioma (PCPG; 182/0), prostate adenocarcinoma (PRAD; 484/19), sarcoma (SARC; 208/30), skin melanoma (SKCM; 412/37), stomach adenocarcinoma (STAD; 401/38), testicular germ cell tumor (TGCT; 149/0), thyroid carcinoma (THCA; 478/48), thymoma (THYM; 123/0), uterine corpus endometrial carcinoma (UCEC; 474/50), uterine carcinosarcoma (UCS; 57/0), and uveal melanoma (UVM; 80/0). Those correlations that were found in at least 2 out of 6 analyses (pancancer exomes/genomes, pancancer mean exomes/genome values across 33 cancers, and cancer-specific exomes/genomes) were considered to be consensus, and the data were used to cluster mutational signatures based on potential underlying factors. The median of the mean normalized *A3A* and *A3B* mRNA levels was used to classify *A3A/B* mRNA into low and high levels, and *APOBEC/ADAR* mutational status was considered to be positive when at least one of the corresponding genes was mutated. An unpaired two-tailed *t* test with Welch's correction (IBM SPSS, v.22) was used to compare the mean number of mutational signatures among paired subgroups, including wild type versus mutated DRR, wild type versus mutated *A3D/G*, wild type versus mutated *ADARB1/ADAR/A3F*, wild type versus mutated *APOBEC/ADAR*, low- versus high-*A1* mRNA level, low- versus high-*A3A/B* mRNA level, and low- versus high-*ADARB1* mRNA level. Gene-E (<https://software.broadinstitute.org/GENE-E/>) was used to cluster *r* heat maps.

Data availability

The data set used for analyses is available upon request from corresponding author.

Received: 8 February 2024; Accepted: 14 June 2024

Published online: 04 July 2024

References

- Conticello, S. G. The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* **9**(229), 221–210 (2008).
- Powell, L. M. *et al.* A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* **50**, 831–840 (1987).
- Niavarani, A. *et al.* APOBEC3A is implicated in a novel class of G-to-A mRNA editing in WT1 transcripts. *PLoS ONE* **10**, e0120089 (2015).
- Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol. cell* **10**, 1247–1253 (2002).
- Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
- Burns, M. B., Temiz, N. A. & Harris, R. S. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Rev. Genet.* **45**, 977–983 (2013).
- Roberts, S. A. *et al.* An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* **45**, 970–976 (2013).
- Alexandrov, L. B. *et al.* The Repertoire of Mutational Signatures in Human Cancer. *bioRxiv*, 322859, <https://doi.org/10.1101/322859> (2018).
- Nik-Zainal, S. *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.* **46**, 487–491 (2014).
- Chan, K. *et al.* An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat. Genet.* **47**, 1067–1072 (2015).
