



Oncogenic Merkel Cell Polyomavirus T Antigen Truncating Mutations are Mediated by APOBEC3 Activity in Merkel Cell Carcinoma

Anni I. Soikkeli^{1,2}, Minna K. Kyläniemi³, Harri Sihto⁴, and Jukka Alinikula¹

ABSTRACT

Merkel cell carcinoma (MCC) is an aggressive skin cancer, which is frequently caused by Merkel cell polyomavirus (MCPyV). Mutations of MCPyV tumor (T) antigens are major pathologic events of virus-positive (MCPyV+) MCCs, but their source is unclear. Activation-induced cytidine deaminase (AID)/APOBEC family cytidine deaminases contribute to antiviral immunity by mutating viral genomes and are potential carcinogenic mutators. We studied the contribution of AID/APOBEC cytidine deaminases to MCPyV large T (LT) truncation events. The MCPyV LT area in MCCs was enriched with cytosine-targeting mutations, and a strong APOBEC3 mutation signature was observed in MCC sequences. *AICDA* and *APOBEC3* expression were detected in the Finnish MCC sample

cohort, and *LT* expression correlated with *APOBEC3H* and *APOBEC3G*. Marginal but statistically significant somatic hypermutation targeting activity was detected in the MCPyV regulatory region. Our results suggest that APOBEC3 cytidine deaminases are a plausible cause of the *LT* truncating mutations in MCPyV+ MCC, while the role of AID in MCC carcinogenesis is unlikely.

Significance: We uncover APOBEC3 mutation signature in MCPyV *LT* that reveals the likely cause of mutations underlying MCPyV+ MCC. We further reveal an expression pattern of APOBECs in a large Finnish MCC sample cohort. Thus, the findings presented here suggest a molecular mechanism underlying an aggressive carcinoma with poor prognosis.

Introduction

Merkel cell carcinoma (MCC) is a rare but aggressive skin cancer mostly affecting the elderly and/or immunocompromised. The carcinoma has poor prognosis and at least two distinct etiologies. Approximately 80% of the tumors are caused by Merkel cell polyomavirus (MCPyV) where the virus is clonally integrated in the genome, have low mutation burden, and lack UV mutation signatures (1–7). Unlike virus-positive (MCPyV+) MCCs, virus-negative (MCPyV–) MCCs have high mutation burden and an UV mutation signature commonly seen in skin cancers (8–10). In addition to mutation accumulation, MCPyV+ and MCPyV– MCC differ in cell morphology and possibly the cell of origin (11, 12).

In MCPyV+ cancers, the MCPyV small tumor antigen (ST) and large tumor antigen (LT) cooperate to promote transformation and growth of the host cells

(reviewed in ref. 13). The ST mediates its effect through interaction with various molecular complexes, which can reduce the activity of tumor suppressor p53, as well as stabilize and increase the expression of the LT (14–16).

MCPyV+ MCCs express highly mutated *LT* with substitutions and deletions frequently leading to premature stop codons and truncation of the protein (17). These mutations inactivate the C-terminal origin-binding domain and the helicase domain of the *LT* required for virus replication. Thus, the truncated *LT* is thought to enable stable integration of MCPyV and inhibit the DNA damage response caused by full-length replicated MCPyV genome (17, 18). The truncated *LT* retains the N-terminal LXCXE motif, which is required for the binding and inhibition of RB tumor suppressor protein along various other factors, thus enabling inappropriate cell-cycle progression and increase in cellular proliferation (19). While the truncating mutations of *LT* are central for the oncogenicity of MCPyV, the cause of these mutations is not known.

MCC was named after the cells it was presumed to arise from, the Merkel mechanoreceptor cells of the skin (20). Since then, it has been discovered that MCCs have a different protein expression pattern (21, 22), differentiation status (23), and location in the skin (24) compared with Merkel cells, calling the cancer-initiating cell type into question. In addition, MCPyV is unable to infect Merkel cells (25) and there are too few Merkel cells in the skin for them to be the cell of origin (26). Fibroblasts (12, 25), keratinocytes (12), and, interestingly, pro/pre-B cells (21) are proposed as the cell of origin of MCC but the issue remains unsettled.

MCPyV+ MCCs express the B-cell identity factor PAX5 and several immunoglobulin (Ig) classes (21, 27). MCCs express TDT, which is involved in

¹Unit of Infection and Immunity, Institute of Biomedicine, University of Turku, Turku, Finland. ²Turku Doctoral Programme of Molecular Medicine, University of Turku, Turku, Finland. ³Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland. ⁴Department of Pathology, University of Helsinki, Helsinki, Finland.

Corresponding Author: Anni I. Soikkeli, Unit of Infection and Immunity, Institute of Biomedicine, University of Turku, Turku 20520, Finland. Phone: 3584-0354-9950; E-mail: anilso@utu.fi

doi: 10.1158/2767-9764.CRC-22-0211

This open access article is distributed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

© 2022 The Authors; Published by the American Association for Cancer Research

antigen receptor gene recombination, and are shown to recombine their *Ig* genes, such as occurs in developing B cells (21). In addition, both MCPyV+ and MCPyV- MCCs are shown to express activation-induced cytidine deaminase (AID), a member of the AID/APOBEC family of cytidine deaminases (6, 28) AID is expressed almost exclusively in B cells, where it initiates somatic hypermutation (SHM) and class-switch recombination (CSR) in *Ig* genes. Being a powerful mutator, it is involved in the development of several cancer types, such as lymphoma, skin cancers (squamous cell carcinoma, basal-cell carcinoma, and melanoma), gastric cancers, and hepatocellular cancers. Particularly the members of APOBEC3 subfamily participate in antiviral defense (29, 30) and mutate viral genomes (31, 32), as well as contribute to carcinogenesis in, for instance, breast, cervical, and lung cancer (8, 29). APOBEC3A, APOBEC3B, and APOBEC3H are most tightly linked to carcinogenesis due to their frequent expression in cancers, nuclear localization, and ability to deaminate cytidines in genomic DNA (31, 33, 34). In addition, polyomaviruses have been shown to induce APOBEC3 expression (35, 36).

In *Ig* genes, AID-induced SHM is targeted to a 250–1,500 bp region downstream of the transcription start site, which codes for the variable domain of antibodies. SHM is targeted to the *Ig* loci by *Ig* enhancers and enhancer-like sequences (37). SHM is known to target several non-*Ig* genes outside the *Ig* locus, a phenomenon which is intimately linked to lymphoma (38–42).

We sought to investigate whether mutations in MCPyV *LT* could arise from an SHM-like cytidine deamination. We found APOBEC3 signature from *LT* area enriched in MCC sequences and APOBEC3 expression in a Finnish MCC sample cohort. We found *AICDA* expression in a subset of Finnish MCC tumors and marginal SHM recruiting activity in the MCPyV regulatory region upstream of *LT* in a B-cell model, but the AID mutation signature was not enriched in *LT* sequences. We conclude that APOBEC3s, rather than AID, are mostly responsible for *LT* mutations in MCC.

Materials and Methods

Cloning Procedure

The sequences tested for SHM targeting activity were cloned to a GFP4-expressing vector either at a *NheI/SpeI* site (upstream position) or at a *BamHI* site (downstream position). The tested sequences were first amplified from MCV-HF (RRID:Addgene_32057; ref. 43) using Q5 High-Fidelity DNA Polymerase (NEB). For downstream cloning, the GFP expression cassette was amplified from the GFP4 vector. Primers were designed according to In-Fusion primer design protocol. Cloning was performed with In-Fusion cloning (Takara) kit according to the manufacturer's protocol. The cloned plasmids were isolated with GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific), and the success of cloning was checked with restriction enzyme digestion. After successful cloning, new plasmid isolations were made with ZymoPure II Maxiprep kit (Zymo research).

Transfection and Cell Culture

The chicken B-cell line DT40 (RRID:CVCL_0249) with modifications (*UNG*^{-/-} *AICDA*^{R^{puro}}; described in ref. 37) was received from David G. Schatz. The cells were authenticated by testing their sensitivity for blasticidin and puromycin after targeted integration of the reporter construct, Western blotting of AID expression and functionally by carrying out the GFP loss assay with control reporters (Fig. 2B). The cells were cultured for approximately 25 days during the assay periods. The cells were cultured at +40°C, 5%

CO₂, 90% humidity. Growth media included RPMI1640 HEPES modification (Sigma) with 10% FBS (HyClone), 1% NCS (Biowest), 1x penicillin-streptomycin antibiotic (Gibco), 1x L-glutamine (Gibco), and 50 μmol/L β-mercaptoethanol. The cells were tested for lack of *Mycoplasma* on January 9, 2020 using PCR *Mycoplasma* test kit I/C (PromoCell).

