TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

SOMATIC HYPERMUTATION TARGETING ACTIVITY OF PROTO-ONCOGENIC ENHANCERS

Anni Soikkeli

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1867 | MEDICA – ODONTOLOGICA | TURKU 2025

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

SOMATIC HYPERMUTATION TARGETING ACTIVITY OF PROTO-ONCOGENIC ENHANCERS

Anni Soikkeli

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1867 | MEDICA – ODONTOLOGICA | TURKU 2025

University of Turku

Faculty of Medicine
Institute of Biomedicine
Medical Microbiology and Immunology
Turku Doctoral Programme of Molecular Medicine (TuDMM)

Supervised by

Docent Jukka Alinikula, PhD
Institute of Biomedicine
University of Turku
Turku, Finland

Professor Emeritus Olli Lassila, MD, PhD
Institute of Biomedicine
University of Turku
Turku, Finland

Reviewed by

Docent Auli Karhu, PhD
Department of Medical and Clinical Genetics
University of Helsinki
Helsinki, Finland

Docent Maria Perdomo, MD, PhD
Department of Virology
University of Helsinki
Helsinki, Finland

Opponent

Docent Eeva Auvinen, PhD
Department of Virology
University of Helsinki
Helsinki, Finland

The originality of this publication has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

ISBN 978-952-02-0098-5 PRINT
ISBN 978-952-02-0099-2 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)
Painosalama, Turku, Finland 2025

To my family and friends

UNIVERSITY OF TURKU
Faculty of Medicine
Institute of Biomedicine
Medical Microbiology and Immunology
ANNI SOIKKELI: Somatic Hypermutation Targeting Activity of Proto-oncogenic Enhancers
Doctoral Dissertation, 154 pp.
Turku Doctoral Programme of Molecular Medicine (TuDMM)
April 2025

ABSTRACT

Somatic hypermutation (SHM) is an essential mechanism for the development of high-affinity antibodies. SHM occurs in germinal center B cells targeting antibody-coding immunoglobulin (*Ig*) genes to increase the affinity of antibodies towards the recognized antigens. SHM is catalyzed by the enzyme AID (activation-induced cytidine deaminase) which is a powerful mutator and is linked to the formation of several types of malignancies. Despite the importance of SHM for efficient immune response, it is a severe threat to genome integrity if targeted outside *Ig* loci. SHM targeting mechanisms are still largely unknown and even less is known about SHM off-targeting. Cis-acting elements in *Ig* loci, termed mutation enhancers, have been found to target SHM to *Ig* genes. In this thesis work, it was investigated if SHM off-targeting relies on mutation enhancers similar to SHM on-targeting.

The SHM targeting activity of viral and endogenic elements was tested and the changes this activity causes to neighboring genes was determined. SHM targeting activity from polyomavirus (SV40, JCPyV and MCPyV) regulatory regions was established and AID/APOBEC-induced mutation accumulation to viral *LT* -area was found. A subset of these mutations caused STOP-codon formation and in the context of SV40, truncated LT-protein expression linked to AID mutagenesis was observed. In the context of MCPyV, expression of truncated LT is tightly linked to formation of Merkel cell carcinoma. Whether truncated LT is relevant for SV40-induced tumorigenesis remains to be elucidated. SHM targeting activity was also found in the *BCL6* enhancer region. Preliminary evidence suggests that inserting a strong SHM targeting element into the *BCL6* enhancer region might increase mutation accumulation to the *BCL6* gene linked to lymphomagenesis, but this finding requires further confirmation.

In conclusion, SHM targeting activity was established outside *Ig* loci in viral and endogenic contexts. The results of this work revealed that SHM targeting activity is a more widespread feature than previously anticipated. SHM off-targeting activity of proto-oncogenic enhancers can predispose to malignant transformation given that the AID enzyme is expressed at some stage of tumor evolution.

KEYWORDS: Somatic hypermutation, AID, APOBEC, polyomavirus, B cell, LT, BCL6, oncogenesis

TURUN YLIOPISTO
Lääketieteellinen tiedekunta
Biolääketieteen laitos
Lääketieteellinen mikrobiologia ja immunologia
ANNI SOIKKELI: Somaattisen hypermutaation kohdentuminen esisyöpägeeneihin
Väitöskirja, 154 s.
Molekyylilääketieteen tohtoriohjelma (TuDMM)
Huhtikuu 2025

TIIVISTELMÄ

Genomiin kertyvien mutaatioiden määrä yritetään normaalisti pitää matalana. Immuunipuolustuksen B-solujen vasta-aineita koodittaviin geeneihin kertyy kuitenkin tarkoituksella mutaatioita ns. somaattisen hypermutaation (SHM) välityksellä. SHM:ta katalysoivan AID-entsyymin aiheuttamat mutaatiot parantavat vasta-aineiden sitoutumista kohteisiinsa, mikä tehostaa immuunipuolustuksen kykyä torjua elimistöön kohdistuvia uhkia. Mekanismi, jolla mutaatiot kohdentuvat juuri vasta-aine-geeneihin, tunnetaan huonosti. Osaltaan tapahtumaa säätelevät tietyt DNA-elementit vasta-aineiden säätelyalueilla. Ajoittain SHM kohdistuu virheellisesti vasta-ainegeenien ulkopuolelle, mikä voi johtaa solusyklin ja DNA:n korjauskoneiston häiriöihin altistaen syövän kehittymiselle. Tässä väitöstutkimuksessa selvitettiin, kohdentuuko SHM vasta-ainegeenien ulkopuolelle virus- ja endogeenisten geenien säätelyalueilta löytyvien elementtien avulla.

Tutkimuksen tulokset osoittivat, että polyoomavirusten (MCPyV, JCPyV ja SV40) säätelyalueet kohdentavat SHM:ta. Voimakkuus vaihteli heikosta erittäin voimakkaaseen. Lisäksi havaittiin, että MCPyV:n *LT*-geenin kertyy APOBEC3-entsyymien aiheuttamia mutaatioita, kun taas SV40:n *LT*-geenin mutaatiot olivat suurelta osin AID:n aiheuttamia. Molemmissa tapauksissa osa näistä mutaatioista aiheutti STOP-kodonien muodostumiseen, mikä johti SV40:n kohdalla lyhentyneen LT-proteiinin ilmentymiseen B-solulinjassa. MCPyV:n lyhentynyt LT aiheuttaa Merkelin solujen karsinoomaa, mutta on epäselvää, onko LT:n lyhentymisellä merkitystä SV40:n kyvylle aiheuttaa kasvaimia.

Tutkimuksessa löydettiin *BCL6*-geenin säätelyalueella useita elementtejä, jotka kohdentavat SHM:ta. Alustavat tulokset viittaavat elementtien kykyyn kohdistaa SHM:ta *BCL6*-geenin ensimmäisen introniin, johon mutaatioiden on toistuvasti näytetty kertyvän lymfoomissa. Tulosten varmistaminen vaatii kuitenkin lisäkokeita.

Tässä tutkimuksessa SHM:n kohdentamisaktiivisuutta löydettiin vasta-ainegeenien ulkopuolelta ja sen todettiin olevan yleisempi ominaisuus genomissa kuin aikaisemmin on ajateltu. Työssä näytetyllä kohdentamisaktiivisuudella voi olla merkitystä erityyppisten syöpien patogeneesissä, edellyttäen AID:n ilmentymistä syövän kehityksen aikana.

AVAINSANAT: Somaattinen hypermutaatio, AID, APOBEC, polyoomavirus, B-solu, LT, BCL6, onkogeenisyys

Table of Contents

Abbreviations

17k	17 kilodalton protein
57kT	57 kilodalton antigen
ABC	Activated B cell like
ac	Acetylation
ADCC	Antibody-dependent cellular cytotoxicity
AICDA	Activation-induced cytidine deaminase (gene)
AID	Activation-induced cytidine deaminase (protein)
AKT	AKT Serine/Threonine Kinase 1
ALL	Acute lymphoblastic leukemia
ALTO	Alternative Large Tumor Open Reading Frame
AML	Acute myeloid leukemia
AP-1/2	Activator protein 1/2
APOBEC	Apolipoprotein B mRNA-editing enzyme, catalytic polypeptide
ATM	ATM Serine/Threonine Kinase
ATR	ATR Serine/Threonine Kinase
BACH2	BTB Domain And CNC Homolog 2
BCC	Basal cell carcinoma
BCL2	B cell lymphoma 2
BCL6	B cell lymphoma 6
BCL7A	B cell lymphoma 7A
BCOR	BCL6 Corepressor
BCR	B cell receptor
BER	Base excision repair
BKPyV	BK polyomavirus
BL	Burkitt's lymphoma
bp	Base pair
BRCA1/2	BRCA1/2 DNA Repair Associated
BTB	Broad-Complex, Tramtrack and Bric a brac
BWA	Burrows-Wheeler aligner
CD	Cluster of differentiation
Cdk	Cyclin-dependent kinase

CDKN1A	Cyclin dependent kinase inhibitor 1A
CDR	Complementary-determining region
CHEK1	Checkpoint Kinase 1
cIgλ	Chicken Igλ enhancer
c-Kit	KIT Proto-Oncogene, Receptor Tyrosine Kinase
CLL	Chronic lymphocytic leukaemia
CMV	Cytomegalo virus
CREBBP	CREB Binding Protein
CRISPR	Clustered regularly interspaced short palindromic repeats
CSR	Class-switch recombination
CTCF	CCCTC-Binding Factor
C-terminal	Carboxyl-terminal
CXCL	C-X-C Motif Chemokine Ligand
CXCR	C-X-C motif Chemokine Receptor
DIVAC	Diversification activator
DLBCL	Diffuse large B cell lymphoma
DNA	Deoxyribonucleic acid
DNTM1	DNA Methyltransferase 1
dox	Doxycycline
E2A	Immunoglobulin enhancer-binding factors E12/E47
EBF1	Early B-cell factor 1
EBNA	Epstein–Barr nuclear antigen 1
E-box	Enhancer box, DNA binding element
EBV	Epstein-Barr virus
ELF1	E74 Like ETS Transcription Factor 1
ERK	Mitogen-Activated Protein Kinase 1
Ets	Erythroblast Transformation Specific
Ets1	ETS Proto-Oncogene 1, Transcription Factor
EXO1	Exonuclease 1
FACT	Facilitates chromatin transcription
FAM72A	Family with Sequence Similarity 72 Member A
FAS	Fas Cell Surface Death Receptor
FBS	Fetal bovine serum
FBXO11	F-Box Protein 11
FL	Follicular lymphoma
FR	Framework region
G4	G-quadruplexes
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GC	Germinal center
GCB	Germinal center B cell like

GFP	Green fluorescent protein
gRNA	Guide RNA
H2/H3	Histone 2/Histone 3
HBV	Hepatitis B virus
HBx	Hepatitis B virus regulatory protein X
HBZ	Tax-1 and HTLV-1 bZIP factor
HCV	Hepatitis C virus
hIgλ	Human Igλ enhancer
HIV-1	Human immunodeficiency virus 1
HMCES	5-hydroxymethylcytosine binding, ES cell specific
HPV	Human papillomavirus
HPyV	Human polyomavirus
HR	Homologous recombination
HTLV-1	Human T-cell Lymphotropic virus type 1
ID3	Inhibitor Of DNA Binding 3
Ig	Immunoglobulin
IgHi	Immunoglobulin heavy chain intronic enhancer
IgL	Immunoglobulin light chain
IRF	Interferon regulatory factor
JAK	Janus Kinase
JCPyV	John Cunningham polyomavirus
K	Lysine in histone complexes
kDa	Kilodalton
KSHV	Kaposi's sarcoma-associated herpesvirus
LANA1	Latent nuclear antigen 1
LPP	LIM Domain Containing Preferred Translocation Partner in Lipoma
LT	Large tumor antigen
MAPK	Mitogen activated protein kinase
MCC	Merkel cell carcinoma
MCPyV	Merkel cell polyomavirus
MDM2	Mouse double minute 2 homolog
me	Methylation
MEF2	Myocyte Enhancer Factor 2
MFI	Mean fluorescence intensity
MHC	Major histocompability complex
MIR-142	MicroRNA-142
MM	Multiple myeloma
MMR	Mismatch repair
mRNA	Messenger RNA
MSH	MutS Homolog

mTOR	Mechanistic Target of Rapamycin Kinase
MYC	MYC Proto-Oncogene, BHLH Transcription Factor
MYD88	MYD88 Innate Immune Signal Transduction Adaptor
NCOR	Nuclear Receptor Corepressor 2
NCRR	Non-coding regulatory region
NCS	Normal chicken serum
NF1	Nuclear factor 1
NF-kB	Nuclear factor kappa B
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
NHL	Non-Hodgkin lymphoma
NLS	Nuclear-localization sequence
NMD	Nonsense-mediated RNA decay
PAX5	Paired Box 5
PCNA	Proliferating Cell Nuclear Antigen
PD-L1	Programmed death ligand 1
PI3	Peptidase Inhibitor 3
PIM1	Serine/Threonine-Protein Kinase Pim-1
PML	Progressive multifocal leukoencephalopathy
Pol II	RNA polymerase II complex
Poz	Poxvirus and Zinc finger
PPA2	Protein phosphatase 2A
PRDM1	PR Domain Zinc Finger Protein 1
PRIMPOL	Primase and DNA Directed Polymerase
RAG	Recombination activating gene
RAS	Rat sarcoma
Rb1	Retinoblastoma 1
REV1	REV1 DNA Directed Polymerase
RHOH	Ras Homolog Family Member H
RNA	Ribonucleic acid
RPKM	Reads Per Kilobase per Million mapped reads
RSV	Rous sarcoma virus
RT	room temperature
RT-qPCR	Real-time quantitative polymerase chain reaction
S	Switch region
SBS	Single base substitution
SCC	Squamous cell carcinomas
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SHM	Somatic hypermutation
SNP	Single nucleotide polymorphism

Sp1	Specificity Protein 1
Spt5	SPT5 Homolog, DSIF Elongation Factor Subunit
ssDNA	Single-stranded DNA
sT	Small tumor antigen
STAT	Signal Transducer and Activator Of Transcription
SV40	Simian vacuolating virus 40
TAD	Topologically associated domain
Tax	Tax Gene Product
TCF3	Transcription Factor 3
TdT	Terminal deoxynucleotidyltransferase
TFBS	Transcription factor binding site
TNF	Tumor necrosis factor
TP53	Tumor protein p53
TPPK	Threonine–proline–proline–lysine motif
TSPyV	Trichodysplasia spinulosa polyomavirus
UMI	Unique molecular identifier
UNG	Uracil-DNA glycosylase
UV	Ultraviolet
VDJ	Variable, diversity and joining gene segments
VP	Viral capsid protein
WHO	World Health Organization
WNT	Wingless-type MMTV integration site family
YY1	Yin and Yang 1
ZCCHC7	Zinc Finger CCHC-Type Containing 7

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Soikkeli A.I., Kyläniemi M. K., Sihto H., Alinikula J**.** Oncogenic Merkel Cell Polyomavirus T Antigen Truncating Mutations are Mediated by APOBEC3 Activity in Merkel Cell Carcinoma. *Cancer Research Communications*, 2022; 2(11): 1344-1354.

II Šenigl F.*, Soikkeli A.I.*, Prost S., Schatz D.G., Slavková M., Hejnar J. Alinikula J**.** The SV40 virus enhancer functions as a somatic hypermutation-targeting element with potential tumorigenic activity. *Tumour Virus Research*, December 2024; Vol 18. 200293
* equal contribution

This dissertation also contains unpublished data.

1 Introduction

Antibodies produced by B cells are among the most important components of the adaptive immune system. Antibodies must recognize a massive number of antigens to provide extensive protection while avoiding self-recognition. Germinal center (GC) B cells adapt to this requirement by undergoing somatic hypermutation (SHM) in their antibody-coding immunoglobulin (Ig) variable (V) genes to optimize the antibodies for better recognition. SHM is catalyzed by an enzyme called activation-induced deaminase (AID) (Muramatsu et al., 1999, 2000; Revy et al., 2000). AID deaminates cytosine bases into uracils in single-stranded DNA. The U:G mismatches are converted into mutations or double-strand breaks by DNA repair mechanisms. AID is a member of the APOBEC family of cytidine deaminases which are powerful mutators and are linked to various human cancers (Siriwardena et al., 2016). While SHM is critical for a functional immune system, failure to correctly target SHM can have severe consequences for genome integrity. SHM off-targeting is most intimately linked to lymphomagenesis due to its predominant occurrence in B cells (Lauring & Basu, 2024). Lymphomas are malignancies of the lymphoid tissue, originating primarily from B or T cells. Most lymphomas are derived from B cells, which highlights the oncogenic potential of SHM and the effector enzyme AID (Küppers, 2005). Although SHM and AID expression are usually limited to GC B cells, AID expression can be induced by inflammatory signals or activated during oncogenesis due to genome dysregulation (Coussens & Werb, 2002). Thus, AID and SHM can cause oncogenic mutations beyond B cell-originating lymphomas, although these events are uncommon.

The current model of SHM targeting is based on AID piggybacking RNA polymerase II (Pol II) during transcription thus accessing the single-stranded DNA substrate and Pol II stalling, which gives a time window for AID to induce mutations (Kodgire et al., 2013; Nambu et al., 2003; Tarsalainen et al., 2022; Willmann et al., 2012). However, Pol II operates genome wide. AID expression occurrence predominantly in GC B cells restricts SHM to these cells. Still, additional layers of regulation must exist to explain why only *Ig* genes and certain off-target genes are targeted by SHM in B cells. An important factor in *Ig* loci is cis-acting DNA elements, mutation enhancers, which target SHM to *IgV* genes (Blagodatski et al.,

2009; Buerstedde et al., 2014). Resolving the SHM targeting mechanism has been an ongoing task for decades and many aspects are still undetermined. In this thesis work, the focus was on finding SHM-targeting elements outside *Ig* loci and investigating the consequences of SHM targeting activity of found elements to their target genes.

Regulatory regions of viruses contain similar features, such as transcription factor binding sites, as genomic enhancers. In addition, many viruses are able to infect B cells and B cells are considered a reservoir cells for several of the polyomavirus family members (Alaribe et al., 2013; Houff & Berger, 2008; Martini et al., 1998; McNees et al., 2019; Tan et al., 2009). Primary infections caused by polyomaviruses are common, resulting in usually mild or asymptomatic respiratory or gastrointestinal infections (Helanterä et al., 2016). After primary infection, polyomaviruses persist in host cells and can reactivate if the host immune defense is compromised (Decaprio & Garcea, 2013). Consequently, polyomaviruses can cause more severe diseases, although these events are rare. Some of these diseases, such as Merkel cell carcinoma and progressive multifocal leukoencephalopathy, are linked to mutations found in polyomavirus genome, but it is unclear what causes these mutations (Shuda et al., 2008; Sunyaev et al., 2009). Thus, in Studies I and II it was investigated if polyomavirus regulatory regions can target SHM and if evidence of AID/APOBEC enzyme-induced mutations can be found from the polyomavirus genome. The concept of viruses targeting SHM has not been studied before, and this work could provide a new mechanism for virus-induced pathogenesis.

The *BCL6* (B cell lymphoma 6 transcription repressor) locus is one of the most frequent sites of SHM off-targeting (Khodabakhshi et al., 2012). AID-induced mutations and translocations in the *BCL6*-gene leading to dysregulation of BCL6 expression are frequently detected particularly in diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL) (Migliazza et al. 1995; Pasqualucci et al. 1998; Shen et al. 1998). However, it is not known how SHM and AID are targeted to *BCL6*. This thesis includes unpublished data indicating that SHM is targeted to the *BCL6* gene by similar genomic elements as in *Ig* locus.

2 Review of the Literature

2.1 B cell development, antibody diversification and somatic hypermutation targeting mechanisms

2.1.1 Early B cell development and antigen-independent antibody diversification

Antibodies were found to be the base of humoral immunity already in 1890 and B cells producing them were found decades later (Cooper et al., 1965). Today the development of B cells is well characterized. Ultimately B cell development aims to the production of antibodies against every encountered antigen. Thus, several genetic mechanisms have evolved to respond to this challenge. B cells develop from hematopoietic stem cells, first to lymphoid precursors and then to immature B cells through intermediate stages in bone marrow. This maturation in an orderly fashion is guided by a set of transcription factors such as PAX5, E2A, EBF1 and Ikaros-family factors (Sigvardsson, 2023). During early development in bone marrow, antibody coding immunoglobulin (*Ig*) genes of B cells undergo V(D)J-recombination, which is the first step to produce massive numbers of antibodies recognizing different antigens. Antibody genes rearrange in V(D)J-recombination from sets of V (variable), D (diversity) and J (joining) segments catalyzed by RAG1 and RAG2 proteins among others (Sigvardsson, 2023). In antibody heavy chains, V and J segments are first joined together followed by D segment joining. In light chains, V and J segments are joined. Only after successful recombination of the heavy chain can developing B cells proceed to light-chain recombination. V(D)J-recombination alone can generate 10^{11} different combinations from VDJ-genes, thus being a massive source of antibody diversity (Abbas et al., 2022). Additional diversity is generated by P-nucleotide addition, exonuclease trimming and non-templated N-nucleotide addition. In addition, the same heavy chain can combine with λ or κ light chains (Abbas et al., 2022).

2.1.2 Germinal center reaction and antigen-dependent antibody diversification catalyzed by AID

The development of B cells continues in the spleen where transitional B cells differentiate to mature B-2 B cells. These cells circulate in blood and lymph and travel to B cell follicles in lymph nodes and spleen. There, antigen-activated B cells migrate to the border between the B cell follicle and the T cell zone. At the border, B cells receive signals from T helper cells, importantly CD40/CD40L, which initiates vigorous proliferation of B cells. Part of these cells migrate to the center of the follicle attracted by follicular dendritic cells. Follicular dendritic cells express CXCL13 -chemokine, which is recognized by B-cell receptor CXCR5. The forming clusters develop into germinal centers (GC) where antibody diversity is further increased in somatic hypermutation (SHM) and class-switch recombination (CSR).

SHM increases the affinity of antibodies towards the recognized antigen by introducing mutations to the *Ig V* region of both heavy and light chains. The *Ig V* regions are responsible for antigen binding. Increased affinity improves antibody neutralization capacity towards pathogens and toxins as well as enhances antibody-mediated cellular responses like antibody-dependent cellular cytotoxicity (ADCC).

CSR alters antibody class (IgM, IgG, IgA, IgD or IgE) to fit the current circumstances. IgM is secreted rapidly since it does not require CSR and commonly exists as a pentamer, providing multiple binding sites. IgA is secreted at mucous membranes; IgG persists long in the body and IgE is specialized in defence against parasites. SHM and CSR are catalyzed by the enzyme AID, encoded by the *AICDA* gene (activation-induced cytidine deaminase). AID is predominantly expressed in B cells and is required for both SHM and CSR (Muramatsu et al., 1999, 2000; Revy et al., 2000). Some species use a process called gene conversion for antibody diversification which is also catalyzed by AID (Tang & Martin, 2007).

GCs consist of two microenvironments called dark zone and light zone (Figure 1). Movement between these zones is guided by chemokine signals. Importantly, CXCL12 expressed by dark zone stroma cells guides proliferating B cells (centroblasts) to the dark zone. Centroblasts recognize CXCL12 with their CXCR4 receptor (Bannard et al., 2013). When the expression of CXCR4 is downregulated, centroblasts migrate to the light zone and halt the proliferation (Bannard et al., 2013).

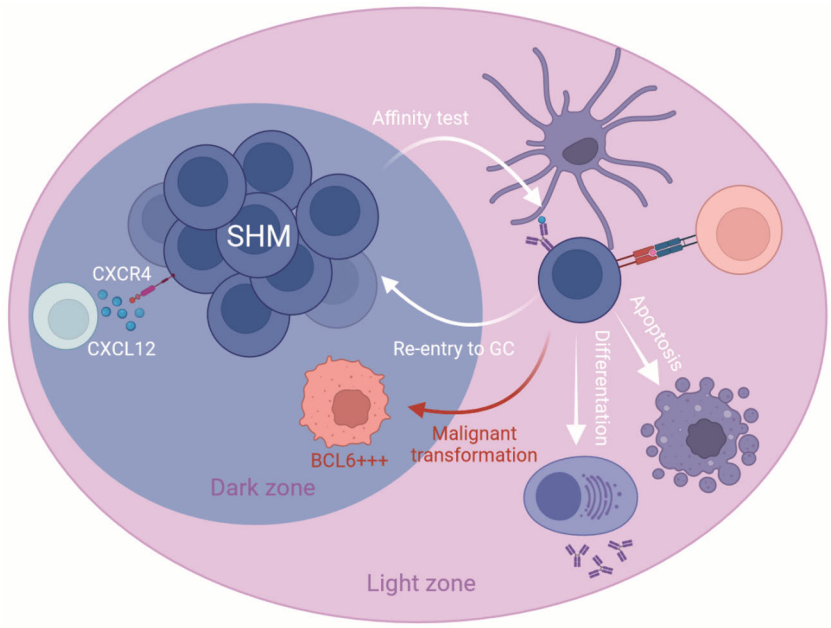

Figure 1. Germinal center reaction. Centroblasts in the dark zone proliferate vigorously and undergo SHM in their *Ig V* genes. Stroma cells secrete CXCL12 chemokine which is recognized by the CXCR4 receptor. This interaction guides the movement of the cells between the dark and the light zone. In the light zone, centrocytes express mutated BCRs on their surface and their affinity is tested by follicular dendritic cells and T helper cells. If the affinity of surface immunoglobulin is sufficient, centrocytes receive survival signals and differentiate into plasma cells or memory B cells. Centrocytes may also re-enter the dark zone and go additional rounds of SHM to achieve higher antibody affinity. If centrocytes do not receive survival signals from helper T cells they undergo apoptosis. If the *BCL6* gene, which allows the continuation of the cell cycle despite DNA damage, is overexpressed B cells may escape from cell cycle control, eventually leading to malignant transformation. Created with BioRender.com.

In the dark zone, the *Ig V* regions of vigorously proliferating centroblasts are mutated by SHM. The *Ig V* gene segments consist of three complementary determining regions (CDRs) where SHM is mainly targeted to and four framework (FR) regions in between CDRs (Schroeder & Cavacini, 2010). SHM mutation frequency in *Ig V* genes is orders of magnitude higher than elsewhere in the genome being around 1 mutation per 1000 base pairs (bp) in cell division (Liu, M. et al., 2008; Liu, M. & Schatz, 2009). In the *Ig V* region, mutations are targeted approximately 150 to 1500 bp downstream from the transcription start site (Liu, M. & Schatz, 2009). AID deaminates cytosines, inducing U:G mismatches to DNA, which can be processed in three ways. First, uracil may be interpreted as thymine by replication machinery, which results in a C>T transition. Second, uracil-DNA glycosylase (UNG) may recognize and remove uracil, followed by error-prone polymerase filling the gap in the base-excision repair (BER) pathway. Third, the mismatch repair (MMR) mechanism may excise longer DNA

segments around the original deamination site, followed by the error-prone polymerase (pol eta) repairing the missing strand. As a result, mutations arise also around the nucleotide originally deaminated by AID. Although AID deaminates cytosines, the type of resulting mutation depends largely on the used repair mechanism. In the physiological context, mutations are more evenly distributed to all four bases, but deficiency of DNA repair proteins UNG, MSH2 or MSH6 produces predominantly C>T or G>A transitions (Gu et al., 2019; Shen, H. M. et al., 2006). Recently, it was discovered that FAM72A can destabilize UNG thus promoting error-prone repair of AID-induced lesion (Feng, Y. et al., 2021; Rogier et al., 2021).

After mutation accumulation caused by SHM in the dark zone, centroblasts enter the GC light zone. There, the affinity of mutated B cell receptors (BCRs) expressed by non-dividing B cells (centrocytes) is tested. Follicular dendritic cells are coated with antigens which centrocytes recognize via mutated BCRs. Antigen is then processed and presented on the surface of centrocytes in MHC II molecules. Helper T cells recognize the antigen presented in MHC II molecules and provide more proliferation and survival signals to those centrocytes that efficiently present antigen. Thus, helper T cells function in selecting centrocytes with high-affinity BCRs.

Centrocytes interacting with follicular helper T cells and dendritic cells also initiate CSR. In CRS, IgM encoded μ-heavy chain is replaced with γ (1, 2, 3 or 4), α (1 and 2) or ε chain, which creates IgG1, IgG2, IgG3 or IgG4; IgA1 or IgA2 or IgE class antibodies, respectively. In addition, IgD is produced from IgM coding messenger RNA (mRNA) by alternative splicing. Cytokine signals guide which heavy chain is appropriate in the current condition. Mechanistically, AID initiates double-strand break by deaminating cytosine at *Ig* switch (S) regions in single-stranded DNA, resulting in uracil, which is not normally present in DNA (Bransteitter et al., 2003; Dunnick et al., 1993; Muramatsu et al., 2000; Storb, 1996). The resulting uracil is then removed by UNG, followed by apurinic/apyrimidinic endonucleases creating a nick to the DNA backbone (Xu et al., 2012). The nick is then processed to a double-strand break by MMR enzymes and ultimately repaired by non-homologous end joining (NHEJ) (Xu et al., 2012). Recently, the chromatin loop extrusion model was proposed to explain how the CSR process is conducted in orderly fashion (Qin & Meng, 2024). In addition, AID-induced abasic site shielding HMCES (5-hydroxymethylcytosine binding, ES cell-specific) -DNA repair factor seems to distinguish SHM from CSR by inhibiting abasic site processing to deletion and further to double-strand break (Wu et al., 2022).

B cells may circulate between dark and light zones several rounds to provide additional affinity selection. The germinal center reaction lasts around two weeks and as a result, future response to the same antigen is more rapid and effective. B cells that have undergone GC reaction ultimately develop into high-affinity antibody producing plasma cells or memory B cells.

2.1.3 Regulation of AID

AID is a member of DNA and RNA editing cytosine deaminase APOBEC family. AID is a powerful mutator with serious oncogenic potential and it is regulated at many levels. AID expression is mostly but not exclusively limited to B cells. AID regulation mechanisms include regulation of transcription and translation, mainly cytoplasmic localization, active transportation in and out of the nucleus and activity regulation through phosphorylation (Basu et al., 2005; Ito et al., 2004; Keim et al., 2013; Rada et al., 2002). AID deaminates single-stranded DNA (ssDNA) substrate, therefore transcription of target gene is required. AID-induced deamination is more frequent at cytidines in WR**C** or WR**C**Y context, whose density in a DNA sequence affects the amount of AID-induced mutations (Liu, M. et al., 2008; Pham et al., 2003; Wang, Y. et al., 2023). In addition to short hotspot sequences, mesoscale features spanning 5-50 bp around mutation site affect AID mutation frequency. AID prefers hotspots surrounded by pyrimidine-pyrimidine sites which makes DNA more flexible and less stacked due to the weak stacking strength of pyrimidines (Wang, Y. et al., 2023). This will presumably optimize the orientation of DNA, allowing the AID catalytic pocket to gain access to cytidines (Wang, Y. et al., 2023).

Certain secondary structures in DNA are preferred by AID. R-loops and G-quadruplexes (G4) are found frequently in *Ig* gene S regions, which are GC-rich with a strong G-bias in the coding strand (Dézé et al., 2023; Duquette et al., 2004; Roy et al., 2008; Yu, K. et al., 2003). R-loop is a stable structure composed of a hybrid template-DNA:RNA-structure and single-stranded DNA non-template strand, the latter of which can be efficiently deaminated by AID (Roy et al., 2008; Yu, K. et al., 2003). G4-structures form often spontaneously in guanine-rich regions of DNA through Hoogsteen hydrogen bonds between stacked planar guanine tetrads (De Magis et al., 2019; Sundquist & Klug, 1989). The precise mechanism of G4s in regulating AID is unclear, but G4s form in vitro to S regions, are suspected to recruit AID and bind AID with 10-fold higher affinity than linear sequences (Qiao et al., 2017; Yewdell et al., 2020; Zheng, S. et al., 2015). In addition, AID binds directly to G4s in RNA, and if this binding is blocked, recruiting AID to DNA in S regions is impaired (Zheng, S. et al., 2015). As R-loops frequently form in G-rich sequences, G4s and R-loops can co-occur and G4 structures can increase R-loop stability, but they can also appear independently (De Magis et al., 2019; Roy et al., 2008). To conclude, direct AID regulation as well as preferred targeting sequences and DNA-structures are important for limiting AID-induced mutations. However, they are not sufficient to explain AID actions predominantly in B cell *Ig* loci.

2.1.4 SHM targeting mechanisms

Despite the importance of SHM regulation to avoid genomic instability, the mechanisms of how SHM is targeted accurately to *Ig* CDRs are largely unknown. However, many features are suggested to associate SHM targeting (Figure 2). Mere binding of AID to DNA does not induce mutations to *Ig V* genes (Matthews et al., 2014). The current model is based on the interaction of AID with Pol II machinery and promotion of Pol II pausing, which creates an extended time frame for the deamination process of AID (Kodgire et al., 2013; Nambu et al., 2003; Tarsalainen et al., 2022; Willmann et al., 2012). AID is recruited to the Pol II complex trough interactions with Spt5 (SPT5 Homolog, DSIF Elongation Factor Subunit) and the RNA exosome complex (Basu et al., 2011; Pavri et al., 2010; Pefanis et al., 2014; Wang, X. et al., 2014). Indeed, Spt5 accumulation correlates with SHM levels in *Ig V* genes and RNA exosome stimulates AID-induced mutations to both DNA strands (Basu et al., 2011; Maul et al., 2014). It has been proposed that three arginine residues of AID promote the AID- Pol II complex to progress from the transcription start site to the gene body (Methot et al., 2018). This prolongs AID–ssDNA interaction, thus "licensing" AID to induce mutations.

Eukaryote DNA is wound around histone proteins forming nucleosomes. Histones can be modified trough acetylation or methylation which regulates gene expression. As AID requires ssDNA substrate to gain access to cytidine, open chromatin and active transcription of the target site are needed. Histone H3.3 modification and Pol II elongation marker FACT (Facilitates Chromatin Transcription) complex are connected to high rates of SHM (Aida et al., 2013; Yu, G. et al., 2021). Histone markers for transcription elongation (H3K36me3, H3K79me2), enhancers (H3K27Ac, H3K4me1) and active promoters (H3K4me3, H3K9Ac) are detected in SHM-susceptible regions (Álvarez-Prado et al., 2018; Begum et al., 2012; Qian et al., 2014; Wang, Q. et al., 2014). In addition, histone H2A and H2B are shown to be monoubiquitinated at loci undergoing SHM (Borchert et al., 2010). Further, the histone chaperone Spt6 was shown to regulate AID targeting through regulation of H3K4me3 and H3K36me3 modifications (Begum et al., 2012). However, these regulation mechanisms and observed markers correlating with SHM are present genome wide. Thus, they are not sufficient to explain why SHM is only targeted to certain genomic locations in B cells.

Importantly, certain Ig-enhancers and Ig-enhancer-like sequences at *Ig* loci, here after referred to as mutation enhancers, target SHM to *Ig* genes (Blagodatski et al., 2009; Buerstedde et al., 2014; Kohler et al., 2012; Kothapalli et al., 2008; McDonald et al., 2013). In previous studies, mutation enhancers have also been referred to as diversification activators (DIVAC) (Blagodatski et al., 2009). Ig mutation enhancers are both required and sufficient for SHM targeting and are conserved in evolution (Blagodatski et al., 2009; Buerstedde et al., 2014). B cell transcription factor binding

sites (TFBS) E-box, IRF, NF-kB, Pu.1, MEF2 and YY1 in Ig mutation enhancers are important for SHM targeting, as deletion of them either individually or in combination decreases or abolishes mutational activity (Buerstedde et al., 2014; Kohler et al., 2012; McDonald et al., 2013). In addition, E2A, MEF2B, Aiolos, and YY1 proteins have been shown to bind their corresponding sites (Dinesh et al., 2020). TFBSs seem to work in a cooperative manner and one TFBS can be compensated with another (Buerstedde et al., 2014; Dinesh et al., 2020).

How mutation enhancers target SHM is still unclear. Mutation enhancers increase Pol II serine 5 and serine 2 phosphorylation in the presence of Spt5 and Pol II stalling in the target gene, and thus seem to promote SHM (Kohler et al., 2012; Tarsalainen et al., 2022). Mutation enhancers can under certain conditions increase the availability of ssDNA in the mutating gene, but this does not always correlate with mutational activity (Tarsalainen et al., 2022). Despite recent advances in studies of SHM targeting, the precise mechanism of action remains unresolved and further investigations are required.

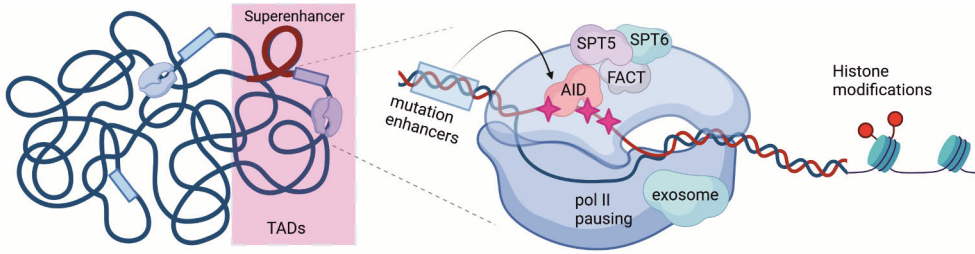

Figure 2. Features associated with SHM targeting. The current model proposes that AID piggybacks RNA polymerase II complex thus gaining access to ssDNA substrate. In addition, Pol II pausing is thought to provide a sufficient time window for AID to deaminate cytidines. The contributing factors in SHM targeting are mutation enhancers, Pol II-associated protein factors Spt5, Spt6, FACT and the RNA exosome, several histone markers, super-enhancer status linked with Pol II collision model, interconnected regions in DNA and topologically associated domains (TADs). Created with BioRender.com.

2.2 Oncogenic potential of SHM

2.2.1 SHM off-targeting

SHM off-targeting outside the *Ig* locus was first described in the *BCL6 (B cell lymphoma 6 transcription repressor)* gene (Pasqualucci et al., 1998; Shen, H. M. et al., 1998). Since then, several other off-target genes have been discovered, most recurrent being c-*MYC*, *PAX5*, *PIM1*, *BACH2*, *FAS*, *CD79A/B*, *RHOH*, *CD95*, *BCL7A*, *MSH6*, *MIR-142* and *ID3* (Gordon et al., 2003; Khodabakhshi et al., 2012; Liu, M. et al., 2008; Müschen et al., 2000; Pasqualucci et al., 2001; Qian et al., 2014).

The number of estimated off-target genes ranges from tens to several hundred (Álvarez-Prado et al., 2018; Hübschmann et al., 2021; Khodabakhshi et al., 2012; Meng et al., 2014; Qian et al., 2014). This estimation is relatively high considering the threat which SHM off-targeting possess to genome integrity. Mutation frequency in off-target genes is lower than in *Ig* genes, which may be due to more stringent monitoring of DNA lesions outside *Ig* loci, but higher than elsewhere in the genome (Bal et al., 2022; Liu, M. et al., 2008; Martin et al., 2018; Pasqualucci et al., 1998). Off-targeting is frequently detected in promoters and 5' regulatory regions which is in line with the on-targeting range of around 2000 base pairs from the transcription start site (Bal et al., 2022; Hübschmann et al., 2021; Leeman-Neill et al., 2023; Liu, M. & Schatz, 2009; Pasqualucci et al., 2003). Currently, four non-Ig mutation enhancers have been identified from *ELF1, MSH6, ZCCHC7* and *BCL6* enhancer regions (Senigl et al., 2019).

SHM-associated epigenetic features in *Ig* loci such as chromatin modifications H3K27ac and H3K36me3, convergent and divergent transcription, super-enhancers, enhancer–promoter connections, RNA exosome-mediated non-coding RNA degradation and RNA polymerase II (pol II) stalling have been implicated to associate in SHM off-targeting (Álvarez-Prado et al., 2018; Basu et al., 2011; Meng et al., 2014; Pefanis et al., 2014; Qian et al., 2014; Wang, Q. et al., 2014). AID off-targets in B cells are located largely in close physical proximity, both in linear DNA and physiological space through long-range interactions (Qian et al., 2014). Experimental evidence suggests that an individual topologically associated domain (TAD) contains more than one SHM off-target gene (Qian et al., 2014). TADs are regions usually defined by CTCF binding sites and cohesion rings, creating a chromatin loop where long-range genomic interactions may occur (Gibcus & Dekker, 2013). Indeed, some TADs in the genome are shown to be more susceptible to SHM (Senigl et al., 2019). SHM-prone TADs possess many characteristics previously described as important for SHM targeting such as epigenetic markers H3K27ac and H3K4me1 and transcription factors E2A, IRF4, PU.1, MEF2B, NF-kB, Aiolos, Ikaros and BCL6 (Senigl et al., 2019). The most significant SHM-susceptible TAD marker in this study was cohesion-loader NIPBL, which is required but not sufficient to alone induce SHM targeting (Senigl et al., 2019). Inserting known mutation enhancer to a non-SHM-susceptible TAD in the genome can turn it into SHM-susceptible (Senigl et al., 2019). Thus, TADs possessing similar factors as *Ig* loci have a role in limiting SHM to certain locations in the genome.