- Lau, P. P., Xiong, W. J., Zhu, H. J., Chen, S. H. & Chan, L. Apolipoprotein B mRNA editing is an intranuclear event that occurs posttranscriptionally coincident with splicing and polyadenylation. *J. Biol. Chem.* **266**, 20550–20554 (1991).
- Bogerd, H. P. *et al.* Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc. Nat. Acad. Sci. United States Am.* **103**, 8780–8785 (2006).
- Burdick, R. C., Hu, W. S. & Pathak, V. K. Nuclear import of APOBEC3F-labeled HIV-1 preintegration complexes. *Proc. Nat. Acad. Sci. United States Am.* **110**, E4780–4789 (2013).
- OhAinle, M., Kerns, J. A., Li, M. M., Malik, H. S. & Emerman, M. Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. *Cell Host Microbe* **4**, 249–259 (2008).
- Ito, S. *et al.* Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc. Nat. Acad. Sci. United States Am.* **101**, 1975–1980 (2004).
- Pasqualucci, L. *et al.* Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412**, 341–346. <https://doi.org/10.1038/35085588> (2001).
- Forconi, F. *et al.* Hairy cell leukemia: at the crossroad of somatic mutation and isotype switch. *Blood* **104**, 3312–3317. <https://doi.org/10.1182/blood-2004-03-0950> (2004).
- Niavarani, A., Shahrabi Farahani, A., Sharafkhan, M. & Rassoulzadegan, M. Pancancer Analysis Identifies Prognostic High-APOBEC1 Expression Level Implicated in Cancer In-Frame Insertions and Deletions. *Carcinogenesis* (2018).
- Middlebrooks, C. D. *et al.* Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors. *Nat. Genet.* **48**, 1330–1338 (2016).
- Kanke, Y. *et al.* Gene aberration profile of tumors of adolescent and young adult females. *Oncotarget* **9**, 6228–6237. <https://doi.org/10.18632/oncotarget.23765> (2018).
- Tsuruoka, N. *et al.* ADAR1 protein induces adenosine-targeted DNA mutations in senescent Bcl6 gene-deficient cells. *J. Biol. Chem.* **288**, 826–836. <https://doi.org/10.1074/jbc.M112.365718> (2013).
- Zheng, Y., Lorenzo, C. & Beal, P. A. DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA. *Nucl. Acids Res.* **45**, 3369–3377. <https://doi.org/10.1093/nar/gkx050> (2017).
- Lindley, R. A., Humbert, P., Larner, C., Akmeemana, E. H. & Pendlebury, C. R. Association between targeted somatic mutation (TSM) signatures and HGS-OvCa progression. *Cancer Med.* **5**, 2629–2640. <https://doi.org/10.1002/cam4.825> (2016).
- Lindley, R. A. & Hall, N. E. APOBEC and ADAR deaminases may cause many single nucleotide polymorphisms curated in the OMIM database. *Mutation Res.* **810**, 33–38. <https://doi.org/10.1016/j.mrfmmm.2018.03.008> (2018).
- Boichard, A., Tsigelny, I. F. & Kurzrock, R. High expression of PD-1 ligands is associated with kataegis mutational signature and APOBEC3 alterations. *Oncoimmunology* **6**, e1284719. <https://doi.org/10.1080/2162402X.2017.1284719> (2017).
- Kimura, M. Evolutionary rate at the molecular level. *Nature* **217**, 624–626 (1968).