The GFP loss assay was performed (as described in ref. 37). First, 12 × 10⁶ cells were transfected with 50 μg of *NotI*-linearized test sequences containing plasmids. For negative control, a plasmid including the GFP reporter without a test sequence was included. Transfection was performed with GenePulser electroporator (Bio-Rad) using settings 0.7 kV, 200 Ω, and 25 μF. Immediately after transfection, cells were transferred to 96-well plates to growth media with 5% extra FBS added. The day after transfection, blasticidin was added at a final concentration of 15 μg/mL to select the transfected cells. Cells were grown for approximately 7 days before primary clones were picked and tested for targeted integration using puromycin selection (final concentration 1 μg/mL). At least one targeted clone per test sequence was then subcloned by limiting dilution and cultured for 12 days before flow cytometry analysis.

Flow Cytometry

Minimum of 12 independent subclones per tested sequence were analyzed using Accuri C6 cytometer (BD Biosciences). Results were then further analyzed with FlowJo software (RRID:SCR_008520) and GraphPad Prism 9 software (RRID:SCR_002798).

Mutation Analysis

A total of 113 MCPyV *LT* sequences from MCC samples and 83 MCPyV *LT* sequences from healthy skin samples submitted to GenBank NCBI Virus database by the year 2020 were analyzed. The sequences varied in length between 65 and 2,885 bases. Sequences were aligned to the reference MCPyV isolate R17b (NCBI:txid 493803) sequence using SnapGene software (RRID:SCR_015052). All single-bp substitutions and insertion/deletion mutations relative to the MCPyV reference genome were included in the analysis. To avoid overrepresentation of MCPyV strain-specific variants, each mutation was calculated as one despite its potential occurrence in multiple sequences. The amount and distribution of single-bp substitutions as well as mutations leading to a stop codon were analyzed. WRC, TCW, and YCC motifs and mutations affecting the underlined cytidines were calculated. These motifs were selected to represent AID, APOBEC3, and UV hotspots, as they are mutually exclusive and thus give a better resolution for mutation analysis and trinucleotide sites were used to achieve better comparability. Mutation frequency was calculated in 200 bp bins. The number of mutations was divided by the number of sequences analyzed. Because some analyzed sequences were truncated or otherwise partial, this bin-based approach prevented bias caused by different number of sequences obtained along the *LT* area. Graphs were made using GraphPad Prism 9 software.

RNA Sequencing

RNA sequencing (RNA-seq) was performed with 82 RNA samples extracted from formalin-fixed, paraffin-embedded (FFPE) tumor tissue blocks. The samples are part of a larger set of tumor samples from a Finnish population-based sample series. RNA was extracted from two FFPE sections per primary MCC tumor using QIAGEN's QIASymphony RNA extraction kit and QIASymphony SP instrument. RNA concentration was measured prior to library preparation. Sequencing was performed in the sequencing unit of the Institute for Molecular

TABLE 1 Summary of mutations detected in *LT* area

<i>LT</i> area	All	<i>P</i>	subst	<i>P</i>	indel	<i>P</i>	No. of seq	No. of seq bases
MCC mutations	318	<0.0001	292	<0.0001	26	0.0002	113	206646
MCC in-frame STOP	39	<0.0001	19	0.0001	20	<0.0001	65	
Control mutations	90		88		2		83	172425
Control in-frame STOP	0		0		0		0	

Medicine Finland. RNA-seq libraries were prepared using QuantSeq 3' mRNA-seq Library prep kit FWD (Lexogen) according to manufacturer's instructions and sequenced with Illumina HiSeq 2500 high output mode, v4 chemistry, and 2 × 100 bp reads. Raw read quality was analyzed with FastQC v0.11.8 (44) and MultiQC v1.7 (45). Raw reads were trimmed with BBDuk (BBMap version 38.87) according to instructions from library kit manufacturer. ERCC control sequences and MCPyV genome were combined to GRCh38 genome and trimmed reads were mapped to genome with STAR 2.7.2 (46) according to library kit manufacturer's instructions. Reads mapping to genes were counted with htseq-count using stranded mode (47). For counts per million (CPM) normalization of read counts, the edgeR package v3.32.1 (48) was used. For annotations of read counts, R package org.Hs.eg.db v3.12.0 (49) was used. Analysis was performed in the CSC Puhti computing environment and in R version 4.5 in macOS 11.5.2. Fold changes of *AICDA* and *APOBECs* were calculated by comparing normalized read counts of MCPyV+ and MCPyV− MCC samples to each other. Heatmap and similarity matrix were generated using Morpheus software provided by Broad Institute (RRID:SCR_017386).

Statistical Analysis

Mann–Whitney *U* test was performed to evaluate statistical significance of the number of mutations, mutation distribution as well as the ratio and type of hotspot mutations between MCC and healthy control sequences. Mann–Whitney *U* test was also performed to evaluate statistical significance of the median GFP loss compared with GFP4 vector without test sequence and for comparison between upstream and downstream constructs. One minus the Spearman rank correlation was used as the calculation method for heatmap hierarchical clustering and Spearman rank correlation for similarity matrices. Fisher exact test was used to calculate statistical differences of *AID/APOBEC*-expressing and nonexpressing MCC samples. Fisher exact test was also used to calculate statistical differences of high and low sun-exposed MCC tumors and their MCPyV status. Statistical analyzes were performed using GraphPad Prism 9 software.

Data Availability Statement

The RNA-seq data are available under the BioProject PRJNA775071 in NCBI Sequence Read Archive (SRA) database.

Results

MCPyV *LT* is Enriched in Cytosine-targeting Mutations in MCC

We compared substitution and insertion/deletion mutations in 83 sequenced healthy skin samples and 113 sequences from published MCC samples. Of the 318 mutations in the MCC samples, 292 (91.8%) were substitutions and 26 (8.2%) were insertions/deletions, while of the 90 mutations in the control sam-

ples, 88 were substitutions (88%) and two insertions/deletions (12%; Table 1). The substitutions occurred mostly at cytosine bases: 238 (81.5%) in the MCC and 49 (55.7%) in the control samples (Fig. 1A).

To determine the distribution of the mutations in the MCPyV T antigen area, we divided the mutations into 200 bp bins by normalizing the number of mutations in each bin area to the number of sequences covering the entire bin. In the control samples, the mutations had a relatively even distribution along the T antigen area (Fig. 1B), with some concentration toward the end of the *LT* gene (peaking at bins 2,200 bp and 2,885 bp). In contrast, the mutation frequency in MCC samples increased after 1,000 bp (bins ending at 1,200 bp, *P* = 0.0089; 1,400 bp, *P* < 0.0001; 1,600 bp, *P* = 0.0088; 1,800 bp, *P* = 0.0370; 2,000 bp, *P* = 0.0014). Because the bins 1,000–1,400 have more cytidines than the rest of the *LT* area (Fig. 1C), we normalized the number of substitutions to the number of given bases in each bin (Fig. 1D). This reveals that most of the substitutions enriched in MCC samples have indeed occurred at cytidines (bins ending at 1,200 bp, *P* = 0.0036; 1,400 bp, *P* < 0.0001; 1,600 bp, *P* = 0.0207; 1,800 bp, *P* = 0.0363; 2,000 bp, *P* = 0.0007; 2,400 bp, *P* = 0.0494) and that the cytidines account for most of the mutations between RB and helicase domains (bins ending at 1,200–2,000 bps).

Given that a truncation mutation in *LT* is a major oncogenic event in MCPyV+ MCC (18) and the fact that a substitution of a C is a likely way to generate a stop codon due to their lack of cytidines, we considered whether the large number of substitutions from C in MCC samples is a result of positive selection of stop codons.

While most (75.7%) of the substitutions targeting Cs were in the region spanning 1,200–2,200 bp (Fig. 1D), 6.5% of all substitutions in the T antigen area caused an in-frame stop codon (Table 1), and the in-frame stop codon distribution (Fig. 1E) explains only a very small proportion of the overall distribution of mutations in MCC samples (Fig. 1B). Instead, out of all insertion and deletion mutations, 77% caused an in-frame stop codon, all of which were in the 1,200–2,200 bp region (Table 1; Fig. 1F). Overall, 57.5% of the MCC sequences carried an in-frame stop codon. Many of the sequences analyzed here were partial, hence the amount of in-frame stop codons detected is likely an underestimation. Nevertheless, these findings recapitulate the previously described increase in premature stop codons in the *LT* (17, 23, 50), but do not explain the bias towards mutations at Cs. It is conceivable that the C bias arises from the activity of *AID/APOBEC* family of cytidine deaminases.