AID off-targeting coincides with super-enhancers. Super-enhancers are large clusters of gene-regulating elements (enhancers), spanning usually over 20 000 bp region (Jia et al., 2019). They have increased ability to bind transcription co-activators and promote transcription. AID off-targets are often cell-type specific, meaning that certain genes are targeted depending on the cell type, rather than the

same genes being targeted universally across all cell types (Qian et al., 2014). However, these off-targets often locate in super-enhancer-rich sites despite the cell type. Thus, SHM targeting is likely based on cell type-specific transcription factors rather than solely B cell-specific factors. Off-target genes often exhibit convergent transcription, which is caused by intragenic super-enhancers (Meng et al., 2014). In the proposed super-enhancer-induced convergent transcription model (Figure 3), one Pol II complex starts transcription from the gene promoter along the gene body, while another Pol II initiates transcription from an intragenic super-enhancer within the gene body, moving towards the gene promoter (Meng et al., 2014). Pol II complexes moving towards each other collide and are not able to pass each other, stalling the elongation (Hobson et al., 2012). This creates a potential time window for AID to access and mutate ssDNA (Meng et al., 2014).

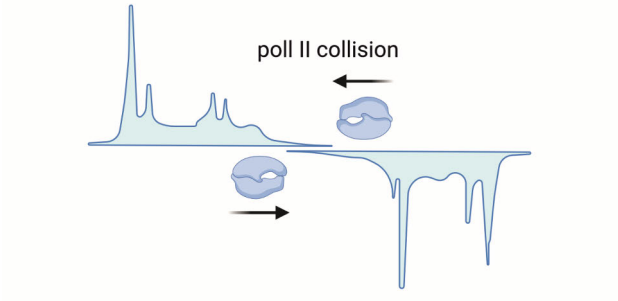

Figure 3. The Pol II collision model. Polymerases starting the transcription simultaneously from the gene promoter and the gene body collide and are unable to get pass each other. This creates time window for AID to induce mutations as transcription is stalled and the ssDNA substrate is freely available. Created with BioRender.com

Super-enhancers coincide with transcriptionally active, open chromatin regions where AID gains access to ssDNA. Indeed, AID-induced mutations are often found in highly transcribed and topologically associated regions, which contain genes responsible for B cell development such as *PAX5, AICDA, BACH2* and *NF-kB* as well as apoptosis and proliferation such as *MYC* and *BCL2* (Qian et al., 2014). In addition to mutating off-target gene bodies, non-coding super-enhancer regions of these genes are commonly mutated (Leeman-Neill et al., 2023; Qian et al., 2014). Divergent, noncoding RNA producing, transcription is also associated with AID off-targeting (Pefanis et al., 2014). However, neither convergent nor divergent transcription is required for efficient SHM targeting by *Ig* mutation enhancers in a reporter system (Tarsalainen et al., 2022). This indicates that inducing convergent or divergent transcription is not the key mechanism of how mutation enhancers promote SHM targeting. Finally, AID preference for a pyrimidine-pyrimidine surrounding was also observed in AID off-targets in diffuse large B cell lymphoma (DLBCL)

data (Wang, Y. et al., 2023). In addition, both R-loop and G4 -structures are associated with genome instabilities (De Magis et al., 2019; Zheng, S. et al., 2015). Together, these findings support the view that SHM on- and off-targeting are regulated by the same mechanisms, but the issue requires further investigation.

2.2.2 Consequences of genome mutations

Impaired AID function and SHM off-targeting are a serious threat to genome instability, which can eventually contribute to cancer onset and progression. In addition to AID, several other factors can induce mutations in the genome. These include other mutagenic enzymes such as APOBECs, ultraviolet (UV) and ionizing radiation, chemicals such as asbestos or tar, reactive oxygen radicals, tobacco smoke, spontaneous deamination and defects in the DNA repair system. Mutational signature profiles are developed to distinguish mutation etiologies from cancers (Alexandrov et al., 2013, 2020; Nik-Zainal et al., 2012). As shown in these works, certain mutation signatures appear in certain malignancies, e.g., APOBEC family enzyme signatures, defective mismatch repair and deamination of 5'methylcytosines appear in B cell-derived non-Hodgkin lymphomas (NHLs), whereas UV-induced signatures predominate in skin melanomas (Alexandrov et al., 2020). In addition, more than one signature can appear in the same type of malignancy (Alexandrov et al., 2020).

AID and other carcinogens introduce point mutations (substitutions, deletions, and insertions) and translocations leading to various outcomes. Point mutations resulting in substitutions can be synonymous, where the codon is altered but the amino acid remains unchanged, or non-synonymous, where the substitution changes the encoded amino acid. Non-synonymous mutations can be nonsense mutations leading to STOP-codon formation or missense mutations, which can lead to non-functional protein formation. Point mutations resulting in insertions or deletions can cause frameshifts and thus premature STOP-codon formation, which may lead to a non-functional or truncated protein. A premature STOP-codon can also trigger nonsense-mediated RNA decay (NMD) which will destroy the produced mRNA and no protein is expressed (Nogueira et al., 2021). Point mutations at splice sites may alter mRNA splicing, causing, for example, missing exons or truncated mRNA. Additionally, mutations in gene regulatory regions may affect protein dosage, leading to excessive or reduced expression or the formation of unstable protein.

Translocations may relocate genes under the control of different regulatory elements, either increasing or reducing their expression. Translocations can also create new fusion genes with functions that differ from the original or disrupt gene function entirely. A special feature of APOBEC family enzymes, including AID, is the ability to induce chromosomal abnormalities through local mutation showers

termed kataegis (Alexandrov et al., 2013; Casellas et al., 2016). Kataegis can cause simultaneous rearrangements in the area of one or more chromosomes, which can have catastrophic consequences for genome integrity (Casellas et al., 2016).

Depending on their effects on gene function, mutations are typically categorized as gain-of-function or loss-of-function. Gain-of-function mutations amplify the normal function of genes, and these mutations are common in proto-oncogenes. Loss-of-function mutations deplete gene function and are common in tumor-suppressor genes. Predominantly, both alleles should be mutated to cause the effect. However, one mutant allele can interfere with the function of the normal allele, or a single normal allele may not be sufficient to maintain gene function. In these cases, mutation of one allele in tumor-suppressor genes is sufficient to cause the effect.

2.2.3 Lymphomas originating from GC B cells

Lymphomas are a very heterogeneous group of cancers ranging from cases that require no acute treatment to highly aggressive cases leading to rapid patient death without treatment. Out of all lymphomas, 95 % arise from B cells, although B and T cells are produced in equal numbers (Küppers, 2005). According to World Health Organization (WHO) there are 19 classified B cell lymphomas and lymphoid proliferations (Alaggio et al., 2022). NHLs are the most common hematological malignancies worldwide and fifth among all cancers (Mlynarczyk et al., 2019; Thandra et al., 2021). In line with this, mature B cell malignancies are the fifth most prevalent cancers in Finland (Pitkäniemi et al., 2024). Out of NHLs, 85 % arise from B cells and most of them have GC origin. Indeed, normal GC B cells resemble many hallmarks of cancer cells including vigorous proliferation and clonal expansion, genome instability, inactivation of tumor suppressor genes and resistance to DNA damage (Mlynarczyk et al., 2019). Importantly, GC B cells undergo SHM and CSR. Thus, GC B cells are naturally at high risk of cancerous progression. NHLs arising from GC B cells are DLBCL, follicular lymphoma (FL) and Burkitt's lymphoma (BL). These malignancies have one of the highest mutation burdens of all tumors (Mlynarczyk et al., 2019). DLBCL is the most prevalent NHL and FL is the second most prevalent (Lackraj et al., 2018; Reddy et al., 2017). DLBCL and BL are aggressive cancers whereas FL is more indolent. FL, however, is uncurable and can progress to DLBCL (Mlynarczyk et al., 2019). AID expression is required in GC-derived NHLs, as AID-deficient mouse models fail to develop these malignancies underlining the oncogenic potential of AID (Pasqualucci et al., 2008).

DLBCLs are divided into germinal center B cell-like (GCB), activated B cell-like (ABC) and unclassifiable subtype (Li, S. et al., 2018). Common genetic alterations in DLBCLs are chromosomal rearrangements of *BCL6*, *BCL2* and *MYC* genes (Li, S. et al., 2018). Additional genes frequently harboring mutations are NF-

kB pathway genes (e.g., MYD88, CD79), B cell development genes (e.g., PAX5, IRF4), histone modifiers (e.g., CREBBP) and transcription factors (e.g., PRDM1) (Meyer et al., 2021; Reddy et al., 2017; Teater & Melnick, 2017). FLs frequently harbor a translocation of the *BCL2* gene, which is relocated under the control of Ig enhancers (Lackraj et al., 2018). Genes highly mutated in FL are involved in chromatin and histone modification, accompanied by gain-of-function mutations of *MEF2B* found in 15 % of cases (Desmots et al., 2019; Meyer et al., 2021; Ying et al., 2013). MEF2 mutations are also common in DLBCLs (Ying et al., 2013). Translocations of *BCL6* also appear in FLs, although less frequently than in DLBCLs (Mlynarczyk et al., 2019).

The hallmark of BL is the translocation of the proto-oncogene *MYC* under the regulation of *Ig* enhancers (Taub et al., 1982). Other commonly mutated genes found in 70 % of BLs are *TCF3,* which encodes the E2A transcription factor, and its regulator *ID3* (Schmitz et al., 2012). *ID3* mutations are typically inactivating and *TCF3* mutations are gain-of-function mutations. BLs are classified into three forms: the African endemic form, which is frequently linked to Epstein-Barr virus (EBV), the sporadic form and the immunodeficiency-related form (Panea et al., 2019). Interestingly, some DLBCLs are also EBV-positive (Swerdlow et al., 2016). In general, many lymphomas are associated with viral agents such as EBV, Kaposi's sarcoma-associated virus (KSHV), Human T-cell Lymphotropic virus type 1 (HTLV1), hepatitis B and C viruses (HBV and HCV) and simian vacuolating virus 40 (SV40) (Machida et al., 2004; Morales-Sánchez & Fuentes-Pananá, 2014; Ren et al., 2018; Vilchez et al., 2002).

2.2.4 BCL6 function in normal and lymphoma B cells

The *BCL6* gene encodes the BCL6 protein, which is a BTB/POZ zinc-finger containing transcriptional repressor originally cloned from DLBCL (Klein & Dalla-Favera, 2008). BCL6 binds to its corepressor proteins SMRT, NCOR and BCOR via the BTB/POZ domain to form effector complexes (Huang, C. et al., 2013).The zinc-finger domain binds to DNA and the third major domain RD2/PEST regulates protein stability and interactions with regulatory proteins (Basso & Dalla-Favera, 2012) BCL6 is expressed in GC B cells as well as in some T cell subpopulations, most notably follicular helper T cells, and in macrophages (Huang, C. et al., 2013; Iqbal et al., 2007). BCL6 is also expressed in non-immune cells like neurons, where it plays a role in neurogenesis (Bonnefont et al., 2019). In B cells, BCL6 is a master regulator of the GC reaction, and BCL6-deficiency prevents GC formation and affinity maturation of antibodies (Huang, C. et al., 2014; Ye et al., 1997). BCL6 complexes repress genes involving DNA damage responses and genes guarding cell cycle checkpoints such as *CDKN1A*, *CHEK1*,

ATR and *TP53,* thus allowing the cell cycle to continue despite accumulating mutations in *Ig* genes induced by SHM (Phan et al., 2005, 2007; Phan & Dalla-Favera, 2004; Ranuncolo et al., 2008). In addition, BCL6 upregulates AID expression by repressing microRNA-155 expression (Basso et al., 2012). BCL6 expression is normally downregulated at the end of GC reaction by transcriptional repressors IRF4 and BLIMP-1 (encoded by *PRDM1*), allowing further differentiation into plasma cells (Polo et al., 2008). Thus, proper function and dosage of BCL6 are essential for a functional immune system (Figure 1).

BCL6 is frequently mutated in GC-derived lymphomas DLBCL, FL and BLs (Li, S. et al., 2018; Panea et al., 2019). In addition, BCL6 involvement is implicated at least in acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), multiple myeloma (MM) and occasionally in solid cancers such as glioma, breast cancer and lung cancer (Cardenas et al., 2017; Hideshima et al., 2009; McLachlan et al., 2022). Oncogenic mutations of *BCL6* are gain-of-function mutations leading to continuous and/or upregulated BCL6 expression allowing cell cycle continuation despite DNA damage. Point mutations and translocations in *BCL6* 5' regulatory region, first intron and first exon disrupt the negative autoregulatory loop, leading to aberrant BCL6 expression (Migliazza et al., 1995; Pasqualucci et al., 2003; Peng et al., 1999; Qian et al., 2014; Wang, X. et al., 2002). In addition, mutations in the *BCL6* super-enhancer that prevents binding of negative regulators (predominantly BLIMP-1 and IRF4) are detected in DLBCLs and normal GC B cells, causing upregulation of BCL6 (Bal et al., 2022; Saito et al., 2007; Shen, J. C. et al., 2019). Translocations placing *BCL6* under the control of strong Ig enhancers are common gain-of-function mutations detected in 30-40 % of DLBCLs (Li, S. et al., 2018). Further, a translocation leading to constitutive *BCL6* expression can repress *PRDM1* expression (Parekh et al., 2007; Shaffer et al., 2000). This will help maintain the GC B cell phenotype and allow additional mutations to accumulate in the genome.

In addition to direct BCL6 mutations, mutations in BCL6 regulating genes can result in aberrant expression indirectly. BCL6 upregulation can be achieved through mutations of FBXO11, impairing FBXO11-induced degradation of BCL6 (Duan et al., 2012). In addition, gain-of-function mutations of the *BCL6* transcription regulator MEF2 or loss-of-function mutations of CREBB and E300 acetyltransferases, which normally block BCL6 interactions with its co-repressors, can deregulate BCL6 expression (Pasqualucci et al., 2011; Ying et al., 2013). Further, a deficiency of AID or methyltransferase 1 (DNTM1) triggers high levels of BCL6 (Jiao et al., 2020). This is achieved via the loss of *BCL6* methylation by DNTM1, which is assisted by AID (Jiao et al., 2020).

GC B cells and GC-derived lymphomas require BCL6 expression (Hatzi et al., 2013). Thus, loss-of-function mutations are less frequently detected than gain-of-

function mutations. In addition, loss-of-function mutations of *BCL6* can lead to oncogenesis due to the loss of BCL6-mediated repression of several oncogenes including *MYC* and *BCL2* (Ci et al., 2009). The requirement of BCL6 expression in lymphoma cells makes BCL6 a tempting therapeutic target. Further, BCL6-deficiency in T cells and macrophages results in life-threatening inflammation at least in mouse models and has other less severe consequences in autoimmunity (Cerchietti & Melnick, 2014; Tai et al., 2023). BCL6-deficiency causes impaired immune response due to the deficiency of high-affinity antibodies and Th2-response (Shaffer et al., 2002). For these reasons, eliminating BCL6 expression entirely is not a suitable treatment option for BCL6-overexpressing DLBCL. Targeting BCL6 co-factors and only partial functions of BCL6 has shown promising results in controlling aberrant BCL6 function (Cardenas et al., 2017; Cerchietti & Melnick, 2014).

2.2.5 AID expression in non-B cell malignancies

AID expression can be induced in non-B cells, although under physiological conditions its expression is limited to activated B cells. Chronic inflammation is considered a risk factor for oncogenesis, even outside the B cell context (Coussens & Werb, 2002). AID expression can be induced by proinflammatory cytokines such as tumor necrosis factor α (TNFα) signaling (Endo et al., 2007, 2008). This triggers NF-kB pathways in non-B cells. AID expression has been detected at least in melanomas, squamous cell carcinomas (SCC), basal cell carcinomas (BCC), Merkel cell carcinomas (MCC), colorectal cancers, T cell lymphomas and lung micro-adenomas as well as hepatocellular carcinomas (Endo et al., 2007, 2008; Machida et al., 2004; Matsushita et al., 2017; Nonaka et al., 2016; Okazaki et al., 2003; Okura et al., 2014; Watabe & Nakamura, 2013). Infectious agents can trigger AID expression, which has been shown in hepatocytes by HCV and in gastric epithelium by Helicobacter pylori (Machida et al., 2004; Matsumoto et al., 2007). HCV is also able to increase SHM rates in lymphomas and hepatocellular carcinomas (Machida et al., 2004). In addition, UV radiation can induce AID expression in human SCCs and melanomas (Nonaka et al., 2016).

2.3 Oncogenic viruses

Viruses highjack cellular replication machinery and can in some rare cases induce cancer by deploying it. In the case of oncogenesis, the viral genome persists in host cells and can be integrated into the host cell genome. Preventing viral replication and entry to the lytic cycle, eventually leading to cell lysis, benefits the virus persistence in host cells.

Currently seven oncogenic human viruses are known: EBV, KSHV, HBV, HCV, HTLV-1, human papillomaviruses (HPVs), and Merkel cell polyomavirus (MCPyV) (Krump & You, 2018). According to the International Agency for Research on Cancer (IARC), also human immunodeficiency virus 1 (HIV-1) and Helicobacter pylori are classified as infectious agents associated with oncogenesis (https://oncogenicviruses2021.iarc.who.int/background/). In addition, SV40 has been associated with human cancers for decades, but the topic remains controversial (Poulin & DeCaprio, 2006). Infectious agents, out of which the majority are viruses, are responsible for approximately 20 % of all cancers (Krump & You, 2018). Interestingly, EBV, KSHV, HCV and SV40 infect B cells, and MCPyV is associated with pro/pre-B cells (Alaribe et al., 2013; Dolcetti et al., 2003; Krump & You, 2018; Machida et al., 2004; McNees et al., 2019; Sauer et al., 2017; zur Hausen et al., 2013).

The mechanism by which oncogenic viruses cause cancer varies. Oncogenic viruses have diverse mechanisms for cell transformation modulating host signalling pathways (NF-kB, WNT/β-catenin, Notch, MAPK, PI3K-AKT-mTOR), immune responses and DNA damage responses. The general feature is inhibition of the tumor suppressor retinoblastoma (RB1) and p53 pathways by viral antigen binding. Small and large tumor antigens (sT and LT) of SV40 and MCPyV, E6 and E7 proteins of HPV, LANA1 protein of KSHV, EBNA proteins of EBV, Tax and HBZ proteins of HTLV1, HBx protein of HBV and polyprotein of HCV regulate p53 and RB1 (Krump & You, 2018; Morales-Sánchez & Fuentes-Pananá, 2014).

2.3.1 Polyomaviruses

Polyomaviruses infect a variety of species including humans, primates and rodents. Currently, 14 polyomaviruses are known to infect humans, two of which also infect animals (Klufah et al., 2021; Torres, 2020). Primary polyomavirus infection is usually mild or asymptomatic, but polyomaviruses persist in the body, may occasionally integrate into the host genome, and inflict more severe consequences in the case of immunosuppression (Decaprio & Garcea, 2013; Helanterä et al., 2016). Currently six human polyomaviruses John Cunningham polyomavirus (JCPyV), BK polyomavirus (BKPyV), Trichodysplasia spinulosa polyomavirus (TSPyV), human polyomavirus 6 (HPyV6), human polyomavirus 7 (HPyV7) and MCPyV are known to cause diseases other than mild infections in humans (Klufah et al., 2021; Prado et al., 2018). In addition, monkey polyomavirus SV40 has been proposed to be associated with several human cancers (Rotondo et al., 2019). JCPyV causes progressive multifocal leukoencephalopathy (PML) whereas BKPyV causes hemorrhagic cystitis and nephropathies in kidney transplant patients under immunocompromised conditions (Prado et al., 2018). TSPyV causes

a rare skin condition characterized by hyperproliferation of skin cells and resulting papules and keratin spines affecting mostly the facial area (van der Meijden et al., 2010). HPyV6 and HPyV7 are linked to pruritic and dyskeratotic dermatosis in organ transplant patients (Ho et al., 2015; Nguyen et al., 2017). MCPyV is the causative agent of MCC in approximately 80 % of cases (Feng, H. et al., 2008; Kassem et al., 2008).

2.3.2 Genomic organization of polyomaviruses

Polyomaviruses are small nonenveloped DNA viruses whose genome consists of ~5000 bp of double-stranded DNA (Prado et al., 2018). The organization of the genome is similar among polyomaviruses. The genome is divided into early, late, and non-coding regulatory regions (Figure 4) and is usually found episomal from host cells. The non-coding regulatory region (NCRR) contains promoters for early and late genes, as well as a replication origin and an enhancer region (Decaprio & Garcea, 2013). Genes of the early region are small tumor antigen (*sT*) and large tumor antigen (*LT*), which are responsible for viral replication, regulation of viral gene expression and hijacking the cellular machinery for virus production (Decaprio & Garcea, 2013; Prado et al., 2018). Structural viral capsid proteins (VP1, VP2 and VP3) are expressed from the late region after early gene expression is downregulated (Decaprio & Garcea, 2013). Some polyomaviruses express additional products from the early region such as middle tumor antigen and 17k tumor antigen as well as VP4 and agnoprotein from the late region and microRNAs from both (Decaprio & Garcea, 2013; Rotondo et al., 2019).

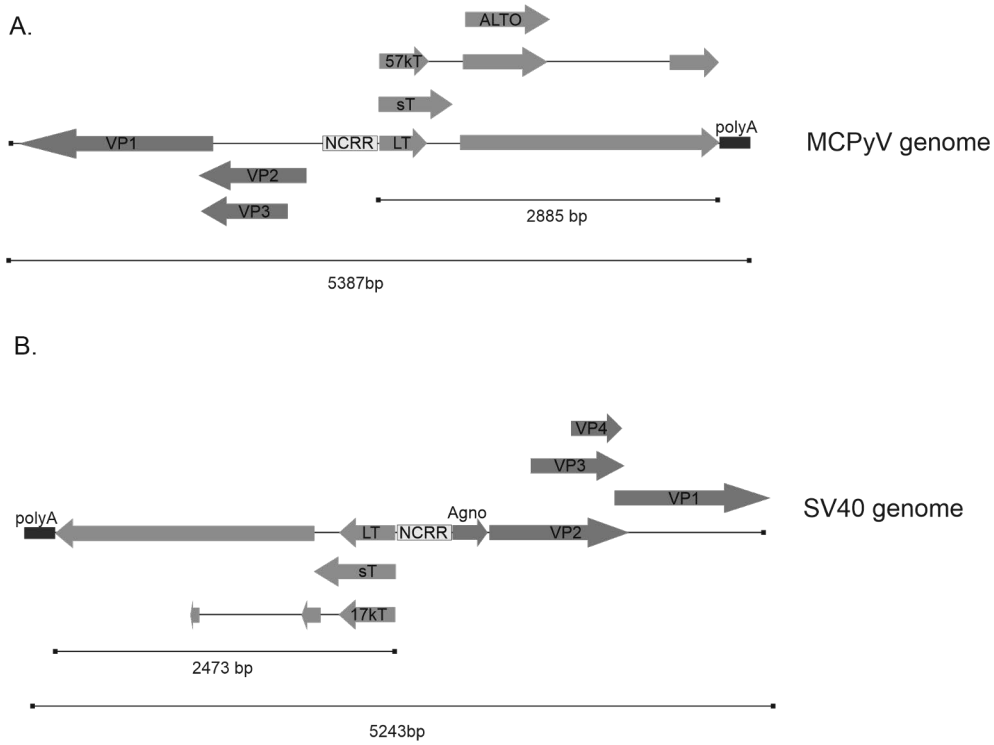

Figure 4. Schematics of MCPyV and SV40 genomes. A. Representation of MCPyV genome. B. Representation of SV40 genome. Protein coding exon regions with the direction of transcription are shown. Proteins encoded from the early region are presented with light grey arrows and proteins encoded from the late region are presented with dark grey arrows. Regulatory regions (NCRR) containing enhancer and promoter sequences are also presented. Sizes of the entire genome and early region are shown. LT: large tumor antigen, sT: small tumor antigen: 17kT: 17 kDa antigen, 57kT: 57 kDa antigen; ALTO: Alto-protein, AGNO: Agno-protein, VP1-4: viral capsid protein 1-4, NCRR: non-coding regulatory region

2.3.3 NCRR of polyomaviruses

In addition to the coding sequences, the polyomavirus genome has an NCRR (Figure 4). The NCRR region is 250–700 bp in length and exhibits variable compositions between different polyomaviruses and their strains (Lednicky & Butel, 1997; Moens et al., 2020; Yang & You, 2020). It orchestrates viral replication and gene expression, making it a major determinant of virus tropism and a regulator of the virus's pathogenic properties (Yang & You, 2020). For example, JCPyV genes under the control of the SV40 NCRR cause SV40-associated choroid plexus tumors while SV40 genes under the control of JCPyV NCRR lead to JCPyV-associated neural tumors and hypomyelination (Feigenbaum et al., 1992). If polyomaviruses integrate to host genome, the site is usually random and the NCRR and *LT* areas are intact (Butel & Lednicky, 2000; Liu, W., MacDonald, et al., 2016). In addition, other viral

proteins like the Tat-protein of HIV-1 may increase the expression of polyomavirus tumor antigens via regulating the binding of cellular transcription factors to the polyomavirus NCRR (Wright et al., 2013).

The NCRR contains binding sites for LT. It has been demonstrated in SV40 that after the initial accumulation of LT, the binding of LT to NCRR starts to inhibit early gene transcription and promote late gene transcription (Brady et al., 1984). The polyomavirus NCRRs contain a variety of binding sites for cellular transcription factors including NF-kB, NF1, AP-1, AP-2, Sp1, Ets1, YY1, p53 and E2F (Moens et al., 2020; Yang & You, 2020). SV40, BKPyV and JCPyV have high sequence homology in other parts of the genome, but only 40 % in their NCRRs (Feigenbaum et al., 1992). The MCPyV NCRR has an even lower sequence homology with SV40, BKPyV and JCPyV regulatory regions (Moens et al., 2020).

Duplication events of NCRR are frequently associated with polyomavirus-caused diseases. In the BKPyV Dunlop strain, the NCRR contains three non-identical P repeat regions when found in kidney transplant patients. The NCRR of JCPyV Mad-1 strain, which has duplication of a 98-bp region, is frequently found in PMLs (Yang & You, 2020). However, in some cases, deleting rather than increasing the number of repeat regions in the regulatory region can increase pathogenicity. This suggests a silencer function of some regions, which was demonstrated in BKPyV (Watanabe & Yoshiike, 1985). Prototypic SV40 strains contain one 72-bp region, which is frequently duplicated. This duplication gives a growth advantage compared to viruses containing one 72-bp region (O'Neill et al., 2003; Schmidt et al., 2016). In addition, the SV40 NCRR has three non-identical 21-bp repeats. The SV40 NCRR laboratory strain 776 contains a duplication of the 72-bp region. In contrast to the NCRR of pathogenic strains of BKPyV and JCPyV, duplication of 72-bp region is uncommon in the SV40 enhancer detected in cancers (Butel & Lednicky, 2000). Indeed, there is evidence showing that NCRR rearrangements increase viral replication and virulence, thus being a disadvantage to the oncogenic process (Prezioso et al., 2020, 2021). In MCPyV NCRR, only one region is shown to duplicate in the Japanese strain (Hashida et al., 2018). Current evidence suggests that MCPyV regulatory region rearrangement does not play a role in the development of MCC, since the NCRR detected in MCCs are similar to the reference strain (Hashida et al., 2018; Passerini et al., 2023; Prezioso et al., 2021). The reference strain is constructed from MCPyV sequences from healthy skin samples.

2.3.4 Polyomavirus tumor antigens target RB1, p53 and PP2A for transformation

Polyomaviruses transform cells through LT and sT proteins. LT is a 700 amino acid protein, the master regulator for host cell transformation in polyomaviruses. LT

contains RB1 and p53 binding sites, which are used to control the host cell cycle, cell proliferation and apoptosis. In addition, LT contains the DnaJ-domain, a helicase domain, threonine–proline–proline–lysine (TPPK) motif, a nuclear-localization sequence (NLS) and a DNA-binding domain (Decaprio & Garcea, 2013). ST is a protein of 200 amino acids encoded from the same gene as LT through alternative splicing (Seneca et al., 2014). ST has a unique region differing from LT in its C-terminal end. The functions of sT are not as well characterized as those of LT, and it is not required for the cellular transformation of all polyomaviruses. However, sT is capable of binding protein phosphatase 2A (PPA2), additionally promoting transformation (Pallas et al., 1992).

LT binds to RB1 via the LXCXE motif (Decaprio & Garcea, 2013; Pipas, 1992). LT is also able to bind RB1-related proteins p130 and p107 through LXCXE (Stubdal et al., 1997). However, the MCPyV LT does not seem to interact with these proteins (Pietropaolo et al., 2020). RB1 regulates the cell cycle at G1/S transition by blocking entry to S phase (Weinberg, 1995). In the G1 phase, unphosphorylated RB1 is bound to E2F family transcription factors, thus blocking E2F from activating genes necessary for the transition to the S phase. RB1 phosphorylation in the G1/S transition is regulated by cyclin D-Cdk4/Cdk6 complexes in early G1 and cyclin E-Cdk2 complex in late G1 (Giacinti & Giordano, 2006). Cyclin/Cdk activity is regulated by mitogenic signals. When LT is bound to RB1, E2F transcription factors are released and the cell cycle may progress to the S phase, enabling cell cycle progression regardless of mitogenic signals and giving the virus control of cell cycle regulation.

LT binds to p53 via a two-part binding site in the helicase domain (Lilyestrom et al., 2006). P53 regulates at least cell entry to apoptosis, cell cycle arrest, DNA repair and senescence (Kruse & Gu, 2009). P53 is normally ubiquitinated by MDM2, which leads to p53 degradation in the absence of signals of cellular stress. When these signals are present, kinases such as ATM, ATR and Chk1/Chk2 activate p53 by phosphorylation. Phosphorylated p53 binds to DNA and activates or represses the expression of target genes, depending on the stress signal. LT binding to p53 can thus dysregulate a variety of cellular functions normally protecting the cell from transformation.

The polyomavirus sT can bind to PP2A subunits A and C (Pallas et al., 1992). PPA2 is a negative regulator for a number of cell survival and proliferation pathways including MEK, ERK and MYC in the RAS signaling cascade, mTOR and AKT in the PI3K-AKT-mTOR-pathway and STAT5 in the JAK/STAT-pathway (Mazhar et al., 2019). PP2A can inhibit MDM2 and activate ATM/ATR, promoting p53-function. As MCPyV transforms cells via a truncated LT, which often lacks p53 binding sites, it is possible that sT binding to PP2A replaces the p53 binding function at least to some extent (Decaprio & Garcea, 2013).

2.3.5 Characteristics of MCPyV

MCPyV was discovered in 2008, and it is the newest human oncogenic virus described (Feng, H. et al., 2008). It was named after Merkel cell carcinoma, the disease in which the virus was first identified. The MCPyV genome is 5387 bp in size, with the early region encoding the sT antigen, LT antigen, 57 kilodalton (57kT) antigen and alternative Large Tumor Open Reading Frame (ALTO) in addition to the late region encoding VP1-3 (Prado et al., 2018; Yang & You, 2020) (Figure 4A). MCPyV is phylogenetically closely related to primate polyomaviruses and has low sequence homology with JCPyV, BKPyV and SV40 (Decaprio & Garcea, 2013; Yang & You, 2020).

2.3.5.1 MCPyV is responsible for the development of most MCCs

MCC is a rare but aggressive skin cancer with a poor prognosis (Schadendorf et al., 2017). MCC patients are usually immunocompromised and elderly. Among immunocompromised patients, over 80 % are MCPyV+ MCCs and with more sensitive detection methods MCPyV-positivity may increase (Rodig et al., 2012; Sahi, 2017). An MCC tumor is usually painless and firm, grows rapidly and has distinct coloring (Schadendorf et al., 2017). Diagnosis is made based on histologically typical undifferentiated small round cell appearance and cytokeratin 20-positivity (Emge & Cardones, 2019; Sahi, 2017). Three histological forms are detected: small cell, trabecular and intermediate, out of which intermediate is the most common (Schadendorf et al., 2017). Treatment options are surgery accompanied by radiation or chemotherapy, depending on the disease stage (Sahi, 2017). MCPyV+ MCCs expressed PD-L1 in 50 % of cases, whereas none of the MCPyV- MCCs expressed PD-L1 (Lipson et al., 2013). PD-L1 blockers, such as avelumab, have shown promising outcomes in the treatment of metastasized forms of MCC (Emge & Cardones, 2019; Sahi, 2017).

MCC was originally described as trabecular carcinoma and later renamed to Merkel cell carcinoma due to morphology and neuroendocrine markers matching those in skin mechano-receptor Merkel cells (DeCaprio, 2017). Currently, Merkel cells are considered an unlikely cell of origin for MCC due to a low number of Merkel cells in the skin, their postmitotic state, different location in the skin compared to tumors and the inability of Merkel cells to be infected by MCPyV (Liu, W., Yang, et al., 2016; Schowalter et al., 2010; Sunshine et al., 2018; Vaigot et al., 1987). The suggested host cells are epidermal keratinocytes, dermal fibroblasts, skin pluripotent stem cells and B cell precursors (pro/pre B cells) (Liu, W., Yang, et al., 2016; Richards et al., 2015; Sunshine et al., 2018; zur Hausen et al., 2013). It has been proposed that MCPyV+ and MCPyV- MCCs originate from different cells, MCPyV+ MCCs from dermal fibroblasts and MCPyV- MCCs from epidermal

keratinocytes (Sunshine et al., 2018). Indeed, MCPyV- MCCs exhibit a distinct UV-mutation signature, a different expression pattern, and a higher mutation load than MCPyV+ MCCs accompanied by inactivating mutations of RB1 and p53 (Goh et al., 2016; Harms, P. W. et al., 2015; Kervarrec et al., 2019). MCPyV+ MCCs are also linked to a more favorable prognosis (Harms, K. L. et al., 2021).

The dermal fibroblast is the only cell type so far found to support productive infection of MCPyV (Liu, W., Yang, et al., 2016). It has also been proposed that MCPyV replication takes place in one cell type and accidentally enters another, perhaps less permissive, nearby cell type where transformation occurs (Liu, W., Yang, et al., 2016; Pietropaolo et al., 2020). Pro/pre-B cells are proposed as the cell-of-origin for MCC due to the gene expression patterns of distinct B-cell markers, including PAX5, TdT, immunoglobulins, c-Kit and RAG1 (Johansson et al., 2019; Sauer et al., 2017; zur Hausen et al., 2013). In addition, rearrangement of *Ig* genes and AID expression have been detected (Matsushita et al., 2017; Watabe & Nakamura, 2013; zur Hausen et al., 2013). Furthermore, idealisib, a drug originally developed to treat B cell malignancies, has been shown to be highly effective in the treatment of MCC (Shiver et al., 2015). Co-occurrence of MCPyV+ MCC and chronic lymphocytic leukemia (CLL) has been established frequently, and MCPyV DNA is found from CLL tumors (Koljonen et al., 2009, 2010; Pantulu et al., 2010; Tadmor et al., 2011). It was also detected that chronic inflammatory disorders increased the risk for MCC (Sahi et al., 2017). Despite the correlation between B cells and MCC, the B-cell origin for either type of MCC has not been proven and the current evidence is against the B-cell origin of MCC (Kervarrec et al., 2019; Liu, W., Yang, et al., 2016; Sunshine et al., 2018).

MCPyV+ MCC express truncated LT (Shuda et al., 2008). Truncation mutations occur at the C-terminus after the RB1-binding site, thus preserving RB-mediated cell cycle control but resulting in the loss helicase domain required for virus replication (Shuda et al., 2008). However, it has been unclear what causes these mutations. Nevertheless, truncation of LT enables stable integration and inhibition of immune response (Li, J. et al., 2013; Shuda et al., 2008). The C-terminal portion causes DNA damage and triggers DNA-damage response, and phosphorylation of the C-terminus induces apoptosis (Li, J. et al., 2013, 2015). LT expression is required for MCC growth and maintenance, and truncation of LT is considered to be a MCC-specific event (Houben et al., 2010, 2012). However, truncated *LT* was detected in healthy skin samples even in individuals without MCC (Pyöriä et al., 2024). This finding suggests that truncation may not be MCC-specific but is probably required for oncogenesis. Nevertheless, LT expression alone is not sufficient for full tumorigenesis (Chang & Moore, 2012). As truncated LT lacks the C-terminal domain including p53 binding sites, p53 actions in MCPyV+ MCC are inhibited in an indirect manner (Houben et al., 2013).

2.3.6 Characteristics of SV40

SV40 was one of the first tumorigenic viruses discovered. It became widely known as a contaminant of the polio vaccine in the 1950s-60s. Components of the polio vaccine were produced in cell lines originating from monkey kidney cells, which unfortunately were infected by SV40. Shortly after contamination, it was discovered that SV40 is tumorigenic in rodent cells and efficiently transforms cultured human cells. (Butel & Lednicky, 2000)

Originally it was suspected that SV40, being a monkey virus, does not circulate among humans, and SV40 detected in the population is solely due to contaminated polio vaccine. This is probably not the case, as SV40 is detected in individuals who have not received the contaminated vaccine and frequent findings of archetypal SV40 strain in the population and in tumors (Butel & Lednicky, 2000; Martini et al., 2007). Cumulative evidence indicates that SV40 does indeed circulate among humans and is perhaps transmitted from human to human (Martini et al., 2007; Rotondo et al., 2019).

The SV40 genome structure is typical for a polyomavirus. It is 5243 bp in size, and the early region encodes the sT antigen, LT antigen, 17k antigen, early leader protein as well as early and late polarity miRNAs. The late region encodes four structural proteins (VP1-4) and an agnoprotein (Figure 4B). (Rotondo et al., 2019)

2.3.6.1 SV40 association to oncogenesis

Evidence of SV40 presence has been found in a variety of cancers. Most frequently SV40 is linked to human mesotheliomas, brain cancers, bone tumors and NHLs (Butel et al., 2003; Carbone et al., 2020; Heinsohn et al., 2011; Huang, H. et al., 1999; Lednicky et al., 1995; Malkin et al., 2001; Martini et al., 2007; Mendoza et al., 1998; Rivera et al., 2008; Shivapurkar et al., 2002; Testa et al., 1998; Thanh et al., 2016; Vilchez & Butel, 2003). SV40 DNA, mRNA and proteins particularly from the *LT* region have been detected in various tumors (Rotondo et al., 2019). Some studies have reported that SV40 DNA and SV40-specific antibodies in patients with lymphoma or mesothelioma were more frequent than in healthy controls or in other tumor types (Mazzoni, Pietrobon, et al., 2017; Shivapurkar et al., 2002; Tognon et al., 2015). Further, rodents display the same tumor types caused by SV40 as humans (e.g. brain, bone and lymphoid tumors) (Martini et al., 2007).

In contrast, some studies show a very rare presence or have not detected evidence of SV40 in mesotheliomas or lymphomas (De Rienzo et al., 2002; Hirvonen et al., 1999; Manfredi et al., 2005; Samaka et al., 2015; Shah, 2007). Before the development of a more specific serological test for SV40 detection, cross-reactivity with JCPyV and BKPyV antigens probably increased the false-positive rate of SV40, potentially leading to their misidentification as SV40 (Viscidi et al., 2003). There

have also been concerns of false-positive PCR results of SV40 as its genome is commonly used in different molecular biology applications (López-Ríos et al., 2004). This claims that the detection of SV40 in cancers might be a misinterpretation. Importantly, the cancer incidence among the population who received the contaminated polio vaccine was not increased which further argues against a significant role of SV40 in human cancers (Poulin & DeCaprio, 2006; Shah, 2007).