27. Jin, S. G., Meng, Y., Johnson, J., Szabo, P. E. & Pfeifer, G. P. Concordance of hydrogen peroxide-induced 8-oxo-guanine patterns with two cancer mutation signatures of upper GI tract tumors. *Sci. Adv.* **8**, 3815. <https://doi.org/10.1126/sciadv.abn3815> (2022).
28. Karihtala, P., Porvari, K. & Kilpivaara, O. Mutational signatures associate with survival in gastrointestinal carcinomas. *Cancer Genom. Proteom.* **19**, 556–569. <https://doi.org/10.21873/cgp.20340> (2022).
29. Levatic, J., Salvadores, M., Fuster-Tormo, F. & Supek, F. Mutational signatures are markers of drug sensitivity of cancer cells. *Nat. Commun.* **13**, 2926. <https://doi.org/10.1038/s41467-022-30582-3> (2022).
30. Chang, J. *et al.* Genomic analysis of oesophageal squamous-cell carcinoma identifies alcohol drinking-related mutation signature and genomic alterations. *Nat. Commun.* **8**, 15290. <https://doi.org/10.1038/ncomms15290> (2017).
31. Liu, C. *et al.* A DNA methylation biomarker of alcohol consumption. *Mol. Psych.* **23**, 422–433. <https://doi.org/10.1038/mp.2016.192> (2018).
32. Schutsky, E. K., Nabel, C. S., Davis, A. K. F., DeNizio, J. E. & Kohli, R. M. APOBEC3A efficiently deaminates methylated, but not TET-oxidized, cytosine bases in DNA. *Nucl. Acids Res.* **45**, 7655–7665. <https://doi.org/10.1093/nar/gkx345> (2017).
33. Christensen, S. *et al.* 5-Fluorouracil treatment induces characteristic T>G mutations in human cancer. *Nat. Commun.* **10**, 4571. <https://doi.org/10.1038/s41467-019-12594-8> (2019).
34. Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407. <https://doi.org/10.1038/ng.3441> (2015).
35. Alexandrov, L. B. *et al.* Mutational signatures associated with tobacco smoking in human cancer. *Science* **354**, 618–622. <https://doi.org/10.1126/science.aag0299> (2016).
36. Diamond, C. P. *et al.* AID, APOBEC3A and APOBEC3B efficiently deaminate deoxycytidines neighboring DNA damage induced by oxidation or alkylation. *Biochim. et Biophys. Acta. General Subj.* **1863**, 129415. <https://doi.org/10.1016/j.bbagen.2019.129415> (2019).
37. Sia, E. A., Kokoska, R. J., Dominska, M., Greenwell, P. & Petes, T. D. Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Mol. Cell. Biol.* **17**, 2851–2858 (1997).
38. TCGA. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
39. Reijns, M. A. M. *et al.* Signatures of TOP1 transcription-associated mutagenesis in cancer and germline. *Nature* **602**, 623–631. <https://doi.org/10.1038/s41586-022-04403-y> (2022).
40. Jimeno, S. *et al.* ADAR-mediated RNA editing of DNA:RNA hybrids is required for DNA double strand break repair. *Nat. Commun.* **12**, 5512. <https://doi.org/10.1038/s41467-021-25790-2> (2021).
41. Steele, E. J. & Lindley, R. A. ADAR deaminase A-to-I editing of DNA and RNA moieties of RNA:DNA hybrids has implications for the mechanism of Ig somatic hypermutation. *DNA Repair* **55**, 1–6. <https://doi.org/10.1016/j.dnarep.2017.04.004> (2017).
42. Nonaka, T. *et al.* Involvement of activation-induced cytidine deaminase in skin cancer development. *J. Clin. Invest.* **126**, 1367–1382. <https://doi.org/10.1172/JCI15122> (2016).
43. Pham, P., Landolph, A., Mendez, C., Li, N. & Goodman, M. F. A biochemical analysis linking APOBEC3A to disparate HIV-1 restriction and skin cancer. *J. Biol. Chem.* **288**, 29294–29304. <https://doi.org/10.1074/jbc.M113.504175> (2013).
44. Butler, K. & Banday, A. R. APOBEC3-mediated mutagenesis in cancer: causes, clinical significance and therapeutic potential. *J. Hematol. Oncol.* **16**, 31. <https://doi.org/10.1186/s13045-023-01425-5> (2023).
45. Wang, Z. *et al.* B cell receptor signaling drives APOBEC3 expression via direct enhancer regulation in chronic lymphocytic leukemia B cells. *Blood Cancer J.* **12**, 99. <https://doi.org/10.1038/s41408-022-00690-w> (2022).
46. Menendez, D., Nguyen, T. A., Snipe, J. & Resnick, M. A. The cytidine deaminase APOBEC3 family is subject to transcriptional regulation by p53. *Mol. Cancer Res. MCR* **15**, 735–743. <https://doi.org/10.1158/1541-7786.MCR-17-0019> (2017).
47. Kondo, S. *et al.* APOBEC3A associates with human papillomavirus genome integration in oropharyngeal cancers. *Oncogene* **36**, 1687–1697. <https://doi.org/10.1038/onc.2016.335> (2017).