MCPyV Regulatory Region Shows Marginal SHM Recruiting Activity

Because of reported *AID* expression in MCC (6, 28), we wanted to test the role of SHM in causing mutations of the MCPyV *LT*. We tested the possibility of the MCPyV genome having *cis*-acting SHM recruiting sequences similar to

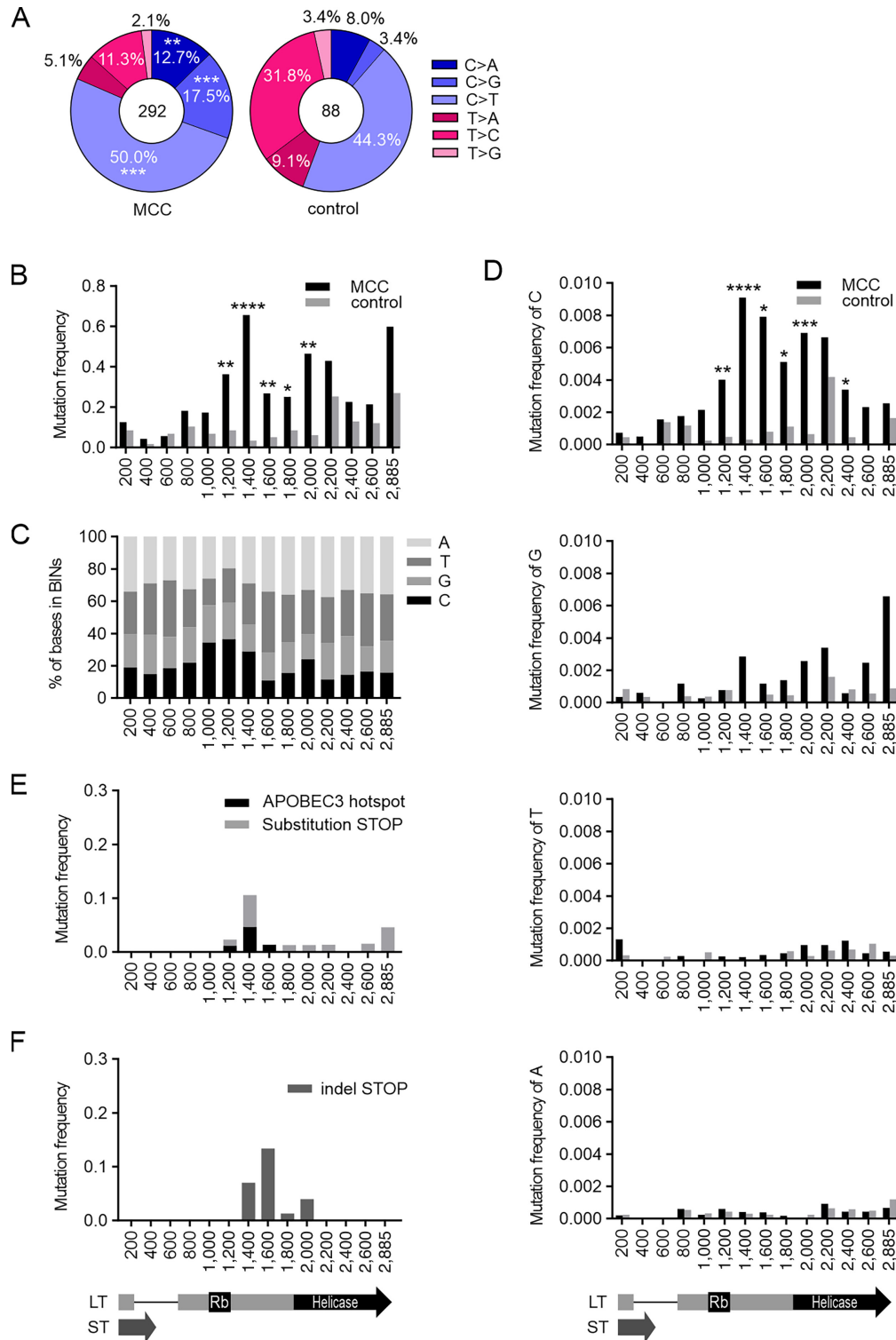


FIGURE 1 Mutations in MCV *LT* area in MCC and control samples. **A**, Substitution mutations in *LT* area in MCC and control samples in six mutation classes: C>A, C>T, C>G and T>A, T>C, T>G. Number of analyzed substitutions is shown in the middle. **B**, Distribution of substitutions along *LT* area is in MCC and control samples. *LT* area is divided to 200 bp bins (final bin 285 bp) and mutation frequency (number of mutations/number of sequences per bin) is shown. A schematic of *ST* and *LT* area is depicted below the graphs. **C**, Base content in *LT* area in each bin. **D**, Distribution of substitutions by mutated base in each bin. **E**, Distribution of in-frame stop codons introduced by substitution mutations in each bin (light gray). Black columns represent the proportion of in-frame stop codons that reside within APOBEC3 TCW hotspots. **F**, Distribution of in-frame stop codons in each bin introduced by insertion/deletion (indel) mutations before or in the indicated bin. Statistical significances between MCC and control mutations are determined using Mann-Whitney *U* test. **** <0.0001 , *** <0.001 , ** <0.01 , * <0.05 .

Downloaded from <http://aacrjournals.org/cancerrescommun/article-pdf/2/11/1344/3221064/crc-22-0211.pdf> by Turku University user on 09 December 2022

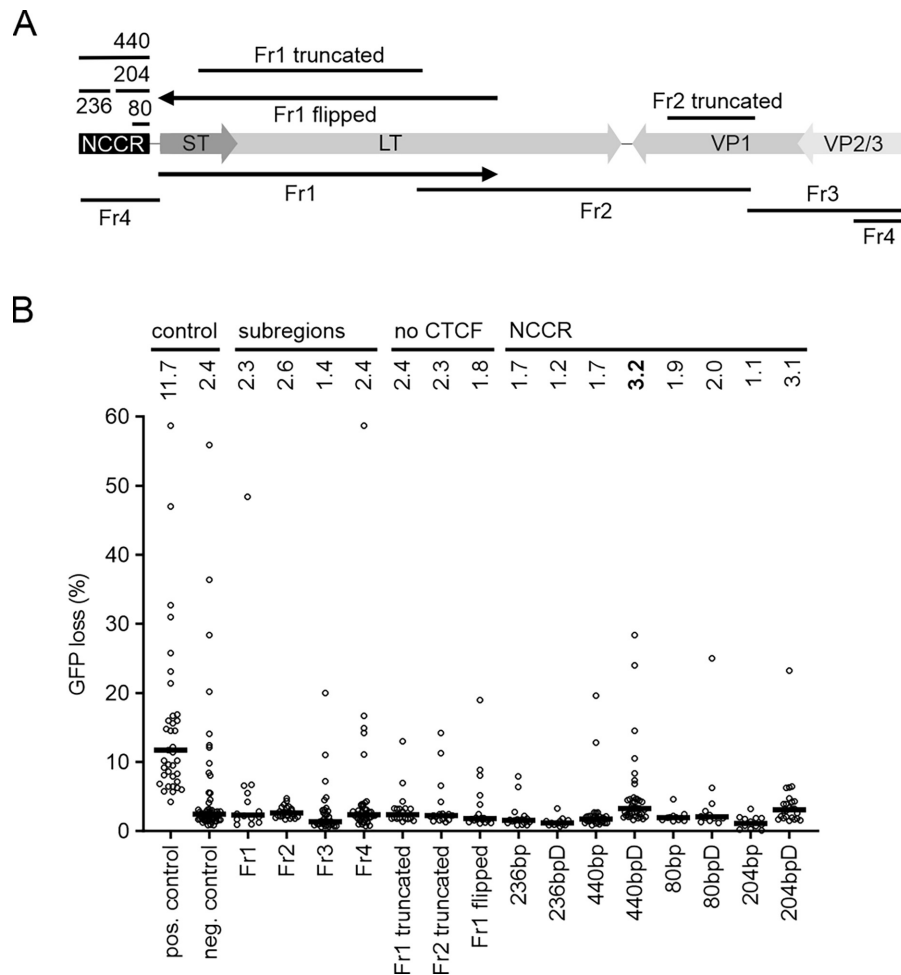


FIGURE 2 SHM targeting activity of MCPyV genome. **A**, Schematic of MCPyV genome and fragments tested in GFP loss assay. Arrows in “Fr1” and “Fr1 flipped” highlight the opposite orientation of these fragments relative to the GFP reporter (37). **B**, SHM targeting activity in MCPyV genome as measured in GFP loss assay. Four subregions (Fr1, Fr2, Fr3, and Fr4) of MCPyV genome as well as truncated Fr1, Fr2, and Fr1 in reverse orientation (Fr1 flipped) were tested (no CTCF). MCPyV NCCR (440 bp) and its 236, 80, and 204 bp subfragments were tested also downstream of the transcription unit (236bpD, 440bpD, 80bpD, and 204bpD). Human Ig lambda enhancer core (37) was used as positive control and the reporter without a test sequence was used as negative control. Median values are shown. Ten datapoints are outside of axis limits. Mann-Whitney *U* test was used to determine statistical significance. *P* values: pos. control <0.0001; Fr1 0.7863; Fr 2 0.4659; Fr3 0.0013; Fr4 0.7144; Fr1 truncated 0.9215; Fr2 truncated 0.7978; MCPyV Fr1 flipped 0.0901; 236bp 0.0009; 236bpD <0.0001; 440bp 0.0004; 440bpD 0.0161; 80bp 0.0672; 80bpD 0.4128; 204bp 0.0001; 204bpD 0.3783.