Explanations for the discrepancy in detecting SV40 may be geographical differences in the administration of SV40-contaminated polio vaccine, differences in the natural spreading of SV40 infection and the proposed hit-and-run mechanism for SV40 tumorigenesis (Baker et al., 2022; Liu, J. et al., 1991; Rotondo et al., 2019). The hit-and-run mechanism assumes that the SV40 LT is required at some point of tumor development but becomes dispensable later and thus is not always detected inside certain groups of cancers. Even if the role of SV40 in human oncogenesis is established in the future, SV40 alone is not sufficient to cause fully progressed cancer, and additional events are required (Jinglan et al., 2009).

The ability of SV40 to cause productive infection varies between hosts and cell types. Cells that are permissive for SV40 infection enable virus replication and support the lytic cycle of the virus eventually leading to cell lysis. In contrast, rodent cells are non-permissive hosts for SV40 and are efficiently transformed by it (Butel & Lednicky, 2000). Non-permissive cells are at higher risk for tumorigenesis because they do not enter the lytic cycle leading to cell death, thus enabling dysregulated cells to persist. However, not all infected non-permissive cells are transformed. In addition, viruses can persist in host cells, regardless of permissiveness, without ever causing transformation.

Humans are semi-permissive hosts for SV40 meaning that the permissiveness depends on cell type. Namely, most cells enter the lytic cycle, but some persist and can be transformed. Indeed, mesothelial cells and human lymphoblastoid B cells, which are most frequently linked to SV40, do not support the lytic cycle efficiently and are transformed frequently (Bocchetta et al., 2000; Dolcetti et al., 2003). Thus, in most human cells, the production of SV40 virions must be restricted for transformation to occur. Mutations in the *LT* C-terminus, leading to truncated LT expression, is one way of achieving this goal. The truncation of LT could predispose SV40 immune evasion and help avoid DNA damage response, as it has been demonstrated for MCPyV (Li, J. et al., 2013; Shuda et al., 2008). Indeed, truncation events have been shown in transformed human cell lines (Gish & Botchan, 1987; Kao et al., 1993). However, the LT C-terminus contains a binding site for p53, and if this site is lost, the function of the p53 protein should be disrupted in another way. Some studies have found evidence of co-occurrence of SV40 LT and p53 mutations in mesotheliomas and osteosarcomas (Heinsohn et al., 2011; Mendoza et al., 1998). However, SV40 LT truncation has not been shown to play a role in human tumors

and conclusive evidence of SV40 as a causative agent in human cancers is still missing. Regardless, SV40 is a widely used tool in laboratory applications. Its sT and LT are routinely used to create immortalized cell lines and the SV40 NCRR is one of the most studied among all regulatory regions (Pipas, 2009; Smith & Shilatifard, 2014).

2.3.7 APOBEC family enzymes

APOBEC cytidine deaminase family comprises of 11 enzymes: APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4 and AID (Salter et al., 2016). APOBECs function in antiviral defense and target at least papillomaviruses, polyomaviruses, hepatitis viruses and human immunodeficiency viruses (Köck & Blum, 2008; Poulain et al., 2020; Sheehy et al., 2002). APOBEC3G was the first APOBEC found to possess antiviral properties against HIV infection. This was accomplished by introducing inactivating G>A mutations to viral DNA encoding the Vif protein, which is required for virus production (Mangeat et al., 2003). Viral infection can trigger interferon signaling, further inducing APOBEC expression (Que et al., 2021; Wang, Z. et al., 2014). APOBEC family member expression is frequent in immune cells (e.g., AID in activated B cells, APOBEC3A in monocytes and macrophages and APOBEC3C-3H in immune centers and peripheral blood cells) (Salter et al., 2016). The majority of APOBECs reside in the cytoplasm likely to protect genomic material from harmful mutations. However, APOBEC3B is mainly nuclear (Lackey et al., 2012). APOBEC3 expression is also shown to be upregulated and APOBEC3 mutational signature is frequent in various malignancies, including breast cancer, head and neck cancers, lung cancer as well as lymphomas (Alexandrov et al., 2013; Burns, Temiz, et al., 2013). APOBEC3A, APOBEC3B and APOBEC3H are most frequently detected in cancers (Burns, Lackey, et al., 2013; Burns, Temiz, et al., 2013; Nik-Zainal et al., 2012; Starrett et al., 2016; Vartanian et al., 2008).

The APOBEC family members have similar structures, APOBEC3B, 3D, 3F and 3G having two zinc domains and the rest having a single domain (Salter et al., 2016; Swanton et al., 2015). The APOBEC family members deaminate cytosines converting them into uracils in RNA and ssDNA substrates of cellular and foreign origin (Salter et al., 2016). However, there is no evidence of enzymatic activity for APOBEC2 and APOBEC4 (Swanton et al., 2015). APOBECs have a general preference to target cytosines at T**C**-dinucleotide or T**C**W-trinucleotide context, and mutations are mainly C>T transitions or C>G transversions (Shi et al., 2017). C>T transitions are presumed to result from the recognition of uracil as thymine during replication, while C>G transversions are thought to arise from BER activity,

followed by error-prone polymerase filling the abasic site (Helleday et al., 2014). More specifically, APOBEC3A is shown to prefer YT**C**R, APOBEC3B AT**C**R, APOBEC3C WY**C**R, APOBEC3D/H T**C**, APOBEC3F TT**C**W, and APOBEC3G CC**C**H (Maiti et al., 2021).

3 Aims

SHM off-targeting is a severe threat to genome integrity, but the mechanisms behind it are still poorly understood. The aim of this thesis was to investigate whether SHM off-targeting relies on SHM targeting genetic elements, similar to SHM on-targeting at *Ig* loci. The goal was to investigate SHM targeting activity of both viral and endogenic genomic elements and to explore the consequences of this activity on neighboring genes. The goal was also to increase understanding of how SHM off-targeting can contribute to oncogenesis.
The specific aims of the thesis were:

1. To investigate if the MCPyV NCRR has SHM targeting activity, and to determine if the AID/APOBEC family enzymes cause mutations to the MCPyV *LT*. The particular interest was on MCPyV+ MCCs, which harbors *LT* truncating mutations leading to oncogenesis, but the origin of these mutations remained unknown.
2. To investigate if the SV40 NCRR has SHM targeting activity, and if this activity can cause AID-induced mutations to accumulate to the SV40 *LT* region. Thus, the aim was also to investigate if SV40 possess MCPyV-like truncation events in the *LT* region, providing a potential mechanism for SV40-induced tumorigenesis.
3. To investigate if the *BCL6* enhancer region contains SHM targeting elements, and if the deletion/insertion of these elements affects the mutation accumulation to the *BCL6* 1st intron, which is frequently mutated in lymphomas, particularly in DLBCLs.

4 Materials and Methods

4.1 Cell lines (I, II, unpublished)

Chicken bursal lymphoma cell line DT40, Human Burkitt's lymphoma cell line Ramos, DLBCL cell lines Mieu and Carnaval as well as human embryonic kidney cell line HEK293 were used to conduct experiments in this thesis. DT40 and Ramos are B cell lines undergoing SHM, thus providing good models to study SHM targeting. Mieu and Carnaval were used as controls to assess the mutation frequency in *BCL6* 1st intron in the NGS experiment. HEK293 cells were used to test the efficacy of guide RNAs in CRISPR/Cas9 experiment. Ramos, Mieu, Carnaval and HEK293 cells were cultured at +37 °C, 5 % CO_2, and 90 % humidity. DT40 cells were cultured at +40 °C, 5 % CO_2, and 90 % humidity. Growth media for Ramos, Mieu and Carnaval cells contained RPMI 1640 HEPES modification (Merck-Sigma) with 10 % FBS (Gibco), 1x Glutamax (Gibco) and 1x penicillin-streptomycin antibiotic (Gibco). Growth media for DT40 cells contained RPMI 1640 HEPES modification (Merck-Sigma) with 10 % FBS (HyClone), 1 % NCS (Biowest), 1x Glutamax (Gibco), 1x penicillin-streptomycin antibiotic (Gibco), and 50 μM β-mercaptoethanol (Gibco). Growth media for HEK293 cells contained DMEM (Merck-Sigma) with 10 % of FBS (Gibco) and 1x Glutamax (Gibco).

4.2 Measurement of SHM targeting activity using GFP loss assay (I, II, unpublished)

In the GFP loss assays, the proportion of GFP-negative cells (GFP loss) in a cell population originating from a single cell is measured. The measured GFP loss is directly proportional to SHM targeting activity. As GFP loss assay has been exploited in every study of this thesis, it will be described below in detail.

4.2.1 Selection of fragments for GFP loss assay

The DNA fragments tested for SHM targeting activity were selected from the polyomavirus genome and NCRR (I, II) or the *BCL6* super-enhancer region

(unpublished). In Study I, the entire MCPyV genome was divided into four fragments (Figure 5A) and measured in the GFP loss assay, since there was no previous research from the subject. The truncated versions of fragments 1 and 2 were generated and fragment 1 was flipped into opposite orientation to study if deleting or moving insulator sequences in these fragments affects the SHM targeting activity. Further, the NCRR of MCPyV was divided into shorter fragments (440 bp, 236 bp, 204 bp and 80 bp) to accurately determine the location of potential SHM targeting activity (Figure 5A). In Study II, the particular NCRR region was directly tested as the location of the SV40 enhancer in the NCRR is well established (Figure 5B). The entire NCRR and part of the Agno-protein sequence were included in JCPyV and BKPyV fragments measured in GFP loss assay (Figure 5C). This was because Agno-sequence contained TFBSs important for SHM targeting.

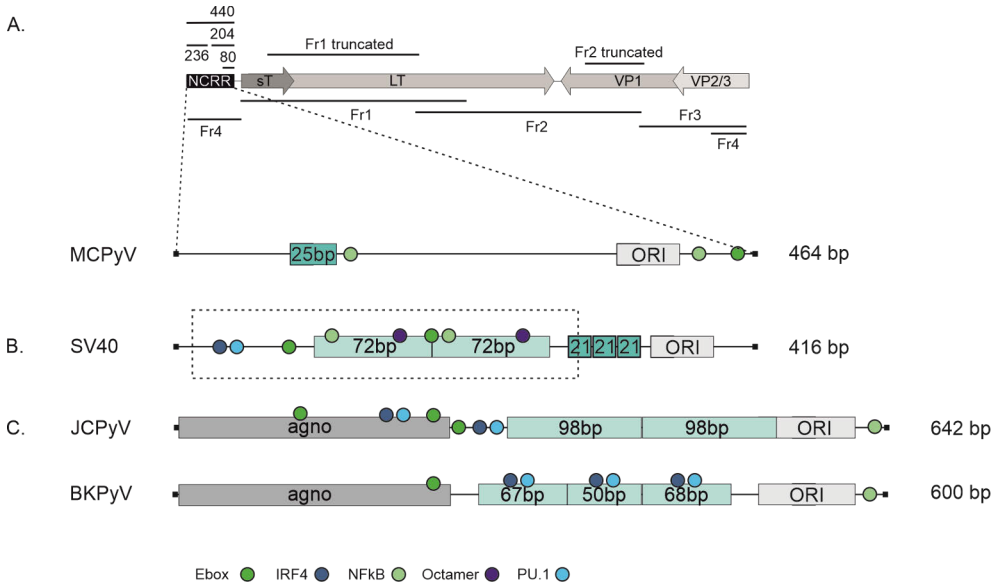

Figure 5. Elements from polyomavirus genomes selected for testing in the GFP loss assay. A. MCPyV genome (strain R17b) including detailed representation of NCRR and fragments tested in GFP loss assay. B. SV40 regulatory region (776 strain, 72bp duplication). The tested enhancer element is indicated with a dashed line box. C. Tested JCPyV (Mad-1 strain) and BKPyV (Dunlop strain) regulatory regions. Key TFBSs in NCRRs associated with SHM are presented. TFBSs are indicated with circles, with the color coding explained in the legend below the figure. The length of the regulatory region (A. and B.) or the tested element (C.) is indicated on the right. Repeat sequences are indicated with rectangles. ORI: origin of replication.

The elements to be measured in GFP loss assay from the *BCL6* enhancer region were selected based on previous data reporting transcription and super-enhancer status, H3K27 acetylation and enhancer element connections to *BCL6* promoter (Qian et al., 2014; Senigl et al., 2019). The locations of elements in the *BCL6* locus are presented in Figure 6.

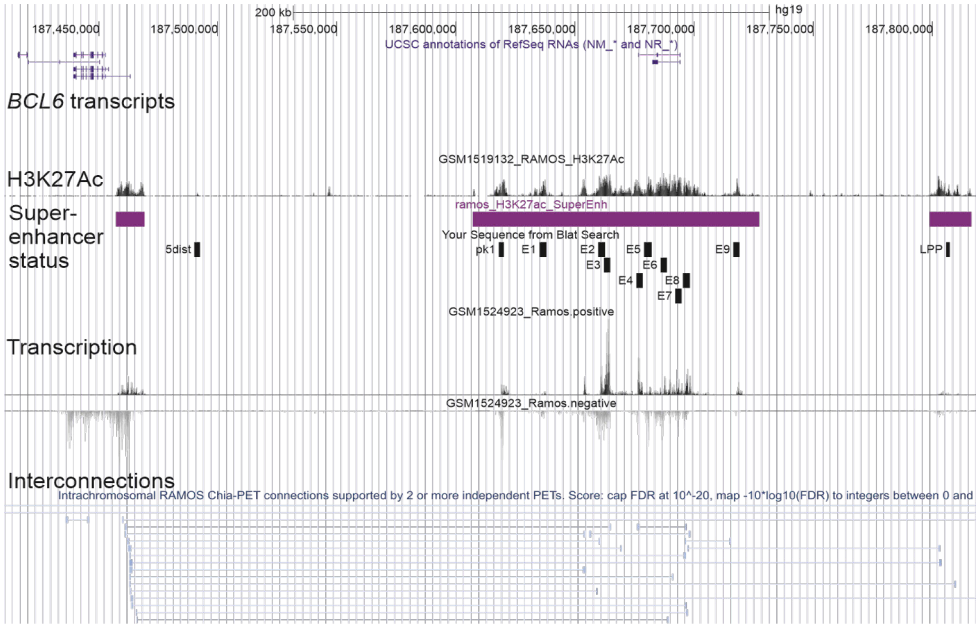

Figure 6. The location of potential SHM targeting elements in the *BCL6* locus in chromosome 3 in hg19 assembly. Selected fragments are indicated with black boxes. Suggested SHM targeting activity indicating factors H3K27 acetylation, transcription (GRO-Seq), interconnections between enhancer and gene (ChIA-PET) as well as super-enhancer status are represented (Qian et al., 2014; Senigl et al., 2019). Genomic coordinates and *BCL6* transcript variants are shown. The figure is a snapshot from the UCSC Genome Browser.

4.2.2 GFP plasmids and DT40 cell line modifications for GFP loss assay

GFP2 and GFP4 reporter vectors were used for measuring GFP loss in DT40 cells. In addition to GFP expression cassette, vectors contain cell line-specific targeting arms on both sides of the cassette. GFP2 cassette is directed towards *Ig* locus where Ig light chain (*IgL*) allele is deleted and replaced by a pyromycin resistance gene. In addition, both *AID* alleles are deleted after which one of them was restored. GFP4 is directed towards the *AID* locus, where both alleles are deleted. One of the alleles

have *AID* gene restored, and the other is replaced by the puromycin resistance gene. The genotypes of modified cell lines used with GFP2 and GFP4 reporters are DT40 $IgL^{-/puro}AID^{R/-}$ and DT40 $UNG^{-/-}AID^{R/puro}$, respectively.

4.2.3 Cloning procedure

Selected DNA-fragments from polyomavirus genomes (I, II) and *BCL6* enhancer region (unpublished) were cloned into GFP expression vectors to SpeI/NheI (GFP2, GFP4) restriction sites when cloning upstream of the GFP expression cassette. The BamHI restriction site was used in the GFP4 vector when cloning downstream of the GFP expression cassette. Selected fragments were cloned to both upstream and downstream positions of GFP expression cassette to assess if part of the seen GFP loss is caused by increased transcription. Selected fragments were amplified from plasmid templates (I, II) or genomic DNA (unpublished) using PCR. Q5 high-fidelity polymerase (NEB) or LongAmp polymerase (NEB) when amplifying long products were used according to the manufacturer's protocols. PCR was conducted with primers containing suitable over-hangs for the cloning procedure, which was performed using In-Fusion cloning kit (Takara). Sequences of cloning primers are provided in Tables 1, 2 and 3. The final cloning primers are a combination of the genome binding part and the suitable In-Fusion overhang, depending on the cloning site. Cloning was performed in chemically competent Stellar *E. coli* bacterial cells (Takara). Plasmid DNA was extracted from single bacterial colonies from ampicillin selective plates and restriction digestion was performed to check the success of cloning.

Table 1. Genome-specific (polyomavirus) In-Fusion cloning primer sequences used in thesis work (I,II).

FRAGMENT	PRIMER ORIENTATION	SEQUENCE	STUDY
MCPyV Fr1	Forward	GGCTCTCTGCAAGCTTTTAGAGAT	I
	Reverse	AGGCCACACCACCCATATAATACA	I
MCPyV Fr2	Forward	CCTCACAAGGCTCATGAGGCT	I
	Reverse	TCCAGATCAGCCCATCAAGGAA	I
MCPyV Fr3	Forward	ATTGCCTCCCACATCTGCAA	I
	Reverse	ATCATCACACTGCTGGCCAAT	I
MCPyV Fr4	Forward	TGCAGCCTAGAGGTAGGAGAT	I
	Reverse	AAGCATGCACCCAGGACCTC	I
MCPyV Fr1 truncated	Forward	AAGCTGCTTTCAAAAGAAGCTGCT	I
	Reverse	GTGGTCGTCTAGCTCATATTCACAAGC	I
MCPyV Fr2 truncated	Forward	AGAGCCTGTCTGAATAGACCCATAG	I
	Reverse	TGGGAGTTAATTCCCCTGATCTTCCT	I
MCPyV NCRR 440bp/236bp	Forward	ATTGGCCAGCAGTGTGATGATG	I
MCPyV NCRR 236bp	Reverse	TAGGCAGCCAAGTTGTGGTTACAT	I
MCPyV NCRR 440bp/204bp	Reverse	GCAGAAAGAGCAGAGGAGCAA	I
MCPyV NCRR80bp/204bp	Forward	TGGCTGCCTAGGTGACTT	I
MCPyV NCRR 80bp	Reverse	GATCTGCCCTTAGATACTGCCTT	I
SV40 NCRR	Forward	TGCAGGGCCTGAAATAACCTCT	II
	Reverse	TAGTGGGACTATGGTTGCTGA	II
JCPyV NCRR	Forward	GAACCTGAAAACACAAAAAAAAACATT	unpublished
	Reverse	TTTAGCTTTTTGCAGCAAA	unpublished
BKPyV NCRR	Forward	GGGATTACTTACCTAGGAGTCT	unpublished
	Reverse	TTTTGCAAAAATTGCAAAAGAAT	unpublished

Table 2. Genome-specific (*BCL6*) In-Fusion cloning primer sequences used in thesis work (unpublished).

FRAGMENT	PRIMER ORIENTATION	SEQUENCE
E1	Forward	CATCAACAGGGAGCTTTGCCTAG
	Reverse	AGAGGCCTGCAAGGGTTATACCAA
E2	Forward	ATGTGCCAGTGTTACCGATGC
	Reverse	AACGTGTATGCCTGTGTCAGAT
E3	Forward	GCTGCCTCTCAGCTTAACCAC
	Reverse	TTGACCAAGCCTCCACACGA
E4	Forward	ACAGGACGAGAAGAGTGAGTCG
	Reverse	GTGATTTGCATTCCCTTCACATCG
E5	Forward	TGCTCTCAAGTCCATATCTGCCTCT
	Reverse	GCTGCCACATTGGAAAGCAC
E6	Forward	ACTTCTCTGGTGAGAGGTTAGGAT
	Reverse	GCAAGTTGTAGTGCCACGGT
E7	Forward	GTCTGGAACTCTCTGCTACTGTGC
	Reverse	ATGATGGTGTCTGGGCTGGAT
E8	Forward	CAGCCAATCAAGAAAGCAGCAAT
	Reverse	CCTGTGTTCTGGGTACTAATCCAG
E9	Forward	GACAGGACAGACGCATACATCTTT
	Reverse	GACACCCACATGATGACGAGTAC
e2f	Forward	GAGAGCCCATGAGTTTAACTGACTT
	Reverse	TGAGTAAGGCTCTCTTGGCACTTTG
5'dist	Forward	CTTCGCCGTACTTCCATCTCTCTA
	Reverse	ATAGAGGTAACTGACATGCCCAAA
pk1	Forward	CTAGGGTGCAGAACCAATGAGT
	Reverse	TTCAGGCCAGCGCGATTTGAA
LPP	Forward	CACAATATTGGTTCAGGTTTATGCT
	Reverse	ACGGTTGTCAGTCTAAGAGTG

Table 3. In-Fusion overhang sequences containing restriction enzyme recognition sites used for cloning fragments to GFP4 vector.

CLONING SITE	OVERHANG ORIENTATION	SEQUENCE
upstream (SpeI)	Forward	TCACGGATCCACTAGT
upstream (NheI)	Reverse	CAGGGAGCAGGCTAGC
downstream (BamHI)	Forward	ATCTTATCATGTCTGGATCC
downstream (BamHI)	Reverse	TCCTGCAGCCCGGGGGATCC

4.2.4 Transfection and assay performance

GFP2 and GFP4 vectors containing cloned DNA-fragments were transfected to DT40 *IgL*$^{-/puro}$*AID*$^{R/-}$ and DT40 *UNG*$^{-/-}$*AID*$^{R/puro}$ cells, respectively, using electroporation. 12 millon cells and 50 µg of NotI or XhoI linearized plasmid were used per reaction. Ethanol precipitation was performed to plasmids before transfection. Electroporation was conducted in 0.4 cm cuvettes using Gene Pulser Xcell machine (BioRad) with settings 700 V, 25 µF and 200 Ω. After electroporation, cells were incubated for 5–8 minutes on ice, after which they were suspended to 10 ml of prewarmed culture media and transferred to a 96-well plate 100 µl/well. The next day, blasticidin was added at a final concentration of 15 µg/ml to select the cells with successful transfection. Cells were incubated for 7–9 days. When primary cell clones appeared visible in plates, single colonies were picked to new 96-well plates and the integration of GFP cassette to the appropriate locus (section 4.2.2) was tested with puromycin selection at a final concentration of 1 µg/ml. Targeted clones are puromycin-sensitive as the integration of the GFP cassette knocks the puromycin resistance gene out of the locus. GFP fluorescence of targeted primary clones was confirmed using flow cytometry. Targeted clones were subcloned to 96-well plates (day 0 of the assay), and after 7–9 days of incubation, at least 12 individual subclones were picked and maintained.

4.2.5 Flow cytometry

At day 12 (GFP4) or 14 (GFP2), the GFP loss of subclones was measured. At least 12 individual subclones per tested fragment were collected (200 µl/well), and the cells were directly analyzed with flow cytometry using either Novocyte (Agilent) or Accuri C6 (BD Bioscience) flow cytometer. In this assay, cells are not preprosessed or stained, as only the proportion of the cells that have lost GFP fluorescence, originating from GFP expression cassette, is significant. Data from 30 000 live cells per subclone were collected. Results were analyzed using FlowJo and GraphPad Prism -softwares.

4.2.6 SV40 *LT* insertion to DT40 genome (II)

To investigate the SV40 NCRRs ability to target mutations to the SV40 *LT* region in Study II, the SV40 early gene transcription unit was introduced to DT40 *IgL*$^{-/puro}$*AID*$^{R/-}$ cell genome using the GFP2 vector system.

The insertion process is described in detail in Study II. The In-Fusion cloning primers for amplifying and cloning SV40 *LT* transcription unit were f orward: GACGGATCCACTAGTggacctgaaataaaagacaaaaagactaaact and reverse: AGGGAGCAGGCTAGCgatccagacatgataagatacattgatgagt.

4.3 Sequencing

4.3.1 Sanger sequencing (II)

In Study II, the *LT* transcription unit was sequenced to determine if AID-induced mutations were accumulated to *LT* area during 12-week culture period. The sequencing procedure is described in detail in Study II. The cloning primers described in section 4.2.6. were used to amplify the SV40 transcription unit from the DT40 cell genome. The transcription unit was cloned into pUC19 vector for sequencing. For determining the mutation accumulation to *LT* area inserted to Ramos and UO-31 cells, the *LT* area was divided into three subregions for sequencing. The sequencing primers are provided in Table 4. Sequencing reactions were prepared according to Eurofins guidelines and sequencing reactions were performed at Eurofins Genomics, Germany.

Table 4. Sequencing primers in Study II.

CELL LINE	REGION	ORIENTATION	SEQUENCE
DT40	STOP codon	Forward	ACACAGGCATAGAGTGTCTGC
Ramos/UO-31	Region 1	Forward	ACAGAGAGGAATCTTTGCAGC
Ramos/UO-31	Region 1	Reverse	AGCAAAGCAAGCAAGAGTTCT
Ramos/UO-31	Region 2	Forward	ACTGCTGACTCTCAACATTCTAC
Ramos/UO-31	Region 2	Reverse	CAACTCCAGCCATCCATTCT
Ramos/UO-31	Region 3	Forward	AATGTTTGGTTCTACAGGCTCT
Ramos/UO-31	Region 3	Reverse	GGGAGGTGTGGGAGGTTT

4.3.2 Mutation analysis of Sanger sequencing data (I, II)

Mutation analysis was performed to determine the causative agent inducing mutations to viral *LT* regions in Studies I and II. Published sequences of MCPyV *LT* MCC (113) and control (83) samples were retrieved from the NCBI database and aligned to the MCPyV reference genome (isolate R17b; NC_010277) using SnapGene software (I). Sequences obtained from SV40 *LT* sequencing were aligned to SV40 reference genome (NC_001669) (II). From Ramos cells, 35 sequences were obtained from the first region (16–810bp), 191 sequences from the second region (694–1521bp) and 190 sequences from the third region (1467–2456bp). From UO-31 cells, 94 sequences were obtained from the first region, 190 sequences from the second region and 192 sequences from the third region. Mutation frequencies, types and occurrence of mutations in AID, APOBEC and UV-radiation hotspots were

determined. The procedure of the mutation analysis is described in detail in studies I and II.

4.3.3 Next-generation sequencing and mutation analysis (unpublished)

To investigate how alterations in the *BCL6* enhancer region affect mutation accumulation to the *BCL6* gene, next-generation sequencing (NGS) was performed (unpublished). The amplicon length was 516 bp (chr3:187744816—187745331 on hg38). Genomic DNA was extracted from *BCL6* enhancer-modified cell lines and control cell lines using gDNA miniprep kit (Zymo Research). After extraction, additional purification with ethanol precipitation was performed. Next, two consecutive PCRs were performed with modified protocol from Yu, G. et al., 2021. The first round of PCR amplifies a 516 bp amplicon from *BCL6* 1st intron and adds 8 bp unique molecular identifiers (UMIs) to the PCR-product. Primer sequences are presented in Table 5. In addition to the parts presented in Table 5, primers contain UMIs with random sequences (NNNNNNNN) between primer backbone and genome-binding sequences. Primers were ordered from Integrated DNA Technologies (IDT). For the second round of PCR, full-length adaptor primers were used (Tables 6 and 7). Primers were designed according to Illumina's "Index Adapters Pooling Guide" and Nextera N505—N506 (i5) as well as N701—N702/N704—N705 (i7) barcodes were used. Each primer pair contained an individual combination of barcodes for distinguishing samples from each other.

For the first PCR, four replicate 40 µl reactions were set up with Q5 polymerase (NEB) according to the manufacturer's protocol. Primer stock concentrations were 5 µM, and 2 µl of each were used. 300 ng of genomic DNA per reaction was used as a template. PCR program was: 1x 5 min at 98 °C, 3x (30 s at 98 °C, 2 min at 67 °C, 1 min at 72 °C), 1x 5 min at 72 °C. PCR reactions were purified using Agencourt AMPure XP Beads (Beckman Coulter) according to the manufacturer's protocol for PCR purification (96-well format), with the exception that 0.9x of beads per reaction volume were used to eliminate >100 bp primer dimers. Each parallel reaction was eluted to 32 µl of elution buffer, and reactions were pooled together before the 2nd PCR.

For each sample, four 40 µl reactions were set up with Q5 polymerase. 25 µl of purified PCR-product from the 1st PCR was used as template, and 3 µl of 5 µM forward and reverse primers were used. PCR program was: 1x 5 min at 98 °C, 20x (30 s at 98 °C, 1 min at 72 °C), 1x 5 min at 72 °C. Reactions were purified, eluted and pooled as described above. Concentrations of purified PCR products were measured with Qubit Fluorometer (Thermo Fisher Scientific). Sample concentrations were adjusted to 4nM and pooled together. NGS library and PhiX-

control were denatured and diluted further to 20 pM according to Illumina's protocol. 550 µl of the library with 45 µl of PhiX-control (7.5 %) was loaded to MiSeq® Reagent Kit v3 (600 cycles) (Illumina). The library was run using MiSeq-sequencer (Illumina).

Table 5. Primers for amplifying the 516 bp amplicon from *BCL6* 1st intron for next-generation sequencing (1st round of PCR) (unpublished).

PRIMER NAME	COMPLEMENTARY TO 2ND PCR PRIMERS	PRIMER BACKBONE	GENOME BINDING PART
BCL6-Int_SHM_F	TCGTCGGCAGCGTC	AGATGTGTATAAGAGACAG	GGAAAGCAGTTTGCAAGCGA
BCL6-Int_SHM_R	GTCTCGTGGGCTCGG	AGATGTGTATAAGAGACAG	GCCGCTGCTCATGATCATTATTT

Table 6. Primers for amplifying the 2nd PCR product from 1st PCR and adding Illumina sequencing indexes (unpublished).

PRIMER NAME	PRIMER BACKBONE	INDEX*	COMPLEMENTARY PART TO 1ST PCR
N50x	AATGATACGGCGACCACCGAGATCTACAC	i5	TCGTCGGCAGCGTC
N70x	CAAGCAGAAGACGGCATACGAGAT	i7	GTCTCGTGGGCTCGG

*see used index in Table 7.

Table 7. Used sequencing indexes and their combinations for distinguishing samples from each other (unpublished).

CELL LINE	ENHANCER MODIFICATION	TREATMENT	I5	I5 BASES IN ADAPTER	I7	I7 BASES IN ADAPTER
Mieu	none	none	N505	GTAAGGAG	N701	TCGCCTTA
Carnaval	none	none	N506	ACTGCATA	N701	TCGCCTTA
Ramos	LPP 1928/-	Doxycycline	N505	GTAAGGAG	N702	CTAGTACG
Ramos	LPP 1928/-	none	N506	ACTGCATA	N702	CTAGTACG
Ramos	LPP -/-	Doxycycline	N505	GTAAGGAG	N704	GCTCAGGA
Ramos	LPP -/-	none	N506	ACTGCATA	N704	GCTCAGGA
Ramos	BCL6e wt/wt	Doxycycline	N505	GTAAGGAG	N705	AGGAGTCC
Ramos	BCL6e wt/wt	none	N506	ACTGCATA	N705	AGGAGTCC

The quality of sequencing reads was checked using FastQC version 0.11.9 in the Chipster web app provided by CSC – IT Center for Science Ltd. Raw sequencing reads were preprocessed with umitools and umicollapse packages to remove duplicate reads caused by PCR amplification based on recognition of UMIs. Next, the sequencing data were processed by using custom-made analysis pipeline (The MiSeq code) developed by Anurupa Yadavalli at David Schatz's laboratory, at Yale University, USA. The MiSeq code utilizes Samtools, Bedtools, Picard, BWA, BBMap, fastq-join, pysamstats, and jvarkit packages. The first round of analysis without UMI recognition was performed using the Puhti supercomputer provided by CSC. Analysis packages were downloaded to Puhti from Bioconda (https://bioconda.github.io/) using the Tykky tool. The final analysis with UMI recognition was performed by Anurupa Yadavalli, as incorporating UMI processing as a part of the MiSeq code required advanced bioinformatical expertise.

The processed results were further analyzed in Excel and figures were made with GraphPad Prism. From each cell line, the overall mutation frequency {(n of mutations/n of reads)/length of the amplicon} was calculated. In addition, mutation frequency per base pair was calculated (n of mutated reads/ n of sequenced reads per position) and median mutation frequency was calculated. The median mutation frequencies were compared between cell lines. Mutation frequencies of AID hotspot mutations were calculated using the same method. The percentages of six different substitution types and insertion/deletion mutations were calculated {(number of certain mutation type /number of all mutations) × 100} in each studied cell line. Mutation frequencies in different alleles were determined by first selecting reads with more than one mutation. These reads were divided into two groups based on whether they appeared on same read as one or more single nucleotide polymorphisms (SNPs) identified in one of the *BCL6* alleles in Ramos cells. These SNPs are T>A at 214 bp, T>C at 223 bp and C>G at 228 bp. After this, the median mutation frequencies per base pair were calculated as described earlier in this section.

4.4 Gene expression studies

4.4.1 Real-time quantitative PCR (unpublished)

BCL6 mRNA expression was determined using RT-qPCR from *BCL6* enhancer mutant cell clones (unpublished). 5 million cells were collected for RNA extraction which was performed with Quick-RNA Miniprep Kit (Zymo Research) according to the manufacturer's protocol. The extracted RNA was converted to cDNA using qScript cDNA SuperMix (Quanta) according to the manufacturer's protocol. Primers for

amplifying the *BCL6* amplicon were forward: TACTCAGATTCTAGCTGTGAGAAC and reverse: GGCAGCGGTCACACTTGTA. Primers for amplifying the *GAPDH* amplicon were forward: ACCTGACCTGCCGTCTAGAAA and reverse: ACCACCTGGTGCTCAGTGTA. For the 20 µl reaction, 0.8 µl of each 10 µM primer, 5 µl of cDNA template at 1:5 dilution and 10 µl of Sensifast 2x master mix (Meridian Bioscience) were used. LightCycler 480 (Roche) instrument was used with the following protocol: 1x (2 min at 95°C) and 50x (15 s at 95°C, 5 s at 60°C, 7 s at 72°C). *BCL6* mRNA expression (Ct value) was compared to *GAPDH* housekeeping gene mRNA expression using the $2^{-\Delta\Delta Ct}$ method. ΔCt (*BCL6* vs. *GAPDH* expression of corresponding cell lines), ΔΔCt (*BCL6* expression) and expression fold changes were calculated (*BCL6* expression relative to $BCL6e^{wt/wt}$ control cell line).

4.4.2 Western blot (II, unpublished)

To evaluate LT protein expression (II), BCL6 protein expression (unpublished) and AID protein expression (II and unpublished) of modified Ramos (II and unpublished) and DT40 cell lines (II) western blot analysis was performed.

Protein extraction and the SDS-PAGE procedure are described in detail in Study II. After the SDS-PAGE, proteins were transferred to nitrocellulose membrane using Xcell blotting system (Thermo Fisher Scientific) or eBlot semidry blotting system (Genescript). The conditions for transfer were 30 V for 1 hour at RT (Xcell) or 9/10-minute program at RT (eBlot). The membranes were blocked with 5 %-BSA-TBS for 2–3 hours at RT. Primary antibody incubations were conducted overnight at +4°C in a rocker (LT, BCL6 and AID) or at RT for 15 min in a rocker (GAPDH). Secondary antibody incubations were performed in the dark at RT in a rocker for 2–3 hours (LT, BCL6 and AID) or 15 minutes (GAPDH). The membranes were imaged using Odyssey Fc machine (LICOR).

Primary antibodies used for western blot in this thesis were αAID 30F12 Rabbit (Cell Signaling) at a 1:1000 dilution, αSV40-LT pAb419 Mouse (Genetex) at a 1:100 dilution, αBCL6 #4242 Rabbit (Cell Signaling) at a 1:1000 dilution and αGAPDH 5G4 Mouse (Hytest) at a 1:10 000 dilution. Used secondary antibodies were IRdye αRabbit 680RD at a 1:2500 dilution for detecting AID and BCL6 expression, IRdye αMouse 800CW at a 1:5000 dilution for detecting LT expression, and IRdye αMouse 680RD at a 1:10000 for detecting GAPDH expression. All the antibodies were diluted with 5 % BSA–TBS.

4.4.3 Analysis of RNA sequencing data (I)

RNA sequencing data of 82 analyzed Finnish MCC tumor samples were obtained from the Rare Cancers Research Group at the University of Helsinki. The details

concerning permissions to collect patient data and tissue samples as well as ethical approval of study protocol regarding these data are described in Sundqvist et al., 2023. The RNA sequencing data are available in the SRA database in NCBI and can be accessed under BioProject PRJNA775071. Data preprocessing was done by Minna Kyläniemi from the Finnish Functional Genomic Centre. For Study I, the fold changes for *AICDA* and *APOBEC* expressions were calculated by comparing normalized read counts of MCPyV+ and MCPyV- MCC samples to each other. Expression levels of *AID/APOBEC* and selected B cell markers were studied. In addition, the expression of DNA repair genes was studied (unpublished). Heatmap and similarity matrix were conducted using Morpheus software (Broad Institute).

RNA sequencing data from Ramos cells were previously produced in our laboratory (unpublished). For this thesis, RPKM values of *AICDA, APOBEC1, APOBEC2, APOBEC3A-3H, APOBEC4, MSH2, MHS6, UNG, REV1, POLH, PRIMPOL, EXO1, PCNA, BRCA1* and *BRCA2* were normalized to *GAPDH* values and the mean of two replicate values was calculated and plotted with standard deviation.

4.4.4 Luciferase assay (unpublished)

Gene expression induced by *BCL6* pk1 and LPP fragments was measured using luciferase assay (unpublished). *BCL6* pk1 and LPP enhancers were cloned to the luciferase vector pGL4.23 at Sal//BamHI site. The luciferase assay procedure is described in Study II.

4.5 Generation of *BCL6* enhancer-modified cell lines (unpublished)

The enhancer deletions/insertions were made to modified Ramos cells with doxycycline (dox)-induced AID expression to achieve better control of the AID expression levels. In addition, the *UNG* gene is also deleted from this cell line to increase the detection of AID-induced mutations in the absence of the BER repair pathway. To investigate if SHM targeting elements found in *BCL6* enhancer affect mutation accumulation to the *BCL6* 1st intron, pk1 or LPP elements were deleted individually from Ramos cell line genome (pk1 2092 bp chr 3: + 187900285—187902376; LPP 1674 bp chr 3: + 188088156—188089829 on hg38). In addition, the previously described SHM targeting element 1928 was inserted into deletion sites (Blagodatski et al., 2009; Kohler et al., 2012).

4.5.1 Design and testing of guide RNAs

Guide RNAs (gRNA) directing the CRISPR/Cas9-system to *BCL6* pk1 and LPP fragment sites were designed using Benchling.com and chop.chop.rc.fas.harvard programs. gBlocks containing gRNAs were ordered from IDT. Both gRNAs at the gBlock were designed to target the same strand in the same orientation according to Canver et al. 2014. gBlocks were cloned to px458 plasmid at the BbsI site with T4 ligase. Plasmids were isolated with GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) and the success of cloning was checked with appropriate restriction enzyme digestion. Cloned gBlocks were transfected into HEK293T cells to test their functionality. Transfection mixture containing 1 µg px458 plasmid with gBlocks and 3ul X-tremeGENE HP transfection reagent in 100 µl of DMEM were incubated at RT for 15 minutes. The mixture was added to two 24-well plate wells of ~ 90 % confluent cells with fresh culture media. After 48 hours of incubation, the success of transfections was evaluated with fluorescence microscopy. Genomic DNA was extracted with genomic DNA extraction kit (Zymo Research) and PCR was used to amplify the genomic area where the deletion was intended. Based on amplicon length seen in gel electrophoresis, the efficiency of gBlocks were determined. The repair template containing the 1928 element was cloned to pUC19-vector with NEBBuilder (NEB) according to the manufacturer's protocol.

4.5.2 Transfection

2 µg of gBlock containing plasmid and repair template containing plasmid were transfected to 2 million Ramos AID-inducible cells. The gBlock plasmid, the repair template and the cells were mixed in 100 µl of SG transfection mixture for cell lines (Lonza), and electroporation was conducted using Nucleofector 4D machine program CL-120. After transfection, cells were transferred to a 6-well plate and cultured for 48 hours in conditioned media before cell sorting. The conditioned media contained 25 % of media collected from cell culture and extra 5 % of FBS in addition to regular culture media.

