

Charles University

Faculty of Science

Molecular and Cellular Biology, Genetics and Virology



Mgr. Martina Slavková

Virus-host cell interactions: Examples of gammaretroviral transcription control and somatic hypermutation targeting of SV40 enhancer

Interakce viru a hostitelské buňky: příklady z kontroly gamaretrovirové transkripce a cílení somatických hypermutací pomocí SV40 enhanceru

Doctoral thesis

Supervisor: RNDr. Jiří Hejnar, CSc.

Consultants: RNDr. Filip Šenigl, Ph.D.; Mgr. Dalibor Miklík, Ph.D.

Laboratory of Viral and Cellular Genetics

Institute of Molecular Genetics of the ASCR, v. v. i.

Prague, 2024

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

Declaration:

I hereby certify that I have written this thesis independently and that I have not used other than the cited sources. This thesis has not been submitted for any other degree or purposes.

November 2024, Prague

.....
Mgr. Martina Slavková

Acknowledgement

Looking back on the years I spend in 'Hejnar's lab', this unexpectedly long and at times bumpy journey has brought many precious moments. As James Norbury writes in his book 'Big Panda and Tiny Dragon', the most important part is not the journey or the destination, but the company. This work would not be possible without many people.

First, I would like to thank all the members, current and former, of the Laboratory of viral and cellular genetics at the Institute of molecular genetics of the Czech academy of sciences, namely to our boss Jiří Hejnar for the opportunity to work in his lab, to Filip Šenigl and Dalibor Miklík for guiding me through the maze of science, and to Dana Kučerová, Ľubomíra Pecnová, and Markéta Reinišová for giving me a helping hand when needed.

Next, I would like to thank my family, who may not understand my work, but are always there to support me.

This study was funded by the Czech Academy of Sciences (Praemium Academiae Award 2018) and the project National Institute of Virology and Bacteriology (Programme EXCELES, ID Project No. LX22NPO5103)-Funded by the European Union-Next Generation EU, the Czech Science Foundation (project 22-30384S), the Finnish Cultural Foundation, the Sigrid Juselius Foundation, the Turku University Foundation, the Finnish Cultural Foundation Kymenlaakso Regional Fund, the Alfred Kordelin Foundation, Cancer Society of Southwest Finland, National Institutes of Health (R01 AI 127642), and also institutional support from the project RVO (68378050).

Abstrakt

Koexistence virů a jejich hostitelských buněk dala vzniknout spletité síti interakcí. Pouze s dobrým porozuměním můžeme tyto interakce elegantně a účinně využít ve výzkumu a k terapeutickým účelům.

Během svého životního cyklu vyžadují retroviry (RV) integraci svého genomu do genomu hostitele. Integrační místo je jedním z faktorů ovlivňujících RV expresi. Integrovaný RV může být stabilně exprimován, epigeneticky umlčen nebo latentní. Již dříve bylo prokázáno, že virus myší leukémie (MLV) je stabilně exprimován po integraci do promotorů a enhancerů aktivních genů. Zde ukazujeme, že tato dlouhodobá stabilní exprese není specifická pouze pro MLV, ale je obecnou charakteristikou gama-RV. Co je ale nejdůležitější, poté, co gama-RV etablojí stabilní expresi, zůstanou stabilně exprimovány i po přesměrování integrací a cílených inzercí do represivního prostředí s laminou asociovaných domén.

Dalším příkladem interakce mezi virem a hostitelem je zkrácení velkého T antigenu (LTA_g) polyomavirů způsobené mutacemi vyvolaných deaminací cytidinu. Ukázali jsme, že enhancer SV40 cílí somatické hypermutace (SHM) a aktivací indukovanou deaminázou (AID) způsobené SHM mohou vést ke zkrácení LTA_g. Zkrácené varianty LTA_g polyomaviru Merkelových buněk jsou tumorogenní a podílejí se na vzniku karcinomu Merkelových buněk. Ačkoliv infekce virem SV40 nebyla dosud přesvědčivě spojena se vznikem nádorů u lidí, zkrácený LTA_g viru SV40 byl nalezen v transformovaných buňkách.

Tyto příklady poukazují na širokou škálu mechanismů (genetických a epigenetických), kterými viry interagují s hostitelem, a kterých lze využít například při návrhu retrovirových vektorů a výzkumu rakoviny.