Ig loci. To do this, we performed a GFP-loss assay in DT40 B cells, which measures SHM recruitment activity (37). We first tested four partly overlapping regions of the MCPyV genome (Fr1, Fr2, Fr3, and Fr4; Fig. 2A). None of these regions showed significant SHM recruitment activity (Fig. 2B). SHM activity in the genome is restricted within topologically associating domains, whose boundaries may prevent the spreading of SHM from one domain to the next (51). Thus, any insulator, such as a CCCTC-binding factor (CTCF) binding site in the test DNA fragment between the potential active SHM-recruiting region and the *GFP* reporter gene, could prevent GFP loss. To address this, we tested subregions of Fr1 and Fr2 that do not contain CTCF sequences (Fr1 truncated and Fr2 truncated, respectively) and tested Fr1 in reverse orientation, which moves the CTCF site away from between the fragment core and the *GFP* transcription unit of the reporter (Fr1 flipped). None of these modified fragments exhibited SHM targeting activity (Fig. 2B).

Because SHM targeting elements in *Ig* loci are found in regulatory regions (enhancers), we also tested the MCPyV noncoding control region (NCCR) more carefully (Fig. 2). We tested NCCR fragments both upstream and downstream of the *GFP* expression cassette, as positioning the element upstream of the reporter affects the transcriptional output but not when positioned downstream (52). Interestingly, the downstream position of the 440 bp fragment (440bpD) showed low but statistically significant (3.2% GFP loss, $P = 0.0161$) SHM targeting activity (Fig. 2B). Further splitting the 440 bp fragment into smaller 80 and 204 bp fragments reduced the targeting activity below detection limit, but similarly to the 440 bp fragment, the 204 bp fragment had more SHM targeting activity in downstream than in upstream position (440bpD: 3.2% and 440bp 1.7%, $P < 0.0001$; 204bpD 3.1% and 204bp 1.1%, $P < 0.0001$). Therefore, we conclude that the NCCR of MCPyV can marginally recruit SHM to a neighboring transcription unit in cells that are capable of SHM.

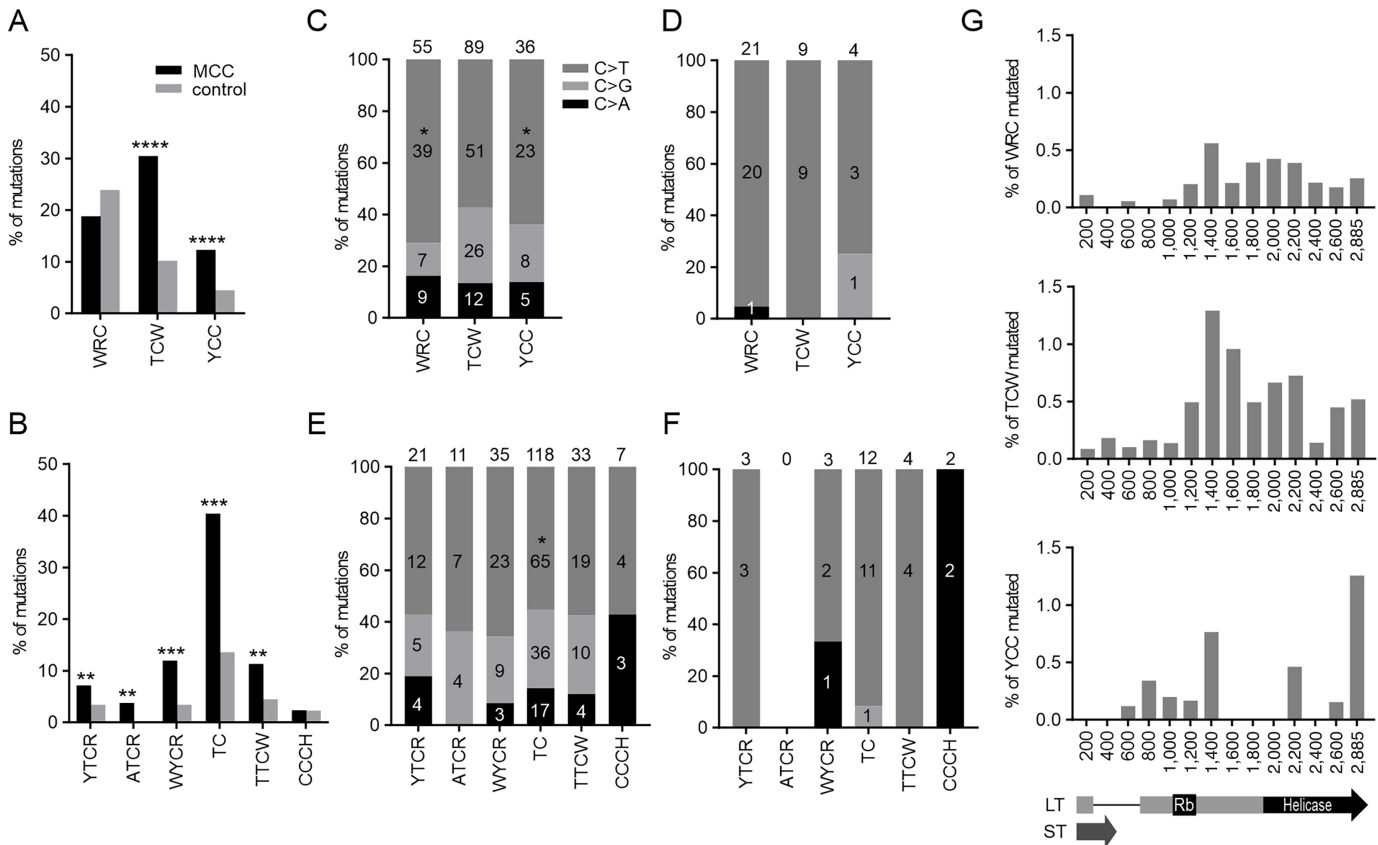


FIGURE 3 AID, APOBEC, and UV hotspot mutations in *LT* area. **A**, The proportion of substitutions in AID (WRC), APOBEC3 (TCW), and UV (YCC) hotspots in sequences from MCC (black) and control (gray) samples. **B**, Proportion of APOBEC3 subfamily hotspot mutations in YTCR, ATCR, WYCR, TC, TTCW, and CCCH in MCC samples and control samples. Colors as in **A**. **C**, The proportion of mutation type (C>T dark gray, C>G light gray and C>A black) in each hotspot in MCC samples. Number of each mutation type as well as total number of mutations in hotspots is indicated. **D**, The proportion of mutation type (C>T, C>G, and C>A) in each hotspot in control samples. Numbers and colors are as in **C**. **E**, The proportion of mutation type (C>T, C>G, and C>A) in each hotspot in MCC samples. Numbers and colors are as in **C**. **F**, The proportion of mutation type (C>T, C>G, and C>A) in each hotspot in control samples. Numbers and colors are as in **C**. **G**, Distribution of mutated WRC, TCW, and YCC hotspots along MCPyV *LT* area. Statistical significances between MCC and control hotspot mutations and their types are determined using Mann–Whitney *U* test. **** <0.0001, *** <0.001, ** <0.01, * <0.05.

Mutations in MCC *LT* Area Concentrate to APOBEC3 Hotspots

To explore the role of AID/APOBEC deaminases in *LT* mutations, we analyzed the immediate sequence context of *LT* substitution mutations for enrichment of AID hotspot sequences (WRC). We found that while 23.9% of mutations were in WRC hotspots in control samples, they were not enriched in MCC samples (18.0%, $P = 0.0630$; Fig. 3A).

As the mutations were clearly accumulated in C bases (Fig. 1), and cytidine deaminase APOBEC3 signature was recently established in MCPyV in the context of TC dinucleotides (36), we analyzed APOBEC3 hotspot mutations in *LT* in the TCW trinucleotide context. The proportion of mutations at APOBEC3 hotspots was increased 3-fold in MCC samples (30.5%) compared with healthy controls (10.2%, $P < 0.0001$; Fig. 3A). Interestingly, 42.9% of the in-frame stop codons caused by substitution mutations at Cs were also APOBEC3 hotspot mutations, out of which 60% were between RB and helicase domains (Fig. 1F). In contrast, none of the in-frame stop codons after this region were in APOBEC3 hotspots.

Individual APOBEC3 enzymes have slightly different specificities in addition to the general TCW motif (30). APOBEC3A prefers YTCR, APOBEC3B ATCR, APOBEC3C WYCR, APOBEC3D/H TC, APOBEC3F TTCW, and APOBEC3G CCCH (30). Therefore, we analyzed APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D/H, APOBEC3F, and APOBEC3G-hotspot mutations in more detail (Fig. 3B). Out of these hotspot motifs, we saw a proportional increase in APOBEC3A (2.1-fold, $P = 0.0019$), APOBEC3B ($P = 0.0056$), APOBEC3C (3.5-fold, $P = 0.0005$), APOBEC3D/H (3.0-fold, $P = 0.0001$), and APOBEC3F (2.5-fold, $P = 0.0025$) hotspot mutations (Fig. 3B). Mutations were not increased in APOBEC3G CCCH hotspots ($P = 0.3066$). However, there is considerable overlap between some of these hotspots, and mutations induced by the AID/APOBEC family are not strictly restricted to these motifs (29).