4.5.3 Cell sorting

Transfected Ramos AID-inducible cells were sorted with FACSAria cell sorter (BD Bioscience), and GFP-positive cells were sorted to a 96- well plate, three cells per well. To improve cell viability, cells were sorted to conditioned culture media.

4.5.4 Screening of targeted cell clones

Single clones were picked around day 16 after GFP-positive cell sorting and clones were cultured until genomic DNA was extracted around day 30 post sort. Clones were screened with PCR using LongAmp polymerase (NEB) according to manufacturer's protocol for CRISPR-induced modifications at *BCL6* enhancer area using primers forward: AGAGCATTTTACAAGTGGCAGAGAGCTT and reverse: TTACAGGTGAAGCCAGCTATTGAGAGAAA for pk1 and forward: TGAAAAGTACCTCACAGGAGAAATCAGA and reverse: CAATTTGCACATCTCCACTTTTCAACTC for LPP. Length of the amplicon indicated if the allele had deletion (pk1 432bp, LPP 660 bp), deletion and repair template insertion (pk1 2400bp; LPP 2660bp) or wild-type (pk1 2523bp; LPP 2334bp). The genotypes of obtained cell lines were $pk1^{1928/-}$, $pk1^{-/-}$, $LPP^{1928/-}$ and $LPP^{-/-}$.

4.5.5 Maintenance of *BCL6* enhancer-modified cell lines

BCL6 enhancer-modified Ramos cell lines were cultured for eight weeks in the presence or absence of doxycycline (800 ng/ml) and histidinol (0.5 mg/ml) to induce and maintain AID expression. During an 8-week culture period, each selected *BCL6* enhancer-modified cell line and parental AID-inducible cell line (hereafter referred to $BCL6^{wt/wt}$) were cultured parallel with and without doxycycline to induce AID expression. Control cell lines Mieu and Carnaval were maintained for 12 weeks.

4.6 Analysis of cellular growth (unpublished)

BCL6-enhancer-modified cell lines were divided to a density of 100 000 cells/ml in three replicate cultures per line. Cell density was monitored at time points of 0 h, 24 h, 48 h, 72 h, 92 h, 98 h, 115 h, 139 h, and 163 h with NovoCyte (Agilent) flow cytometer. Results from cytometry were analyzed with FlowJo software.

4.7 Statistical analyses (I, II, unpublished)

In the GFP loss assay, statistical significance between the median GFP loss of tested enhancers and the negative control, or between two tested enhancers (two groups), was calculated using the Mann–Whitney U-test. The Mann–Whitney U-test was also used to determine statistical significance between the medians of mean fluorescence intensity. Statistical significance of differences between MCC and control mutations (distribution, type, STOP codon formation and hotspot) was determined using the same test. In the luciferase assay, the statistical significances of enhancers relative to luciferase activity normalized to positive control were

calculated using unpaired t-test. Fisher's exact test was used to evaluate the statistical significance of WR**C** mutations compared to overall C-targeting mutations. Fisher's exact test was also used to calculate the statistical significance of differences between AID/APOBEC-expressing and non-expressing MCC samples. Further, Fisher's exact test was used to evaluate statistically significant differences of high and low sun-exposed MCC tumors and their MCPyV status. One minus the Spearman rank correlation was used to calculate the heatmap hierarchical clustering, and the Spearman rank correlation was used for generation of similarity matrices. The Mann–Whitney U-test was used to calculate the statistical significances between mutation frequencies in the *BCL6* 1st intron in different cell lines and alleles. All the graphs and statistical analyses in this thesis were conducted using GraphPad Prism-software.

5 Results

5.1 Polyomavirus NCRRs have SHM targeting activity

In Studies I and II, the polyomavirus's ability to target SHM was investigated. Thus, the SHM targeting activity was measured with GFP loss assay from the whole genome (I), the NCRR (I, II) or subfragments of NCRR (I) (Figure 5). The p-values presented in tables in this chapter represent comparison to negative control.

In Study I, the entire MCPyV genome (strain R17b) was tested (Table 8). Weak, but statistically significant SHM targeting activity was observed in a 440 bp subfragment when it was cloned downstream of the GFP expression cassette, compared to negative control and the same fragment cloned upstream ($p=0.0161$ and $p<0.0001$, respectively). Interestingly, downstream cloned 204 bp fragment, which is a subfragment of the 440bp fragment, also exhibited significantly higher SHM targeting activity compared to its upstream-cloned counterpart ($p<0.0001$).

Table 8. GFP loss (median) of MCPyV genome and NCRR measured in DT40 cell line in GFP4 reporter (I).

FRAGMENT	GFP LOSS	P-VALUE	STUDY
MCPyV Fr1	2.3 %	0.7863	I
MCPyV Fr2	2.6 %	0.4659	I
MCPyV Fr3	1.4 %	0.0013	I
MCPyV Fr4	2.4 %	0.7144	I
MCPyV Fr1 truncated	2.4 %	0.9215	I
MCPyV Fr2 truncated	2.3 %	0.7978	I
MCPyV Fr1 flipped	1.8 %	0.0901	I
MCPyV NCRR 440bp upstream	1.7 %	0.0004	I
MCPyV NCRR 440bp downstream	3.2 %	0.0161	I
MCPyV NCRR 236bp upstream	1.7 %	0.0009	I
MCPyV NCRR 236bp downstream	1.2 %	<0.0001	I
MCPyV NCRR 204bp upstream	1.1 %	0.0001	I
MCPyV NCRR 204bp downstream	3.1 %	0.3783	I
MCPyV NCRR 80bp upstream	1.9 %	0.0672	I
MCPyV NCRR 80bp downstream	2.0 %	0.4128	I
Positive control (hIgλ)	11.7 %	<0.0001	I
Negative control	2.4 %		I

In Study II, the SV40 NCRR (776 strain) was found to have strong SHM targeting activity compared to negative control in DT40 and Ramos B cells ($p<0.0001$) (Tables 9 and 10). The magnitude was comparable to Ig enhancers (Tables 9 and 10). Increasing the number of 72-bp repeats increased also the SHM targeting activity of the SV40 NCRR in DT40 cells in GFP2 (II) and GFP4 reporter (Table 9) and in Ramos cells in GFP7 reporter (Table 10). Deletion of transcription factor IRF-Ets (IRF4, PU.1), NF-kB and/or E-box binding sites (IPEN) decreased SHM targeting activity compared to non-mutated SV40 NCRR (Table 9). Surprisingly, the SV40 NCRR has SHM targeting activity beyond the B cell context in renal cell and breast carcinoma, epithelial and fibroblast cell lines (Table 10), which has not been shown for Ig enhancers (II). The difference compared to negative control was statistically significant in HEK293T ($p=0.0071$) and MDA-MB-231 cells ($p=0.0022$). Cloning the SV40 NCRR downstream of the GFP expression cassette decreased SHM targeting activity compared to the upstream position ($p=0.0142$) (Table 9). However, the SHM targeting activity of SV40 NCRR in the downstream position is significantly higher than in negative control ($p <0.0001$).

In addition, the SHM targeting activity of JCPyV (Mad-1 strain) and BKPyV (Dunlop strain) NCRRs were measured (unpublished). The JCPyV NCRR had statistically significant SHM targeting activity compared to the negative control ($p<0.0001$) (Table 9). BKPyV did not show SHM targeting activity (Table 9).

Table 9. GFP loss (median) of SV40 NCRR and modifications as well as JCPyV and BKPyV NCRR measured in DT40 cells in GFP4 reporter (II and unpublished).

FRAGMENT	GFP LOSS	P-VALUE	STUDY
SV40 NCRR (2x72bp)	47.5 %	<0.0001	II
SV40 NCRR (4x72bp)	74.3 %	<0.0001	II
SV40 NCRR (2x72bp) IPEN mutant	0.28 %	0.0005	II
SV40 NCRR (2x72bp) NF-kB mutant	7.56 %	<0.0001	II
SV40 NCRR (2x72bp) E-box1 mutant	33.8 %	<0.0001	II
SV40 NCRR (2x72bp) E-box2 mutant	47.6 %	<0.0001	II
SV40 NCRR (2x72bp) E-box1+2 mutant	27.4 %	<0.0001	II
Positive control (hIgλ)	20.2 %	<0.0001	II
Negative control	1.89 %		II
SV40 NCRR (2x72bp) downstream	12.4 %	<0.0001	unpublished
JCPyV NCRR	7.29 %	<0.0001	unpublished
BKPyV NCRR	1.98 %	0.4149	unpublished
Negative control	1.99 %		unpublished

Table 10. GFP loss (median) of SV40 NCRR measured in Ramos and other human cell lines in GFP7 reporter (Study II).

FRAGMENT	GFP LOSS	P-VALUE	STUDY
SV40 NCRR (4x72bp) in Ramos cells	1.76 %	<0.0001	II
SV40 NCRR (2x72bp) in Ramos cells	0.86 %	<0.0001	II
SV40 NCRR (1x72bp) in Ramos cells	0.17 %	<0.0001	II
Positive control (IgHi) in Ramos cells	1.18 %	<0.0001	II
Negative control	0.02 %		II
SV40 NCRR (4x72bp) in UO-31 cells	1.4 %	0.0590	II
Negative control	0.26 %		II
SV40 NCRR (4x72bp) in HEK293T	1.99 %	0.0071	II
Negative control	1.02 %		II
SV40 NCRR (4x72bp) in NIH 3T3	3.24 %	0.2222	II
Negative control	1.11 %		II
SV40 NCRR (4x72bp) in MDA-MB-231	0.74 %	0.0022	II
Negative control	0.04 %		II
SV40 NCRR (4x72bp) in U2OS	0.66 %	0.0909	II
Negative control	0.21 %		II

It should be noted that GFP2, GFP4, and GFP7 differ in their experimental designs and sensitivities, and thus the GFP loss values measured from different vectors are not directly comparable. The GFP4 vector system is more sensitive compared to GFP2, as GFP4 contains a hypermutation targeting sequence (HTS) within the GFP gene, which is enriched with potential STOP codons. If mutations occur in the HTS, they are likely to result in a complete loss of GFP fluorescence in that cell clone. This is because the HTS contains sequences designed to form STOP codons due to transition mutations. The UNG gene is deleted from the DT40 cell line used with the GFP4 reporter, as deletion of UNG in DT40 cells causes a strong bias toward transition mutations (Saribasak et al., 2005). The GFP7 reporter also contains an HTS similarly prone to STOP codon formation due to cytidine-targeting mutations (Senigl et al., 2019), but this assay is commonly conducted in UNG-proficient Ramos cells. In addition, GFP7 is introduced into cells via a lentivirus reporter, which inserts the reporter at random sites in the genome, whereas GFP2 and GFP4 are targeted to a fixed locus which might affect the reporter mutability.

5.2 SV40 NCRR, but not MCPyV, JCPyV and BKPyV NCRRs increases mean fluorescence intensity in GFP reporter system

To investigate the relationship between mutation targeting activity and transcription of GFP gene, the mean fluorescence intensity (MFI) was determined. GFP loss represents the percentage of cells that have lost the GFP fluorescence due to accumulated mutations. The MFI reflects the *GFP* gene transcription and is determined from the GFP-positive population in the GFP loss measurement. It should be noted that features other than transcription affect the MFI, thus transcription is not equivalent to MFI. Nevertheless, MFI can give some approximate of enhancers ability to promote transcription. Unlike GFP loss values, MFI values from experiments measured using different cytometers are not comparable. Thus, these values should only be compared with values presented in the same table. The p-values presented in tables in this chapter represent comparison to negative control, except for SV40 NCRR E-box mutants where the comparison is towards 2x72bp repeat SV40 NCRR.

For the fragments in Study I, MFI was higher than in negative control in 80bp and 204 bp upstream cloned fragments (Table 11). The MFI was lower than in the negative control for all other NCRR fragments (Table 11). The MFI was lower in downstream than upstream cloned fragments for the 80 bp ($p<0.0001$), 204 bp ($p=0.0031$) and 440 bp ($p=0.2203$) fragments but vice versa in the 236 bp ($p<0.0001$) fragment (Table 11). These results suggest, that MCPyV NCRR fragments are not very strong transcription enhancers at least in the GFP4 reporter system in B cells. However, the 204 bp fragment might have some effect on transcription.

Table 11. MFI (median) of MCPyV NCRR fragments (I) measured in DT40 cell line (unpublished).

FRAGMENT	MFI	P-VALUE	REPORTER
MCPyV NCRR 440bp upstream	20608	<0.0001	GFP4
MCPyV NCRR 440bp downstream	18950	<0.0001	GFP4
MCPyV NCRR 236bp upstream	19458	<0.0001	GFP4
MCPyV NCRR 236bp downstream	21356	0.001	GFP4
MCPyV NCRR 204bp upstream	27785	<0.0001	GFP4
MCPyV NCRR 204bp downstream	23526	0.3268	GFP4
MCPyV NCRR 80bp upstream	24970	0.1985	GFP4
MCPyV NCRR 80bp downstream	21072	0.0004	GFP4
Negative control	24221		GFP4

In Study II, increasing the number of 72-bp repeats in the SV40 NCRR was associated with increase in MFI in Ramos cells (Table 12). SV40 NCRR with four 72-bp repeats showed significantly higher MFI compared to SV40 NCRR with two 72-bp repeats (p=0.0002). However, there was no difference between SV40 NCRR with two 72-bp repeats and SV40 NCRR with one 72-bp repeat (p=0.8510).

Table 12. MFI (median) of SV40 NCRRs with different number of 72-bp repeats measured in Ramos cells (II).

FRAGMENT	MFI	P-VALUE	REPORTER
SV40 NCRR (4x72bp)	3158	<0.0001	GFP7
SV40 NCRR (2x72bp)	2584	<0.0001	GFP7
SV40 NCRR (1x72bp)	2596	<0.0001	GFP7
Negative control	2264		GFP7

Compared to the negative control, the MFI was significantly lower when the SV40 NCRR was cloned downstream of the GFP expression cassette (p<0.0001) (Table 13). The MFI of NF-kB-mutated SV40 NCRR was higher than in the negative control or the SV40 NCRR in the downstream position (p <0.0001) (Table 13). Thus, the deletion of NF-kB did not affect MFI. Compared to the non-mutated SV40 NCRR, there were no differences of MFI in E-box mutants (Table 14). Thus, the deletion of E-boxes did not affect MFI. However, deletion of IPEN TFBSs abolished SHM targeting activity, as well as the ability to drive gene expression measured using luciferase assay (II).

Compared to the negative control, JCPyV and BKPyV showed a statistically significant decrease in MFI (p= 0.0108 and p <0.0001, respectively) (Table 13). Similarly to MCPyV NCRR fragments, JCPyV and BKPyV do not seem to promote transcription of the GFP reporter in B cells.

Table 13. MFI (median) of SV40 NCRR cloned in downstream orientation, SV40 NF-kB mutant, JCPyV NCRR and BKPyV NCRR measured in DT40 cells (unpublished).

FRAGMENT	MFI	P-VALUE	REPORTER
SV40 NCRR (2x72bp) downstream	77308	<0.0001	GFP4
SV40 NCRR (2x72bp) NF-kB mutant	90767	>0.9999	GFP4
JCPyV NCRR	83697	0.0108	GFP4
BKPyV NCRR	78599	<0.0001	GFP4
Negative control	87277		GFP4

Table 14. MFI (median) of SV40 NCRR E-box mutants (II) measured in DT40 cells (unpublished).

FRAGMENT	MFI	P-VALUE	REPORTER
SV40 NCRR (2x72bp)	166		GFP4
SV40 NCRR (2x72bp) E-box1 mutant	157	0.5007	GFP4
SV40 NCRR (2x72bp) E-box2 mutant	164	0.7826	GFP4
SV40 NCRR (2x72bp) E-box1+2 mutant	172	0.5697	GFP4

5.3 Mutation pattern and frequency in polyomavirus *LT* region

To further investigate if the SHM targeting activity of polyomavirus NCRRs causes mutation accumulation to neighboring genes, the mutation pattern and distribution of the viral *LT* region, which is adjacent to the NCRR region, were analyzed (Figure 4). In Studies I and II, the mutations were found to be distributed along the *LT* region. In Study I, the mutation frequency of the MCPyV *LT* peaks strongly between 1200–2200 bp and at the end of the *LT* at bin 2885 bp in the MCC sequences, but not in control sequences. This is likely due to the selection of certain mutations to tumor sequences. In addition, the overrepresentation of C bases at the 1000–1400 bp region might cause an increase in C-targeting mutations in this area (I). In Study II, the mutation frequency of the SV40 *LT* in Ramos and UO-31 cells is highest at 600–1600 bp. In addition, mutation frequency is also high at the end of the *LT* region at bin 2473 bp in Ramos cells (II). Due to the different experimental settings in Studies I and II, the mutation distribution in Study I is skewed by mutation selection during tumor development. Substitutions from cytosines comprise 74.9 % of all mutations in the MCPyV *LT* in MCC sequences whereas, in MCPyV *LT* control sequences, SV40 *LT* sequences in Ramos and UO-31 samples, the proportion of C-targeting substitutions ranges from 40 % to 54.4 % (Figure 7).

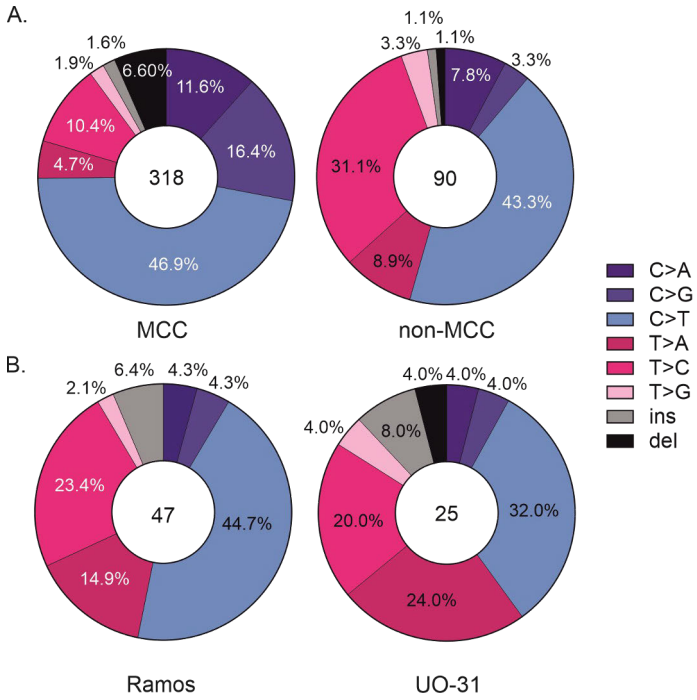

Figure 7. Mutation type distribution in MCPyV and SV40 *LT* regions. A. Mutations in MCPyV *LT* region from MCC and non-MCC samples (control) (I). B. Mutations in SV40 *LT* region in Ramos and UO-31 cells. Inside the circle is the total number of mutations. The percentage of each mutation type is shown. 12 different substitution types were converted to six classes which are indicated in legend.

According to mutation signature profiles, mutations caused by APOBEC enzymes are predominantly C>T or C>G in the T**C**W context (SBS2 and SBS13), whereas mutations caused by AID are often C>T in the WR**C** context (SBS84) (Alexandrov et al., 2013, 2020). AID also indirectly induces frequent T>A and T>C mutations (SBS85) (Alexandrov et al., 2020). Therefore, the mutation accumulation in these hotspots was determined in Studies I and II. Furthermore, in Study I, mutations in UV hotspots (Y**C**C) were analyzed in MCC and control samples, as MCC is a skin cancer in which UV-radiation is a likely mutagenic agent. These results revealed that mutations are enriched in T**C**W and Y**C**C hotspots in MCC samples compared to control samples (I). Enrichment was also detected in four out of five APOBEC3 subfamily hotspots (I). In contrast, more mutations were detected at AID hotspots (WR**C**) in control samples than in MCC samples (Figure 8), although the difference was not statistically significant (I). In Study II, the SV40 *LT* showed a statistically significant amount of C targeting mutations at WR**C** hotspots both in Ramos and UO-31 cells ($p=0.0025$ and $p=0.0012$, respectively). The proportion of SV40 *LT* mutations in T**C**W was almost as high as in WR**C** in Ramos

(38.1 % of C targeting mutations in T**C**W vs 42.9 % in WR**C**), but not in UO-31 (60.0 % vs 10.0 %, respectively) (Figure 8).

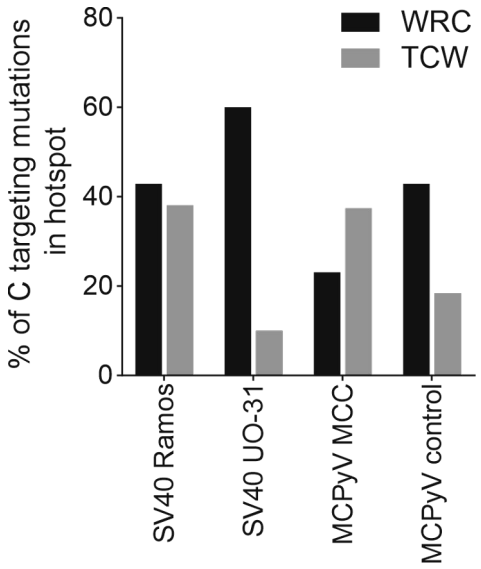

Figure 8. AID (WR**C**) and APOBEC3 (T**C**W) hotspot mutations in MCPyV and SV40 *LT* region.

The DNA repair pathway utilized influences the mutation profiles induced by AID and APOBECs (Petljak & Maciejowski, 2020; Pilzecker & Jacobs, 2019). Therefore, the expression of repair genes from the Finnish MCC samples (I) and Ramos RNA sequencing datasets (unpublished) were investigated. Selected DNA repair genes are important in MMR, BER and homology-directed (HR) repair pathways as well as in SHM-associated repair (Petljak & Maciejowski, 2020; Pilzecker & Jacobs, 2019). Analysis of the DNA repair gene expressions in Finnish MCC samples showed widespread expression of *MSH2, MSH6, UNG, BRCA1, BRCA2, POLH, and PCNA* in these tumor samples (Figure 9A). This analysis revealed that there is a statistically significant positive correlation of *LT* with *UNG* and *BRCA2* (r=0.23, p=0.038 and r=0.28, p=0.011, respectively) and a significant negative correlation with *REV1* (r=-0.24, p=0.030) (Figure 9B). Unpublished RNA sequencing data revealed that Ramos cells express all the DNA repair genes examined in this thesis (Figure 10A). Thus, the studied MCC tumor samples (I) and Ramos cells (II) do not have major deficiencies in the studied DNA repair genes.

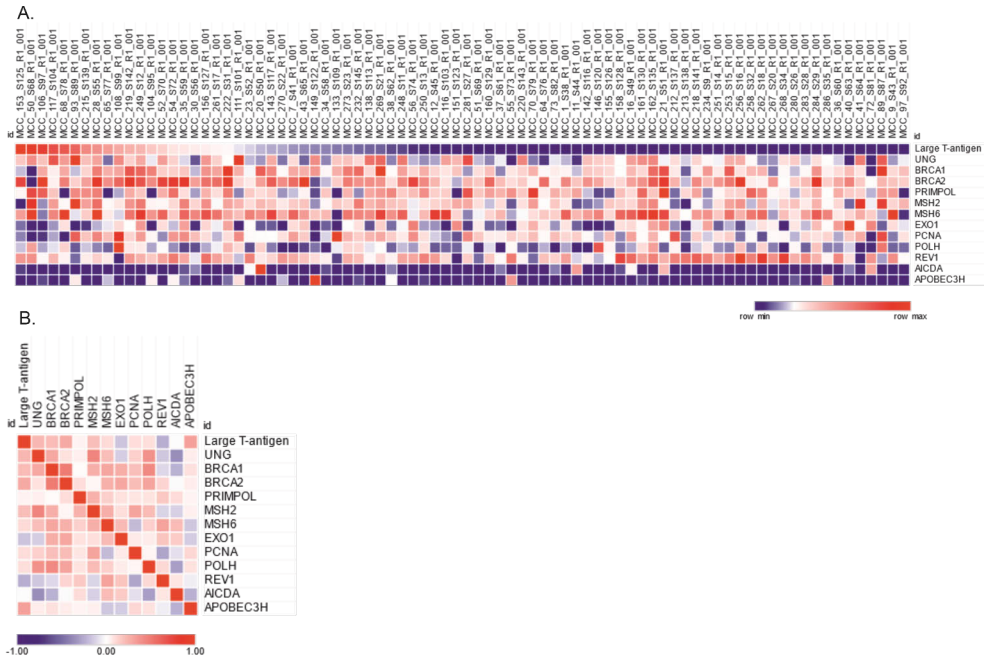

Figure 9. Expression of selected DNA repair genes in Finnish MCC samples. A. Heat map presenting the relative expression of selected DNA repair genes with *LT, AICDA* and *APOBEC3H* expression in individual tumor samples. Samples are arranged according to *LT* expression level. B. Similarity matrix of DNA repair and *LT, AICDA* and *APOBEC3H* expression.

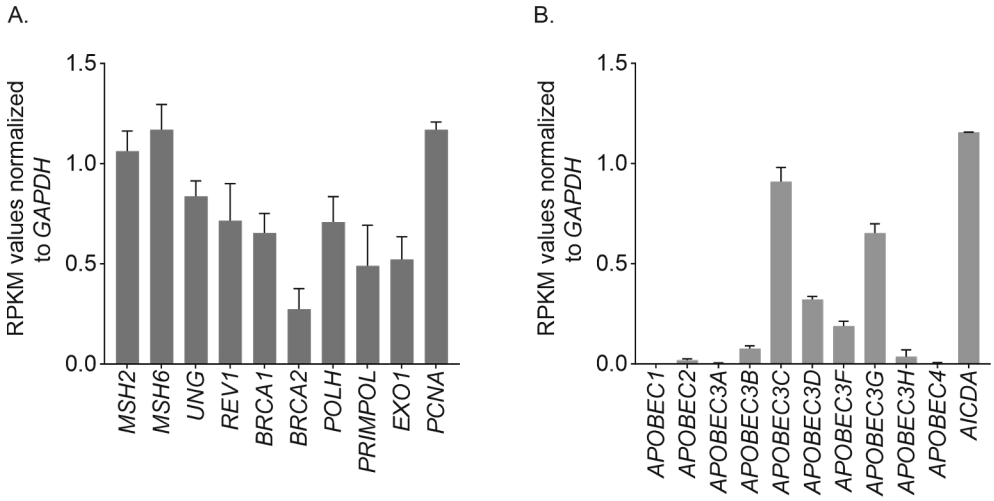

Figure 10. Expression of DNA repair genes and genes encoding AID/APOBEC enzymes in Ramos cells (unpublished). A. Expression of selected DNA repair genes plotted by RPKM values. B. Expression of AID/APOBEC coding genes plotted by RPKM values. RPKM values were normalized to *GAPDH* expression. The mean of two samples with standard deviation is shown.

Since truncation mutations in the *LT* are important for MCPyV-induced tumorigenesis, and a truncated LT may help preserve viral DNA in host cells, mutations in MCPyV and SV40 *LT* that induce a STOP-codon, either at mutation site or via frameshift mutations, were analyzed. In Study I, STOP-codon mutation frequency was increased at the 1200–2000 bp region in MCC MCPyV *LT* (I), which is in line with previously reported truncation events (Li, J. et al., 2013; Schmitt et al., 2012; Shuda et al., 2008). In addition, STOP-codon-forming mutations co-occurred with APOBEC T**C**W hotspot mutations, implicating that APOBECs are responsible for *LT* truncating mutations seen in MCPyV+ MCCs (I).

In Study II, SV40 *LT* STOP-codon-forming mutations were identified in the 800–1000 bp, 1600–1800 bp and 2473 bp regions in Ramos cells, and in the 1000 bp, 1400 bp and 2473 bp regions in UO-31 cells, establishing that truncation mutations occur in the SV40 *LT* in both cell types.

5.4 AID/APOBEC enzyme activity in *LT* region contributes to truncated LT formation

In Study I, *LT* and *AID/APOBEC* expression levels from RNA sequencing data obtained from Finnish MCC samples were determined. In this cohort, 61.0 % of samples expressed MCPyV *LT*. As the purpose of this study was to investigate viral NCRRs function in mutation accumulation, the main interest was in MCPyV+ MCC samples. All AID/APOBEC family members were expressed in this sample cohort, although *APOBEC1, APOBEC3B* and *APOBEC4* levels were very low. *AID* expression was similar in both MCPyV+ and MCPyV- MCCs (32.0 % vs 34.4 %). *APOBEC3A, APOBEC3C, APOBEC3D, APOBEC3G, APOBEC3H and APOBEC4* were expressed more frequently in MCPyV+ than in MCPyV- MCCs, and the difference was statistically significant for *APOBEC3H* ($p=0.0024$). In addition, a statistically significant correlation between *LT* and *APOBEC3G* expression (Spearman correlation $r=0.25$, $p=0.025$) as well as *LT* and *APOBEC3H* expression ($r=0.32$, $p=0.004$) were detected. The correlation between T cell, B cell and germinal center markers and AID/APOBECs was observed (I), which most likely reflects the infiltration of lymphocytes to tumors. These results further support the view that the mutations seen in MCPyV *LT* in MCCs are induced by APOBEC3 enzymes.

In Study II, the SV40 *LT* transcription unit was integrated into the genome of Ramos, DT40 and UO-31 cells to investigate mutation accumulation in the *LT*. LT expression was confirmed using western blot and truncated form of LT was detected in DT40 cells (II). AID expression was confirmed using western blot, quantitative PCR and immunohistochemistry (II). Quantitative PCR revealed that the *AICDA* expression in UO-31 cells was lower than *AICDA* expression in Ramos cells (II). RNA sequencing data from Ramos cells revealed that these cells express all

*APOBEC*s, except for *APOBEC1* (Figure 10B). The expression of all APOBECs was lower than AID expression. *APOBEC2* expression was 1.7 % of *AICDA* expression, *APOBEC3A* 0.3 %, *APOBEC3B* 6.6 %, *APOBEC3C* 78.7 %, *APOBEC3D* 27.8 %, *APOBEC3F* 16.3 %, *APOBEC3G* 56.5 %, *APOBEC3H* 3.1 % and *APOBEC4* 0.3 %. Together with the sequencing results from Study II, these results suggests that the SHM targeting activity of the SV40 NCRR can guide AID to introduce mutations into the SV40 *LT*. Further, these mutations can occasionally lead to the expression of a truncated LT. Additionally, there is evidence of APOBEC-induced mutations in the SV40 *LT*.

5.5 *BCL6* enhancer has SHM targeting activity which is distributed across large enhancer region

To elucidate the SHM off-targeting capacity of endogenous enhancers, the SHM targeting activity of elements from the endogenous *BCL6* locus was investigated (unpublished). The investigated fragments (Figure 6) were cloned downstream of the GFP cassette to avoid overestimation of SHM targeting activity potentially caused by transcription. Statistically significant SHM targeting activity compared to negative control was found in enhancer elements E3 (GFP loss 6.48 %, $p<0.0001$), E4 (GFP loss 5.99 %, $p<0.0001$), E6 (GFP loss 2.55 %, $p= 0.0188$), E7 (GFP loss 2.29 %, $p= 0.0064$), e2f/pk1 (GFP loss 2.33 %/2.27 %, $p= 0.0078/0.0243$), LPP (GFP loss 2.85 %, $p<0.0001$) and 5'dist (GFP loss 2.27 %, $p= 0.0248$) (Figure 11). These findings suggest that the SHM targeting activity is distributed across a large enhancer region of approximately 300 000 bp, based on UCSC Genome Browser.

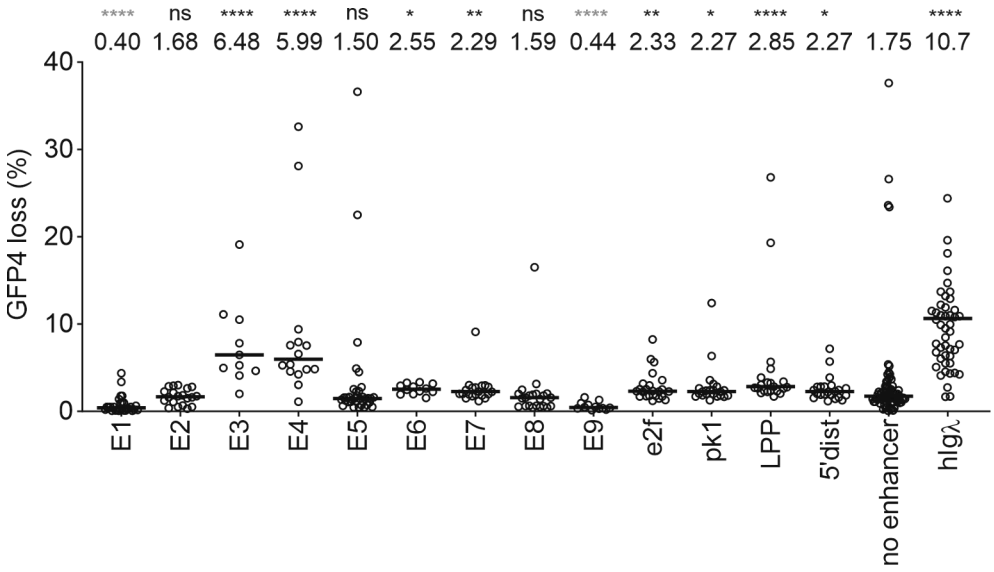

Figure 11. SHM targeting activity of *BCL6* enhancer fragments E1-E9, e2f, pk1, LPP and 5'dist measured with GFP loss assay (unpublished). The percentage of GFP loss is presented in y-axis. The median GFP loss value and the statistical significance of medians compared to negative control are presented above the graph. A grey asterisk indicates that the difference is significantly lower compared to the negative control. Data is pooled from several GFP loss experiments. Mann-Whitney U-test was used to calculate statistical significance. Negative control is the GFP4 vector without tested DNA element and positive control contains the human Igλ cloned into the GFP4 vector. Data points outside axis limits: E3 1, E4 1, E7 3, e2f 1, LPP 2, 5'dist 1, no enhancer 1, hIgλ 8.

The SHM targeting activity of pk1 and LPP was measured also in the upstream position (GFP loss pk1 upstream 3.21 % vs downstream 2.27 % p=0.0328; LPP upstream 1.75 % vs downstream 2.85 % p=0.0574) (Figure 12). Compared to the negative control, only LPP in the downstream position showed statistically significant SHM targeting activity (p=0.0165) providing an additional example where the enhancer element shows SHM targeting activity in the downstream position but not in the upstream position (Figure 12). The GFP loss of pk1 cloned upstream did not reach statistical significance when compared to the negative control, although its median value was higher than that of LPP in the downstream position. However, the percentage of GFP loss of pk1 was significant when the experiments were pooled (Figure 11). Thus, pk1 and LPP possess weak but statistically significant SHM targeting activity in DT40 cells. Pk1 has previously been shown exhibit SHM targeting activity previously in Ramos cells (Senigl et al., 2019). The SHM targeting activity of positive control hIgλ was significantly higher in the upstream position compared to the downstream position (GFP loss upstream 42.1 % vs downstream 10.5 %, p=0.0245) (Figure 12).

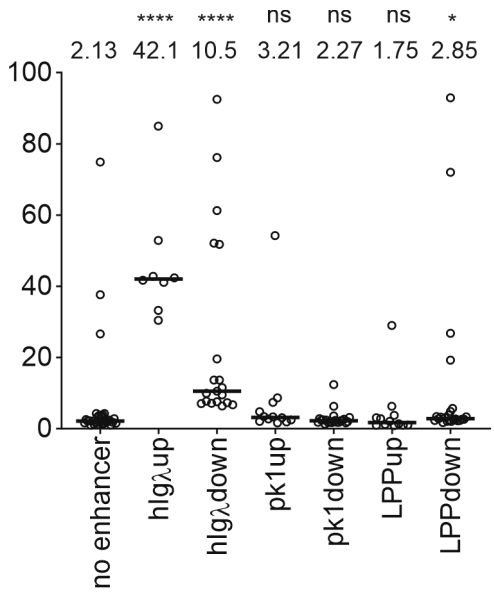

Figure 12. GFP loss of pk1 and LPP enhancers (unpublished). GFP loss of pk1, LPP and Igλ measured in the upstream and downstream positions. The percentage of GFP loss is presented in y-axis. The median GFP loss value and the statistical significance of medians compared to negative control are presented above the graph. The negative control is the empty GFP4 vector without tested DNA element and the positive control contains the human Igλ cloned into the GFP4 vector. Statistical significances were calculated using Mann-Whitney U-test.

5.6 MFI values and results from luciferase assay indicate that pk1 and LPP enhancers are weak transcription enhancers

Table 15 presents GFP loss values from an individual experiment, but these values differ from those in Figure 11 because, while GFP loss values performed with different flow cytometers are comparable, MFI values are not. This experiment was selected, since it represents the highest number of tested elements in the same experiment. It should be kept in mind that MFI is not equivalent to transcription but can provide some approximation of it, as explained in section 5.2. The correlation between GFP loss and MFI for *BCL6* enhancer elements was variable. E3 and E4, which exhibited the highest GFP loss, also had relatively high MFIs (Table 15). In contrast, E1, which showed very low GFP loss, had the highest MFI in this group, while 5'dist, which had some GFP loss, displayed a very low MFI (Table 15).

Table 15. GFP loss (median) of *BCL6* enhancers with corresponding MFI values measured in DT40 cell line (unpublished).

FRAGMENT	GFP LOSS	MFI
E1	0.22 %	149923
E2	2.06 %	81787
E3	4.65 %	102186
E4	6.43 %	138533
E5	1.64 %	91040
E7	2.82 %	81055
5'dist	2.45 %	59550
e2f	2.87 %	87163
pk1	1.93 %	75851
LPP	3.05 %	90161
hIgλ	51.8 %	102992
negative control	1.84 %	83026

The MFIs of hIgλ in both upstream and downstream positions were significantly higher than in the negative control ($p=0.0007$ and $p=0.0004$, respectively), supporting the view that hIgλ also functions as a strong transcription enhancer (Table 16). All pk1 and LPP MFI values were lower compared to negative control and the difference was statistically significant for all except for LPP in the downstream position (pk1 upstream, $p<0.0001$; pk1 downstream, $p=0.0026$; LPP upstream, $p=0.0002$ and LPP downstream, $p=0.43152$) (Table 16). MFI was significantly higher for elements in the downstream position compared to the upstream position for both pk1 and LPP (pk1 upstream vs downstream, $p=0.0177$; LPP upstream vs downstream, $p=0.0004$) (Table 16).

Table 16. GFP loss (median) of pk1, LPP and hIgλ enhancers with corresponding MFI values measured in DT40 cell line (unpublished).

FRAGMENT	GFP LOSS	MFI
pk1 upstream	3.21 %	76051
pk1 downstream	2.27 %	80713
LPP upstream	1.75 %	78092
LPP downstream	2.85 %	91860
hIgλ upstream	42.1 %	115604
hIgλ downstream	10.5 %	101449
negative control	2.13 %	92296

Luciferase assay provided further evidence of pk1 and LPP not being very strong transcription enhancers, as the fold change of relative luciferase activity compared to the negative control was 0.63 for pk1 and 0.34 for LPP (Figure 13). The fold change compared to empty vector was 4.23 for cIgλE which is an enhancer from the chicken *Ig* locus (Figure 13).

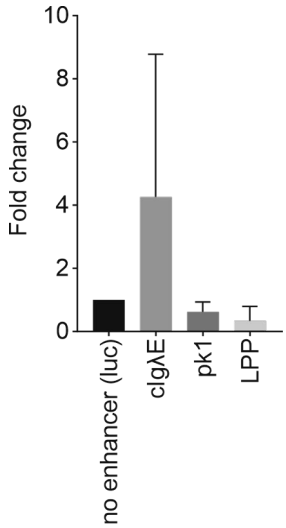

Figure 13. Luciferase activity of pk1 and LPP enhancers (unpublished). Luciferase activity fold change of pk1 and LPP over the empty vector. The mean of two measurements is shown with standard deviation. Chicken Igλ enhancer cloned into luciferase vector is a positive control and empty luciferase vector is a negative control.

5.7 Characterization of *BCL6* enhancer-modified Ramos AID-inducible cells

To investigate whether the SHM targeting elements found in the *BCL6* enhancer affect mutation accumulation in the *BCL6* 1st intron, pk1 and LPP elements were deleted individually from the Ramos cell line genome. Additionally, the previously described SHM targeting element 1928 was inserted into deletion sites (Blagodatski et al., 2009; Kohler et al., 2012). The cell lines studied were Ramos $pk1^{1928/-}$, $pk1^{-/-}$, $LPP^{1928/-}$, $LPP^{-/-}$ and $BCL6^{wt/wt}$.