Klíčová slova: gamaretrovirus, integrační místo, cílený knock-in, genová exprese; SV40, velký T antigen, AID, somatické hypermutace

Abstract

The coexistence of viruses and their host cells has given rise to an intricate web of interactions. Only with a good understanding of these interactions can we progress to smarter and more efficient deployment of viral properties in research and therapeutic fields.

During their life cycle, retroviruses (RV) require integration of their genome into the host genome. The integration site is one of the factors determining RV expression and may also act negatively towards epigenetic silencing of the provirus and latency. The murine leukemia virus (MLV) was previously shown to be stably expressed upon integration into promoters and enhancers of active genes. Here we show that this long-term stability is not specific to MLV, but it is a general characteristic of gamma-RV. Most importantly, after they establish stable expression, they remain stably expressed even after integration retargeting and targeted insertions in the repressive environment of lamina-associated domains.

Another example of virus-host interaction is the truncation of large T antigen (LTA_g) of polyomaviruses caused by cytidine deamination-induced mutations. We showed that SV40 enhancer possesses somatic hypermutation (SHM) targeting activity and activation induced deaminase (AID)-induced SHM may lead to truncated LTA_g. Such truncated variants of LTA_g have been proven as tumorigenic in Merkel cell polyomavirus causing Merkel cell carcinoma. Although SV40 infection has not been persuasively linked to human cancer, truncated SV40 LTA_g was found in transformed cells.

These examples show the wide range of mechanisms (genetic and epigenetic) by which viruses interact with the host and which can be used, for example, in the design of retroviral vectors or in cancer research.

Keywords: gammaretrovirus, integration site, targeted knock-in, gene expression; SV40, large T antigen, AID, somatic hypermutation

Contents

Abbreviations	9
1. Literature overview	14
1.1 Retroviruses	14
1.1.1 Introduction of retroviruses	14
1.1.2 Retroviral classification	15
1.1.3 Impact of retroviral infections on humans and animals	15
1.1.4 Integration preferences of retroviruses	16
1.1.5 Retargeting of proviral integrations	17
1.1.6 Targeted knock-in	20
1.1.7 Integration site-dependent proviral expression	21
1.2. Polyomaviruses	23
1.2.1 Introduction of polyomaviruses	23
1.2.2 Classification and representatives of polyomaviruses	23
1.2.3 Large T antigen	23
1.2.4 Polyomavirus-induced tumorigenesis	24
1.3 AID-mediated somatic hypermutations	26
1.4 Seizing control of viral properties	28
1.5 Functional organization of the genome	29
2. Aims	30
3. Materials and methods	31
3.1 Benchtop methods	31
3.2 Cell-culture methods	39
3.3 Bioinformatic methods	42
4. Results	44
4.1 Gammaretroviral vectors remain transcriptionally stable even after retargeting and targeted insertion in the cells of different origin	44
4.1.1 Stable expression of MLV-derived vectors with retargeting Bin INs	44
4.1.2 Retargeted integration sites are present in stably expressed population	46

4.1.3 Stable expression after retargeted integration is a general feature of gammaretroviruses.....	49
4.1.4 Stability of expression almost randomly distributed γ RV promoters in the context of α RV SIN vectors	51
4.1.5 Gammaretroviruses display stable expression even after targeted insertion into LADs	53
4.1.6 FeLV-derived vector's targeted knock-in into the Jurkat cell line	56
4.2 The SV40 enhancer as an AID-mediated somatic hypermutations targeting element	58
4.2.1 SV40 enhancer possesses SHM targeting activity in the B cell lines	58
4.2.2 Enhancer targeted SHM activity is retained in non-B cells	62
4.2.3 Truncated large T antigen of SV40 as a result of mutations in SV40 LT region in the B and non-B cell lines	65
5. Discussion	68
6. Conclusions	72
Contribution to the publications.....	73
Oligo tables	74
References	82

Abbreviations

AAVS1	adeno-associated integration site 1
AID	activation-induced cytidine deaminase
AIDS	acquired immunodeficiency syndrome
ALLINI	allosteric IN inhibitor
APE	apurinic/aprimidinic endonucleases
APOBEC	apolipoprotein B mRNA editing enzyme, catalytic polypeptide
AS	ASLV-derived SIN vectors
ASLV	avian sarcoma leukosis virus
BET	bromodomain and extra-terminal domain
Bin	BET-independent
BKPyV	BK polyomavirus
BLV	bovine leukemia virus
BsR	blasticidin resistance
CA	capsid
CAR	chimeric antigen receptor
Cas9	CRISPR-associated protein 9
CBX1	chromodomain of heterochromatin-binding protein 1 β
CCD	core catalytic domain
CCR5	CC motif chemokine receptor 5
CDYL	chromodomain of Y-like protein
ChIP	chromatin immunoprecipitation
CN	copy number
CPSF6	cleavage and polyadenylation specificity factor 6
CrERV	cervid endogenous gammaretrovirus
CRISPR	clustered regularly interspaced short palindromic repeat