UV radiation is also a probable source of C mutations in skin cancers, and UV mutation patterns are observed in MCPyV– MCCs, but not in MCPyV+ MCCs (7). UV radiation causes frequent C>T and CC>TT mutations at dipyrimidine sites (53, 54). Because TCC or CCC trinucleotides are mutated most frequently in the single-base mutation signatures SBS7a and SBS7b of the

TABLE 2 Percentage of *AID/APOBEC* expressing samples and fold change of RNA expression level in MCPyV+ and MCPyV– groups

	<i>AICDA</i>	<i>APOBEC1</i>	<i>APOBEC2</i>	<i>APOBEC3A</i>	<i>APOBEC3B</i>	<i>APOBEC3C</i>	<i>APOBEC3D</i>	<i>APOBEC3E</i>	<i>APOBEC3G</i>	<i>APOBEC3H</i>	<i>APOBEC4</i>
MCPyV+	32.0	0.0	14.0	46.0	2.0	94.0	20.0	76.0	92.0	42.0	4.0
MCPyV–	34.4	3.1	21.9	37.5	6.3	96.9	18.8	90.6	93.8	9.4	3.1
<i>P</i>	>0.9999	0.3902	0.3812	0.4984	0.5574	>0.9999	>0.9999	0.1432	>0.9999	0.0024	>0.9999
Fold change MCPyV+/ MCPyV–	0.7	0.0	0.3	1.2	0.2	1.2	3.4	0.8	1.5	4.0	1.7

COSMIC database, we used *YCC* here as a UV hotspot motif. The difference in the proportion of mutations in *YCC* was 2.7-fold in healthy versus MCC samples (4.5% and 12.3%, respectively, $P < 0.0001$), but these mutations were infrequent (Fig. 3A). There were only three double-base substitutions at CC dinucleotides in the MCC samples: CC>AT, CC>GT, and GG>AA. It should be noted that using *YCC* as UV radiation hotspot may underestimate the number of UV-induced mutations. Using tumor location information combined to the viral status of the Finnish MCC samples (assessed by qPCR; refs. 55, 56), we determined that 35.6% of MCPyV+ MCCs appear in areas of body where low sun exposure is expected, whereas only 20.6% of MCPyV– MCCs appear at low sun-exposed areas. However, the difference was not statistically significant (Fisher exact test, $P = 0.1320$).

Analysis of the proportions of C>A, C>G, and C>T substitutions at each hotspot (Fig. 3C–F) showed that in the context of WRC and YCC hotspots, C>T substitutions were predominant, while in *APOBEC3* hotspots also C>G substitutions were common. This is in line with the mutation type distribution described for *AID*, *APOBEC*, and UV radiation (10, 29, 39, 42) and further implicates these factors as mutators in MCC. Note, that these classifications are not definitive, as several factors influence which base is incorporated to the site of initial lesion.

Curiously, *AID* and *APOBEC3* hotspots were mainly mutated after RB binding site (starting from 1,200 bp bin), whereas UV hotspot mutations seemed to have accumulated in the beginning of the *LT* area (up until 1,400 bp bin) and at the very end of the *LT* (at 2885 bp bin; Fig. 3G). Overall, *APOBEC3* had the highest hotspot mutation rate and the best resemblance with overall C-mutation distribution, enforcing the notion of *APOBEC3* signature in the *LT*.

Finnish MCCs Express *AICDA* and *APOBEC3s*

To investigate the expression of *AID/APOBEC* family proteins in MCC tumors, we analyzed the expression levels of *APOBECs* and *AICDA*, the gene encoding for *AID*, from 3' Tag RNA-seq data of 82 Finnish MCC tumors. 61.0% of the MCC samples expressed MCPyV *LT*, while 39.0% had undetectable levels of *LT* expression in this sample cohort. On the basis of *LT* expression, the tumors were divided into MCPyV+ and MCPyV– MCCs to detect whether the expression of *AID/APOBEC* family members is linked to *LT* expression. The 3' RNA-seq probably underestimates the number of MCPyV+ MCC samples, as in previous studies using other methods to analyze the same cohort, the number of MCPyV+ MCC samples was higher (27, 55, 57).

AICDA expression levels were 1.4-fold higher, and the proportion of *AICDA*-expressing samples was higher in the MCPyV– group (34.4%) than in the MCPyV+ group (32.0%; Table 2).

APOBEC1 and *APOBEC4* were expressed sparsely (Table 2), and their expression levels were very low. *APOBEC2* levels were higher (3.1-fold), and *APOBEC2* was more frequently expressed in the MCPyV– group than in the MCPyV+ group (21.9% and 14.0%, respectively). However, neither *APOBEC2* nor *APOBEC4* are known to have deaminase activity (29).

More tumors in the MCPyV+ group expressed *APOBEC3A* (46.0% vs. 37.5%), *APOBEC3D* (20.0% vs. 18.8%), and *APOBEC3H* (42.0% vs. 9.4%) than in the MCPyV– group, and their mRNA expression of was stronger (1.2-fold, 3.4-fold, and 4.0-fold, respectively; Fig. 4A). The majority of tumors expressed *APOBEC3C*, *APOBEC3E*, and *APOBEC3G*, and the expression of *APOBEC3C* and *APOBEC3G* was increased in MCPyV+ tumors (1.2- and 1.5-fold, respectively). A very small percentage of all MCCs in this sample cohort expressed *APOBEC3B*.

We performed hierarchical clustering for *AID/APOBEC* and *LT* expression data (Fig. 4A). The *LT* expression clustered together with *APOBEC3* subfamily, excluding *APOBEC3B*. *APOBEC3H*, and *APOBEC3G* expression correlated with *LT* expression (Fig. 4B; Spearman correlation *APOBEC3H* $r = 0.32$, $P = 0.004$; *APOBEC3G* $r = 0.25$, $P = 0.025$), suggesting a possible mechanistic link between MCPyV infection and *APOBEC3G/APOBEC3H* expression.

APOBEC3B excluded, *APOBEC3s* and *LT* coexpressed with markers of T lymphocytes (*CD8A*, *CD8B*, *CD4*, *CD3G*, and *granzyme B*; Fig. 4C). Also, *IFN γ* (*IFNG*), and *IFNG* regulating long noncoding RNA (*IFNG-AS*) coexpressed with these *APOBEC3s*. This is in line with the presence and potential contribution of tumor-infiltrating lymphocytes for *APOBEC3* expression in MCC tumors (36, 58). Interestingly, we found coexpression of B-lymphocyte markers (*CD19* and *MS4A1*, encoding for *CD20*) with most of these *APOBEC3s*. This suggests that as well as T lymphocytes, B lymphocytes also infiltrate MCC tumors.

In addition, germinal center B-cell markers (*BCL6*, *Ki67*, *CD4*, *ICOS*, and *PDCDI* encoding for PD-1) were expressed, which could indicate formation of tertiary lymphoid structures. Therefore, we cannot exclude tumor-infiltrating B lymphocytes rather than malignant cells as the cellular source of *AID* expression in the tumors.

Overall, our findings indicate clear *APOBEC3* signature and *APOBEC3* expression in MCPyV+ MCC tumors.

Discussion

The human *AID/APOBEC* gene family of cytosine deaminases codes for 11 proteins with various functions: *AID*, *APOBEC1*, *APOBEC2*, *APOBEC3A*,