48. Saito, Y. *et al.* Involvement of APOBEC3B in mutation induction by irradiation. *J. Radiat. Res.* **61**, 819–827. <https://doi.org/10.1093/jrr/rraa069> (2020).
49. Kocakavuk, E. *et al.* Radiotherapy is associated with a deletion signature that contributes to poor outcomes in patients with cancer. *Nat. Genet.* **53**, 1088–1096. <https://doi.org/10.1038/s41588-021-00874-3> (2021).
50. Botvinnik, A. *et al.* APOBEC3G rescues cells from the deleterious effects of DNA damage. *FEBS J.* **288**, 6063–6077. <https://doi.org/10.1111/febs.16025> (2021).
51. Tong, Y. *et al.* Radiosensitization to γ -ray by functional inhibition of APOBEC3G. *Int. J. Mol. Sci.* **23**(9), 5069. <https://doi.org/10.3390/ijms23095069> (2022).
52. Chang, M. T. *et al.* Small-cell carcinomas of the bladder and lung are characterized by a convergent but distinct pathogenesis. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **24**, 1965–1973. <https://doi.org/10.1158/1078-0432.CCR-17-2655> (2018).
53. Takizawa, M., Nakano, M., Fukami, T. & Nakajima, M. Decrease in ADAR1 expression by exposure to cigarette smoke enhances susceptibility to oxidative stress. *Toxicol. Lett.* **331**, 22–32. <https://doi.org/10.1016/j.toxlet.2020.05.019> (2020).
54. Koutros, S. *et al.* Targeted deep sequencing of bladder tumors reveals novel associations between cancer gene mutations and mutational signatures with major risk factors. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **27**, 3725–3733. <https://doi.org/10.1158/1078-0432.CCR-20-4419> (2021).
55. Anant, S. *et al.* ARCD-1, an apobec-1-related cytidine deaminase, exerts a dominant negative effect on C to U RNA editing. *Am. J. Physiol. Cell Physiol.* **281**, C1904–1916. <https://doi.org/10.1152/ajpcell.2001.281.6.C1904> (2001).
56. Cenci, C. *et al.* Down-regulation of RNA editing in pediatric astrocytomas: ADAR2 editing activity inhibits cell migration and proliferation. *J. Biol. Chem.* **283**, 7251–7260. <https://doi.org/10.1074/jbc.M708316200> (2008).
57. Chen, C. X. *et al.* A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. *RNA* **6**, 755–767. <https://doi.org/10.1017/s1355838200000170> (2000).
58. Wittkopp, C. J., Adolph, M. B., Wu, L. L., Chelico, L. & Emerman, M. A single nucleotide polymorphism in human APOBEC3C enhances restriction of lentiviruses. *PLoS Pathog.* **12**, e1005865. <https://doi.org/10.1371/journal.ppat.1005865> (2016).
59. Beyer, U. *et al.* Rare ADAR and RNASEH2B variants and a type I interferon signature in glioma and prostate carcinoma risk and tumorigenesis. *Acta Neuropathol.* **134**, 905–922. <https://doi.org/10.1007/s00401-017-1774-y> (2017).
60. Eisenberg, E. & Levanon, E. Y. A-to-I RNA editing - immune protector and transcriptome diversifier. *Nat. Rev. Gene.* **19**, 473–490. <https://doi.org/10.1038/s41576-018-0006-1> (2018).
61. Sebastiani, P. *et al.* RNA editing genes associated with extreme old age in humans and with lifespan in *C. elegans*. *PLoS ONE* **4**, e8210. <https://doi.org/10.1371/journal.pone.0008210> (2009).
62. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31. <https://doi.org/10.1186/s13059-016-0893-4> (2016).
63. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucl. Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).

Acknowledgements

I am truly thankful to Prof. Silvestro G Conticello (Istituto di Fisiologia Clinica, Italy), Dr. Maria Secrier (Division of Biosciences, University College London, UK), Prof. Franca Fraternali (Institute of Structural and Molecular

Biology, University College London, UK), Dr. Diako Ebrahimi (Texas Biomedical Research Institute, US), and Dr. Joseph CF Ng (University College London, UK), who kindly shared their ideas with me and critically reviewed this article before submission.

Author contributions

A.N. conceptualized the study, collected and analyzed the data, and drafted the manuscript.

Funding

This study was supported by a grant from Digestive Disease Research Institute, Tehran University of Medical Sciences.

Competing interests

The author declares no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-64986-6>.

Correspondence and requests for materials should be addressed to A.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024