The features potentially affecting mutation accumulation in the *BCL6* 1st intron that were not related to the deleted/inserted SHM targeting elements were determined. Therefore, BCL6 expression, AID expression and the rate of cellular growth in $BCL6e^{wt/wt}$, $pk1^{1928/-}$, $pk1^{-/-}$, $LPP^{1928/-}$ and $LPP^{-/-}$, both in doxycycline-treated (dox) and untreated Ramos cell lines, were measured. The rate of cell division was used as a measurement of cellular growth. The rate of cell division was similar across all cell lines (Figure 14A). However, the $pk1^{1928/-}$ cells, regardless of

doxycycline-treatment status, grew slightly slower than the other cell lines (Figure 14A). In addition, the growth of the LPP$^{1928/-}$ untreated cell line began to decline more rapidly after 115 hours of culture compared to the other cell lines (Figure 14A).

BCL6 mRNA expression levels in BCL6e$^{wt/wt}$, pk1$^{1928/-}$, pk1$^{-/-}$, LPP$^{1928/-}$ and LPP$^{-/-}$ cell lines were measured using quantitative PCR. Compared to the BCL6e$^{wt/wt}$ control cell line, the LPP$^{1928/-}$ dox cell line showed 2-fold *BCL6* mRNA expression, LPP$^{1928/-}$ control 1.5-fold expression, LPP$^{-/-}$ dox 1.4-fold expression, LPP $^{-/-}$ control 1.2-fold expression and BCL6e$^{wt/wt}$ dox showed 1.4-fold expression (Figure 14B). Thus, doxycycline-treated cell lines with a deleted LPP enhancer expressed higher levels of *BCL6* mRNA compared to their untreated controls. In contrast, pk1 deleted cell lines had lower *BCL6* mRNA expression levels compared to the BCL6e$^{wt/wt}$ control cell line, with only minor differences between doxycycline-treated and untreated cell lines. Nevertheless, all cell lines expressed *BCL6* mRNA, which is crucial, as transcription of the target gene is a prerequisite for AID-induced mutations.

AID expression during the 8-week culture period was observed in all modified cell lines. Figure 14C shows AID, GAPDH and BCL6 expression after four weeks of culturing. The cell lines exhibited some variation in AID expression levels. AID expression levels were normalized to GAPDH expression, and compared to Ramos wild-type cells pk1$^{1928/-}$ expressed AID 6.3-fold, pk1$^{-/-}$ 4.4-fold, BCL6e$^{wt/wt}$ 5.8-fold, LPP$^{1928/-}$ 28-fold and LPP$^{-/-}$ 3.4-fold. Compared to BCL6e$^{wt/wt}$ cell line, pk1$^{1928/-}$ expressed AID 1.1-fold, pk1$^{-/-}$ 0.8-fold, LPP$^{1928/-}$ 4.8-fold and LPP$^{-/-}$ 0.6-fold. BCL6 protein expression was detected in all cell lines using western blot (Figure 14C). These results indicate that two isoforms of BCL6 are expressed in Ramos cells, consistent with previous findings that at least two protein isoforms of BCL6 are known (Shen, Y. et al., 2008).

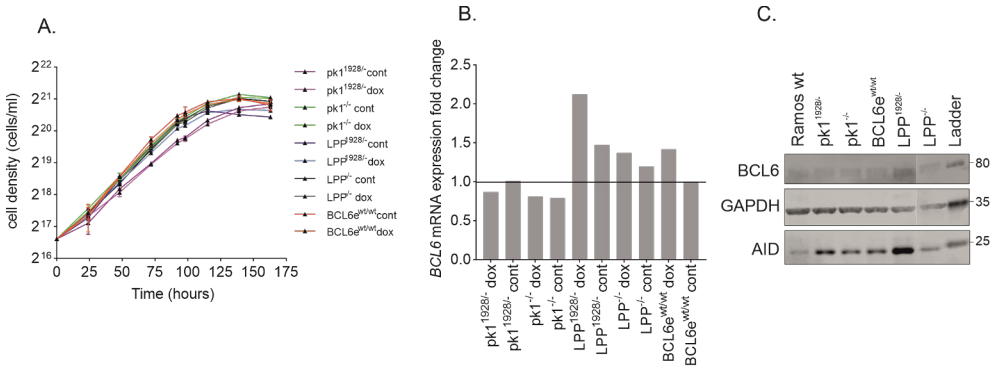

Figure 14. Cellular growth as well as BCL6 and AID expression of *BCL6*-modified Ramos cell lines (unpublished). A. The rate of cell division in *BCL6* modified cell lines. Each data point represents the mean cell density from three replicate cultures with standard deviation at time points: 0 h, 24 h, 48 h, 72 h, 92 h, 98 h, 115 h, 139 h, or 163 h. Cell density was measured by flow cytometry. B. Fold change in *BCL6* mRNA expression compared to the BCL6e$^{wt/wt}$ control cell line, measured by quantitative PCR. C. BCL6, GAPDH and AID protein expression in *BCL6*-modified cell lines and wild-type Ramos cell line after four weeks of culture, detected by western blot.

5.8 AID-dependent mutations accumulate to *BCL6* 1st intron in Ramos cells

To investigate how the deletion or insertion of enhancers in the *BCL6* enhancer region affects mutation accumulation in the *BCL6* gene, an amplicon from the *BCL6* 1st intron was sequenced. This amplicon was selected from an area frequently mutated in DLBCLs and normal GC B cells (Migliazza et al., 1995; Pasqualucci et al., 1998, 2003; Shen, J. C. et al., 2019). Although the deleted pk1 and LPP are located 160–340 kb away from the 1st intron (approximation from UCSC Genome Browser), SHM targeting can occur through interconnections between the enhancer and the 1st intron (Hübschmann et al., 2021; Qian et al., 2014; Senigl et al., 2019).

For pilot experiments, the number of cell lines was limited to assess sufficient sequencing depth for detecting mutations. Thus, the analysis was conducted using the LPP$^{1928/-}$, LPP$^{-/-}$, BCL6e$^{wt/wt}$, Carnaval and Mieu cell lines. The read count decreased after 460 bp, so mutations beyond this point were excluded from further analysis. Mutations were detected in all sequenced cell lines (Table 17). The overall mutation frequency was higher in all doxycycline-treated Ramos cell lines compared to their untreated controls and the DLBCL cell lines (Table 17). A similar trend was seen in mutation frequencies at AID hotspots (Table 17).

Table 17. Mutation frequencies in *BCL6* 1st intron in Ramos *BCL6*-modified cell lines and DLBCL cell lines (unpublished).

CELL LINE	OVERALL MUT.FREQ.	OVERALL MUT.FREQ. IN AID HOTSPOTS	MEDIAN MUT.FREQ.	MEDIAN MUT. FREQ. IN AID HOTSPOTS
Mieu	1.98×10^{-4}/bp	2.75×10^{-5} /bp	1.39×10^{-4}/bp	1.51×10^{-4}/bp
Carnaval	2.31×10^{-4}/bp	2.46×10^{-5}/bp	2.02×10^{-4}/bp	2.84×10^{-4}/bp
Ramos LPP$^{1928/-}$ dox	4.63×10^{-4}/bp	9.43×10^{-5}/bp	2.97×10^{-4}/bp	7.41×10^{-4}/bp
Ramos LPP$^{1928/-}$ cont	2.48×10^{-4}/bp	1.88×10^{-5}/bp	1.43×10^{-4}/bp	1.52×10^{-4}/bp
Ramos LPP$^{-/-}$ dox	3.13×10^{-4}/bp	5.47×10^{-5}/bp	2.46×10^{-4}/bp	4.88×10^{-4}/bp
Ramos LPP$^{-/-}$ cont	2.41×10^{-4}/bp	1.89×10^{-5}/bp	2.0×10^{-4}/bp	2.0×10^{-4}/bp
Ramos BCL6e$^{wt/wt}$ dox	3.13×10^{-4}/bp	4.37×10^{-5}/bp	2.8×10^{-4}/bp	3.28×10^{-4}/bp
Ramos BCL6e$^{wt/wt}$ cont	2.91×10^{-4}/bp	1.65×10^{-5}/bp	2.03×10^{-4}/bp	1.71×10^{-4}/bp

In doxycycline-treated cell lines, mutation frequencies per base pair were higher in the regions 1–170 bp and 320–460 bp when visually assessed compared to untreated and the DLBCL cell lines (Figure 15). Three SNPs were detected in one of the *BCL6* alleles in Ramos cells through Sanger sequencing, corresponding to positions with high mutation frequency peaks in this data set. The SNPs identified were T>A at 214 bp, T>C at 223 bp and C>G at 228 bp, based on base pair numbering in the examined amplicon (Figure 15).

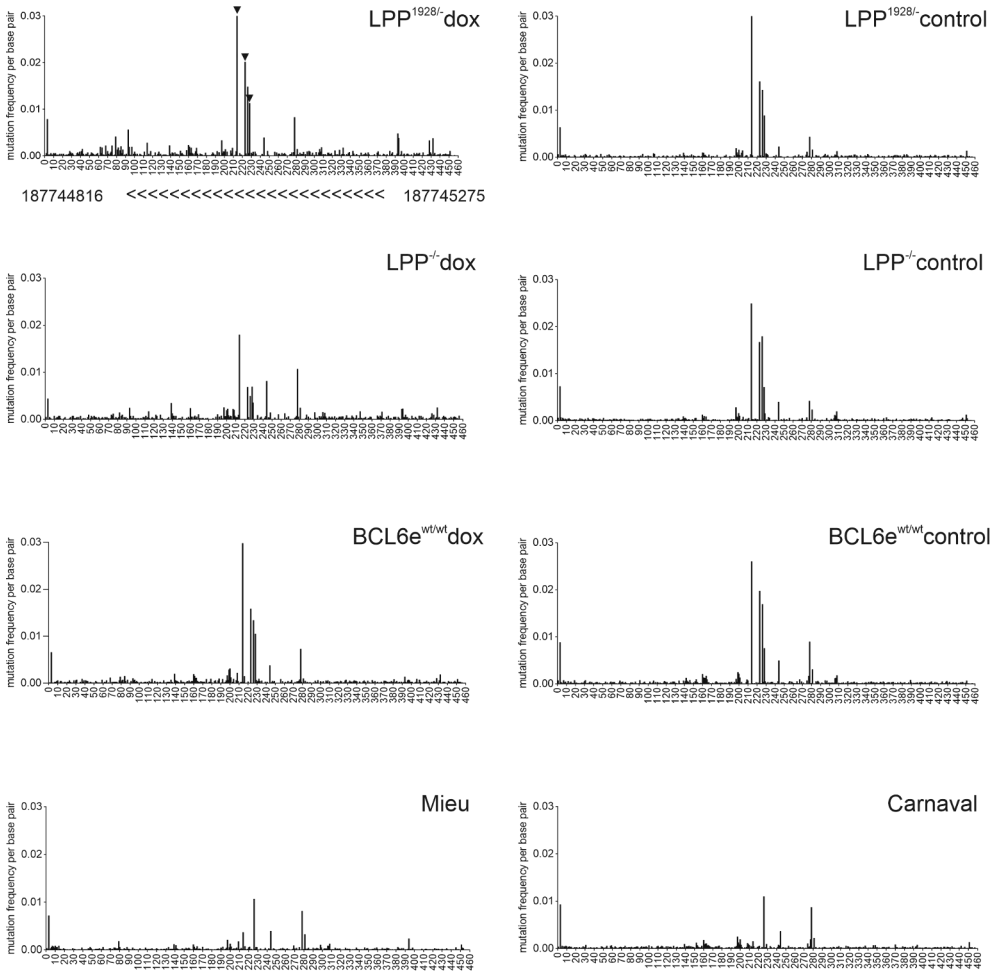

Figure 15. Mutation frequency of individual sequenced base pairs in the *BCL6* 1st intron in Ramos cells. *BCL6* enhancer-modified cell lines and DLBCL cell lines (unpublished). Genomic coordinates and locations of three SNPs (T>A at 214 bp, T>C at 223 bp and C>G at 228 bp) in Ramos cells as well as the direction of transcription are shown. SNPs are indicated with arrows.

The median mutation frequency calculated from mutation frequencies at individual base positions, was higher in all doxycycline-treated Ramos cell lines ($LPP^{1928/-}$ 2.97×10^{-4}/bp vs 1.43×10^{-4}/bp; $LPP^{-/-}$ 2.46×10^{-4}/bp vs 2.0×10^{-4}/bp and $BCL6e^{wt/wt}$ 2.8×10^{-4}/bp vs 2.03×10^{-4}/bp) compared to their untreated controls and the difference was statistically significant for $LPP^{1928/-}$ ($p<0.0001$) and $LPP^{-/-}$ ($p=0.0070$) (Figure 16 and Table 17). The median mutation frequencies in Mieu and Carnaval were 1.39×10^{-4}/bp and 2.02×10^{-4}/bp, respectively (Table 17). Together with the overall mutation frequencies, the rate of median mutation frequencies in

DLBCL cell lines corresponds to the frequencies in Ramos cells. Compared to the doxycycline-treated BCL6e$^{wt/wt}$ cell line both DLBCL lines exhibited slightly lower mutation frequencies, with the difference compared to Mieu being statistically significant ($p<0.0001$) (Figure 16). Additionally, the median mutation frequency in AID hotspots was significantly higher in all doxycycline-treated Ramos cell lines compared to their untreated controls (7.41×10^{-4}/bp vs 1.52×10^{-4}/bp, $p<0.0001$ in LPP$^{1928/-}$, 4.88×10^{-4}/bp vs 2.0×10^{-4}/bp, $p=0.0003$ in LPP$^{-/-}$ and 3.28×10^{-4}/bp vs 1.71×10^{-4}/bp, $p<0.0001$ in BCL6e$^{wt/wt}$) (Figure 16 and Table 17).

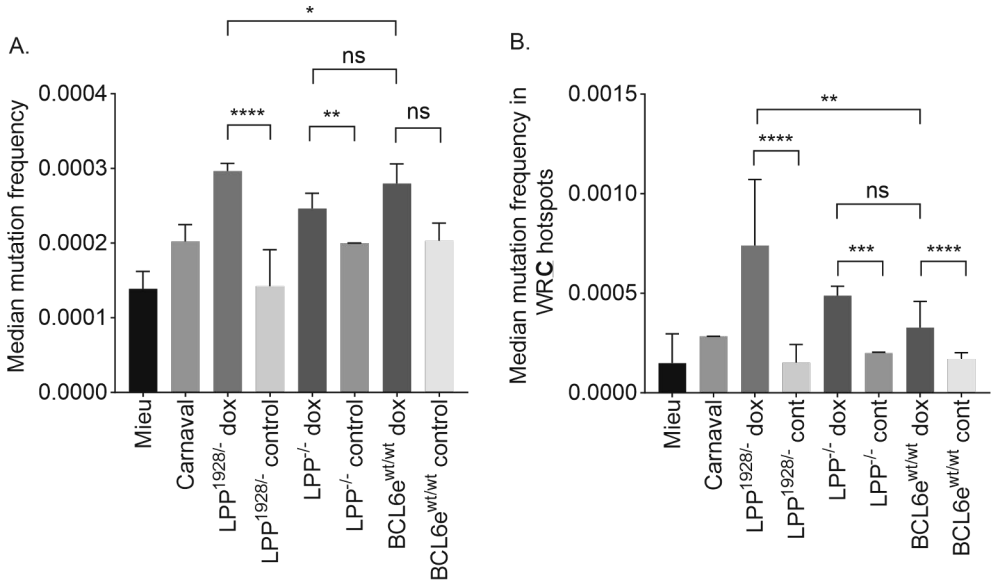

Figure 16. Median mutation frequency in the *BCL6* 1st intron in *BCL6* modified Ramos cell lines and DLBCL cell lines (unpublished). Median mutation frequency (overall and in AID hotspots), calculated from individual mutation frequencies per base pair, is plotted on the y-axis. Statistical significances for the most relevant comparisons between cell lines are indicated. 95 % confidence intervals are presented.

The proportion of C>T(G>A) mutations was higher in all doxycycline-treated cells (LPP$^{1928/-}$ 44.6 %, LPP$^{-/-}$ 39.5%, BCL6e$^{wt/wt}$ 30.2%) compared to their untreated control cell lines (LPP$^{1928/-}$ 15.6 %, LPP$^{-/-}$ 15.9 %, BCL6e$^{wt/wt}$ 12.9 %) and compared to Mieu (24.6 %) and Carnaval (20.6 %) (Figure 17). C>T(G>A) mutations (in doxycycline-treated Ramos cell lines) and T>C (A>G) mutations (in other cell lines) were the most prominent mutation types (Figure 17), which is consistent with the findings of Shen et al. on mutation type distribution in the *BCL6* 1st intron in normal human B cells (Shen, J. C. et al., 2019). These results indicate that AID-dependent mutations accumulate in the *BCL6* 1st intron in Ramos cells. Although the Ramos cell line used in this experiment is UNG-deficient, which may influence the mutation

pattern, other key BER and MMR repair pathway genes are expressed, at least at the RNA level (Figure 10A).

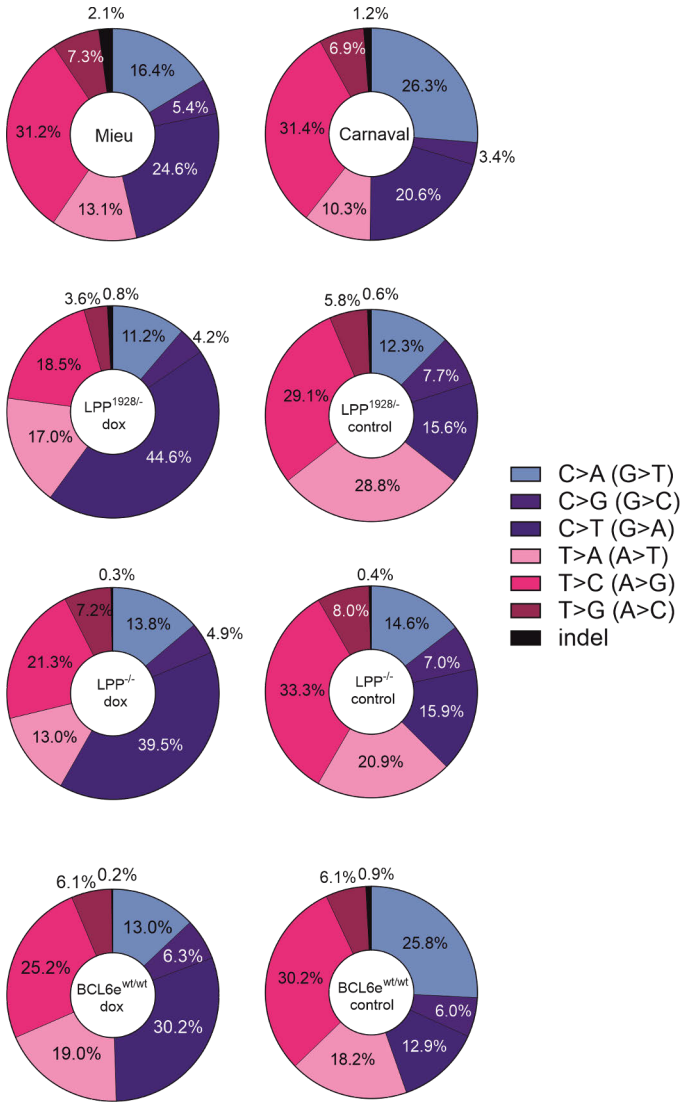

Figure 17. Proportions on mutation types detected in *BCL6* modified Ramos cell lines, Mieu and Carnaval (unpublished). Mutations are categorized into six substitution classes C>A (G>T), C>G (G>C), C>T (G>A), T>A (A>T), T>C (A>G), T>G (A>C) and a class containing insertions and deletions (indel).

In the LPP$^{1928/-}$ doxycycline-treated cell line, the overall mutation frequency was 1.5x higher than in the BCL6e$^{wt/wt}$ doxycycline-treated cell line. In addition, the LPP$^{1928/-}$ doxycycline-treated cell line exhibited a significantly higher median

mutation frequency compared to the BCL6e$^{wt/wt}$ doxycycline-treated cell line (2.97 × 10^{-4}/bp vs 2.8 × 10^{-4}/bp, p=0.0262) (Figure 16A). The median mutation frequency at AID hotspots was also significantly higher in the LPP$^{1928/-}$ doxycycline-treated cell line compared to the BCL6e$^{wt/wt}$ doxycycline-treated cell line (7.41 × 10^{-4}/bp vs 3.28 × 10^{-4}/bp, p=0.0061) (Figure 16B). These results suggest that the SHM targeting element 1928 increases mutation accumulation in the *BCL6* 1st intron. However, since the LPP$^{1928/-}$ doxycycline-treated cell line showed 4.8-fold AID expression compared to the BCL6e$^{wt/wt}$ doxycycline-treated cell line (Figure 14C), the observed differences could be due to higher AID expression.

The overall mutation frequency in LPP$^{-/-}$ doxycycline-treated cell lines was the same as in BCL6e$^{wt/wt}$ dox cells (3.13 × 10^{-4}/bp). The median mutation frequency in the LPP$^{-/-}$ doxycycline-treated cell line was lower compared to the BCL6e$^{wt/wt}$ doxycycline-treated cell line, but the difference was not statistically significant (2.46 × 10^{-4}/bp vs 2.8 × 10^{-4}/bp, p=0.7287) (Figure 16A). The LPP$^{-/-}$ doxycycline-treated cell line exhibited 0.6-fold AID expression compared to the BCL6e$^{wt/wt}$ doxycycline-treated cell line (Figure 14C). Mutation frequency in AID hotspots was higher in the LPP$^{-/-}$ dox cells compared to BCL6e$^{wt/wt}$ cells (5.47 × 10^{-5}/bp vs 4.37 × 10^{-5}/bp, respectively). Additionally, the median mutation frequency in AID hotspots was higher in the LPP$^{-/-}$ doxycycline-treated cell line compared to the BCL6e$^{wt/wt}$ cell line, although the difference was not statistically significant (4.88 × 10^{-4}/bp vs 3.28 × 10^{-4}/bp, p=0.5694) (Figure 16B). Thus, further sequencing and analysis are needed to confirm whether the deletion of the LPP enhancer affects mutation frequency in the *BCL6* 1st intron.

5.9 Mutation frequency between alleles in Ramos LPP$^{1928/-}$ cell line

To overcome the issue of varying AID levels between different cell lines, the differences in mutation frequencies between alleles were investigated. In LPP$^{1928/-}$ cell line, one allele had the LPP enhancer deleted, while the other one had the 1928 element inserted into the LPP deletion site. The method for distinguishing mutations originating from different alleles is described in detail in section 4.3.3, but it is based on SNPs that Ramos cells carry in one of the *BCL6* alleles. This approach revealed that in the non-SNP allele, the median mutation frequency and the median mutation frequency in AID hotspots were higher compared to the other allele (2.18 × 10^{-4}/bp vs 1.85 × 10^{-4}/bp, p=0.3074 and 4.44 × 10^{-4}/bp vs 1.50 × 10^{-4}/bp p=0.0003, respectively) (Table 18 and Figure 18). To confirm that the observed difference was not a cell line-specific feature, the median mutation frequencies in different alleles were analyzed in LPP$^{1928/-}$ control, BCL6e$^{wt/wt}$ dox and BCL6e$^{wt/wt}$ control cell lines. In the LPP$^{1928/-}$ control cell line, the median mutation frequency in the SNP allele

was significantly higher than in the non-SNP allele (1.4×10^{-4}/bp vs 1.2×10^{-4}/bp, p=0.0100), but the difference between median mutation frequencies in AID hotspots was not statistically significant (1.4×10^{-4}/bp vs 1.2×10^{-4}/bp, p=0.3333) (Table 18 and Figure 18). In the BCL6e$^{wt/wt}$ dox and BCL6e$^{wt/wt}$ control cell lines, the overall median mutation frequency and the AID hotspot median mutation frequency between alleles were very close to each other, and differences were not statistically significant (overall median frequency BCL6e$^{wt/wt}$ dox, p=0.6506; overall median frequency Ramos BCL6e$^{wt/wt}$ control, p=0.3083; median frequency in AID hotspots BCL6e$^{wt/wt}$ dox, p=0.5359; median frequency in AID hotspots BCL6e$^{wt/wt}$ control, p=0.5182) (Table 18 and Figure 18). In conclusion, the higher overall median mutation frequency and the higher median AID hotspot mutation frequency in the non-SNP allele compared to the SNP allele in the LPP$^{1928/-}$ doxycycline-treated cell line suggest that the SHM targeting element 1928 may increase mutation accumulation in this allele. However, based on this experiment, the location of the 1928 in the non-SNP allele could not be confirmed. Nevertheless, the marginal and mostly non-significant differences observed between alleles in the LPP$^{1928/-}$ control, BCL6e$^{wt/wt}$ dox and BCL6e$^{wt/wt}$ control cell lines support the hypothesis that the higher mutation frequency in the non-SNP allele could be influenced by 1928.

Table 18. Mutation frequencies in different alleles in the *BCL6* 1st intron in Ramos LPP$^{1928/-}$ and BCL6e$^{wt/wt}$ cell lines (unpublished).

CELL LINE	MEDIAN MUT.FREQ. SNP ALLELE	MEDIAN MUT. FREQ. NON-SNP ALLELE	MEDIAN MUT.FREQ. IN AID HOTSPOTS SNP ALLELE	MEDIAN MUT. FREQ. IN AID HOTSPOTS NON-SNP ALLELE
Ramos LPP$^{1928/-}$ dox	1.85×10^{-4}/bp	2.18×10^{-4}/bp	1.50×10^{-4}/bp	4.44×10^{-4}/bp
Ramos LPP$^{1928/-}$ cont	1.4×10^{-4}/bp	1.2×10^{-4}/bp	1.4×10^{-4}/bp	1.2×10^{-4}/bp
Ramos BCL6e$^{wt/wt}$ dox	1.67×10^{-4}/bp	1.66×10^{-4}/bp	1.54×10^{-4}/bp	1.64×10^{-4}/bp
Ramos BCL6e$^{wt/wt}$ cont	1.06×10^{-4}/bp	1.05×10^{-4}/bp	1.0×10^{-4}/bp	0.99×10^{-4}/bp

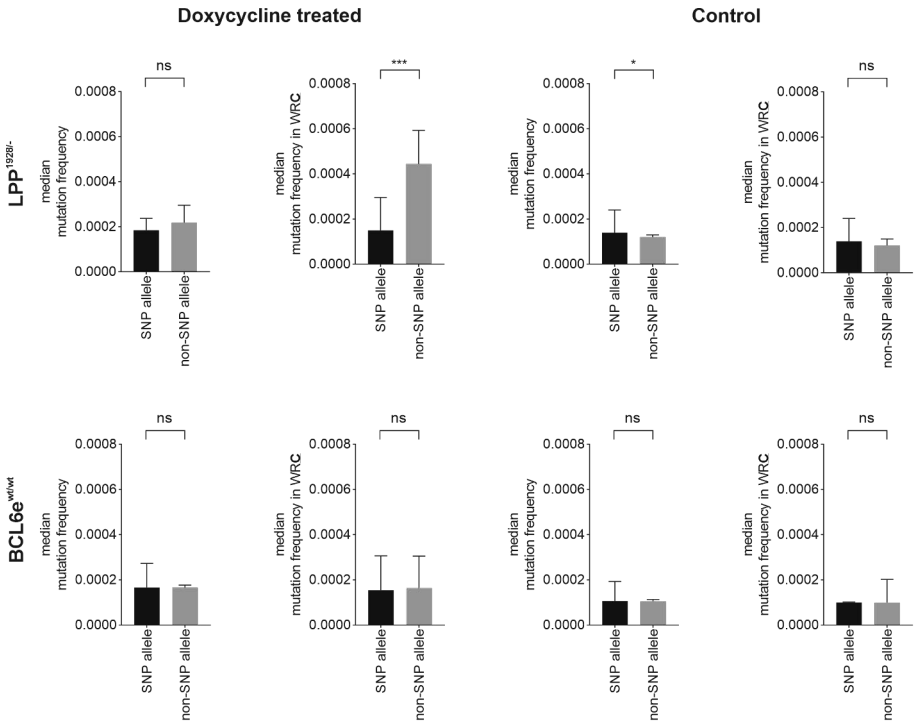

Figure 18. Median mutation frequency and mutation frequency in AID hotspots in two alleles in *BCL6* 1st intron in LPP$^{1928/-}$ and BCL6e$^{wt/wt}$ doxycycline-treated and non-treated control cell lines (unpublished). 95 % confidence intervals and statistical significance between differences in allele mutation frequencies are shown.

6 Discussion

6.1 Transcription and mutation targeting activity are difficult to fully distinguish from each other

Since AID requires ssDNA substrate for deamination, increased transcription of the target gene might also increase the mutation accumulation. To distinguish SHM targeting activity from transcription activity, the GFP loss assay was designed. In the GFP reporter systems, an RSV or CMV promoter drives *GFP* gene expression meaning that the insertion of a tested DNA fragment should not affect *GFP* transcription. However, if the tested fragment includes a transcription enhancer or promoter, it could influence the rate of transcription. Cloning the tested fragment downstream of the *GFP* gene should reduce this effect. The expected result would be that when the fragment is cloned upstream, the SHM targeting activity, as measured by GFP loss, would be higher compared to when the fragment is cloned downstream of the *GFP* gene.

The SV40 NCRR caused a 3.8-fold, and the hIgλ enhancer caused a 4-fold GFP loss when tested in the upstream vs downstream position. In addition, there was a 6.2-fold and a 5-fold GFP loss in the downstream position compared to the negative control, respectively. For the hIgλ enhancer, the MFI values were higher in the upstream compared to the downstream position, and higher in the downstream position compared to the negative control. The MFI of the SV40 NCRR cloned downstream was lower than the MFI of the negative control. The luciferase assay provided further evidence that SV40 NCRR enhanced transcription in B cells (II). In conclusion, these results implicate that the SV40 NCRR and hIgλ increase transcription of the GFP reporter in B cells. This is supported by various studies describing SV40 and Ig enhancers functioning as traditional enhancers (reviewed in Smith and Shilatifard 2014).

Conversely, in a previous study where the hIgλ enhancer was tested in the GFP4 reporter, the cloning site did not affect the GFP loss (Buerstedde et al., 2014). However, the measurements of this fragment in Buerstedde et al. study were based on a single experiment, and this discrepancy might be explained by experiment-to-experiment variation. Additionally, the GFP4 reporter is more sensitive than the GFP2 and GFP7 reporters (Buerstedde et al., 2014), and this increased sensitivity

may help explain the occasional variation observed in GFP loss measurements of the same fragment, particularly for strong SHM-targeting enhancers. In other words, the GFP4 reporter might sometimes be too sensitive for accurately measuring SHM targeting activity. Despite the challenges in measuring strong transcription enhancers with the sensitive GFP4 reporter, GFP loss remains a useful and effective method for assessing SHM targeting activity.

Other viral and genomic fragments tested in this work showed intermediate SHM targeting activity at best, and primarily lower MFI values than the negative control. Collectively, these results indicate that for an element to exhibit strong SHM targeting activity, it must also efficiently promote transcription.

Increasing the number of SV40 NCRRs in a reporter increases the MFI as well as GFP loss, which is in line with previous studies showing that the SV40 enhancer promotes transcription in B cell lines (Davidson et al., 1986; Schirm et al., 1987). Further, deletion of IPEN TFBSs from the SV40 NCRR abolishes both SHM targeting activity and its ability to drive gene expression in the luciferase assay, indicating that these TFBSs are essential for both SHM-targeting and transcription. Deletion of E-box1 and NF-kB decreased SHM targeting activity but did not affect MFI. Indeed, there is evidence suggesting that transcription and mutation enhancers may rely on different TFBSs for these operations (Dinesh et al., 2020). However, the issue is complicated by the RSV promoter driving transcription in the GFP reporter. Therefore, no definite conclusions about the role of NF-kB and E-box1 in the SV40 NCRRs ability to drive transcription can be drawn based on solely on MFI. Collectively, these results suggest that mutation and transcription enhancer activities cannot be entirely separated.

The explanation for why some mutation enhancers exhibit low SHM targeting in the downstream but not upstream position remains to be determined. As some downstream-cloned enhancers (hIgλ, BCL6 E3 and E4) exhibit significantly higher MFI than the negative control, the increase in transcription is a possible, yet unlikely, explanation. Indeed, the SV40 and certain Ig enhancers have been shown to increase transcription of their target genes even when positioned downstream (Banerji et al., 1983; Schirm et al., 1987). This is typically accomplished by chromatin looping over a longer distance (Pachano et al., 2022). In the GFP reporter, however, the distance between the tested fragment and the target gene is a few thousands base pairs at most. Therefore, it seems unlikely that these fragments promote transcription from the downstream position via chromatin looping in the GFP reporter. However, changing the position of the tested fragment may lead to undetermined alterations to Pol II machinery, which could affect SHM targeting activity and transcription, as both processes are understood to operate through this machinery (Kodgire et al. 2013; Nambu et al. 2003; Willmann et al. 2012). Since this phenomenon is observed in mutation enhancers from both viral and human genomes, it is likely not species-

specific. As discussed in Study I, altering the position of the tested fragment could also shift any potential insulator sequences that might block SHM targeting activity and cause differences in GFP loss when cloned upstream or downstream. While the possibility of these results being an artifact cannot be entirely excluded, this seems unlikely, as only certain mutation enhancers and their subfragments consistently exhibit this phenomenon in this work. Ultimately, the location of mutation enhancers, whether upstream or downstream of their target gene, might not be as a significant factor in SHM targeting in a physiological context, due to long-range genomic interactions, e.g., via chromatin looping.

6.2 SHM targeting activity of polyomavirus NCRRs might benefit viral life cycle

In this thesis work, the SHM targeting activity of polyomavirus NCRRs was established (I and II). This activity relied on TFBSs (II), as has also been shown for Ig enhancers (Buerstedde et al., 2014; Kohler et al., 2012; McDonald et al., 2013). In contrast to Ig enhancers, the SV40 NCRR exhibited SHM targeting activity beyond B cells (II). If SHM targeting activity were merely disadvantageous to the virus, it would likely have been eliminated through evolution. In some cases, viruses may exploit SHM targeting activity for their own evolution and preservation in the host cells, either as episomes or integrated to host genome. In addition, mutations can benefit viruses beyond preservation. For example, JCPyV acquires the ability to evade neutralizing antibodies through mutations in VP genes (Ray et al., 2015).

Because SHM targeting activity relies on TFBSs, it is plausible that certain enhancers incidentally contain the necessary TFBSs and other features favorable for SHM targeting. As demonstrated in Study II and by McDonald et al., 2013, even a single point mutation in a critical TFBS can significantly influence the mutation enhancer's ability to target SHM. Thus, viruses whose genome is mutated relatively rapidly can easily gain or lose SHM targeting activity. Indeed, mutations and rearrangements in polyomavirus regulatory regions are shown to alter viral tropism and increase their pathogenicity (Yang & You, 2020). However, only certain genes are SHM off-targets (Álvarez-Prado et al., 2018; Hübschmann et al., 2021; Khodabakhshi et al., 2012; Meng et al., 2014; Qian et al., 2014). Hence, it seems unlikely that a genomic or viral mutational enhancer would frequently arise through random mutations. In addition, viruses with double-stranded DNA genomes, including polyomaviruses, exhibit the lowest mutation rates among all viruses (Sanjuán & Domingo-Calap, 2016). Further, polyomaviruses often persist in host cells without causing any issues in immunocompetent individuals (Helanterä et al., 2016). Thus, rather than being advantageous to polyomaviruses, SHM targeting activity may simply be an incidental occurrence with no significant role in the viral

life cycle. As the concept of SHM modifying viral genomes is new, further investigation is needed.

6.3 Mutation signatures in polyomavirus *LT* reflect AID/APOBEC activity and functional DNA repair pathways

Based on mutation patterns and distribution as well as AID/APOBEC expression in MCC samples, the majority of mutations observed in *LT* in MCC are likely caused by APOBECs, rather than AID (I). In contrast, mutations detected in SV40 *LT* closely resembled AID signature (II). This is in line with results from the GFP loss experiments where the SV40 NCRR exhibited strong while the MCPyV NCRR showed weak SHM targeting activity. However, although the AID mutation signature was identified in MCPyV *LT*, these mutations were not enriched in MCPyV+ MCCs. Further, the APOBEC3 signature was detected in SV40 *LT* in Ramos cells but not in UO-31 cells (II).

APOBEC3 family enzymes play a crucial role in viral defense (Siriwardena et al., 2016). Thus, it is not surprising to find the APOBEC3 mutation signature in polyomavirus *LT*. APOBEC3s have been found to mutate for example HPV, poxviruses, HIV and HSV-1 (Herpes simplex virus 1) (Forni et al., 2023; Shapiro et al., 2021; Sheehy et al., 2002; Vartanian et al., 2008; Vieira et al., 2014; Wang, Z. et al., 2014). Additionally, APOBEC3 expression and its mutation signature have been directly linked to polyomavirus infection, which can activate APOBEC3 expression via interferon signaling (Que et al., 2021; Shapiro et al., 2021; Starrett et al., 2019; Verhalen et al., 2016). UV signature in *LT* from skin cancer samples is also an expected finding (I). However, AID mutation signature is perhaps less expected finding, although infections with EBV and HCV have been shown to induce AID expression (Machida et al., 2004). One possible reason for the infrequent detection of the AID mutational signature is that AID expression is largely restricted to GC B cells.

Although the mutational signatures of JCPyV and BKPyV were not examined in this thesis, both viruses seem to be targets of APOBEC3 enzymes. JCPyV has been shown to upregulate APOBEC3A and APOBEC3B expression, and an APOBEC3A footprint has been detected in the JCPyV genome (Poulain et al., 2020). Similarly, an APOBEC3 mutational signature, along with APOBEC3 expression, has been detected in BKPyV-infected kidneys (Peretti et al., 2018). While the JCPyV NCRR exhibited SHM targeting activity and the BKPyV NCRR did not, the presence of an AID mutational signature in the genomes of JCPyV and BKPyV remains to be determined.

In Studies I and II, mutation signatures were used to distinguish the actions of three C-targeting mutators: AID, APOBEC3 and UV radiation. However, these signatures can sometimes overlap, and multiple factors can affect the same cells. One key factor affecting the observed mutation profile is the DNA repair pathway utilized. There are distinct signatures for deficiencies of MMR, BER and HR-mediated pathways (Alexandrov et al., 2020). Regarding the mutation signatures detected in Study I, major deficiencies in expressed DNA repair genes on the RNA expression level in the MCC data were not detected. However, significant negative correlation between *REV1* and *LT* was established.

REV1 or another error-prone polymerase has been proposed to contribute to APOBEC-induced C>G transversions (Helleday et al., 2014), suggesting that the relatively low proportion of detected C>G mutations in Study I may be explained by the absence of REV1. However, since the APOBEC3 mutation signatures SBS2 (C>T) and SBS13 (C>G) typically appear together, with their relative proportions varying (Petljak & Maciejowski, 2020), the mutation type distribution observed in this study is unlikely to result solely from low REV1 expression.

The significant correlation between *LT* and *MSH2* expression aligns with previous findings showing MSH2 and MSH6 deficiencies particularly in MCPyV- MCCs (Gambichler et al., 2021). In general, the universal expression of MMR, BER, and HR -associated genes, along with the significant correlation between *LT* and *MSH2*/*BRCA2* expression, suggests that MCPyV+ MCC might have a highly functioning DNA repair machinery. This finding is consistent with the better prognosis and lower mutation burden observed in MCPyV+ compared to MCPyV- MCCs (Harms, K. L. et al., 2021; Harms, P. W. et al., 2015).

In Study II, the majority of mutations detected in the SV40 *LT* in Ramos were C>T and T>A/T>C mutations, which aligns with the widespread expression of *UNG, MSH2, MSH6, PRIMPOL* and other SHM-associated DNA repair genes. This mutation pattern is also consistent with previously observed AID-induced mutation distributions (Gu et al., 2019; Pilzecker et al., 2016; Pilzecker & Jacobs, 2019; Shen, H. M. et al., 2006; Siriwardena et al., 2016). Additionally, in Study II, the renal cell carcinoma cell line UO-31 exhibited a higher proportion of T>A and T>C mutations in SV40 *LT* compared to Ramos cells. This suggests that in UO-31 cells, AID-induced lesions may be more frequently repaired through the MMR pathway than in Ramos cells. Although DNA repair gene expression in UO-31 cells was not examined in Study II, the MMR pathway is usually active in renal cell carcinoma cells (Leach et al., 2002).