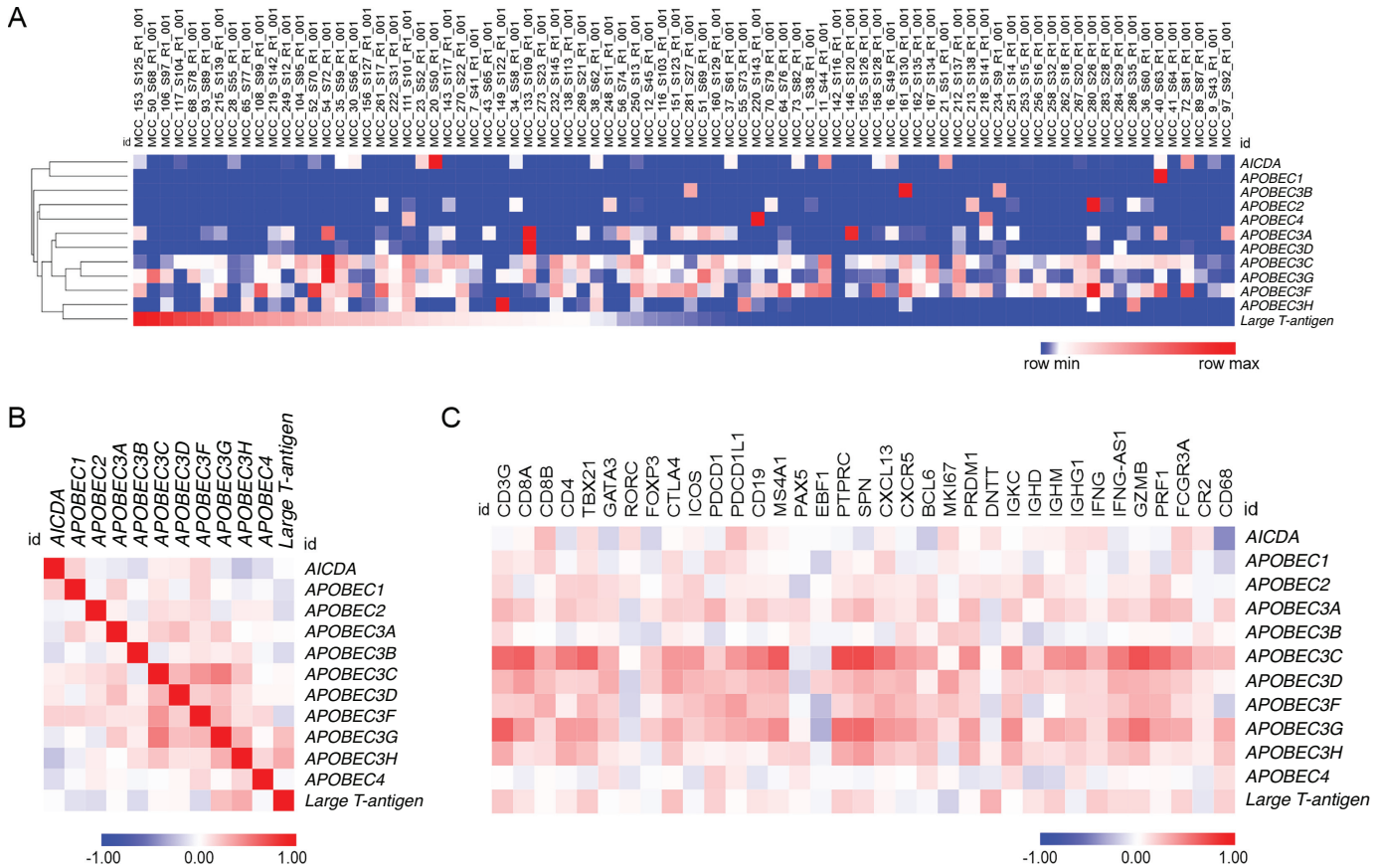


FIGURE 4 *AICDA* and *APOBEC* expression in Finnish MCC sample cohort. **A**, Heatmap of *AID/APOBEC* expression across individual tumors. Hierarchical clustering was performed with one minus Spearman rank correlation method. MCC samples are arranged according to *LT* expression level. **B**, Similarity matrix of *AICDA*, *APOBEC*, and *LT* expression. *APOBEC3H* and *APOBEC3G* are significantly coexpressed with *LT* (Spearman rank correlation *APOBEC3H* $r = 0.32$, $P = 0.004$ and *APOBEC3G* $r = 0.25$, $P = 0.025$). **C**, Similarity matrix of selected *T lymphocyte*, *B lymphocyte*, *NK-cell*, and *follicular dendritic cell markers* and their coexpression with *AICDA*, *APOBECs*, and *LT*.

APOBEC3B, *APOBEC3C*, *APOBEC3D*, *APOBEC3E*, *APOBEC3G*, *APOBEC3H*, and *APOBEC4* (refs. 29, 59; reviewed in ref. 60). *APOBEC2* and *APOBEC4* do not appear to catalyze cytosine deamination causing genomic mutations and *AID* drives SHM and CSR by deaminating single-stranded cytidines in *Ig*. The 7-member *APOBEC3* subfamily contributes to restriction of viral infections with both deaminase-independent mechanisms as well as through cytosine deamination (60). *APOBEC3G* was first found in defense against HIV-1, where virion-packed *APOBEC3s* can introduce G-to-A-mutations, viral DNA degradation, and lethal coding mutations of the genome-integrated virus (61). In addition to HIV-1, *APOBECs* are associated in defense against several cancer-associated viruses such as human papilloma virus, hepatitis B virus, and human T lymphotropic virus 1 (60, 62). *APOBEC3s* have low expression in a range of host cell types, but their expression is greatly increased upon viral infection and cytokine-induced stimuli. Not surprisingly, *APOBECs* and viruses have been engaged in an arms race of immune evasion evident in polymorphism and copy-number variation in *APOBEC3* gene as well as selective pressure on viral genomes sequences, which can be seen as a “footprint” in the genomes of a wide range of viral species, including *Polyomaviridae* (63).

While beneficial for antiviral immunity, the *AID/APOBEC* are known to induce genomic mutations and chromosomal aberrations with oncogenic outcomes in lymphocytes as well as in solid tissue cancers (reviewed in ref. 29). However, the contribution of *APOBECs* in cancer via mutations of a viral genome has remained poorly understood.

In this study, we found *APOBEC3* deaminase mutation signature in the MCC *LT* area. Up to 43% of in-frame stop codon-causing substitutions occurred in *APOBEC3* hotspots with the mutational outcome expected for *APOBEC3s*, and *APOBEC3* was widely expressed in MCPyV+ MCC tumors. Thus, our analysis implicates *APOBEC3s* as the primary mutators of *LT*.

APOBEC3s are largely agreed to target *TCW* motifs (29, 64). Focusing the analysis on this hotspot avoids mistakenly counting in UV-induced mutations, which frequently target pyrimidine dimers (54). CC>TT substitutions were absent in *LT*, and 35% of the Finnish MCPyV+ MCC tumors were in low sun exposure areas of the body. Thus, our findings do not support the role of UV in *LT* mutations.

Using a recently specified target motif preference for individual *APOBEC3* subfamily members (30), we could not discriminate a specific *APOBEC3* as the

Downloaded from <http://aacrjournals.org/cancerrescommun/article-pdf/2/11/1344/3221064/cre-22-0211.pdf> by Turku University user on 09 December 2022

primary mutator in the *LT*. As expected, due to a loose definition and overlap with other hotspots, the TC dinucleotides suggested as APOBEC3D/H hotspot (30) were most frequently mutated, agreeing with the conclusion drawn from characterizing mutations in MCPyV by using the TC dinucleotide as APOBEC3 hotspot (36).

MCPyV can induce APOBEC3A, APOBEC3B, and APOBEC3G expression in cell models (35, 36), where at least APOBEC3B can be upregulated by IFN γ (36). APOBEC3B has been the primary suspect out of the APOBEC family in cancers (65), but its expression was very rare in our relatively large MCC tumor dataset and did not significantly correlate with *LT* expression. Nevertheless, the host cell may have expressed *APOBEC3B* during initial MCPyV infection or earlier in tumor development before the tumor removal, because APOBEC expression is naturally episodic and fluctuating (66). Thus, the involvement of APOBEC3B cannot be entirely excluded. Nevertheless, we found coexpression of other *APOBEC3s* with *INFG* and *INFG* regulating noncoding RNA and our findings are in line with the IFN γ -induced APOBEC3 expression.

In contrast to *APOBEC3B*, the expression of *APOBEC3A* and *APOBEC3H* was widespread. There is accumulating evidence that APOBEC3A and 3H are equally or even more potent mutators in cancers than APOBEC3B (67–69). *APOBEC3A* and *APOBEC3H* were expressed in a larger proportion of Finnish MCPyV+ MCC samples (46.0% and 42.0%, respectively) than MCPyV– MCC samples (37.5% and 9.0%, respectively). The difference was statistically significant for *APOBEC3H*, which also showed statistically significant coexpression with *LT*. In addition, specific hotspot mutations (at YT CR and TC) were frequent, and also both APOBEC3A and APOBEC3H can mutate TCW (67–69). It should be noted that the mutation pattern and RNA expression data in our study came from different sources, making it possible, although unlikely, that the different genetic background of the subjects obscured our findings.

APOBEC3C, *APOBEC3F*, and *APOBEC3G* were expressed in the majority of the MCC samples. *APOBEC3C* and *APOBEC3G* levels were higher in MCPyV+ MCCs compared with MCPyV– MCCs, and *APOBEC3G* showed statistically significant coexpression with *LT*, which could implicate an immune response to MCPyV infection and make APOBEC3C and APOBEC3G potential *LT* mutators. Because *APOBEC3D* was expressed only marginally, and along with APOBEC3G and APOBEC3F, mainly reside in the cytoplasm (59), our data do not strongly support them as primary mutators.

Given the contribution of APOBEC3s in genomic cancer mutations and viral mutations in antiviral defense, our findings strongly suggest that APOBEC3 subfamily, particularly APOBEC3A, APOBEC3H, APOBEC3C, and to smaller extent APOBEC3G, contribute to the *LT* mutation signature and potentially premature stop codon formation.

AID is expressed in activated B lymphocytes and its association in B-lymphocyte cancers is well established (70) and has been implicated in skin cancers such as melanoma (71). We also detected *AICDA* expression in a subset of Finnish MCC samples, where the expression was stronger and slightly more frequently expressed in MCPyV– MCC samples, conforming with a previous finding (6). We also did not find AID mutation signatures enriched in MCC samples over control.