Ct	C-terminal tail
CTD	C-terminal domain
d2GFP	GFP fused to the destabilization domain
dC	deoxycytidine
dCt	deleted C-terminal tail
ddPCR	droplet digital PCR
DDR	DNA damage response
DIVAC	diversification activator
DoT	domain target
dpi	days post infection
dpn	days post nucleofection
dT	deoxythymidine
dU	deoxyuridine
E2	tethering domain of the human papilloma virus E2 protein
EGFP	enhanced green fluorescence protein
Enh	enhancer
ET	extra-terminal
FACT	facilitates chromatin transcription
FAIRE	formaldehyde-assisted isolation of regulatory elements
FeLV	feline leukemia virus
gRNA	guide RNA
H3K36me2/3	di- and tri-methylation of lysine 36 on histone H3
HDGF	hepatoma-derived growth factor
HIV	human immunodeficiency virus
HRP-2	HDGF family-related protein 2
HTLV-1	human T-cell leukemia virus 1

HTS	hypermuation target sequence
IBD	integrase binding domain
Ig	immunoglobulin
IgHi	Ig heavy intronic
IL2RG	IL-2 receptor common γ -chain gene
IN	integrase
INLAI	IN-LEDGF/p75 allosteric inhibitor
IP	immunoprecipitation
IPEN	mutated IRF-Ets, e-boxes and NF κ B
IRES	internal ribosome entry site
JCPyV	JC polyomavirus
KoRV	koala retrovirus
KRAB	Krüppel-associated box
LAD	lamina-associated domain
LANA	N-terminal end of Kaposi sarcoma's latency associated nuclear antigen
LEDGF	lens epithelium-derived growth factor
LEDGIN	LEDGF/p75-IN inhibitor
LTA _g	large T antigen
LTR	long terminal repeat
MA	matrix
MCC	Merkel cell carcinoma
MCPyV	Merkel cell polyomavirus
MFI	mean fluorescence intensity
MINI	multimeric IN inhibitor
MLV	murine leukemia virus
MOI	multiplicity of infection

MoMLV	Moloney MLV
MoMSV	Moloney murine sarcoma virus
mut	mutation
NC	nucleocapsid
NCRR	non-coding regulatory region
OCT4	octamer-binding transcription factor 4
PIC	pre-integration complex
PP2A	protein phosphatase 2A
pRb	retinoblastoma protein
PyV	polyomavirus
RH	RNaseH
RSV	Rous sarcoma virus
RT	reverse transcriptase
RV	retrovirus
SFFV	spleen focus-forming virus
SHM	somatic hypermutation
SIN	self-inactivating
SIV	simian immunodeficiency virus
SNV	spleen necrosis virus
SRLV	small ruminant lentivirus
SV40	simian vacuolating virus 40
SV40e	SV40 enhancer
T2A	self-cleaving T2A peptide
TALEN	transcription activator-like effector nuclease
TFBS	transcription factor binding site
TRAC	T-cell receptor α constant

TRIM28	tripartite motif containing 28
Tss	transcription start site
U3	3' unique
UNG	uracil DNA glycosylase
VP	viral protein
VSV-G	vesicular stomatitis virus G glycoprotein
W390A	tryptophan 390 to alanine
WPRE	woodchuck hepatitis virus posttranscriptional regulatory element
wt	wild-type
ZFN	zinc-finger nuclease
ZFP	zinc-finger protein

1. Literature overview

1.1 Retroviruses

1.1.1 Introduction of retroviruses

Retroviruses are enveloped RNA viruses of the *Retroviridae* family. After fusion of viral and cellular membranes, a capsid enters cytoplasm. Every capsid carries two copies of the viral genome. Both copies serve as a template for reverse transcription, which is initiated in cytoplasm, and give rise to single double-stranded DNA. Retroviral DNA is transported into the nucleus, where it is integrated into the host genome. Integration is accomplished by viral enzyme, integrase (IN), and is an essential step of the retroviral life cycle. Integrated retroviral DNA, so called provirus, may then be transcribed by the host RNA polymerase II, so the transcripts have 5' caps and 3' polyA sequences like host mRNAs. After translation of viral RNAs into structural proteins, enzymes, and envelope proteins, genomic RNAs, unspliced transcripts, are packaged into new viral particles. Assembled particles are then released from the infected cells, and after maturation, they can undergo another cycle of infection. Schematic depiction of the retroviral life cycle is in Figure 1.