6.4 Lymphocytes and kidney cells are potential sites for AID/APOBEC-induced mutation accumulation to the polyomavirus genome

For the SHM targeting activity of polyomavirus NCRRs to have physiological significance, polyomaviruses must infect cells that express AID/APOBEC enzymes and support SHM. Ultimately, AID and APOBEC3 expression can be induced in any cell type by inflammatory signals and the viral infection (Endo et al., 2007, 2008; Que et al., 2021; Wang, Z. et al., 2014). However, B lymphocytes are most prone for AID expression and SHM. In addition, AID/APOBEC mutation accumulation to polyomavirus *LT* in B cells was shown in studies I and II. Indeed, SV40 and JCPyV are known to infect B cells (Alaribe et al., 2013; Dolcetti et al., 2003; Houff & Berger, 2008; Monaco et al., 1996). In addition, bone marrow and lymphoid cells are frequently proposed as reservoir cells for polyomaviruses (Martini et al., 1998; McNees et al., 2019; Tan et al., 2009).

Although MCPyV is linked to B cells through the proposed B-cell origin of MCC, current evidence largely contradicts this view (Kervarrec et al., 2019; Liu, W., Yang, et al., 2016; Sunshine et al., 2018). Consistent with this, the results of Study I do not support SHM-targeting mechanisms as a major factor in pathogenic mutation accumulation. However, there is evidence of MCPyV presence in leukocytes and monocytes in healthy individuals (Mazzoni, Rotondo, et al., 2017; Mertz et al., 2010). Given that immune cells frequently express AID/APOBEC enzymes (Salter et al., 2016), these cells may serve as potential reservoir for MCPyV. Both the AID and APOBEC3 signatures could arise from either the reservoir cells or the cell-of-origin of MCC. Future research is needed to determine where oncogenic APOBEC3-induced mutations are introduced into *LT* in MCPyV+ MCCs.

Kidney cells function as reservoirs for SV40, JCPyV and BKPyV (Li, R. M. et al., 2002; Peretti et al., 2018; Sunyaev et al., 2009). BKPyV is closely associated with the kidneys, as it resides in the kidney epithelium during latency and causes nephropathies in kidney transplant patients (Prado et al., 2018; Zhou et al., 2023). MCPyV has also been detected in a subset of renal cell carcinomas (Mobaraki et al., 2024). In Study II, it was demonstrated that UO-31 cells support SHM targeting activity of the SV40 NCRR, exhibit endogenous AID expression and accumulate mutations at AID hotspots in the SV40 *LT*. Thus, in addition to lymphocytes, kidneys represent another potential site for SHM-induced mutation accumulation in the polyomavirus *LT*s.

The SV40 NCRR exhibited SHM targeting activity in a variety of cells (II), which is likely linked to its TFBS composition. The SV40 NCRR can probably bind factors that are expressed widely across cell types. These factors may compensate for some B-cell-specific factors that support SHM targeting in B cells. Therefore,

mutations may accumulate in SV40 *LT* beyond lymphocytes and kidney cells given that AID-expression is induced.

6.5 Potential effects of SHM targeting activity of polyomavirus NCRRs beyond *LT* region

In Studies I and II, the focus was on mutation accumulation in the viral *LT* region. However, the SV40 regulatory region promotes bidirectional transcription, with early and late gene transcription being induced from different areas (Gidoni et al., 1985). This raises the question of whether the mutation enhancer properties are also bidirectional. Indeed, mutations in polyomavirus VP genes can alter viral tropism, enabling the virus to infect new cell types. This has been established for at least SV40 and JCPyV (Gorelik et al., 2011; Magaldi et al., 2012; Sunyaev et al., 2009). In the case of SV40, the accumulation of mutations in the viral genome, resulting in altered tropism and replication capacity, could explain why SV40 has been detected in multiple different malignancies. These features are particularly important for JCPyV, whose mutations in VP genes and NCRR contribute to the pathogenesis of PML. This occurs through mutated VP, which increases binding to the surface of oligodendrocytes in the brain, and mutated NCRR, which enhances the replication capacity of the viral genome in the cells (Gorelik et al., 2011; Sunyaev et al., 2009; Yang & You, 2020). It has been proposed that JCPyV uses B cells as a route to the brain, and reactivates under immunosuppression, eventually leading to PML (Yang & You, 2020). If JCPyV resides in B cells or in other SHM-prone cells, mutations in VP genes could be SHM-induced, since the JCPyV NCRR exhibits SHM targeting activity. Indeed, the current view is that mutations predisposing PML development takes place in the host (Iida et al., 1993; Seppälä et al., 2017). Therefore, investigating whether JCPyV harbors an AID mutation signature would be of great interest.

In general, the ability of mutation enhancers to target mutations bidirectionally has not been studied. However, the cloning of tested elements upstream and downstream of the expression cassette provided some indication of this potential. One explanation for why the MCPyV 440bp (I) and *BCL6* LPP (unpublished) enhancers exhibit SHM targeting activity only in the downstream position could be the unidirectional nature of SHM targeting. In contrast, the SV40 and hIgλ enhancers demonstrate SHM targeting also in the downstream position, suggesting they might be capable of bidirectional targeting. In addition to targeting SHM to viral genes, viral mutation enhancers might also direct mutations to host genes, potentially leading to various consequences. If viral enhancers are inserted into the genome, they could convert TADs into SHM-susceptible regions, as shown for Ig mutation enhancers (Senigl et al., 2019).

6.6 The role of polyomavirus SHM targeting activity in human diseases

In this work, it was revealed that SHM targeting activity appears to be a relatively common feature of polyomavirus NCRR regions, though the biological significance of this phenomenon requires further investigation. The SV40 NCRR exhibited strong SHM targeting activity in a variety of cells and truncated LT expression in DT40 cells was observed (II), making it the most compelling example of biologically significant mutation accumulation. However, the truncation events in protein level appear to be rare as they were not detected in Ramos or UO-31 cells regardless of the observed mutation accumulation to *LT* (II). In addition, the effects of truncated LT to cellular transformation were not studied in this thesis. Furthermore, conflicting evidence exists regarding the role of SV40 in human oncogenesis (Poulin & DeCaprio, 2006; Shah, 2007).

In contrast, MCPyV and JCPyV NCRRs showed weak and intermediate SHM targeting activity, respectively, but both have a well-established role in human pathogenicity (Decaprio & Garcea, 2013). In addition to its role as the causative agent of PML, there is some evidence suggesting JCPyV may also contribute to tumorigenesis (Ahye et al., 2020; Del Valle et al., 2004; Zheng, H. C. et al., 2022). Since the GFP4 reporter is highly sensitive in detecting SHM targeting activity, the observed activity of MCPyV and JCPyV might lead to an overestimation of potential mutation accumulation in vivo. Therefore, the biological significance of an enhancer's ability to target mutations should always be critically evaluated. To establish the relevance of SHM targeting activity in viral regulatory regions to human pathogenicity, it would be essential to identify a virus with both considerable SHM targeting activity and a clear role in human disease. Nevertheless, the SHM targeting activity of viral enhancers presents a surprising finding, introducing a novel potential mechanism for viral oncogenicity and pathogenicity.

6.7 Preliminary outcomes of the BCL6 study

In this study, it was demonstrated that the genomic *BCL6* enhancer region contains multiple SHM targeting elements. The mutation enhancers identified in this study exhibited 21.2 % to 60.6 % of GFP loss compared to human Igλ enhancer from *Ig* loci. This rate of SHM targeting activity aligns with the observed AID-induced mutation burden in off-target sites which is lower than in *Ig* loci (Bal et al., 2022; Liu, M. et al., 2008; Martin et al., 2018; Pasqualucci et al., 1998).

The SHM targeting activity of the pk1 enhancer was demonstrated in DT40 cells and previously shown in Ramos cells (Senigl et al., 2019). However, the SHM targeting activity of the pk1 enhancer was relatively lower in DT40 than in Ramos cells when compared to Ig enhancers (21.2 % and 63 %, respectively). This suggests

that there may be differences in ability of mutation enhancers to target SHM between cell lines, although the feature is believed to be evolutionarily conserved (Buerstedde et al., 2014). The transcription factor content and their binding to TFBSs might differ between cell types. Investigating the TFBS content and transcription factor binding of *BCL6* mutation enhancers could provide mechanistic insights into these differences. Differences in the expression levels or activity of AID between DT40 and Ramos cells could potentially affect the rate of SHM. In study II, chicken, human, and mouse AID induced slightly different rates of GFP loss when measured in the GFP7 reporter in DT40, suggesting some differences in AID activity between species. However, AID regulation is complex, and the mere binding of AID to DNA does not correlate with levels of SHM (Feng, Y. et al., 2020; Matthews et al., 2014). Thus, the expression levels of AID alone may not be sufficient to explain the differences in SHM targeting activity of same enhancer in different cells.

AID-dependent mutations accumulate in the *BCL6* 1st intron in Ramos cells. This aligns with studies showing mutations in the *BCL6* 1st intron in DLBCL cells, normal GC cells and Ramos cells (Migliazza et al., 1995; Pasqualucci et al., 1998, 2003; Qian et al., 2014; Shen, J. C. et al., 2019; Wang, X. et al., 2002). However, confirming the mutability of *BCL6* in Ramos cells was essential for future studies, as there is also a report stating that *BCL6* does not mutate in Ramos cells (Ronai et al., 2007). These conflicting findings are likely due to insufficient sequencing depth in Ronai et al., as mutations in the *BCL6* locus are rarer compared to the *Ig* locus (Martin et al., 2018). The mutation frequency observed in this thesis work were in the same range as previously reported mutation frequencies in DLBCLs and normal GC B cells (Migliazza et al., 1995; Pasqualucci et al., 1998; Shen, J. C. et al., 2019). Mutation frequencies in Mieu and Carnaval were lower than those in Ramos doxycycline-treated cells, which might be due to lower, endogenous AID levels, although this was not measured. In addition, the *UNG*-deficiency in the doxycycline-inducible Ramos cell line may increase the mutation accumulation due to the impaired BER pathway. The deletion of *UNG* is expected to facilitate the detection of AID-induced mutations, but it can simultaneously increase the occurrence of background mutations in control cells, as BER does not exclusively repair AID-induced mutations.

The *BCL6* expression in LPP$^{1928/-}$ doxycycline-treated cell line was 1.5x higher than in other cell lines. Furthermore, the *BCL6* mRNA expression was higher in all doxycycline-treated cell lines compared to their corresponding control cell lines. Thus, it cannot be entirely ruled out that the rate of transcription explains part of the observed differences in mutation frequencies between cell lines.

Preliminary evidence was found that the SHM targeting element 1928 might increase both overall and AID hotspot mutation frequencies, but definitive confirmation of this finding is still lacking. In the BCL6e$^{wt/wt}$ cell line, mutation

frequencies were equal between alleles, regardless of AID expression, which is consistent with the previous study reporting that mutations in *BCL6* in DLBCL are predominantly biallelic (Migliazza et al., 1995). Since mutation frequencies were also equal between alleles in the $LPP^{1928/-}$ control cell line, it is unlikely that the observed difference is a specific feature of the $LPP^{1928/-}$ cell line. Together, these observations suggest that mutations in Ramos cells tend to distribute relatively evenly between alleles. However, a more sophisticated method for distinguishing mutations from different alleles might be required.

It is possible that deleting the LPP or pk1 enhancer alone does not significantly affect mutation frequency in vivo, or that the effect may not be detectable. Indeed, differences caused by the deletion of mutation enhancers might be challenging to observe, as the mutation frequency in *BCL6* is inherently low (Liu, M. et al., 2008; Martin et al., 2018; Pasqualucci et al., 1998). Furthermore, it is possible that pk1 and LPP enhancers have biologically significant SHM targeting activity, but they may target mutations outside the *BCL6* 1st intron. Further studies that involve deleting *BCL6* enhancers with stronger SHM targeting activity (E3 and E4) will provide better conditions for detecting differences in mutation frequencies. In addition, deleting multiple elements or the entire *BCL6* enhancer could help achieve this goal. However, since GC B cells, including Ramos cells, are typically dependent on BCL6 expression (Hatzi et al., 2013), deletions that abolish BCL6 expression are unsuitable for this study unless BCL6 expression is simultaneously rescued in some manner.

The overall mutation frequency in the $LPP^{-/-}$ and $BCL6^{wt/wt}$ doxycycline-treated cell lines was the same, implying that deleting the LPP enhancer does not affect mutation accumulation in the *BCL6* 1st intron. However, the mutation frequency in AID hotspots was marginally higher in $LPP^{-/-}$ dox cells compared to $BCL6e^{wt/wt}$ dox cells, and mutation frequency per base pair was also slightly higher in certain regions of the amplicon. This observation is most likely due to normal variation. However, increased sequencing depth is necessary to determine whether LPP deletion plays a role in mutation accumulation in the *BCL6* 1st intron. Additionally, it would be interesting to investigate whether deleting the LPP enhancer affects mutation accumulation in the *LPP* gene, another proto-oncogene located adjacent to the *BCL6* locus and a potential SHM off-target site (Hübschmann et al., 2021; Leeman-Neill et al., 2023).

Collectively, these results suggest that SHM targeting in genomic off-targets could operate through mutation enhancers, similar to the mechanism observed in the *Ig* loci, although further confirmation is still required.

6.8 Prospects of understanding the role of SHM targeting in the *BCL6* gene

BCL6 mutations are most frequently detected in lymphomas, particularly in DLBCL and FL (Migliazza et al., 1995; Mlynarczyk et al., 2019; Pasqualucci et al., 2003), although they have also been observed in leukemias, and more rarely, in solid tumors (Cardenas et al., 2017; Hideshima et al., 2009; McLachlan et al., 2022). The primary targets of BCL6 are genes responsible for DNA damage detection, cell cycle regulation and cellular differentiation, and its dysregulation is tightly linked to oncogenesis (Phan et al., 2005, 2007; Phan & Dalla-Favera, 2004; Ranuncolo et al., 2008). Thus, understanding the mechanisms behind this dysregulation is crucial. While DLBCL generally responds well to current treatments, 30 % of patients either do not respond to treatment or experience relapses (Leppä et al., 2019; Li, S. et al., 2018). The current first-line treatment is R-CHOP, a combination of rituximab (which targets the CD20 receptor expressed by B cells), prednisone (an anti-inflammatory glucocorticoid), cyclophosphamide, doxorubicin and vincristine (cell cycle blockers). R-CHOP is used to treat all types of DLBCL, despite the variety of subtypes (Morin et al., 2022; Steen et al., 2021). Therefore, more accurate classification of DLBCLs based on underlying driver mutations could guide the selection and development of more targeted treatment options. Ultimately, confirming that mutations in the *BCL6* 1st intron are targeted by DNA elements located in the enhancer region would provide a mechanistic proof-of-concept for endogenic SHM off-targeting. This model would have wider implications for other SHM off-target genes with oncogenic properties. Furthermore, it would demonstrate that SHM off-targeting relies, at least in part, on the same mechanisms as SHM on-targeting. Such information would contribute to our understanding of SHM as a phenomenon and provide tools for further developing applications based on high-affinity antibodies, such as vaccines and biological drugs.

7 Conclusions

In this thesis work, SHM targeting activity from viral and endogenic genomic enhancers was established. Evidence supporting the relevance of SHM targeting activity in certain oncogenic processes was identified, though further studies are needed to confirm these findings. The results highlight the potential threat of SHM and AID/APOBEC enzymes to genomic integrity. However, uncovering the precise mechanisms of SHM off-targeting requires additional research. Based on this work, the following conclusions can be drawn:

- SHM targeting activity and transcriptional promotion by certain enhancers are not equal features, but these functions cannot be entirely separated. This applies to both viral and genomic mutation enhancers.
- Polyomavirus NCRRs from SV40, JCPyV and MCPyV exhibit SHM targeting activity with varying intensities, suggesting that SHM targeting is far more widespread phenomenon than previously expected.
- AID and APOBEC3 enzymes induce mutations in the polyomavirus *LT* region.
- B cells and kidney cells are potential sites for polyomavirus mutation accumulation, as these cells express AID, support SHM and serve as polyomavirus reservoirs.
- The MCPyV NCRR has weak SHM targeting activity, but APOBEC3 enzymes are more likely responsible for MCC-related mutations in the MCPyV *LT* region.
- The SV40 NCRR exhibits strong SHM targeting activity in B cells and other cell types.
- AID-induced mutations, including some that introduce STOP codons, accumulate in the SV40 *LT* region in Ramos and UO-31 cells. This can result in the expression of a truncated LT, as demonstrated in DT40 cells. The broader relevance of this finding in tumorigenesis and human oncogenesis remains to be determined.

- The JCPyV NCRR has intermediate SHM targeting activity. Whether this activity contributes to mutation accumulation in neighboring genes remains to be investigated.
- The *BCL6* enhancer exhibits SHM targeting activity throughout the enhancer region.
- In Ramos cells, the mutation load in the *BCL6* 1st intron increases in the presence of AID.
- Preliminary data suggest that inserting a strong SHM targeting element into the LPP deletion site in the *BCL6* enhancer increases mutation accumulation in the *BCL6* 1st intron. However, further investigation is needed to confirm this finding.

Acknowledgements

This doctoral thesis was carried out at the Institute of Biomedicine, University of Turku in the Turku Doctoral Programme of Molecular Medicine (TuDMM).

First and foremost, I would like to express my sincere gratitude to my supervisors, Docent Jukka Alinikula and Professor Olli Lassila, for providing me with the opportunity to carry out this thesis work. Jukka, your genuine enthusiasm and insights into science are truly admirable. Thank you for keeping your office door open for (lengthy) discussions and always having time to help. I'm grateful for the relaxed atmosphere where we can agree to disagree. Olli, thank you for warmly welcoming me into your research team and guiding me on the path of doctoral thesis work in Jukka's group. Your deep expertise in immunology and can-do attitude have been inspirational.

I wish to thank the reviewers, Docent Auli Karhu and Docent Maria Perdomo, for providing constructive and valuable comments, resulting in a greatly improved thesis. I would like to express my gratitude to Docent Eeva Auvinen for accepting the invitation to be the opponent for this work. My warm thank you also to the follow-up committee members Professor Marko Salmi and Docent Pieta Mattila for helpful advice and support during the process.

I want to acknowledge all my co-authors and collaborators for their contributions. I wish to gratefully thank Professor David Schatz and his research group for sharing their expertise and research methods about SHM targeting. Special thanks to Dr Anurupa Yadavalli for providing bioinformatic expertise. I wish to thank Dr Filip Šenigl and his team for their insights and experimental input regarding the SV40 project. A warm thank you to Adjunct professor Harri Sihto and the Rare Cancers Research Group for providing the extensive MCC data set and knowledge of this subject to be used in this thesis work.

I want to thank all the past and present members of the SHM group. I wish to express particular gratitude to Dr Minna Kyläniemi for collaborating and coauthoring as well as for providing the help and support especially in the early stage of this process. My warmest thank you to Alina Tarsalainen for the invaluable peer support, sharing projects, refreshing coffee breaks, and beyond. Special thanks to Alina for the grammar check of this thesis. My deepest gratitude goes to Ann Sofie

Wierda not only for professional support but, importantly, the emotional one. You and our venting sessions have kept me sane through these years.

My colleagues during the years at Medisiina D and Mikro, thank you for creating a warm and relaxed environment to work in. Especially I want to thank the ladies in the Autoimmune lab at Mikro. Your impressive expertise combined with how you instantly took me as part of the Autoimmune family were key factors in my decision to stay in the field of immunology.

My dear friends Emma, Maisa, Maiju and Kreetta, thank you for always being there for me and for your unwavering support, which extends to every aspect of life. Kreetta, particular thanks to you for cosuffering this journey and for the countless lunch dates throughout the years. Special thanks to our extended family Aikku, Topias and the kids for providing counterbalance to work. Whether it is midsummer in Kustavi, a holiday sailing trip, or a dinner on a regular Saturday, time spent with you is always a source of joy and relaxation.

I'm extremely grateful to my parents Tuula and Risto and to my brother Simo for believing in me, supporting me and allowing me to go my own way. Thank you for always welcoming me back home to recharge and for keeping the sauna ready to be heated. Finally, Henry, thank you for all the encouragement and patience during this process. Thank you for giving me a decent match in our debates (both scientific and non-scientific ones), making me laugh every day, and for being my safe haven.

This study would not have been possible without financial support from the Turku University Foundation, the Maud Kuistila Memorial Foundation, the Finnish Cultural Foundation and the Kymenlaakso Regional Fund, the Alfred Kordelin Foundation, the Orion Research Foundation, the Cancer Foundation Finland sr, the Cancer Society of Southwest Finland, the Sigrid Juselius Foundation, and TuDMM. All the funders are gratefully acknowledged.

March 2025
Anni Soikkeli

References

Abbas, A. K., Lichtman, A. H., & Pillai, S. (2022). *Cellular and Molecular Immunology* (Tenth edit). Elsevier.

Ahye, N., Bellizzi, A., May, D., & Wollebo, H. S. (2020). The role of the jc virus in central nervous system tumorigenesis. *International Journal of Molecular Sciences*, *21*(17), 1–23. https://doi.org/10.3390/ijms21176236

Aida, M., Hamad, N., Stanlie, A., Begum, N. A., & Honjo, T. (2013). Accumulation of the FACT complex, as well as histone H3.3, serves as a target marker for somatic hypermutation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(19), 7784–7789. https://doi.org/10.1073/pnas.1305859110

Alaggio, R., Amador, C., Anagnostopoulos, I., Attygalle, A. D., Araujo, I. B. de O., Berti, E., Bhagat, G., Borges, A. M., Boyer, D., Calaminici, M., Chadburn, A., Chan, J. K. C., Cheuk, W., Chng, W. J., Choi, J. K., Chuang, S. S., Coupland, S. E., Czader, M., Dave, S. S., … Xiao, W. (2022). The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia*, *36*(7), 1720–1748. https://doi.org/10.1038/s41375-022-01620-2

Alaribe, F. N., Mazzoni, E., Rigolin, G. M., Rizzotto, L., Maniero, S., Pancaldi, C., Manfrini, M., Martini, F., & Tognon, M. G. (2013). Extended lifespan of normal human B lymphocytes experimentally infected by SV40 or transfected by SV40 large T antigen expression vector. *Leukemia Research*, *37*(6), 681–689. https://doi.org/10.1016/j.leukres.2013.02.003

Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., Wu, Y., Boot, A., Covington, K. R., Gordenin, D. A., Bergstrom, E. N., Islam, S. M. A., Lopez-Bigas, N., Klimczak, L. J., McPherson, J. R., Morganella, S., Sabarinathan, R., Wheeler, D. A., Mustonen, V., Boutros, P., … Yu, W. (2020). The repertoire of mutational signatures in human cancer. *Nature*, *578*(7793), 94–101. https://doi.org/10.1038/s41586-020-1943-3

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A. J. R., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Børresen-Dale, A. L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjörd, J. E., Foekens, J. A., … Stratton, M. R. (2013). Signatures of mutational processes in human cancer. *Nature*, *500*(7463), 415–421. https://doi.org/10.1038/nature12477

Álvarez-Prado, Á. F., Pérez-Durán, P., Pérez-García, A., Benguria, A., Torroja, C., de Yébenes, V. G., & Ramiro, A. R. (2018). A broad atlas of somatic hypermutation allows prediction of activation-induced deaminase targets. *Journal of Experimental Medicine*, *215*(3), 761–771. https://doi.org/10.1084/jem.20171738

Baker, S. C., Mason, A. S., Slip, R. G., Skinner, K. T., Macdonald, A., Masood, O., Harris, R. S., Fenton, T. R., Periyasamy, M., Ali, S., & Southgate, J. (2022). Induction of APOBEC3-mediated genomic damage in urothelium implicates BK polyomavirus (BKPyV) as a hit-and-run driver for bladder cancer. *Oncogene*, *41*(15), 2139–2151. https://doi.org/10.1038/s41388-022-02235-8

Bal, E., Kumar, R., Hadigol, M., Holmes, A. B., Hilton, L. K., Loh, J. W., Dreval, K., Wong, J. C. H., Vlasevska, S., Corinaldesi, C., Soni, R. K., Basso, K., Morin, R. D., Khiabanian, H., Pasqualucci,

L., & Dalla-Favera, R. (2022). Super-enhancer hypermutation alters oncogene expression in B cell lymphoma. *Nature*, *607*(7920), 808–815. https://doi.org/10.1038/s41586-022-04906-8

Banerji, J., Olson, L., & Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell*, *33*(3), 729–740.

Bannard, O., Horton, R. M., Allen, C. D. C., An, J., Nagasawa, T., & Cyster, J. G. (2013). Germinal center centroblasts transition to a centrocyte phenotype according to a timed program and depend on the dark zone for effective selection. *Immunity*, *39*(5), 912–924. https://doi.org/10.1016/j.immuni.2013.08.038

Basso, K., & Dalla-Favera, R. (2012). Roles of BCL6 in normal and transformed germinal center B cells. *Immunological Reviews*, *247*(1), 172–183. https://doi.org/10.1111/j.1600-065X.2012.01112.x

Basso, K., Schneider, C., Shen, Q., Holmes, A. B., Setty, M., Leslie, C., & Riccardo, D. F. (2012). BCL6 positively regulates AID and germinal center gene expression via repression of miR-155. *Journal of Experimental Medicine*, *209*(13), 2455–2465. https://doi.org/10.1084/jem.20121387

Basu, U., Chaudhuri, J., Alpert, C., Dutt, S., Ranganath, S., Li, G., Schrum, J. P., Manis, J. P., & Alt, F. W. (2005). The AID antibody diversification enzyme is regulated by protein kinase a phosphorylation. *Nature*, *438*(7067), 508–511. https://doi.org/10.1038/nature04255

Basu, U., Meng, F. L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wasserman, C. R., Wesemann, D. R., Januszyk, K., Gregory, R. I., Deng, H., Lima, C. D., & Alt, F. W. (2011). The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell*, *144*(3), 353–363. https://doi.org/10.1016/j.cell.2011.01.001

Begum, N. A., Stanlie, A., Nakata, M., Akiyama, H., & Honjo, T. (2012). The histone chaperone Spt6 is required for activation-induced cytidine deaminase target determination through H3K4me3 regulation. *Journal of Biological Chemistry*, *287*(39), 32415–32429. https://doi.org/10.1074/jbc.M112.351569

Blagodatski, A., Batrak, V., Schmidl, S., Schoetz, U., Caldwell, R. B., Arakawa, H., & Buerstedde, J. M. (2009). A cis-acting diversification activator both necessary and sufficient for AID-mediated hypermutation. *PLoS Genetics*, *5*(1), 1–11. https://doi.org/10.1371/journal.pgen.1000332

Bocchetta, M., Di Resta, I., Powers, A., Fresco, R., Tosolini, A., Testa, J. R., Pass, H. I., Rizzo, P., & Carbone, M. (2000). Human mesothelial cells are unusually susceptible to simian virus 40-mediated transformation and asbestos cocarcinogenicity. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(18), 10214–10219. https://doi.org/10.1073/pnas.170207097

Bonnefont, J., Tiberi, L., van den Ameele, J., Potier, D., Gaber, Z. B., Lin, X., Bilheu, A., Herpoel, A., Velez Bravo, F. D., Guillemot, F., Aerts, S., & Vanderhaeghen, P. (2019). Cortical Neurogenesis Requires Bcl6-Mediated Transcriptional Repression of Multiple Self-Renewal-Promoting Extrinsic Pathways. *Neuron*, *103*(6), 1096-1108.e4. https://doi.org/10.1016/j.neuron.2019.06.027

Borchert, G. M., Holton, N. W., Edwards, K. A., Vogel, L. A., & Larson, E. D. (2010). Histone H2A and H2B are monoubiquitinated at AID-targeted loci. *PLoS ONE*, *5*(7). https://doi.org/10.1371/journal.pone.0011641

Brady, J., Bolen, J. B., Radonovich, M., Salzman, N., & Khoury, G. (1984). Stimulation of simian virus 40 late gene expression by simian virus 40 tumor antigen. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(7 I), 2040–2044. https://doi.org/10.1073/pnas.81.7.2040

Bransteitter, R., Pham, P., Scharfft, M. D., & Goodman, M. F. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(7), 4102–4107. https://doi.org/10.1073/pnas.0730835100

Buerstedde, J. M., Alinikula, J., Arakawa, H., McDonald, J. J., & Schatz, D. G. (2014). Targeting Of Somatic Hypermutation By immunoglobulin Enhancer And Enhancer-Like Sequences. *PLoS Biology*, *12*(4). https://doi.org/10.1371/journal.pbio.1001831

Burns, M. B., Lackey, L., Carpenter, M. A., Rathore, A., Land, A. M., Leonard, B., Refsland, E. W., Kotandeniya, D., Tretyakova, N., Nikas, J. B., Yee, D., Temiz, N. A., Donohue, D. E., Mcdougle, R. M., Brown, W. L., Law, E. K., & Harris, R. S. (2013). APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*, *494*(7437), 366–370. https://doi.org/10.1038/nature11881

Burns, M. B., Temiz, N. A., & Harris, R. S. (2013). Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nature Genetics*, *45*(9), 977–983. https://doi.org/10.1038/ng.2701

Butel, J., & Lednicky, J. (2000). Response to more about: cell and molecular biology of simian virus 40: implications for human infections and disease. *Journal of the National Cancer Institute*, *92*(6), 496–497. https://doi.org/10.1093/jnci/92.6.496

Butel, J., Vilchez, R., Jorgensen, J., & Kozinetz, C. (2003). Association between SV40 and non-Hodgkin's lymphoma. *Leukemia and Lymphoma*, *44*(SUPPL. 3). https://doi.org/10.1080/10428190310001623784

Canver, M. C., Bauer, D. E., Dass, A., Yien, Y. Y., Chung, J., Masuda, T., Maeda, T., Paw, B. H., & Orkin, S. H. (2014). Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/cas9 nuclease system in mammalian cells. *Journal of Biological Chemistry*, *289*(31), 21312–21324. https://doi.org/10.1074/jbc.M114.564625

Carbone, M., Gazdar, A., & Butel, J. (2020). SV40 and human mesothelioma. *Translational Lung Cancer Research*, *9*(Suppl 1), S47–S59. https://doi.org/10.21037/tlcr.2020.02.03

Cardenas, M. G., Oswald, E., Yu, W., Xue, F., Jr, A. D. M., & Melnick, A. M. (2017). The Expanding Role of the BCL6 Oncoprotein as a Cancer Therapeutic Target. *Clinical Cancer Research*, *23*(4), 885–893. https://doi.org/10.1158/1078-0432.CCR-16-2071

Casellas, R., Basu, U., Yewdell, W. T., Chaudhuri, J., Robbiani, D. F., & Di Noia, J. M. (2016). Mutations, kataegis and translocations in B cells: understanding AID promiscuous activity. *Nature Reviews. Immunology*, *16*(3), 164–176. https://doi.org/10.1038/nri.2016.2

Cerchietti, L., & Melnick, A. (2014). Targeting Bcl-6 in Diffuse Large B-Cell Lymphoma : *Expert Rev Hematol*, *6*(4), 343–345. https://doi.org/10.1586/17474086.2013.826928.TARGETING

Chang, Y., & Moore, P. S. (2012). Merkel Cell Carcinoma: A Virus-Induced Human Cancer. *Annual Review of Pathology: Mechanisms of Disease*, *7*(1), 123–144. https://doi.org/10.1146/annurev-pathol-011110-130227

Ci, W., Polo, J. M., Cerchietti, L., Shaknovich, R., Wang, L., Shao, N. Y., Ye, K., Farinha, P., Horsman, D. E., Gascoyne, R. D., Elemento, O., & Melnick, A. (2009). The BCL6 transcriptional program features repression of multiple oncogenes in primary B cells and is deregulated in DLBCL. *Blood*, *113*(22), 5536–5548. https://doi.org/10.1182/blood-2008-12-193037

Cooper, M., Peterson, R., & Good, R. (1965). Delineation of the Thymic and Bursal Lymphoid Systems in the Chicken. *Nature*, *205*(4967), 143–146.

Coussens, L. M., & Werb, Z. (2002). Inflammation and Cancer. *Nature*, *420*(December), pages 860–867. https://doi.org/10.1016/B978-0-12-374279-7.17002-X

Davidson, I., Fromental, C., Augereau, P., Wildeman, A., Zenke, M., & Chambon, P. (1986). Cell-type specific protein binding to the enhancer of simian virus 40 in nuclear extracts. *Nature*, *323*(6088), 544–548. https://doi.org/10.1038/323544a0

De Magis, A., Manzo, S. G., Russo, M., Marinello, J., Morigi, R., Sordet, O., & Capranico, G. (2019). DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(3), 816–825. https://doi.org/10.1073/pnas.1810409116

De Rienzo, A., Tor, M., Sterman, D. H., Aksoy, F., Albelda, S. M., & Testa, J. R. (2002). Detection of SV40 DNA sequences in malignant mesothelioma specimens from the United States, but not from Turkey. *Journal of Cellular Biochemistry*, *84*(3), 455–459. https://doi.org/10.1002/jcb.10058

DeCaprio, J. A. (2017). Merkel cell polyomavirus and Merkel cell carcinoma. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 372, Issue 1732). Royal Society Publishing. https://doi.org/10.1098/rstb.2016.0276

Decaprio, J. A., & Garcea, R. L. (2013). A cornucopia of human polyomaviruses. *Nature Reviews Microbiology*, *11*(4), 264–276. https://doi.org/10.1038/nrmicro2992

Del Valle, L., Enam, S., Lara, C., Miklossy, J., Khalili, K., & Gordon, J. (2004). Primary Central Nervous System Lymphoma Expressing the Human Neurotropic Polyomavirus, JC Virus, Genome. *Journal of Virology*, *78*(7), 3462–3469. https://doi.org/10.1128/jvi.78.7.3462-3469.2004

Desmots, F., Roussel, M., Pangault, C., Llamas-Gutierrez, F., Pastoret, C., Guiheneuf, E., Le Priol, J. ôme, Camara-Clayette, V., Caron, G., Henry, C., Belaud-Rotureau, M. A., Godmer, P., Lamy, T., Jardin, F., Tarte, K., Ribrag, V., & Fest, T. (2019). Pan-HDAC inhibitors restore PRDM1 response to IL21 in CREBBP-mutated follicular lymphoma. *Clinical Cancer Research*, *25*(2), 735–746. https://doi.org/10.1158/1078-0432.CCR-18-1153

Dézé, O., Ordanoska, D., Rossille, D., Miglierina, E., Laffleur, B., & Cogné, M. (2023). Unique repetitive nucleic acid structures mirror switch regions in the human IgH locus. *Biochimie*, *214*, 167–175. https://doi.org/10.1016/j.biochi.2023.08.017

Dinesh, R. K., Barnhill, B., Ilanges, A., Wu, L., Michelson, D. A., Senigl, F., Alinikula, J., Shabanowitz, J., Hunt, D. F., & Schatz, D. G. (2020). Transcription factor binding at Ig enhancers is linked to somatic hypermutation targeting. *European Journal of Immunology*, *50*(3), 380–395. https://doi.org/10.1002/eji.201948357

Dolcetti, R., Martini, F., Quaia, M., Gloghini, A., Vignocchi, B., Cariati, R., Martinelli, M., Carbone, A., Boiocchi, M., & Tognon, M. (2003). Simian virus 40 sequences in human lymphoblastoid B-cell lines. *J Virol*, *77*(2), 1595–1597.

Duan, S., Cermak, L., Pagan, J. K., Rossi, M., Martinengo, C., Di Celle, P. F., Chapuy, B., Shipp, M., Chiarle, R., & Pagano, M. (2012). FBXO11 targets BCL6 for degradation and is inactivated in diffuse large B-cell lymphomas. *Nature*, *481*(7379), 90–94. https://doi.org/10.1038/nature10688

Dunnick, W., Hertz, G. Z., Scappino, L., & Gritzmacher, C. (1993). Dna sequences at immunoglobulin switch region recombination sites. *Nucleic Acids Research*, *21*(9), 2285. https://doi.org/10.1093/nar/21.9.2285

Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F., & Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes and Development*, *18*(13), 1618–1629. https://doi.org/10.1101/gad.1200804

Emge, D. A., & Cardones, A. R. (2019). Updates on Merkel Cell Carcinoma. *Dermatologic Clinics*, *37*(4), 489–503. https://doi.org/10.1016/j.det.2019.06.002

Endo, Y., Marusawa, H., Kinoshita, K., Morisawa, T., Sakurai, T., Okazaki, I. M., Watashi, K., Shimotohno, K., Honjo, T., & Chiba, T. (2007). Expression of activation-induced cytidine deaminase in human hepatocytes via NF-κB signaling. *Oncogene*, *26*(38), 5587–5595. https://doi.org/10.1038/sj.onc.1210344

Endo, Y., Marusawa, H., Kou, T., Nakase, H., Fujii, S., Fujimori, T., Kinoshita, K., Honjo, T., & Chiba, T. (2008). Activation-Induced Cytidine Deaminase Links Between Inflammation and the Development of Colitis-Associated Colorectal Cancers. *Gastroenterology*, *135*(3), 889–898. https://doi.org/10.1053/j.gastro.2008.06.091

Feigenbaum, L., Hinrichs, S. H., & Jay, G. (1992). JC virus and simian virus 40 enhancers and transforming proteins: role in determining tissue specificity and pathogenicity in transgenic mice. *Journal of Virology*, *66*(2), 1176–1182. https://doi.org/10.1128/jvi.66.2.1176-1182.1992

Feng, H., Shuda, M., Chang, Y., & Moore, P. S. (2008). Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science*, *319*(5866), 1096–1100. https://doi.org/10.1126/science.1152586

Feng, Y., Li, C., Stewart, J. A., Barbulescu, P., Desivo, N. S., Álvarez-Quilón, A., Pezo, R. C., Perera, M. L. W., Chan, K., Tong, A. H. Y., Mohamad-Ramshan, R., Berru, M., Nakib, D., Li, G., Kardar, G. A., Carlyle, J. R., Moffat, J., Durocher, D., Di Noia, J. M., … Martin, A. (2021). FAM72A antagonizes UNG2 to promote mutagenic repair during antibody maturation. *Nature*, *600*(7888), 324–328. https://doi.org/10.1038/s41586-021-04144-4

Feng, Y., Seija, N., Di Noia, J. M., & Martin, A. (2020). AID in Antibody Diversification: There and Back Again. *Trends in Immunology*, *41*(7), 586–600. https://doi.org/10.1016/j.it.2020.04.009

Forni, D., Cagliani, R., Pozzoli, U., & Sironi, M. (2023). An APOBEC3 Mutational Signature in the Genomes of Human-Infecting Orthopoxviruses. *MSphere*, *8*(2). https://doi.org/10.1128/msphere.00062-23

Gambichler, T., Rached, N. A., Tannapfel, A., Becker, J. C., Vogt, M., Skrygan, M., Wieland, U., Silling, S., Susok, L., Stücker, M., Meyer, T., Stockfleth, E., Junker, K., Käfferlein, H. U., Brüning,

T., & Lang, K. (2021). Expression of mismatch repair proteins in merkel cell carcinoma. *Cancers*, *13*(11), 1–10. https://doi.org/10.3390/cancers13112524

Giacinti, C., & Giordano, A. (2006). RB and cell cycle progression. *Oncogene*, *25*(38), 5220–5227. https://doi.org/10.1038/sj.onc.1209615

Gibcus, J. H., & Dekker, J. (2013). The Hierarchy of the 3D Genome. *Molecular Cell*, *49*(5), 773–782. https://doi.org/10.1016/j.molcel.2013.02.011

Gidoni, D., Kadonaga, J. T., Barrera-Saldana, H., Takahashi, K., Chambon, P., & Tjian, R. (1985). Bidirectional SV40 transcription mediated by tandem Sp1 binding interactions. *Science*, *230*(4725), 511–517. https://doi.org/10.1126/science.2996137

Gish, W. R., & Botchan, M. R. (1987). Simian virus 40-transformed human cells that express large T antigens defective for viral DNA replication. *Journal of Virology*, *61*(9), 2864–2876. https://doi.org/10.1128/jvi.61.9.2864-2876.1987

Goh, G., Walradt, T., Markarov, V., Blom, A., Riaz, N., Doumani, R., Stafstrom, K., Moshiri, A., Yelistratova, L., Levinsohn, J., Chan, T. A., Nghiem, P., Lifton, R. P., & Choi, J. (2016). Mutational landscape of MCPyV-positive and MCPyV-negative merkel cell carcinomas with implications for immunotherapy. *Oncotarget*, *7*(3), 3403–3415. https://doi.org/10.18632/oncotarget.6494

Gordon, M. S., Kanegai, C. M., Doerr, J. R., & Wall, R. (2003). Somatic hypermutation of the B cell receptor genes B29 (Igβ, CD79b) and mb1 (Igα, CD79a). *Proceedings of the National Academy of Sciences of the United States of America*, *100*(7), 4126–4131. https://doi.org/10.1073/pnas.0735266100

Gorelik, L., Reid, C., Testa, M., Brickelmaier, M., Bossolasco, S., Pazzi, A., Bestetti, A., Carmillo, P., Wilson, E., McAuliffe, M., Tonkin, C., Carulli, J. P., Lugovskoy, A., Lazzarin, A., Sunyaev, S., Simon, K., & Cinque, P. (2011). Progressive multifocal leukoencephalopathy (PML) development is associated with mutations in JC virus capsid protein VP1 that change its receptor specificity. *Journal of Infectious Diseases*, *204*(1), 103–114. https://doi.org/10.1093/infdis/jir198

Gu, X., Booth, C. J., Liu, Z., & Strout, M. P. (2019). *AID-associated DNA repair pathways regulate malignant transformation in a murine model of BCL6-driven diffuse large B-cell lymphoma*. *127*(1), 102–113. https://doi.org/10.1182/blood-2015-02-628164.The

Harms, K. L., Zhao, L., Johnson, B., Wang, X., Carskadon, S., Palanisamy, N., Rhodes, D. R., Mannan, R., Vo, J. N., Choi, J. E., Chan, M. P., Fullen, D. R., Patel, R. M., Siddiqui, J., Ma, V. T., Hrycaj, S., McLean, S. A., Hughes, T. M., Bichakjian, C. K., … Harms, P. W. (2021). Virus-positive merkel cell carcinoma is an independent prognostic group with distinct predictive biomarkers. *Clinical Cancer Research*, *27*(9), 2494–2504. https://doi.org/10.1158/1078-0432.CCR-20-0864

Harms, P. W., Vats, P., Verhaegen, M. E., Robinson, D. R., Wu, Y. M., Dhanasekaran, S. M., Palanisamy, N., Siddiqui, J., Cao, X., Su, F., Wang, R., Xiao, H., Kunju, L. P., Mehra, R., Tomlins, S. A., Fullen, D. R., Bichakjian, C. K., Johnson, T. M., Dlugosz, A. A., & Chinnaiyan, A. M. (2015). The distinctive mutational spectra of polyomavirus-negative merkel cell carcinoma. *Cancer Research*, *75*(18), 3720–3727. https://doi.org/10.1158/0008-5472.CAN-15-0702

Hashida, Y., Higuchi, T., Matsui, K., Shibata, Y., Nakajima, K., Sano, S., & Daibata, M. (2018). Genetic Variability of the Noncoding Control Region of Cutaneous Merkel Cell Polyomavirus: Identification of Geographically Related Genotypes. *Journal of Infectious Diseases*, *217*(10), 1601–1611. https://doi.org/10.1093/infdis/jiy070

Hatzi, K., Jiang, Y., Huang, C., Garrett-Bakelman, F., Gearhart, M. D., Giannopoulou, E. G., Zumbo, P., Kirouac, K., Bhaskara, S., Polo, J. M., Kormaksson, M., MacKerell, A. D., Xue, F., Mason, C. E., Hiebert, S. W., Prive, G. G., Cerchietti, L., Bardwell, V. J., Elemento, O., & Melnick, A. (2013). A hybrid mechanism of action for BCL6 in B cells defined by formation of functionally distinct complexes at enhancers and promoters. *Cell Reports*, *4*(3), 578–588. https://doi.org/10.1016/j.celrep.2013.06.016

Heinsohn, S., Scholz, R., & Kabisch, H. (2011). SV40 and p53 as team players in childhood lymphoproliferative disorders. *Int J Oncol*, *38*(5), 1307–1317. https://doi.org/10.3892/ijo.2011.967

Helanterä, I., Sadeghi, M., Lautenschlager, I., Hedman, K., & Auvinen, E. (2016). Polyoomavirukset sairauksien aiheuttajina. *Duodecim; Lääketieteellinen Aikakauskirja*, *132*(5), 439–445.