We found very weak, yet statistically significant, SHM targeting activity in the MCPyV NCCR. Therefore, it is possible that the SHM targeting activity possessed by the MCPyV NCCR could target AID-mediated SHM to the adjacent T

antigen area (or other nearby genes), similar to the enhancers in *Ig* loci. As SHM is a property of germinal-center B lymphocytes, the scenario of AID inducing carcinogenic mutations in MCPyV is more likely if MCC indeed derived from activated skin-resident B lymphocytes. Expression of B-lymphocyte markers in MCC (reviewed in 72) fit better for pro/pre-B lymphocyte stage. Thus, we conclude that AID-induced SHM is unlikely the primary mechanism for *LT* mutations, but its role in MCC carcinogenesis cannot be entirely excluded. Because there is a correlation between APOBEC mutations and DNA replication (29) and the NCCR also has the replication of origin for the MCPyV, it is conceivable that the mutations causing the GFP loss in our reporter system result from APOBEC activity in the DT40 B cells used to carry out the assay. This remains to be addressed experimentally.

Tumors are often infiltrated with immune cells (73), and T lymphocytes, macrophages, and natural killer cells were detected more frequently in MCPyV+ MCCs than in MCPyV– MCCs, demonstrating immune cell infiltration also in MCC (58). The role of B lymphocytes in tumor environment has not been studied as extensively as other immune cells, but there is evidence showing the importance of tumor-infiltrating B lymphocytes (74). We did observe expression of B-cell markers including *AICDA* in addition to T-lymphocyte markers in the RNA-seq data. It remains to be investigated whether a high number of tumor-infiltrating B lymphocytes could also explain the B-lymphocyte marker expression in MCC.

In conclusion, our findings support the view where APOBEC3 enzymes mutate MCPyV *LT* area and that this mutational activity is a major cause of premature stop codon formation and thus MCC carcinogenesis.

Authors' Disclosures

A.I. Soikkeli reports grants from Sigrid Juselius Foundation, The Finnish Cultural Foundation-The Kymenlaakso Regional Fund, Turku University Foundation, and The Maud Kuistila Memorial Foundation during the conduct of the study. J. Alinikula reports grants from Sigrid Juselius Foundation, Cancer Society of South-West Finland. The Emil Aaltonen Foundation, and Jane and Aatos Erkko Foundation during the conduct of the study. No disclosures were reported by other authors.

Authors' Contributions

A.I. Soikkeli: Formal analysis, funding acquisition, investigation, visualization, methodology, writing-original draft, writing-review and editing. **M.K. Kyläniemi:** Formal analysis, investigation, methodology, writing-review and editing. **H. Sihto:** Resources, data curation, funding acquisition, investigation, methodology, writing-review and editing. **J. Alinikula:** Conceptualization, resources, formal analysis, supervision, funding acquisition, investigation, visualization, methodology, project administration, writing-review and editing.

Acknowledgments

The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources, Virve Koljonen and Tom Böhling for collection of tumor samples, Alina Tarsalainen for language revision, Nea Tulonen and Ann Sofie Wierda for technical assistance, and Annamari Ranki for helpful insights. This work was financially supported by Sigrid Juselius Foundation (J. Alinikula, A.I. Soikkeli), The Finnish Cultural Foundation-The Kymenlaakso Regional Fund (A.I. Soikkeli), Cancer Society of South-West Finland

(J. Alinikula), Turku University Foundation (A.I. Soikkeli), The Maud Kuis-tila Memorial Foundation (A.I. Soikkeli), The Emil Aaltonen Foundation (J. Alinikula), Turku Doctoral Programme of Molecular Medicine (A.I. Soikkeli), Jane and Aatos Erkko Foundation (J. Alinikula), and University of Helsinki Funds (H. Sihto).

Note

Supplementary data for this article are available at Cancer Research Communications Online (<https://aacrjournals.org/cancerrescommun/>).

Received May 25, 2022; revised August 17, 2022; accepted September 28, 2022; published first November 04, 2022.

References

- Becker JC, Houben R, Ugurel S, Trefzer U, Pföhler C, Schrama D. MC polyomavirus is frequently present in Merkel cell carcinoma of European patients. *J Invest Dermatol* 2009;129: 248-50.
- Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science* 2008;319: 1096-100.
- Kassem A, Schöpflin A, Diaz C, Weyers W, Stickeler E, Werner M, et al. Frequent detection of merkel cell polyomavirus in human merkel cell carcinomas and identification of a unique deletion in the VP1 gene. *Cancer Res* 2008;68: 5009-13.
- Foulongne V, Dereure O, Kluger N, Molès JP, Guillot B, Segondy M. Merkel cell polyomavirus DNA detection in lesional and nonlesional skin from patients with Merkel cell carcinoma or other skin diseases. *Br J Dermatol* 2010;162: 59-63.
- Kervarrec T, Tallet A, Miquelstorena-Standley E, Houben R, Schrama D, Gambichler T, et al. Morphologic and immunophenotypic features distinguishing Merkel cell polyomavirus-positive and negative Merkel cell carcinoma. *Mod Pathol* 2019;32: 1605-16.
- Matsushita M, Iwasaki T, Nonaka D, Kuwamoto S, Nagata K, Kato M, et al. Higher expression of activation-induced cytidine deaminase is significantly associated with Merkel cell Polyomavirus-negative Merkel cell carcinomas. *Yonago Acta Med* 2017;60: 145-53.
- Harms PW, Vats P, Verhaegen ME, Robinson DR, Wu YM, Dhanasekaran SM, et al. The distinctive mutational spectra of polyomavirus-negative merkel cell carcinoma. *Cancer Res* 2015;75: 3720-7.
- Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature* 2013;500: 415-21.
- Brash DE. UV signature mutations. *Photochem Photobiol* 2015;91: 15-26.
- Chatterjee N, Walker GC. Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen* 2017;58: 235-63.
- Kuwamoto S, Higaki H, Kanai K, Iwasaki T, Sano H, Nagata K, et al. Association of Merkel cell polyomavirus infection with morphologic differences in Merkel cell carcinoma. *Hum Pathol* 2011;42: 632-40.
- Sunshine JC, Jahchan NS, Sage J, Choi J. Are there multiple cells of origin of Merkel cell carcinoma? *Oncogene* 2018;37: 1409-16.
- Ahmed MM, Cushman CH, Decaprio JA. Merkel cell polyomavirus: oncogenesis in a stable genome. *Viruses* 2022;14: 58.
- Park DE, Cheng J, Berrios C, Montero J, Cortés-Cros M, Ferretti S, et al. Dual inhibition of MDM2 and MDM4 in virus-positive Merkel cell carcinoma enhances the p53 response. *Proc Natl Acad Sci U S A* 2019;116: 1027-32.
- Kwon HJ, Shuda M, Feng H, Camacho CJ, Moore PS, Chang Y. Merkel cell polyomavirus small T antigen controls viral replication and oncoprotein expression by targeting the cellular ubiquitin ligase SCF Fbw7. *Cell Host Microbe* 2013;14: 125-35.
- Dye KN, Welcker M, Clurman BE, Roman A, Galloway DA. Merkel cell polyomavirus Tumor antigens expressed in Merkel cell carcinoma function independently of the ubiquitin ligases Fbw7 and β -TrCP. *PLoS Pathog* 2019;15: e1007543.
- Shuda M, Feng H, Kwon HJ, Rosen ST, Gjoerup O, Moore PS, et al. T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus. *Proc Natl Acad Sci U S A* 2008;105: 16272-7.
- Li J, Wang X, Diaz J, Tsang SH, Buck CB, You J. Merkel cell polyomavirus large T antigen disrupts host genomic integrity and inhibits cellular proliferation. *J Virol* 2013;87: 9173-88.
- Hesbacher S, Pfitzer L, Wiedorfer K, Angermeyer S, Borst A, Haferkamp S, et al. RB1 is the crucial target of the Merkel cell polyomavirus Large T antigen in Merkel cell carcinoma cells. *Oncotarget* 2016;7: 32956-68.
- Chang Y, Moore PS. Merkel cell carcinoma: a virus-induced human cancer. *Annu Rev Pathol* 2012;7: 123-44.
- Hausen AZ, Rennspiess D, Winneppinckx V, Speel EJ, Kurz AK. Early B-cell differentiation in merkel cell carcinomas: clues to cellular ancestry. *Cancer Res* 2013;73: 4982-7.
- Chteinberg E, Sauer CM, Rennspiess D, Beumers L, Schiffelers L, Eben J, et al. Neuroendocrine key regulator gene expression in merkel cell carcinoma. *Neoplasia* 2018;20: 1227-35.
- Liu W, MacDonald M, You J. Merkel cell polyomavirus infection and Merkel cell carcinoma. *Curr Opin Virol* 2016;20: 20-7.
- Calder KB, Smoller BR. New insights into Merkel cell carcinoma. *Adv Anat Pathol* 2010;17: 155-61.
- Liu W, Yang R, Payne AS, Schowalter RM, Spurgeon ME, Lambert PF, et al. Identifying the target cells and mechanisms of merkel cell polyomavirus infection. *Cell Host Microbe* 2016;19: 775-87.
- Schowalter RM, Pastrana DV, Pumphrey KA, Moyer AL, Buck CB. Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. *Cell Host Microbe* 2010;7: 509-15.
- Johansson B, Sahi H, Koljonen V, Böhlting T. The expression of terminal deoxynucleotidyl transferase and paired box gene 5 in Merkel cell carcinomas and its relation to the presence of Merkel cell polyomavirus DNA. *J Cutan Pathol* 2019;46: 26-32.
- Watabe R, Nakamura M. Expression of activation-induced cytidine deaminase in Merkel cell carcinoma with lymph-node metastasis. *Eur J Dermatol* 2013;23: 539-40.
- Siriwardena S, Chen K, Bhagwat AS. The functions and malfunctions of AID/APOBEC family deaminases: the known knowns and the known unknowns. *Chem Rev* 2016;116: 12688-710.
- Maiti A, Hou S, Schiffer CA, Matsuo H. Interactions of APOBEC3s with DNA and RNA. *Curr Opin Struct Biol* 2021;67: 195-204.
- Vartanian J-P, Guétard D, Henry M, Wain-Hobson S. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science* 2008;230: 230-3.
- Wang Z, Wakae K, Kitamura K, Aoyama S, Liu G, Koura M, et al. APOBEC3 deaminases induce hypermutation in human papillomavirus 16 DNA upon beta interferon stimulation. *J Virol* 2014;88: 1308-17.
- Shi K, Carpenter MA, Banerjee S, Shaban NM, Kurahashi K, Salamango DJ, et al. Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat Struct Mol Biol* 2017;24: 131-9.
- Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* 2013;45: 970-6.
- Verhalen B, Starrett GJ, Harris RS, Jiang M. Functional upregulation of the DNA cytosine deaminase APOBEC3B by polyomaviruses. *J Virol* 2016;90: 6379-86.