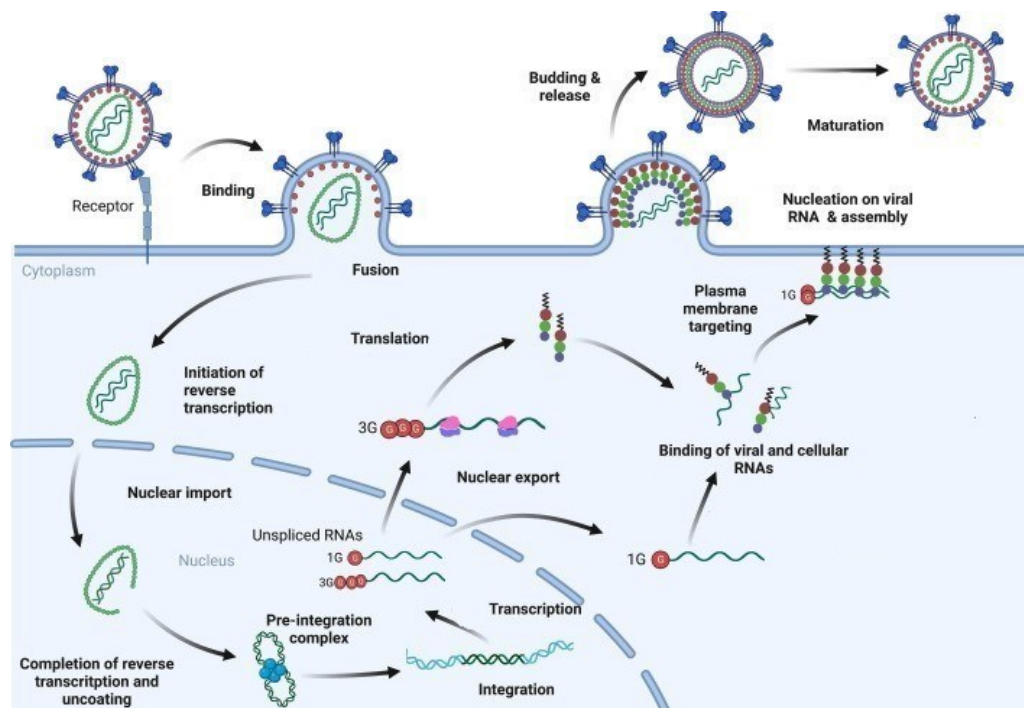


Figure 1. Retroviral life cycle. First, the virus binds to cell-surface enhancers, the viral membrane fuse with cellular membrane, and the viral capsid gets into cytoplasm. After initiation of reverse transcription in cytoplasm, the viral capsid is transported into nucleus, where reverse transcription is completed and capsid is uncoated. Viral DNA with bound integrase, the preintegration complex, is tethered to distinct genomic sites, when the viral DNA is integrated into the host genome by activity of integrase. Integrated viral DNA, so-called proviruses, then may be transcribed. Viral mRNA is transported from the nucleus, translated and viral proteins bind genomic RNA. After new viral particles are assembled on plasmatic membrane, they might be released, then viral particles mature. The figure was adapted from Chameettachal et al (2023).

The proviral genome consists of the identical long terminal repeats (LTRs) which flank the internal coding sequence. The 5'LTR acts as a promoter and enhancer, while the 3' LTR acts as a transcription terminator. The coding part of the provirus comprises of *gag*, structural proteins of matrix (MA), capsid (CA), and nucleocapsid (NC), *pro*, the protease, *pol*, proteins with enzymatic activities like the reverse transcriptase (RT), RNaseH (RH), and IN, and *env*, envelope proteins. More complex retroviruses also encode accessory genes.

1.1.2 Retroviral classification

The *Retroviridae* family contains two subfamilies, *Orthoretrovirinae* and *Spumaretrovirinae*. Retroviruses of *Orthoretrovirinae* subfamily are further divided into six genera: *Alpha-* (α), *Beta-* (β), *Gamma-* (γ), *Delta-* (δ), *Epsilon-* (ϵ) -retrovirus (RV), and *Lentivirus*.