Helleday, T., Eshtad, S., & Nik-Zainal, S. (2014). Mechanisms underlying mutational signatures in human cancers. *Nature Reviews Genetics*, *15*(9), 585–598. https://doi.org/10.1038/nrg3729

Hideshima, T., Ph, D., Richardson, P., & Ph, D. (2009). Bcl6 as a Novel Therapeutic Target in Multiple Myeloma (MM). *BLOOD*, *114*(22), 295. https://doi.org/10.1182/blood.V114.22.295.295

Hirvonen, A., Mattson, K., Karjalainen, A., Ollikainen, T., Tammilehto, L., Hovi, T., Vainio, H., Pass, H. I., Resta, I. Di, Carbone, M., & Linnainmaa, K. (1999). Simian virus 40 (SV40)-like DNA sequences not detectable in Finnish mesothelioma patients not exposed to SV40-contaminated polio vaccines. *Molecular Carcinogenesis*, *26*(2), 93–99. https://doi.org/10.1002/(SICI)1098-2744(199910)26:2<93::AID-MC4>3.0.CO;2-Z

Ho, J., Jedrych, J. J., Feng, H., Natalie, A. A., Grandinetti, L., Mirvish, E., Crespo, M. M., Yadav, D., Fasanella, K. E., Proksell, S., Kuan, S. F., Pastrana, D. V., Buck, C. B., Shuda, Y., Moore, P. S., & Chang, Y. (2015). Human polyomavirus 7-associated pruritic rash and viremia in transplant recipients. *Journal of Infectious Diseases*, *211*(10), 1560–1565. https://doi.org/10.1093/infdis/jiu524

Hobson, D. J., Wei, W., Steinmetz, L. M., & Svejstrup, J. Q. (2012). RNA Polymerase II Collision Interrupts Convergent Transcription. *Molecular Cell*, *48*(3), 365–374. https://doi.org/10.1016/j.molcel.2012.08.027

Houben, R., Adam, C., Baeurle, A., Hesbacher, S., Grimm, J., Angermeyer, S., Henzel, K., Hauser, S., Elling, R., Bröcker, E., Gaubatz, S., Becker, J., & Schrama, D. (2012). An intact retinoblastoma protein-binding site in Merkel cell polyomavirus large T antigen is required for promoting growth of Merkel cell carcinoma cells. *International Journal of Cancer*, *130*(4), 847–856. https://doi.org/10.1002/ijc.26076

Houben, R., Dreher, C., Angermeyer, S., Borst, A., Utikal, J., Haferkamp, S., Peitsch, W. K., Schrama, D., & Hesbacher, S. (2013). Mechanisms of p53 restriction in merkel cell carcinoma cells are independent of the merkel cell polyoma virus T antigens. *Journal of Investigative Dermatology*, *133*(10), 2453–2460. https://doi.org/10.1038/jid.2013.169

Houben, R., Shuda, M., Weinkam, R., Schrama, D., Feng, H., Chang, Y., Moore, P. S., & Becker, J. C. (2010). Merkel Cell Polyomavirus-Infected Merkel Cell Carcinoma Cells Require Expression of Viral T Antigens. *Journal of Virology*, *84*(14), 7064–7072. https://doi.org/10.1128/jvi.02400-09

Houff, S. A., & Berger, J. R. (2008). The bone marrow, B cells, and JC virus. *Journal of NeuroVirology*, *14*(5), 341–343. https://doi.org/10.1080/13550280802348222

Huang, C., Gonzalez, D. G., Cote, C. M., Jiang, Y., Hatzi, K., Teater, M., Dai, K., Hla, T., Haberman, A. M., & Melnick, A. (2014). The BCL6 RD2 Domain Governs Commitment of Activated B Cells to Form Germinal Centers. *Cell Reports*, *8*(5), 1497–1508. https://doi.org/10.1016/j.celrep.2014.07.059

Huang, C., Hatzi, K., & Melnick, A. (2013). Lineage-specific functions of Bcl-6 in immunity and inflammation are mediated by distinct biochemical mechanisms. *Nature Immunology*, *14*(4), 380–388. https://doi.org/10.1038/ni.2543

Huang, H., Reis, R., Yonekawa, Y., Lopes, J. M., Kleihues, P., & Ohgaki, H. (1999). Identification in human brain tumors of DNA sequences specific for SV40 large T antigen. *Brain Pathology*, *9*(1), 33–42. https://doi.org/10.1111/j.1750-3639.1999.tb00207.x

Hübschmann, D., Kleinheinz, K., Wagener, R., Bernhart, S. H., López, C., Toprak, U. H., Sungalee, S., Ishaque, N., Kretzmer, H., Kreuz, M., Waszak, S. M., Paramasivam, N., Ammerpohl, O., Aukema, S. M., Beekman, R., Bergmann, A. K., Bieg, M., Binder, H., Borkhardt, A., … Siebert, R. (2021). Mutational mechanisms shaping the coding and noncoding genome of germinal center derived B-cell lymphomas. *Leukemia*, *35*(7), 2002–2016. https://doi.org/10.1038/s41375-021-01251-z

Iida, T., Kitamura, T., Guo, J., Taguchi, F., Aso, Y., Nagashima, K., & Yogo, Y. (1993). Origin of JC polyomavirus variants associated with progressive multifocal leukoencephalopathy. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(11), 5062–5065. https://doi.org/10.1073/pnas.90.11.5062

Iqbal, J., Greiner, T. C., Patel, K., Dave, B. J., Smith, L., Ji, J., Wright, G., Sanger, W. G., Pickering, D. L., Jain, S., Horsman, D. E., Shen, Y., Fu, K., Weisenburger, D. D., Hans, C. P., Campo, E., Gascoyne, R. D., Rosenwald, A., Jaffe, E. S., … Chan, W. C. (2007). Distinctive patterns of BCL6

molecular alterations and their functional consequences in different subgroups of diffuse large B-cell lymphoma. *Leukemia*, *21*(11), 2332–2343. https://doi.org/10.1038/sj.leu.2404856

Ito, S., Nagaoka, H., Shinkura, R., Begum, N., Muramatsu, M., Nakata, M., & Honjo, T. (2004). Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(7), 1975–1980. https://doi.org/10.1073/pnas.0307335101

Jia, Y., Chng, W. J., & Zhou, J. (2019). Super-enhancers: Critical roles and therapeutic targets in hematologic malignancies. *Journal of Hematology and Oncology*, *12*(1), 1–17. https://doi.org/10.1186/s13045-019-0757-y

Jiao, J., Lv, Z., Zhang, P., Wang, Y., Yuan, M., Yu, X., Otieno Odhiambo, W., Zheng, M., Zhang, H., Ma, Y., & Ji, Y. (2020). AID assists DNMT1 to attenuate BCL6 expression through DNA methylation in diffuse large B-cell lymphoma cell lines. *Neoplasia (United States)*, *22*(3), 142–153. https://doi.org/10.1016/j.neo.2020.01.002

Jinglan, L., Kaur, G., Zhawar, V. K., Zimonjic, D. B., Popescu, N. C., Kandpal, R. P., & Athwal, R. S. (2009). Role of SV40 integration site at chromosomal interval 1q21.1 in immortalized CRL2504 cells. *Cancer Research*, *69*(19), 7819–7825. https://doi.org/10.1158/0008-5472.CAN-09-1003

Johansson, B., Sahi, H., Koljonen, V., & Böhling, T. (2019). 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. *Journal of Cutaneous Pathology*, *46*(1), 26–32. https://doi.org/10.1111/cup.13372

Kao, C., Hauser, P., Reznikoff, W. S., & Reznikoff, C. A. (1993). Simian virus 40 (SV40) T-antigen mutations in tumorigenic transformation of SV40-immortalized human uroepithelial cells. *Journal of Virology*, *67*(4), 1987–1995. https://doi.org/10.1128/jvi.67.4.1987-1995.1993

Kassem, A., Schöpflin, A., Diaz, C., Weyers, W., Stickeler, E., Werner, M., & Zur Hausen, A. (2008). Frequent detection of merkel cell polyomavirus in human merkel cell carcinomas and identification of a unique deletion in the VP1 gene. *Cancer Research*, *68*(13), 5009–5013. https://doi.org/10.1158/0008-5472.CAN-08-0949

Keim, C., Kazadi, D., Rothschild, G., & Basu, U. (2013). Regulation of AID, the B-cell genome mutator. *Genes and Development*, *27*(1), 1–17. https://doi.org/10.1101/gad.200014.112

Kervarrec, T., Samimi, M., Guyétant, S., Sarma, B., Chéret, J., Blanchard, E., Berthon, P., Schrama, D., Houben, R., & Touzé, A. (2019). Histogenesis of Merkel cell carcinoma: A comprehensive review. *Frontiers in Oncology*, *9*(JUN), 1–13. https://doi.org/10.3389/fonc.2019.00451

Khodabakhshi, A. H., Morin, R. D., Fejes, A. P., Mungall, A. J., Mungall, K. L., Bolger-Munro, M., Johnson, N. A., Connors, J. M., Gascoyne, R. D., Marra, M. A., Birol, I., & Jones, S. J. M. (2012). Recurrent targets of aberrant somatic hypermutation in lymphoma. *Oncotarget*, *3*(11), 1308–1319. https://doi.org/10.18632/oncotarget.653

Klein, U., & Dalla-Favera, R. (2008). Germinal centres: Role in B-cell physiology and malignancy. *Nature Reviews Immunology*, *8*(1), 22–33. https://doi.org/10.1038/nri2217

Klufah, F., Mobaraki, G., Liu, D., Alharbi, R. A., Kurz, A. K., Speel, E. J. M., Winnepenninckx, V., & zur Hausen, A. (2021). Emerging role of human polyomaviruses 6 and 7 in human cancers. *Infectious Agents and Cancer*, *16*(1), 1–12. https://doi.org/10.1186/s13027-021-00374-3

Köck, J., & Blum, H. E. (2008). Hypermutation of hepatitis B virus genomes by APOBEC3G, APOBEC3C and APOBEC3H. *Journal of General Virology*, *89*(5), 1184–1191. https://doi.org/10.1099/vir.0.83507-0

Kodgire, P., Mukkawar, P., Ratnam, S., Martin, T. E., & Storb, U. (2013). Changes in RNA polymerase II progression influence somatic hypermutation of Ig-related genes by AID. *Journal of Experimental Medicine*, *210*(7), 1481–1492. https://doi.org/10.1084/jem.20121523

Kohler, K. M., McDonald, J. J., Duke, J. L., Arakawa, H., Tan, S., Kleinstein, S. H., Buerstedde, J.-M., & Schatz, D. G. (2012). Identification of Core DNA Elements That Target Somatic Hypermutation. *The Journal of Immunology*, *189*(11), 5314–5326. https://doi.org/10.4049/jimmunol.1202082

Koljonen, V., Kukko, H., Pukkala, E., Sankila, R., Böhling, T., Tukiainen, E., Sihto, H., & Joensuu, H. (2009). Chronic lymphocytic leukaemia patients have a high risk of Merkel-cell polyomavirus DNA-positive Merkel-cell carcinoma. *British Journal of Cancer*, *101*(8), 1444–1447. https://doi.org/10.1038/sj.bjc.6605306

Koljonen, V., Kukko, H., Tukiainen, E., Böhling, T., Sankila, R., Joensuu, H., & Pukkala, E. (2010). Second cancers following the diagnosis of Merkel cell carcinoma: A nationwide cohort study. *Cancer Epidemiology*, *34*(1), 62–65. https://doi.org/10.1016/j.canep.2009.12.007

Kothapalli, N., Norton, D. D., & Fugmann, S. D. (2008). Cutting Edge: A cis -Acting DNA Element Targets AID-Mediated Sequence Diversification to the Chicken Ig Light Chain Gene Locus . *The Journal of Immunology*, *180*(4), 2019–2023. https://doi.org/10.4049/jimmunol.180.4.2019

Krump, N. A., & You, J. (2018). Molecular mechanisms of viral oncogenesis in humans. In *Nature Reviews Microbiology* (Vol. 16, Issue 11). https://doi.org/10.1038/s41579-018-0064-6

Kruse, J. P., & Gu, W. (2009). Modes of p53 Regulation. *Cell*, *137*(4), 609–622. https://doi.org/10.1016/j.cell.2009.04.050

Küppers, R. (2005). MECHANISMS OF B-CELL LYMPHOMA PATHOGENESIS. *NATURE REVIEWS | CANCER*, *5*(April). https://doi.org/10.1038/nrc1589

Lackey, L., Demorest, Z. L., Land, A. M., Hultquist, J. F., Brown, W. L., & Harris, R. S. (2012). APOBEC3B and AID have similar nuclear import mechanisms. *Journal of Molecular Biology*, *419*(5), 301–314. https://doi.org/10.1016/j.jmb.2012.03.011

Lackraj, T., Goswami, R., & Kridel, R. (2018). Pathogenesis of follicular lymphoma. *Best Practice and Research: Clinical Haematology*, *31*(1), 2–14. https://doi.org/10.1016/j.beha.2017.10.006

Lauring, M. C., & Basu, U. (2024). Somatic hypermutation mechanisms during lymphomagenesis and transformation. *Current Opinion in Genetics and Development*, *85*(Figure 1), 102165. https://doi.org/10.1016/j.gde.2024.102165

Leach, F. S., Koh, M., Sharma, K., McWilliams, G., Talifero-Smith, L. T., Codd, A., Olea, R., & Elbahloul, O. (2002). Mismatch repair gene mutations in renal cell carcinoma. *Cancer Biology and Therapy*, *1*(5), 530–536. https://doi.org/10.4161/cbt.1.5.171

Lednicky, J., & Butel, J. (1997). Tissue culture adaptation of natural isolates of simian virus 40: Changes occur in viral regulatory region but not in carboxy-terminal domain of large T-antigen. *Journal of General Virology*, *78*(7), 1697–1705. https://doi.org/10.1099/0022-1317-78-7-1697

Lednicky, J., Garcea, R., Bergsagel, D., & Butel, J. (1995). Natural Simian virus 40 strains are present in human choroid plexus and ependymoma tumors. In *Virology* (Vol. 212, Issue 2, pp. 710–717). https://doi.org/10.1006/viro.1995.1529

Leeman-Neill, R. J., Song, D., Bizarro, J., Wacheul, L., Rothschild, G., Singh, S., Yang, Y., Sarode, A. Y., Gollapalli, K., Wu, L., Zhang, W., Chen, Y., Lauring, M. C., Whisenant, D. E., Bhavsar, S., Lim, J., Swerdlow, S. H., Bhagat, G., Zhao, Q., … Basu, U. (2023). Noncoding mutations cause super-enhancer retargeting resulting in protein synthesis dysregulation during B cell lymphoma progression. *Nature Genetics*, *55*(12), 2160–2174. https://doi.org/10.1038/s41588-023-01561-1

Leppä, S., Meriranta, L., Pasanen, A., & Jyrkkiö, S. (2019). *Diffuusin suurisoluisen B-solulymfooman nykyhoito*. 1185–1192.

Li, J., Diaz, J., Wang, X., Tsang, S. H., & You, J. (2015). Phosphorylation of Merkel Cell Polyomavirus Large Tumor Antigen at Serine 816 by ATM Kinase Induces Apoptosis in Host Cells. *Journal of Biological Chemistry*, *290*(3), 1874–1884. https://doi.org/10.1074/jbc.M114.594895

Li, J., Wang, X., Diaz, J., Tsang, S. H., Buck, C. B., & You, J. (2013). Merkel Cell Polyomavirus Large T Antigen Disrupts Host Genomic Integrity and Inhibits Cellular Proliferation. *Journal of Virology*, *87*(16), 9173–9188. https://doi.org/10.1128/jvi.01216-13

Li, R. M., Branton, M. H., Tanawattanacharoen, S., Falk, R. A., Jennette, J. C., & Kopp, J. B. (2002). Molecular identification of SV40 infection in human subjects and possible association with kidney disease. *Journal of the American Society of Nephrology*, *13*(9), 2320–2330. https://doi.org/10.1097/01.ASN.0000028249.06596.CF

Li, S., Young, K. H., & Medeiros, L. J. (2018). Diffuse large B-cell lymphoma. *Pathology*, *50*(1), 74–87. https://doi.org/10.1016/j.pathol.2017.09.006

Lilyestrom, W., Klein, M. G., Zhang, R., Joachimiak, A., & Chen, X. S. (2006). Crystal structure of SV40 large T-antigen bound to p53: Interplay between a viral oncoprotein and a cellular tumor suppressor. *Genes and Development*, *20*(17), 2373–2382. https://doi.org/10.1101/gad.1456306

Lipson, E. J., Vincent, J. G., Loyo, M., Kagohara, L. T., Luber, B. S., Wang, H., Xu, H., Nayar, S. K., Wang, T. S., Sidransky, D., Anders, R. A., Topalian, S. L., & Taube, J. M. (2013). PD-L1 expression in the Merkel cell carcinoma microenvironment: association with inflammation, Merkel cell polyomavirus and overall survival. *Cancer Immunology Research*, *1*(1), 54–63. https://doi.org/10.1158/2326-6066.CIR-13-0034

Liu, J., Li, H., Nomura, K., Dofuku, R., & Kitagawa, T. (1991). Cytogenetic analysis of hepatic cell lines derived from SV40-T antigen gene-harboring transgenic mice. *Cancer Genetics and Cytogenetics*, *55*(2), 207–216. https://doi.org/10.1016/0165-4608(91)90079-A

Liu, M., Duke, J. L., Richter, D. J., Vinuesa, C. G., Goodnow, C. C., Kleinstein, S. H., & Schatz, D. G. (2008). Two levels of protection for the B cell genome during somatic hypermutation. *Nature*, *451*(7180), 841–845. https://doi.org/10.1038/nature06547

Liu, M., & Schatz, D. G. (2009). Balancing AID and DNA repair during somatic hypermutation. *Trends in Immunology*, *30*(4), 173–181. https://doi.org/10.1016/j.it.2009.01.007

Liu, W., MacDonald, M., & You, J. (2016). Merkel cell polyomavirus infection and Merkel cell carcinoma. *Current Opinion in Virology*, *20*, 20–27. https://doi.org/10.1016/j.coviro.2016.07.011

Liu, W., Yang, R., Payne, A. S., Schowalter, R. M., Spurgeon, M. E., Lambert, P. F., Xu, X., Buck, C. B., & You, J. (2016). Identifying the Target Cells and Mechanisms of Merkel Cell Polyomavirus Infection. *Cell Host and Microbe*, *19*(6), 775–787. https://doi.org/10.1016/j.chom.2016.04.024

López-Ríos, F., Illei, P. B., Rusch, V., & Ladanyi, M. (2004). Evidence against a role for SV40 infection in human mesotheliomas and high risk of false-positive PCR results owing to presence of SV40 sequences in common laboratory plasmids. *Lancet*, *364*(9440), 1157–1166. https://doi.org/10.1016/S0140-6736(04)17102-X

Machida, K., Cheng, K. T. N., Sung, V. M. H., Shimodaira, S., Lindsay, K. L., Levinet, A. M., Lai, M. Y., & Lai, M. M. C. (2004). Hepatitis C virus induces a mutator phenotype: Enhanced mutations of immunoglobulin and protooncogenes. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(12), 4262–4267. https://doi.org/10.1073/pnas.0303971101

Magaldi, T. G., Buch, M. H. C., Murata, H., Erickson, K. D., Neu, U., Garcea, R. L., Peden, K., Stehle, T., & DiMaio, D. (2012). Mutations in the GM1 Binding Site of Simian Virus 40 VP1 Alter Receptor Usage and Cell Tropism. *Journal of Virology*, *86*(13), 7028–7042. https://doi.org/10.1128/jvi.00371-12

Maiti, A., Hou, S., Schiffer, C. A., & Matsuo, H. (2021). Interactions of APOBEC3s with DNA and RNA. *Current Opinion in Structural Biology*, *67*, 195–204. https://doi.org/10.1016/j.sbi.2020.12.004

Malkin, D., Chilton-Macneill, S., Meister, L. A., Sexsmith, E., Diller, L., & Garcea, R. L. (2001). Tissue-specific expression of SV40 in tumors associated with the Li-Fraumeni syndrome. *Oncogene*, *20*(33), 4441–4449. https://doi.org/10.1038/sj.onc.1204583

Manfredi, J. J., Dong, J., Liu, W. J., Resnick-Silverman, L., Qiao, R., Chahinian, P., Saric, M., Gibbs, A. R., Phillips, J. I., Murray, J., Axten, C. W., Nolan, R. P., & Aaronson, S. A. (2005). Evidence against a role for SV40 in human mesothelioma. *Cancer Research*, *65*(7), 2602–2609. https://doi.org/10.1158/0008-5472.CAN-04-2461

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., & Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature*, *424*(6944), 99–103. https://doi.org/10.1038/nature01709

Martin, O. A., Garot, A., Le Noir, S., Aldigier, J.-C., Cogné, M., Pinaud, E., & Boyer, F. (2018). Detecting Rare AID-Induced Mutations in B-Lineage Oncogenes from High-Throughput Sequencing Data Using the Detection of Minor Variants by Error Correction Method. *The Journal of Immunology*, *201*(3), 950–956. https://doi.org/10.4049/jimmunol.1800203

Martini, F., Corallini, A., Balatti, V., Sabbioni, S., Pancaldi, C., & Tognon, M. (2007). Simian virus 40 in humans. *Infectious Agents and Cancer*, *2*(1), 1–12. https://doi.org/10.1186/1750-9378-2-13

Martini, F., Dolcetti, R., Gloghini, A., Iaccheri, L., Carbone, A., Boiocchi, M., & Tognon, M. (1998). Simian-virus-40 footprints in human lymphoproliferative disorders of HIV- and HIV+ patients. *International Journal of Cancer*, *78*(6), 669–674. https://doi.org/10.1002/(SICI)1097-0215(19981209)78:6<669::AID-IJC1>3.0.CO;2-B

Matsumoto, Y., Marusawa, H., Kinoshita, K., Endo, Y., Kou, T., Morisawa, T., Azuma, T., Okazaki, I., Honjo, T., & Chiba, T. (2007). *Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium*. *13*(4), 470–476. https://doi.org/10.1038/nm1566

Matsushita, M., Iwasaki, T., Nonaka, D., Kuwamoto, S., Nagata, K., Kato, M., Kitamura, Y., & Hayashi, K. (2017). Higher expression of activation-induced cytidine deaminase is significantly associated with Merkel cell Polyomavirus-negative Merkel cell carcinomas. *Yonago Acta Medica*, *60*(3), 145–153. https://doi.org/10.33160/yam.2017.09.002

Matthews, A. J., Husain, S., & Chaudhuri, J. (2014). Binding of AID to DNA Does Not Correlate with Mutator Activity. *The Journal of Immunology*, *193*(1), 252–257. https://doi.org/10.4049/jimmunol.1400433

Maul, R. W., Cao, Z., Venkataraman, L., Giorgetti, C. A., Press, J. L., Denizot, Y., Du, H., Sen, R., & Gearhart, P. J. (2014). Spt5 accumulation at variable genes distinguishes somatic hypermutation in germinal center B cells from ex vivo-activated cells. *Journal of Experimental Medicine*, *211*(11), 2297–2306. https://doi.org/10.1084/jem.20131512

Mazhar, S., Taylor, S. E., Sangodkar, J., & Narla, G. (2019). Targeting PP2A in cancer: Combination therapies. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1866*(1), 51–63. https://doi.org/10.1016/j.bbamcr.2018.08.020

Mazzoni, E., Pietrobon, S., Bilancia, M., Vinante, F., Rigo, A., Ferrarini, I., D'Agostino, A., Casali, M. V., Martini, F., & Tognon, M. (2017). High prevalence of antibodies reacting to mimotopes of Simian virus 40 large T antigen, the oncoprotein, in serum samples of patients affected by non-Hodgkin lymphoma. *Cancer Immunology, Immunotherapy*, *66*(9), 1189–1198. https://doi.org/10.1007/s00262-017-2008-9

Mazzoni, E., Rotondo, J. C., Marracino, L., Selvatici, R., Bononi, I., Torreggiani, E., Touzé, A., Martini, F., & Tognon, M. G. (2017). Detection of Merkel cell polyomavirus DNA in serum samples of healthy blood donors. *Frontiers in Oncology*, *7*(NOV). https://doi.org/10.3389/fonc.2017.00294

McDonald, J. J., Alinikula, J., Buerstedde, J.-M., & Schatz, D. G. (2013). A Critical Context-Dependent Role for E Boxes in the Targeting of Somatic Hypermutation. *The Journal of Immunology*, *191*(4), 1556–1566. https://doi.org/10.4049/jimmunol.1300969

McLachlan, T., Matthews, W. C., Jackson, E. R., Staudt, D. E., Douglas, A. M., Findlay, I. J., Persson, M. L., Duchatel, R. J., Mannan, A., Germon, Z. P., & Dun, M. D. (2022). B-cell Lymphoma 6 (BCL6): From Master Regulator of Humoral Immunity to Oncogenic Driver in Pediatric Cancers. *Molecular Cancer Research : MCR*, *20*(12), 1711–1723. https://doi.org/10.1158/1541-7786.MCR-22-0567

McNees, A., Harrigal, L., Kelly, A., Minard, C., Wong, C., & Butel, J. (2019). Viral microRNA effects on persistent infection of human lymphoid cells by polyomavirus SV40. In *PLoS ONE* (Vol. 13, Issue 2). https://doi.org/10.1371/journal.pone.0192799.g001

Mendoza, S. M., Konishi, T., & Miller, C. W. (1998). Integration of SV40 in human osteosarcoma DNA. *Oncogene*, *17*(19), 2457–2462. https://doi.org/10.1038/sj.onc.1202179

Meng, F. L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K. R., Meyers, R. M., Amor, C., Wasserman, C. R., Neuberg, D., Casellas, R., Nussenzweig, M. C., Bradner, J. E., Liu, X. S., & Alt, F. W. (2014). Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell*, *159*(7), 1538–1548. https://doi.org/10.1016/j.cell.2014.11.014

Mertz, K. D., Junt, T., Schmid, M., Pfaltz, M., & Kempf, W. (2010). Inflammatory monocytes are a reservoir for merkel cell polyomavirus. *Journal of Investigative Dermatology*, *130*(4), 1146–1151. https://doi.org/10.1038/jid.2009.392

Methot, S., Litzler, L. C., Subramani, P. G., Eranki, A. K., Fifield, H., Patenaude, A. M., Gilmore, J. C., Santiago, G. E., Bagci, H., Côté, J. F., Larijani, M., Verdun, R. E., & Di Noia, J. M. (2018). A licensing step links AID to transcription elongation for mutagenesis in B cells. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-03387-6

Meyer, S. N., Koul, S., & Pasqualucci, L. (2021). Mouse Models of Germinal Center Derived B-Cell Lymphomas. *Frontiers in Immunology*, *12*(August), 1–19. https://doi.org/10.3389/fimmu.2021.710711

Migliazza, A., Martinotti, S., Chen, W., Fusco, C., Ye, B., Knowles, D. M., Offit, K., Chaganti, R. S. K., & Dalla-Favera, R. (1995). Frequent somatic hypermutation of the 5′ noncoding region of the BCL6 gene in B-cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(26), 12520–12524. https://doi.org/10.1073/pnas.92.26.12520

Mlynarczyk, C., Fontán, L., & Melnick, A. (2019). Germinal center-derived lymphomas: The darkest side of humoral immunity. *Immunological Reviews*, *288*(1), 214–239. https://doi.org/10.1111/imr.12755

Mobaraki, G., Shi, S., Liu, D., Smits, K. M., Severens, K., Lommen, K., Rennspiess, D., Speel, E. J. M., Winnepenninckx, V., Klufah, F., Samarska, I., & Hausen, A. zur. (2024). Mapping of Human Polyomavirus in Renal Cell Carcinoma Tissues. *International Journal of Molecular Sciences*, *25*(15). https://doi.org/10.3390/ijms25158213

Moens, U., Prezioso, C., & Pietropaolo, V. (2020). Genetic diversity of the noncoding control region of the novel human polyomaviruses. *Viruses*, *12*(12). https://doi.org/10.3390/v12121406

Monaco, M. C., Atwood, W. J., Gravell, M., Tornatore, C. S., & Major, E. O. (1996). JC virus infection of hematopoietic progenitor cells, primary B lymphocytes, and tonsillar stromal cells: implications for viral latency. *Journal of Virology*, *70*(10), 7004–7012. https://doi.org/10.1128/jvi.70.10.7004-7012.1996

Morales-Sánchez, A., & Fuentes-Pananá, E. M. (2014). Human viruses and cancer. *Viruses*, *6*(10), 4047–4079. https://doi.org/10.3390/v6104047

Morin, R. D., Arthur, S. E., & Hodson, D. J. (2022). Molecular profiling in diffuse large B-cell lymphoma: why so many types of subtypes? *British Journal of Haematology*, *196*(4), 814–829. https://doi.org/10.1111/bjh.17811

Muramatsu, M., Kazuo, K., Sidonia, F., Shuichi, Y., Yoichi, S., & Tasuku, H. (2000). Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme. *Cell*, *102*(5), 553–563.

Muramatsu, M., Sankaranand, V. S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N. O., & Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *Journal of Biological Chemistry*, *274*(26), 18470–18476. https://doi.org/10.1074/jbc.274.26.18470

Müschen, M., Re, D., Jungnickel, B., Diehl, V., Rajewsky, K., & Küppers, R. (2000). Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction. *Journal of Experimental Medicine*, *192*(12), 1833–1839. https://doi.org/10.1084/jem.192.12.1833

Nambu, Y., Sugai, M., Gonda, H., Lee, C. G., Katakai, T., Agata, Y., Yokota, Y., & Shimizu, A. (2003). Transcription-Coupled Events Associating with Immunoglobulin Switch Region Chromatin. *Science*, *302*(5653), 2137–2140. https://doi.org/10.1126/science.1092481

Nguyen, K. D., Lee, E. E., Yue, Y., Stork, J., Pock, L., North, J. P., Vandergriff, T., Cockerell, C., Hosler, G. A., Pastrana, D. V., Buck, C. B., & Wang, R. C. (2017). Human polyomavirus 6 and 7 are associated with pruritic and dyskeratotic dermatoses. *Journal of the American Academy of Dermatology*, *76*(5), 932-940.e3. https://doi.org/10.1016/j.jaad.2016.11.035

Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L. A., Menzies, A., Martin, S., Leung, K., Chen, L., Leroy, C., Ramakrishna, M., Rance, R., Lau, K. W., Mudie, L. J., … Stratton, M. R. (2012). Mutational

processes molding the genomes of 21 breast cancers. *Cell*, *149*(5), 979–993. https://doi.org/10.1016/j.cell.2012.04.024

Nogueira, G., Fernandes, R., García-Moreno, J. F., & Romão, L. (2021). Nonsense-mediated RNA decay and its bipolar function in cancer. *Molecular Cancer*, *20*(1), 1–19. https://doi.org/10.1186/s12943-021-01364-0

Nonaka, T., Toda, Y., Hiai, H., Uemura, M., Nakamura, M., Yamamoto, N., Asato, R., Hattori, Y., Bessho, K., Minato, N., & Kinoshita, K. (2016). Involvement of activation-induced cytidine deaminase in skin cancer development. *Journal of Clinical Investigation*, *126*(4), 1367–1382. https://doi.org/10.1172/JCI81522

O'Neill, F. J., Greenlee, J. E., & Carney, H. (2003). The archetype enhancer of simian virus 40 DNA is duplicated during virus growth in human cells and rhesus monkey kidney cells but not in green monkey kidney cells. *Virology*, *310*(1), 173–182. https://doi.org/10.1016/S0042-6822(03)00116-8

Okazaki, I. M., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., & Honjo, T. (2003). Constitutive expression of AID leads to tumorigenesis. *Journal of Experimental Medicine*, *197*(9), 1173–1181. https://doi.org/10.1084/jem.20030275

Okura, R., Yoshioka, H., Yoshioka, M., Hiromasa, K., Nishio, D., & Nakamura, M. (2014). *Letter to the Editor Expression of AID in malignant melanoma with BRAF V600E mutation*. 2013–2014. https://doi.org/10.1111/exd.12402

Pachano, T., Haro, E., & Rada-Iglesias, A. (2022). Enhancer-gene specificity in development and disease. *Development (Cambridge, England)*, *149*(11). https://doi.org/10.1242/dev.186536

Pallas, D. C., Weller, W., Jaspers, S., Miller, T. B., Lane, W. S., & Roberts, T. M. (1992). The third subunit of protein phosphatase 2A (PP2A), a 55-kilodalton protein which is apparently substituted for by T antigens in complexes with the 36- and 63-kilodalton PP2A subunits, bears little resemblance to T antigens. *Journal of Virology*, *66*(2), 886–893. https://doi.org/10.1128/jvi.66.2.886-893.1992

Panea, R. I., Love, C. L., Shingleton, J. R., Reddy, A., Bailey, J. A., Moormann, A. M., Otieno, J. A., Ong'echa, J. M., Oduor, C. I., Schroeder, K. M. S., Masalu, N., Chao, N. J., Agajanian, M., Major, M. B., Fedoriw, Y., Richards, K. L., Rymkiewicz, G., Miles, R. R., Alobeid, B., … Dave, S. S. (2019). The whole-genome landscape of Burkitt lymphoma subtypes. *Blood*, *134*(19), 1598–1607. https://doi.org/10.1182/blood.2019001880

Pantulu, N. D., Pallasch, C. P., Kurz, A. K., Kassem, A., Frenzel, L., Sodenkamp, S., Kvasnicka, H. M., Wendtner, C. M., & Zur Hausen, A. (2010). Detection of a novel truncating Merkel cell polyomavirus large T antigen deletion in chronic lymphocytic leukemia cells. *Blood*, *116*(24), 5280–5284. https://doi.org/10.1182/blood-2010-02-269829

Parekh, S., Polo, J. M., Shaknovich, R., Juszczynski, P., Lev, P., Ranuncolo, S. M., Yin, G., Klein, U., Cattoretti, G., Dalla Favera, R., Shipp, M. A., & Melnick, A. (2007). BCL6 programs lymphoma cells for survival and differentiation through distinct biochemical mechanisms. *Blood*, *110*(6), 2067–2074. https://doi.org/10.1182/blood-2007-01-069575

Pasqualucci, L., Bhagat, G., Jankovic, M., Compagno, M., Smith, P., Muramatsu, M., Honjo, T., Iii, H. C. M., Nussenzweig, M. C., & Dalla-favera, R. (2008). *AID is required for germinal center – derived lymphomagenesis*. *40*(1), 108–112. https://doi.org/10.1038/ng.2007.35

Pasqualucci, L., Dominguez-Sola, D., Chiarenza, A., Fabbri, G., Grunn, A., Trifonov, V., Kasper, L. H., Lerach, S., Tang, H., Ma, J., Rossi, D., Chadburn, A., Murty, V. V., Mullighan, C. G., Gaidano, G., Rabadan, R., Brindle, P. K., & Dalla-Favera, R. (2011). Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature*, *471*(7337), 189–196. https://doi.org/10.1038/nature09730

Pasqualucci, L., Migliazza, A., Basso, K., Houldsworth, J., Chaganti, R. S. K., & Dalla-favera, R. (2003). Plenary paper Mutations of the BCL6 proto-oncogene disrupt its negative autoregulation in diffuse large B-cell lymphoma. *Blood*, *101*(8), 2914–2923. https://doi.org/10.1182/blood-2002-11-3387.Supported

Pasqualucci, L., Migliazza, A., Fracchiolla, N., William, C., Neri, A., Baldini, L., Chaganti, R. S. K., Klein, U., Küppers, R., Rajewsky, K., & Dalla-Favera, R. (1998). BCL-6 mutations in normal germinal center B cells: Evidence of somatic hypermutation acting outside Ig loci. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(20), 11816–11821. https://doi.org/10.1073/pnas.95.20.11816

Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R. S. K., Küppers, R., & Dalla-Favera, R. (2001). *Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas*. *412*(July), 341–346.