36. Que L, Li Y, Dainichi T, Kukimoto I, Nishiyama T, Nakano Y, et al. Interferon-gamma induced APOBEC3B contributes to Merkel cell polyomavirus genome mutagenesis in Merkel cell carcinoma. *J Invest Dermatol* 2021;142: 1793-1803.
37. Buerstedde JM, Alinikula J, Arakawa H, McDonald JJ, Schatz DG. Targeting of somatic hypermutation by immunoglobulin enhancer and enhancer-like sequences. *PLoS Biol* 2014;12: e1001831.
38. Bahjat M, Guikema JEJ. The complex interplay between DNA injury and repair in enzymatically induced mutagenesis and DNA damage in B lymphocytes. *Int J Mol Sci* 2017;18: 1876.
39. Álvarez-Prado ÁF, Pérez-Durán P, Pérez-García A, Benguria A, Torroja C, de Yébenes VG, et al. A broad atlas of somatic hypermutation allows prediction of activation-induced deaminase targets. *J Exp Med* 2018;215: 761-71.
40. Pasqualucci L, Dalla-Favera R. Genetics of diffuse large B-cell lymphoma. *Blood* 2018;131: 2307-19.
41. Liu M, Duke JL, Richter DJ, Vinuesa CG, Goodnow CC, Kleinstein SH, et al. Two levels of protection for the B cell genome during somatic hypermutation *Nature* 2008;451: 841-5.
42. Methot SP, Di Noia JM. Molecular mechanisms of somatic hypermutation and class switch recombination. *Adv Immunol* 2017;133: 37-87.
43. Feng H, Kwun HJ, Liu X, Gjoerup O, Stolz DB, Chang Y, et al. Cellular and viral factors regulating Merkel cell polyomavirus replication. *PLoS One* 2011;6: e22468.
44. Simon A. FastQC: a quality control tool for high throughput sequence data; 2010.
45. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016;32: 3047-8.
46. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29: 15-21.
47. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;31: 166-9.
48. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2009;26: 139-40.
49. Carlson M. org.Hs.eg.db: genome wide annotation for human. R package version 3.8.2; 2019.
50. Laude HC, Jonchère B, Maubec E, Carlotti A, Marinho E, Couturaud B, et al. Distinct merkel cell polyomavirus molecular features in tumour and non tumour specimens from patients with merkel cell carcinoma. *PLoS Pathog* 2010;6: e1001076.
51. Senigl F, Maman Y, Dinesh RK, Alinikula J, Seth RB, Pecnova L, et al. Topologically associated domains delineate susceptibility to somatic hypermutation. *Cell Rep* 2019; 29: 3902-15.
52. Tarsalainen A, Maman Y, Meng F-L, Kyläniemi MK, Soikkeli A, Budzyńska P, et al. Ig enhancers increase RNA polymerase II stalling at somatic hypermutation target sequences. *J Immunol* 2022;208: 143-54.
53. Luo JL, Hergenbahn M, Koomagi R, Yang Q, Hollstein M, Tong WM, et al. UV-induced DNA damage and mutations in hupki (human p53 knock-in) mice recapitulate p53 hotspot alterations in sun-exposed human skin. *Cancer Res* 2001;61: 8158-63.
54. Pfeifer GP, You YH, Besaratinia A. Mutations induced by ultraviolet light. *Mutat Res* 2005;571: 19-31.
55. Sihto H, Kukko H, Koljonen V, Sankila R, Böhling T, Joensuu H. Clinical factors associated with merkel cell polyomavirus infection in merkel cell carcinoma. *J Natl Cancer Inst* 2009;101: 938-45.
56. Sundqvist B, Sihto H, von Willebrand M, Böhling T, Koljonen V. LRIG1 is a positive prognostic marker in Merkel cell carcinoma and Merkel cell carcinoma expresses epithelial stem cell markers. *Virchows Arch* 2021;479: 1197-1207.
57. Sihto H, Kukko H, Koljonen V, Sankila R, Böhling T, Joensuu H. Merkel cell polyomavirus infection, large T antigen, retinoblastoma protein and outcome in Merkel cell carcinoma. *Clin Cancer Res* 2011;17: 4806-13.
58. Sihto H, Beohling T, Kavola H, Koljonen V, Salmi M, Jalkanen S, et al. Tumor infiltrating immune cells and outcome of merkel cell carcinoma: a population-based study. *Clin Cancer Res* 2012;18: 2872-81.
59. Salter JD, Bennett RP, Smith HC. The APOBEC protein family: united by structure, divergent in function. *Trends Biochem Sci* 2016;41: 578-94.
60. Harris RS, Dudley JP. APOBECs and virus restriction. *Virology* 2015;479-480: 131-45.
61. Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 2002;418: 646-50.
62. Zapatka M, Borozan I, Brewer DS, Iskar M, Grundhoff A, Alawi M, et al. The landscape of viral associations in human cancers. *Nat Genet* 2020;52: 320-30.
63. Poulain F, Lejeune N, Willemart K, Gillet NA. Footprint of the host restriction factors APOBEC3 on the genome of human viruses. *PLoS Pathog* 2020;16: e1008718.
64. Chan K, Gordenin DA. Clusters of multiple mutations: incidence and molecular mechanisms. *Annu Rev Genet* 2015;49: 243-67.
65. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet* 2013;45: 977-83.
66. Petljak M, Alexandrov LB, Brummel JS, Price S, Wedge DC, Grossmann S, et al. Characterizing mutational signatures in human cancer cell lines reveals episodic APOBEC mutagenesis. *Cell* 2019;176: 1282-94.
67. Petljak M, Maciejowski J. Molecular origins of APOBEC-associated mutations in cancer. *DNA Repair* 2020;94: 102905.
68. Cortez LM, Brown AL, Dennis MA, Collins CD, Brown AJ, Mitchell D, et al. APOBEC3A is a prominent cytidine deaminase in breast cancer. *PLoS Genet* 2019;15: e1008545.
69. Starrett GJ, Luengas EM, McCann JL, Ebrahimi D, Temiz NA, Love RP, et al. The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. *Nat Commun* 2016;7: 12918.
70. Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, et al. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J Biol Chem* 1999;274: 18470-6.
71. Okura R, Yoshioka H, Yoshioka M, Hiromasa K, Nishio D, Nakamura M. Expression of AID in malignant melanoma with BRAF V600E mutation *Exp Dermatol* 2014;23: 347-8.
72. Sauer CM, Haugg AM, Chteinberg E, Rennspiess D, Winnepenninckx V, Speel EJ, et al. Reviewing the current evidence supporting early B-cells as the cellular origin of Merkel cell carcinoma. *Crit Rev Oncol Hematol* 2017;116: 99-105.
73. Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol* 2013;14: 1014-22.
74. Laumont CM, Banville AC, Gilardi M, Hollern DP, Nelson BH. Tumour-infiltrating B cells: immunological mechanisms, clinical impact and therapeutic opportunities. *Nat Rev Cancer* 2022;22: 414-30.