Passerini, S., Prezioso, C., Babini, G., Ferlosio, A., Cosio, T., Campione, E., Moens, U., Ciotti, M., & Pietropaolo, V. (2023). Detection of Merkel Cell Polyomavirus (MCPyV) DNA and Transcripts in Merkel Cell Carcinoma (MCC). *Pathogens*, *12*(7), 1–8. https://doi.org/10.3390/pathogens12070894

Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., San-Martin, B. R., Barreto, V., Nieland, T. J., Root, D. E., Casellas, R., & Nussenzweig, M. C. (2010). Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell*, *143*(1), 122–133. https://doi.org/10.1016/j.cell.2010.09.017

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Chao, J., Rabadan, R., Economides, A. N., & Basu, U. (2014). Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature*, *514*(7522), 389–393. https://doi.org/10.1038/nature13580

Peng, H.-Z., Du, M.-Q., Koulis, A., Aiello, A., Dogan, A., Pan, L.-X., & Isaacson, P. G. (1999). Nonimmunoglobulin Gene Hypermutation in Germinal Center B Cells. *Blood*, *93*(7), 2167–2172. https://doi.org/10.1182/blood.v93.7.2167

Peretti, A., Geoghegan, E. M., Pastrana, D. V., Smola, S., Feld, P., Sauter, M., Lohse, S., Ramesh, M., Lim, E. S., Wang, D., Borgogna, C., FitzGerald, P. C., Bliskovsky, V., Starrett, G. J., Law, E. K., Harris, R. S., Killian, J. K., Zhu, J., Pineda, M., … Buck, C. B. (2018). Characterization of BK Polyomaviruses from Kidney Transplant Recipients Suggests a Role for APOBEC3 in Driving In-Host Virus Evolution. *Cell Host and Microbe*, *23*(5), 628-635.e7. https://doi.org/10.1016/j.chom.2018.04.005

Petljak, M., & Maciejowski, J. (2020). Molecular origins of APOBEC-associated mutations in cancer. *DNA Repair*, *94*(June), 102905. https://doi.org/10.1016/j.dnarep.2020.102905

Pham, P., Bransteitter, R., Petruska, J., & Goodman, M. F. (2003). Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature*, 103–107.

Phan, R. T., & Dalla-Favera, R. (2004). The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells. *Nature*, *432*. https://doi.org/10.1038/nature03148

Phan, R. T., Saito, M., Basso, K., Niu, H., & Dalla-Favera, R. (2005). BCL6 interacts with the transcription factor Miz-1 to suppress the cyclin-dependent kinase inhibitor p21 and cell cycle arrest in germinal center B cells. *Nature Immunology*, *6*(10), 1054–1060. https://doi.org/10.1038/ni1245

Phan, R. T., Saito, M., Kitagawa, Y., Means, A. R., & Dalla-Favera, R. (2007). Genotoxic stress regulates expression of the proto-oncogene Bcl6 in germinal center B cells. *Nature Immunology*, *8*(10), 1132–1139. https://doi.org/10.1038/ni1508

Pietropaolo, V., Prezioso, C., & Moens, U. (2020). Merkel cell polyomavirus and merkel cell carcinoma. *Cancers*, *12*(7), 1–38. https://doi.org/10.3390/cancers12071774

Pilzecker, B., Buoninfante, O. A., Pritchard, C., Blomberg, O. S., Huijbers, I. J., Van Den Berk, P. C. M., & Jacobs, H. (2016). PrimPol prevents APOBEC/AID family mediated DNA mutagenesis. *Nucleic Acids Research*, *44*(10), 4734–4744. https://doi.org/10.1093/nar/gkw123

Pilzecker, B., & Jacobs, H. (2019). Mutating for good: DNA damage responses during somatic hypermutation. *Frontiers in Immunology*, *10*(MAR), 1–13. https://doi.org/10.3389/fimmu.2019.00438

Pipas, J. M. (1992). Common and unique features of T antigens encoded by the polyomavirus group. *Journal of Virology*, *66*(7), 3979–3985. https://doi.org/10.1128/jvi.66.7.3979-3985.1992

Pipas, J. M. (2009). SV40: Cell transformation and tumorigenesis. *Virology*, *384*(2), 294–303. https://doi.org/10.1016/j.virol.2008.11.024

Pitkäniemi, J., Malila, N., Heikkinen, S., & Seppä, K. (2024). *Syöpä 2022. Tilastoraportti Suomen syöpätilanteesta*. Suomen Syöpäyhdistys.

Polo, J. M., Ci, W., Licht, J. D., & Melnick, A. (2008). Reversible disruption of BCL6 repression complexes by CD40 signaling in normal and malignant B cells. *Blood*, *112*(3), 644–651. https://doi.org/10.1182/blood-2008-01-131813

Poulain, F., Lejeune, N., Willemart, K., & Gillet, N. A. (2020). Footprint of the host restriction factors APOBEC3 on the genome of human viruses. *PLoS Pathogens*, *16*(8), 1–30. https://doi.org/10.1371/JOURNAL.PPAT.1008718

Poulin, D. L., & DeCaprio, J. A. (2006). Is there a role for SV40 in human cancer? *Journal of Clinical Oncology*, *24*(26), 4356–4365. https://doi.org/10.1200/JCO.2005.03.7101

Prado, J. C. M., Monezi, T. A., Amorim, A. T., Lino, V., Paladino, A., & Boccardo, E. (2018). Human polyomaviruses and cancer: An overview. *Clinics*, *73*. https://doi.org/10.6061/clinics/2018/e558s

Prezioso, C., Carletti, R., Obregon, F., Piacentini, F., Manicone, A. M., Soda, G., Moens, U., Di Gioia, C., & Pietropaolo, V. (2021). Evaluation of Merkel Cell Polyomavirus DNA in Tissue Samples from Italian Patients with Diagnosis of MCC. *Viruses*, *13*(1). https://doi.org/10.3390/v13010061

Prezioso, C., Obregon, F., Ambroselli, D., Petrolo, S., Checconi, P., Rodio, D. M., Coppola, L., Nardi, A., De Vito, C., Sarmati, L., Andreoni, M., Palamara, A. T., Ciotti, M., & Pietropaolo, V. (2020). Merkel cell polyomavirus (MCPyV) in the context of immunosuppression: Genetic analysis of noncoding control region (NCCR) Variability among a HIV-1-positive population. *Viruses*, *12*(5). https://doi.org/10.3390/v12050507

Pyöriä, L., Pratas, D., Toppinen, M., Simmonds, P., Hedman, K., Sajantila, A., & Perdomo, M. F. (2024). Intra-host genomic diversity and integration landscape of human tissue-resident DNA virome. *Nucleic Acids Research*, *52(21)*(October), 13073–13093.

Qian, J., Wang, Q., Dose, M., Pruett, N., Kieffer-Kwon, K. R., Resch, W., Liang, G., Tang, Z., Mathé, E., Benner, C., Dubois, W., Nelson, S., Vian, L., Oliveira, T. Y., Jankovic, M., Hakim, O., Gazumyan, A., Pavri, R., Awasthi, P., … Casellas, R. (2014). B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell*, *159*(7), 1524–1537. https://doi.org/10.1016/j.cell.2014.11.013

Qiao, Q., Wang, L., Meng, F.-L., Hwang, J. K., Alt, F. W., & Wu, H. (2017). Article AID Recognizes Structured DNA for Class Switch Recombination. *Molecular Cell*, *67*, 361–373. https://doi.org/10.1016/j.molcel.2017.06.034

Qin, Y., & Meng, F. L. (2024). Taming AID mutator activity in somatic hypermutation. *Trends in Biochemical Sciences*, 1–11. https://doi.org/10.1016/j.tibs.2024.03.011

Que, L., Li, Y., Dainichi, T., Kukimoto, I., Nishiyama, T., Nakano, Y., Shima, K., Suzuki, T., Sato, Y., Horike, S., Aizaki, H., Watashi, K., Kato, T., Aly, H. H., Watanabe, N., Kabashima, K., Wakae, K., & Muramatsu, M. (2021). Interferon-gamma induced APOBEC3B contributes to Merkel cell polyomavirus genome mutagenesis in Merkel cell carcinoma. In *Journal of Investigative Dermatology*. Society for Investigative Dermatology. https://doi.org/10.1016/j.jid.2021.12.019

Rada, C., Jarvis, J. M., & Milstein, C. (2002). AID-GFP chemirec protein increases hypermutation og Ig genes with no evidence of nuclear localization. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(10), 7003–7008. https://doi.org/10.1073/pnas.092160999

Ranuncolo, S. M., Polo, J. M., & Melnick, A. (2008). BCL6 represses CHEK1 and suppresses DNA damage pathways in normal and malignant B-cells. *Blood Cells, Molecules, and Diseases*, *41*(1), 95–99. https://doi.org/10.1016/j.bcmd.2008.02.003

Ray, U., Cinque, P., Gerevini, S., Longo, V., Lazzarin, A., Schippling, S., Martin, R., Buck, C. B., & Pastrana, D. V. (2015). JC polyomavirus mutants escape antibody-mediated neutralization. *Science Translational Medicine*, *7*(306). https://doi.org/10.1126/scitranslmed.aab1720

Reddy, A., Zhang, J., Davis, N. S., Moffitt, A. B., Love, C. L., Waldrop, A., Leppa, S., Pasanen, A., Meriranta, L., Karjalainen-Lindsberg, M. L., Nørgaard, P., Pedersen, M., Gang, A. O., Høgdall, E., Heavican, T. B., Lone, W., Iqbal, J., Qin, Q., Li, G., … Dave, S. S. (2017). Genetic and

Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell*, *171*(2), 481-494.e15. https://doi.org/10.1016/j.cell.2017.09.027

Ren, W., Ye, X., Su, H., Li, W., Liu, D., Pirmoradian, M., Wang, X., Zhang, B., Zhang, Q., Chen, L., Nie, M., Liu, Y., Meng, B., Huang, H., Jiang, W., Zeng, Y., Li, W., Wu, K., Hou, Y., … Pan-Hammarström, Q. (2018). Genetic landscape of hepatitis B virus-associated diffuse large B-cell lymphoma. *Blood*, *131*(24), 2670–2681. https://doi.org/10.1182/blood-2017-11-817601

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Lagelouse, R., Gennery, A., Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A. G., Brousse, N., Muramatsu, M., Notarangelo, L. D., Kinoshita, K., Honjo, T., … Durandy, A. (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). *Cell*, *102*(5), 565–575. https://doi.org/10.1016/S0092-8674(00)00079-9

Richards, K. F., Guastafierro, A., Shuda, M., Toptan, T., Moore, P. S., & Chang, Y. (2015). Merkel cell polyomavirus T antigens promote cell proliferation and inflammatory cytokine gene expression. *Journal of General Virology*, *96*(12), 3532–3544. https://doi.org/10.1099/jgv.0.000287

Rivera, Z., Strianese, O., Bertino, P., Yang, H., Pass, H., & Carbone, M. (2008). The relationship between simian virus 40 and mesothelioma. *Current Opinion in Pulmonary Medicine*, *14*(4), 316–321. https://doi.org/10.1097/MCP.0b013e3283018220

Rodig, S. J., Cheng, J., Wardzala, J., DoRosario, A., Scanlon, J. J., Laga, A. C., Martinez-Fernandez, A., Barletta, J. A., Bellizzi, A. M., Sadasivam, S., Holloway, D. T., Cooper, D. J., Kupper, T. S., Wang, L. C., & DeCaprio, J. A. (2012). Improved detection suggests all Merkel cell carcinomas harbor Merkel polyomavirus. *Journal of Clinical Investigation*, *122*(12), 4645–4653. https://doi.org/10.1172/JCI64116

Rogier, M., Moritz, J., Robert, I., Lescale, C., Heyer, V., Abello, A., Martin, O., Capitani, K., Thomas, M., Thomas-Claudepierre, A. S., Laffleur, B., Jouan, F., Pinaud, E., Tarte, K., Cogné, M., Conticello, S. G., Soutoglou, E., Deriano, L., & Reina-San-Martin, B. (2021). Fam72a enforces error-prone DNA repair during antibody diversification. *Nature*, *600*(7888), 329–333. https://doi.org/10.1038/s41586-021-04093-y

Ronai, D., Iglesias-Ussel, M. D., Fan, M., Li, Z., Martin, A., & Scharff, M. D. (2007). Detection of chromatin-associated single-stranded DNA in regions targeted for somatic hypermutation. *Journal of Experimental Medicine*, *204*(1), 181–190. https://doi.org/10.1084/jem.20062032

Rotondo, J. C., Mazzoni, E., Bononi, I., Tognon, M., & Martini, F. (2019). Association Between Simian Virus 40 and Human Tumors. *Frontiers in Oncology*, *9*(July), 1–19. https://doi.org/10.3389/fonc.2019.00670

Roy, D., Yu, K., & Lieber, M. R. (2008). Mechanism of R-Loop Formation at Immunoglobulin Class Switch Sequences. *Molecular and Cellular Biology*, *28*(1), 50–60. https://doi.org/10.1128/mcb.01251-07

Sahi, H. (2017). *Merkelinsolukarsinooma – mitä uutta? 11*, 2365–2371.

Sahi, H., Sihto, H., Artama, M., Koljonen, V., Böhling, T., & Pukkala, E. (2017). History of chronic inflammatory disorders increases the risk of Merkel cell carcinoma, but does not correlate with Merkel cell polyomavirus infection. *British Journal of Cancer*, *116*(2), 260–264. https://doi.org/10.1038/bjc.2016.391

Saito, M., Gao, J., Basso, K., Kitagawa, Y., Smith, P. M., Bhagat, G., Pernis, A., Pasqualucci, L., & Dalla-Favera, R. (2007). A Signaling Pathway Mediating Downregulation of BCL6 in Germinal Center B Cells Is Blocked by BCL6 Gene Alterations in B Cell Lymphoma. *Cancer Cell*, *12*(3), 280–292. https://doi.org/10.1016/j.ccr.2007.08.011

Salter, J. D., Bennett, R. P., & Smith, H. C. (2016). The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends in Biochemical Sciences*, *41*(7), 578–594. https://doi.org/10.1016/j.tibs.2016.05.001

Samaka, R. M., Aiad, H. A., Kandil, M. A., Asaad, N. Y., & Holah, N. S. (2015). The prognostic role and relationship between E2F1 and SV40 in diffuse large B-cell lymphoma of Egyptian patients. *Analytical Cellular Pathology*, *2015*. https://doi.org/10.1155/2015/919834

Sanjuán, R., & Domingo-Calap, P. (2016). Mechanisms of viral mutation. *Cellular and Molecular Life Sciences*, *73*(23), 4433–4448. https://doi.org/10.1007/s00018-016-2299-6

Saribasak, H., Saribasak, N. N., Ipek, F. M., Ellwart, J. W., Arakawa, H., & Buerstedde, J.-M. (2005). Uracil DNA Glycosylase Disruption Blocks Ig Gene Conversion and Induces Transition Mutations. *Journal of Immunology*, *176*(1), 365–371.

Sauer, C. M., Haugg, A. M., Chteinberg, E., Rennspiess, D., Winnepenninckx, V., Speel, E. J., Becker, J. C., Kurz, A. K., & zur Hausen, A. (2017). Reviewing the current evidence supporting early B-cells as the cellular origin of Merkel cell carcinoma. *Critical Reviews in Oncology/Hematology*, *116*, 99–105. https://doi.org/10.1016/j.critrevonc.2017.05.009

Schadendorf, D., Lebbé, C., zur Hausen, A., Avril, M. F., Hariharan, S., Bharmal, M., & Becker, J. C. (2017). Merkel cell carcinoma: Epidemiology, prognosis, therapy and unmet medical needs. *European Journal of Cancer*, *71*, 53–69. https://doi.org/10.1016/j.ejca.2016.10.022

Schirm, S., Jiricny, J., & Schaffner, W. (1987). The SV40 enhancer can be dissected into multiple segments, each with a different cell type specificity. *Genes & Development*, *1*(1), 65–74. https://doi.org/10.1101/gad.1.1.65

Schmidt, K., Keiser, S., Günther, V., Georgiev, O., Hirsch, H. H., Schaffner, W., & Bethge, T. (2016). Transcription enhancers as major determinants of SV40 polyomavirus growth efficiency and host cell tropism. *Journal of General Virology*, *97*(7), 1597–1603. https://doi.org/10.1099/jgv.0.000487

Schmitt, M., Wieland, U., Kreuter, A., & Pawlita, M. (2012). C-terminal deletions of Merkel cell polyomavirus large T-antigen, a highly specific surrogate marker for virally induced malignancy. *International Journal of Cancer*, *131*(12), 2863–2868. https://doi.org/10.1002/ijc.27607

Schmitz, R., Young, R. M., Ceribelli, M., Jhavar, S., Xiao, W., Zhang, M., Wright, G., Shaffer, A. L., Hodson, D. J., Buras, E., Liu, X., Powell, J., Yang, Y., Xu, W., Zhao, H., Kohlhammer, H., Rosenwald, A., Kluin, P., Müller-Hermelink, H. K., … Staudt, L. M. (2012). Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature*, *490*(7418), 116–120. https://doi.org/10.1038/nature11378

Schowalter, R. M., Pastrana, D. V., Pumphrey, K. A., Moyer, A. L., & Buck, C. B. (2010). Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. *Cell Host and Microbe*, *7*(6), 509–515. https://doi.org/10.1016/j.chom.2010.05.006

Schroeder, H. W., & Cavacini, L. (2010). Structure and function of immunoglobulins. *Journal of Allergy and Clinical Immunology*, *125*(2 SUPPL. 2), S41–S52. https://doi.org/10.1016/j.jaci.2009.09.046

Seneca, N. T. M., Sáenz Robles, M. T., & Pipas, J. M. (2014). Removal of a small C-terminal region of JCV and SV40 large T antigens has differential effects on transformation. *Virology*, *468–470*(412), 47–56. https://doi.org/10.1016/j.virol.2014.07.038

Senigl, F., Maman, Y., Dinesh, R. K., Alinikula, J., Seth, R. B., Pecnova, L., Omer, A. D., Rao, S. S. P., Weisz, D., Buerstedde, J. M., Aiden, E. L., Casellas, R., Hejnar, J., & Schatz, D. G. (2019). Topologically Associated Domains Delineate Susceptibility to Somatic Hypermutation. *Cell Reports*, *29*(12), 3902-3915.e8. https://doi.org/10.1016/j.celrep.2019.11.039

Seppälä, H., Virtanen, E., Saarela, M., Laine, P., Paulín, L., Mannonen, L., Auvinen, P., & Auvinen, E. (2017). Single-molecule sequencing revealing the presence of distinct JC polyomavirus populations in patients with progressive multifocal leukoencephalopathy. *Journal of Infectious Diseases*, *215*(6), 889–895. https://doi.org/10.1093/infdis/jiw399

Shaffer, A. L., Rosenwald, A., & Staudt, L. M. (2002). *LYMPHOID MALIGNANCIES : THE DARK SIDE OF B-CELL DIFFERENTIATION*. *2*(December), 920–932. https://doi.org/10.1038/nri953

Shaffer, A. L., Yu, X., He, Y., Boldrick, J., Chan, E. P., & Staudt, L. M. (2000). BCL-6 represses genes that function in lymphocyte differentiation, inflammation, and cell cycle control. *Immunity*, *13*(2), 199–212. https://doi.org/10.1016/S1074-7613(00)00020-0

Shah, K. V. (2007). SV40 and human cancer: A review of recent data. *International Journal of Cancer*, *120*(2), 215–223. https://doi.org/10.1002/ijc.22425

Shapiro, M., Krug, L. T., & MacCarthy, T. (2021). Mutational pressure by host APOBEC3s more strongly affects genes expressed early in the lytic phase of herpes simplex virus-1 (HSV-1) and human polyomavirus (HPyV) infection. *PLoS Pathogens*, *17*(4), 1–27. https://doi.org/10.1371/journal.ppat.1009560

Sheehy, A. M., Gaddis, N. C., Choi, J. D., & Malim, M. H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature*, *418*(6898), 646–650. https://doi.org/10.1038/nature00939

Shen, H. M., Peters, A., Baron, B., Zhu, X., & Storb, U. (1998). Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science*, *280*(5370), 1750–1752. https://doi.org/10.1126/science.280.5370.1750

Shen, H. M., Tanaka, A., Bozek, G., Nicolae, D., & Storb, U. (2006). Somatic Hypermutation and Class Switch Recombination in Msh6−/−Ung−/− Double-Knockout Mice. *The Journal of Immunology*, *177*(8), 5386–5392. https://doi.org/10.4049/jimmunol.177.8.5386

Shen, J. C., Kamath-Loeb, A. S., Kohrn, B. F., Loeb, K. R., Preston, B. D., & Loeb, L. A. (2019). A high-resolution landscape of mutations in the BCL6 super-enhancer in normal human B cells. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(49), 24779–24785. https://doi.org/10.1073/pnas.1914163116

Shen, Y., Ge, B., Ramachandrareddy, H., McKeithan, T., & Chan, W. C. (2008). Alternative splicing generates a short BCL6 (BCL6S) isoform encoding a compact repressor. *Biochemical and Biophysical Research Communications*, *375*(2), 190–193. https://doi.org/10.1016/j.bbrc.2008.07.116

Shi, K., Carpenter, M. A., Banerjee, S., Shaban, N. M., Kurahashi, K., Salamango, D. J., McCann, J. L., Starrett, G. J., Duffy, J. V., Demir, Ö., Amaro, R. E., Harki, D. A., Harris, R. S., & Aihara, H. (2017). Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nature Structural and Molecular Biology*, *24*(2), 131–139. https://doi.org/10.1038/nsmb.3344

Shivapurkar, N., Harada, K., Reddy, J., Scheuermann, R. H., Xu, Y., Mckenna, R. W., Milchgrub, S., Kroft, S. H., Feng, Z., & Gazdar, A. F. (2002). Presence of simian virus 40 DNA sequences in human lymphomas. *The Lancet*, *359*, 851–852.

Shiver, M. B., Mahmoud, F., & Ling, G. (2015). Response to Idelalisib in a Patient with Stage IV Merkel-Cell Carcinoma. *New England Journal of Medicine*, *373*(16), 1580–1582. https://doi.org/10.1056/nejmc1510015

Shuda, M., Feng, H., Kwun, H. J., Rosen, S. T., Gjoerup, O., Moore, P. S., & Chang, Y. (2008). T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus. *PNAS*, *105*(42), 16272–16277.

Sigvardsson, M. (2023). Transcription factor networks link B-lymphocyte development and malignant transformation in leukemia. *Genes and Development*, *37*(15–16), 703–723. https://doi.org/10.1101/gad.349879.122

Siriwardena, S., Chen, K., & Bhagwat, A. S. (2016). The Functions and Malfunctions of AID/APOBEC Family Deaminases: the known knowns and the known unknowns. *Chemical Reviews*, *116*(20), 12688. https://doi.org/10.1021/acs.chemrev.6b00296.The

Smith, E., & Shilatifard, A. (2014). Enhancer biology and enhanceropathies. *Nature Structural and Molecular Biology*, *21*(3), 210–219. https://doi.org/10.1038/nsmb.2784

Starrett, G. J., Luengas, E. M., McCann, J. L., Ebrahimi, D., Temiz, N. A., Love, R. P., Feng, Y., Adolph, M. B., Chelico, L., Law, E. K., Carpenter, M. A., & Harris, R. S. (2016). The DNA cytosine deaminase APOBEC3H haplotype i likely contributes to breast and lung cancer mutagenesis. *Nature Communications*, *7*(May), 1–13. https://doi.org/10.1038/ncomms12918

Starrett, G. J., Serebrenik, A. A., Roelofs, P. A., Mccann, J. L., Verhalen, B., Jarvis, M. C., Stewart, T. A., Law, E. K., Krupp, A., Jiang, M., Martens, J. W. M., Cahir-mcfarland, E., Span, P. N., &

Harris, S. (2019). *Polyomavirus T Antigen Induces APOBEC3B Expression Using an LXCXE-Dependent and TP53-Independent Mechanism*. *10*(1), 1–14.

Steen, C. B., Luca, B. A., Esfahani, M. S., Azizi, A., Sworder, B. J., Nabet, B. Y., Kurtz, D. M., Liu, C. L., Khameneh, F., Advani, R. H., Natkunam, Y., Myklebust, J. H., Diehn, M., Gentles, A. J., Newman, A. M., & Alizadeh, A. A. (2021). The landscape of tumor cell states and ecosystems in diffuse large B cell lymphoma. *Cancer Cell*, *39*(10), 1422-1437.e10. https://doi.org/10.1016/j.ccell.2021.08.011

Storb, U. (1996). The molecular basis of somatic hypermutation of immunoglobulin genes. *Current Opinion in Immunology*, *8*(2), 206–214. https://doi.org/10.1016/S0952-7915(96)80059-8

Stubdal, H., Zalvide, J., Campbell, K. S., Schweitzer, C., Roberts, T. M., & DeCaprio, J. A. (1997). Inactivation of pRB-Related Proteins p130 and p107 Mediated by the J Domain of Simian Virus 40 Large T Antigen. *Molecular and Cellular Biology*, *17*(9), 4979–4990. https://doi.org/10.1128/mcb.17.9.4979

Sundquist, W. I., & Klug, A. (1989). Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, *342*(December), 825–829.

Sundqvist, B. Z., Kilpinen, S. K., Böhling, T. O., Koljonen, V. S. K., & Sihto, H. J. (2023). Transcriptomic analyses reveal three distinct molecular subgroups of Merkel cell carcinoma with differing prognoses. *International Journal of Cancer*, *152*(10), 2099–2108. https://doi.org/10.1002/ijc.34425

Sunshine, J. C., Jahchan, N. S., Sage, J., & Choi, J. (2018). Are there multiple cells of origin of Merkel cell carcinoma? In *Oncogene* (Vol. 37, Issue 11, pp. 1409–1416). Springer US. https://doi.org/10.1038/s41388-017-0073-3

Sunyaev, S. R., Lugovskoy, A., Simon, K., & Gorelik, L. (2009). Adaptive mutations in the JC virus protein capsid are associated with progressive multifocal leukoencephalopathy (PML). *PLoS Genetics*, *5*(2). https://doi.org/10.1371/journal.pgen.1000368

Swanton, C., McGranahan, N., Starrett, G. J., & Harris, R. S. (2015). APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discovery*, *5*(7), 704–712. https://doi.org/10.1158/2159-8290.CD-15-0344

Swerdlow, S. H., Campo, E., Pileri, S. A., Lee Harris, N., Stein, H., Siebert, R., Advani, R., Ghielmini, M., Salles, G. A., Zelenetz, A. D., & Jaffe, E. S. (2016). The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*, *127*(20), 2375–2390. https://doi.org/10.1182/blood-2016-01-643569

Tadmor, T., Aviv, A., & Polliack, A. (2011). Merkel cell carcinoma, chronic lymphocytic leukemia and other lymphoproliferative disorders: An old bond with possible new viral ties. *Annals of Oncology*, *22*(2), 250–256. https://doi.org/10.1093/annonc/mdq308

Tai, Y., Sakaida, Y., Kawasaki, R., Kanemaru, K., Akimoto, K., Brombacher, F., Ogawa, S., Nakamura, Y., & Harada, Y. (2023). Foxp3 and Bcl6 deficiency synergistically induces spontaneous development of atopic dermatitis-like skin disease. *International Immunology*, *35*(9), 423–435. https://doi.org/10.1093/intimm/dxad018

Tan, C. S., Dezube, B. J., Bhargava, P., Autissier, P., Wüthrich, C., Miller, J., & Koralnik, I. J. (2009). Detection of JC virus DNA and proteins in the bone marrow of HIV-positive and HIV-negative patients: Implications for viral latency and neurotropic transformation. *Journal of Infectious Diseases*, *199*(6), 881–888. https://doi.org/10.1086/597117

Tang, E. S., & Martin, A. (2007). Immunoglobulin gene conversion: Synthesizing antibody diversification and DNA repair. *DNA Repair*, *6*(11), 1557–1571. https://doi.org/10.1016/j.dnarep.2007.05.002

Tarsalainen, A., Maman, Y., Meng, F.-L., Kyläniemi, M. K., Soikkeli, A., Budzyńska, P., McDonald, J. J., Šenigl, F., Alt, F. W., Schatz, D. G., & Alinikula, J. (2022). Ig Enhancers Increase RNA Polymerase II Stalling at Somatic Hypermutation Target Sequences. *The Journal of Immunology*, *208*(1), 143–154. https://doi.org/10.4049/jimmunol.2100923

Taub, R., Kirsch, I., Morton, C., Lenoir, G., Swan, D., Tronick, S., Aaronson, S., & Leder, P. (1982). Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(24 I), 7837–7841. https://doi.org/10.1073/pnas.79.24.7837

Teater, M., & Melnick, A. (2017). Untangling the Web of Lymphoma Somatic Mutations. *Cell*, *171*(2), 270–272. https://doi.org/10.1016/j.cell.2017.09.031

Testa, J. R., Carbone, M., Hirvonen, A., Khalili, K., Krynska, B., Linnainmaa, K., Pooley, F. D., Rizzo, P., Rusch, V., & Xiao, G. H. (1998). A multi-institutional study confirms the presence and expression of simian virus 40 in human malignant mesotheliomas. *Cancer Research*, *58*(20), 4505–4509.

Thandra, K., Barsouk, A., Saginala, K., Padala, S., Barsouk, A., & Prashanth, R. (2021). Epidemiology of non-Hodgkin's lymphomas. *Medical Sciences*, *9*(5). https://doi.org/10.1159/000411699

Thanh, T. D., Van Tho, N., Lam, N. S., Dung, N. H., Tabata, C., & Nakano, Y. (2016). Simian virus 40 may be associated with developing malignant pleural mesothelioma. *Oncology Letters*, *11*(3), 2051–2056. https://doi.org/10.3892/ol.2016.4174

Tognon, M., Luppi, M., Corallini, A., Taronna, A., Barozzi, P., Rotondo, J. C., Comar, M., Casali, M. V., Bovenzi, M., D'Agostino, A., Vinante, F., Rigo, A., Ferrarini, I., Barbanti-Brodano, G., Martini, F., & Mazzoni, E. (2015). Immunologic evidence of a strong association between non-Hodgkin lymphoma and simian virus 40. *Cancer*, *121*(15), 2618–2626. https://doi.org/10.1002/cncr.29404

Torres, C. (2020). Evolution and molecular epidemiology of polyomaviruses. *Infection, Genetics and Evolution*, *79*(December 2019), 104150. https://doi.org/10.1016/j.meegid.2019.104150

Vaigot, P., Pisani, A., Darmon, Y. M., & Ortonne, J. P. (1987). The majority of epidermal Merkel cells are non-proliferative: A quantitative immunofluorescence analysis. *Acta Dermato-Venereologica*, *67*(6), 517–541. https://doi.org/10.2340/0001555567517520

van der Meijden, E., Janssens, R. W. A., Lauber, C., Bavinck, J. N. B., Gorbalenya, A. E., & Feltkamp, M. C. W. (2010). Discovery of a new human polyomavirus associated with Trichodysplasia Spinulosa in an immunocompromized patient. *PLoS Pathogens*, *6*(7), 1–10. https://doi.org/10.1371/journal.ppat.1001024

Vartanian, J., Guétard, D., Henry, M., & Wain-Hobson, S. (2008). Evidence for Editing of Human Papillomavirus DNA by APOBEC3 in Benign and Precancerous Lesions. *Science*, *230*(April), 230–233.

Verhalen, B., Starrett, G. J., Harris, R. S., & Jiang, M. (2016). Functional Upregulation of the DNA Cytosine Deaminase APOBEC3B by Polyomaviruses. *Journal of Virology*, *90*(14), 6379–6386. https://doi.org/10.1128/jvi.00771-16

Vieira, V. C., Leonard, B., White, E. A., Starrett, G. J., Temiz, N. A., Lorenz, L. D., Lee, D., Soares, M. A., Lambert, P. F., Howley, P. M., & Harris, R. S. (2014). Human papillomavirus E6 triggers upregulation of the antiviral and cancer genomic DNA deaminase APOBEC3B. *MBio*, *5*(6), 1–8. https://doi.org/10.1128/mBio.02234-14

Vilchez, R., & Butel, J. (2003). SV40 in human brain cancers and non-Hodgkin's lymphoma. *Oncogene*, *22*(33), 5164–5172. https://doi.org/10.1038/sj.onc.1206547

Vilchez, R., Madden, C., Kozinetz, C., Halvorson, S., White, Z., Jorgensen, J., Finch, C., & Butel, J. (2002). Association between simian virus 40 and non-Hodgkin lymphoma. *Lancet*, *359*(9309), 817–823. https://doi.org/10.1016/S0140-6736(02)07950-3

Viscidi, R. P., Rollison, D. E. M., Viscidi, E., Clayman, B., Rubalcaba, E., Daniel, R., Major, E. O., & Shah, K. V. (2003). Serological cross-reactivities between antibodies to simian virus 40, BK virus, and JC virus assessed by virus-like-particle-based enzyme immunoassays. *Clinical and Diagnostic Laboratory Immunology*, *10*(2), 278–285. https://doi.org/10.1128/CDLI.10.2.278-285.2003

Wang, Q., Oliveira, T., Jankovic, M., Silva, I. T., Hakim, O., Yao, K., Gazumyan, A., Mayer, C. T., Pavri, R., Casellas, R., Nussenzweig, M. C., & Robbiani, D. F. (2014). Epigenetic targeting of

activation-induced cytidine deaminase. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(52), 18667–18672. https://doi.org/10.1073/pnas.1420575111

Wang, X., Fan, M., Kalis, S., Wei, L., & Scharff, M. D. (2014). A source of the single-stranded DNA substrate for activation-induced deaminase during somatic hypermutation. *Nature Communications*, *5*(May), 1–11. https://doi.org/10.1038/ncomms5137

Wang, X., Li, Z., Naganuma, A., & Ye, B. H. (2002). Negative autoregulation of BCL-6 is bypassed by genetic alterations in diffuse large B cell lymphomas. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(23), 15018–15023. https://doi.org/10.1073/pnas.232581199

Wang, Y., Zhang, S., Yang, X., Hwang, J. K., Zhan, C., Lian, C., Wang, C., Gui, T., Wang, B., Xie, X., Dai, P., Zhang, L., Tian, Y., Zhang, H., Han, C., Cai, Y., Hao, Q., Ye, X., Liu, X., … Meng, F. L. (2023). Mesoscale DNA feature in antibody-coding sequence facilitates somatic hypermutation. *Cell*, *186*(10), 2193-2207.e19. https://doi.org/10.1016/j.cell.2023.03.030

Wang, Z., Wakae, K., Kitamura, K., Aoyama, S., Liu, G., Koura, M., Monjurul, A. M., Kukimoto, I., & Muramatsu, M. (2014). APOBEC3 Deaminases Induce Hypermutation in Human Papillomavirus 16 DNA upon Beta Interferon Stimulation. *Journal of Virology*, *88*(2), 1308–1317. https://doi.org/10.1128/jvi.03091-13

Watabe, R., & Nakamura, M. (2013). Expression of activation-induced cytidine deaminase in Merkel cell carcinoma with lymph-node metastasis. In *European journal of dermatology : EJD* (Vol. 23, Issue 4, pp. 539–540). https://doi.org/10.1684/ejd.2013.2078

Watanabe, S., & Yoshiike, K. (1985). Decreasing the number of 68-base-pair tandem repeats in the BK virus transcriptional control region reduces plaque size and enhances transforming capacity. *Journal of Virology*, *55*(3), 823–825. https://doi.org/10.1128/jvi.55.3.823-825.1985

Weinberg, R. A. (1995). The retinoblastoma protein and cell cycle control. *Cell*, *81*(3), 323–330. https://doi.org/10.1016/0092-8674(95)90385-2

Willmann, K. L., Milosevic, S., Pauklin, S., Schmitz, K. M., Rangam, G., Simon, M. T., Maslen, S., Skehel, M., Robert, I., Heyer, V., Schiavo, E., Reina-San-Martin, B., & Petersen-Mahrt, S. K. (2012). A role for the RNA pol II-associated PAF complex in AID-induced immune diversification. *Journal of Experimental Medicine*, *209*(11), 2099–2111. https://doi.org/10.1084/jem.20112145

Wright, C. A., Nance, J. A., & Johnson, E. M. (2013). Effects of Tat proteins and Tat mutants of different human immunodeficiency virus type 1 clades on glial JC virus early and late gene transcription. *Journal of General Virology*, *94*(PART3), 514–523. https://doi.org/10.1099/vir.0.047902-0

Wu, L., Shukla, V., Yadavalli, A. D., Dinesh, R. K., Xu, D., Rao, A., & Schatz, D. G. (2022). HMCES protects immunoglobulin genes specifically from deletions during somatic hypermutation. *Genes and Development*, *36*(7), 433–450. https://doi.org/10.1101/gad.349438.122

Xu, Z., Zan, H., Pone, E. J., Mai, T., & Casali, P. (2012). Immunoglobulin class-switch DNA recombination: Induction, targeting and beyond. *Nature Reviews Immunology*, *12*(7), 517–531. https://doi.org/10.1038/nri3216

Yang, J. F., & You, J. (2020). Regulation of polyomavirus transcription by viral and cellular factors. *Viruses*, *12*(10), 1–18. https://doi.org/10.3390/v12101072

Ye, B. H., Cattoretti, G., Shen, Q., Zhang, J., Hawe, N., Waard, R. De, Leung, C., Nouri-shirazi, M., Orazi, A., Chaganti, R. S. K., Rothman, P., Stall, A. M., Pandolfi, P., & Dalla-Favera, R. (1997). The BCL-6 proto-oncogene controls germinal-centre formation and Th2-type inflammation. *Nature Genetics*, *16*.

Yewdell, W. T., Kim, Y., Chowdhury, P., Lau, C. M., Smolkin, R. M., Belcheva, K. T., Fernandez, K. C., Cols, M., Yen, W. F., Vaidyanathan, B., Angeletti, D., McDermott, A. B., Yewdell, J. W., Sun, J. C., & Chaudhuri, J. (2020). A Hyper-IgM Syndrome Mutation in Activation-Induced Cytidine Deaminase Disrupts G-Quadruplex Binding and Genome-wide Chromatin Localization. *Immunity*, *53*(5), 952-970.e11. https://doi.org/10.1016/j.immuni.2020.10.003

Ying, C. Y., Dominguez-Sola, D., Fabi, M., Lorenz, I. C., Hussein, S., Bansal, M., Califano, A., Pasqualucci, L., Basso, K., & Dalla-Favera, R. (2013). MEF2B mutations lead to deregulated

expression of the oncogene BCL6 in diffuse large B cell lymphoma. *Nature Immunology*, *14*(10), 1084–1092. https://doi.org/10.1038/ni.2688

Yu, G., Zhang, Y., Gupta, V., Zhang, J., MacCarthy, T., Duan, Z., & Scharff, M. D. (2021). The role of HIRA-dependent H3.3 deposition and its modifications in the somatic hypermutation of immunoglobulin variable regions. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(50), 1–11. https://doi.org/10.1073/pnas.2114743118

Yu, K., Chedin, F., Hsieh, C. L., Wilson, T. E., & Lieber, M. R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nature Immunology*, *4*(5), 442–451. https://doi.org/10.1038/ni919

Zheng, H. C., Xue, H., Jin, Y. Z., Jiang, H. M., & Cui, Z. G. (2022). The Oncogenic Effects, Pathways, and Target Molecules of JC Polyoma Virus T Antigen in Cancer Cells. *Frontiers in Oncology*, *12*(March), 1–10. https://doi.org/10.3389/fonc.2022.744886

Zheng, S., Vuong, B. Q., Vaidyanathan, B., Lin, J. Y., Huang, F. T., & Chaudhuri, J. (2015). Non-coding RNA Generated following Lariat Debranching Mediates Targeting of AID to DNA. *Cell*, *161*(4), 762–773. https://doi.org/10.1016/j.cell.2015.03.020

Zhou, X., Zhu, C., & Li, H. (2023). BK polyomavirus: latency, reactivation, diseases and tumorigenesis. *Frontiers in Cellular and Infection Microbiology*, *13*(September), 1–9. https://doi.org/10.3389/fcimb.2023.1263983

zur Hausen, A., Rennspiess, D., Winnepenninckx, V., Speel, E. J., & Kurz, A. K. (2013). Early B-Cell differentiation in merkel cell carcinomas: Clues to cellular ancestry. *Cancer Research*, *73*(16), 4982–4987. https://doi.org/10.1158/0008-5472.CAN-13-0616

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

ISBN 978-952-02-0098-5 PRINT
ISBN 978-952-02-0099-2 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)

Painosalama, Turku, Finland 2025