Among α RVs belong the first ever-discovered retrovirus Rous sarcoma virus (RSV), which is a part of the avian sarcoma leukosis viruses (ASLV) group. This genus mainly contains viruses infecting birds.

The most studied γ RV is the murine leukemia virus (MLV), a causing agent of leukemia in mice. Another studied γ RVs in this thesis are the feline leukemia virus (FeLV), the koala retrovirus (KoRV), the spleen focus-forming virus (SFFV), the spleen necrosis virus (SNV), and the cervid endogenous gammaretrovirus (CrERV).

The δ RV genus is represented by the human T-cell leukemia virus 1 (HTLV-1) or the bovine leukemia virus (BLV).

The *Lentivirus* genera contain the simian immunodeficiency virus (SIV), which gave rise to the well-known human immunodeficiency virus 1 (HIV-1) and HIV-2, the causing agents of acquired immunodeficiency syndrome (AIDS), or the small ruminant lentiviruses (SRLV).

1.1.3 Impact of retroviral infections on humans and animals

Retroviruses are causing agents of many pathologies in humans and animals and pose significant burden for health care systems, e.g. HTLV-1 and HIV, financial losses in raising companion or farm animals, e.g. FeLV, ALV, SRLV, and BLV, or obstacles in conservation, e.g. KoRV.

In 2012, it was estimated that 5-10 million people worldwide are infected with the HTLV-1, however, there is not enough data from highly populated regions, so the number of infected people might be much higher (Gessain and Cassar 2012).

According to published data from the Czech National Institute of Public Health, since 2012 the increase of the newly diagnosed cases of HIV+ is more than 200 people (253 in 2023), predominantly men (84.5% in 2023), every year. In total, at the end of the year 2023, there were 4619 HIV+ people living in the Czech Republic. Worldwide, it was estimated that there are 39.9 million HIV+ people at the end of 2023 with 1.3 million new cases in 2023 (<https://www.hiv.gov/>).

In different regions, 1-20% of cats are infected with the FeLV. Infected cats may develop various FeLV-associated diseases (Parisi et al. 2023). Currently, there is no treatment available, only isolation from other cats or vaccination provide protection.

Significant financial losses in animal farming are caused by increased mortality rates after ALV infection of poultry (reviewed in Fandiño et al. 2023), decreased yields in milk production in SRLV positive sheep and goats (Davies et al. 2023), or increased mortality rates after BLV infection in cattle (reviewed in Lv et al. 2024).

In an example of wild animals or animals held at captivity at zoos, KoRV infection represents threat to health and conservation of koalas (*Phascolarctos cinereus*), which are listed as an endangered species. Increased expression of the KoRV is associated with increase of

leukemia, lymphomas and chlamydial disease in infected animals in comparison to healthy individuals (Tarlinton et al. 2005; W. Xu et al. 2013).

1.1.4 Integration preferences of retroviruses

Integration of retroviral DNA is an essential step in the retroviral life cycle. It is mediated by the viral IN, which creates a complex with retroviral DNA, so called pre-integration complex (PIC). The process of integration into the host genome is then a two-step reaction involving 3' end processing of viral DNA, when two terminal bases are removed from both 3' ends, and strand transfer, which makes a provirus an indivisible part of the host genome (Engelman et al. 1991; Brown et al. 1989).

Integrations are not completely random. Different retroviruses exhibit distinct global and local integration preferences. First, PICs are tethered to distinct parts of the genome with integrations occurring near the site of tethering. This PIC tethering is mostly mediated by interaction between the IN and chromatin-binding proteins (Lewinski et al. 2006). At the site of tethering, local bending caused by DNA wrapping around histones (Müller and Varmus 1994; Pruss et al. 1994) and DNA lesions increase frequency of integrations (Senavirathne et al. 2023), while tightly packed chromatin obstructs integrations (Michieletto et al. 2019). On the sequence level, retroviral INs prefer or disfavor certain nucleotides at the cleavage site (Miklík et al. 2023).

The HIV preferentially integrates into transcription units of active genes (Mitchell et al. 2004; Elleder et al. 2002). Integration of HIV is orchestrated by interaction between several viral components and cellular cofactors. First, the viral CA protein binds to the cleavage and polyadenylation specificity factor 6 (CPSF6), which traffics reverse transcription complexes into nuclear speckles (Sowd et al. 2016; Achuthan et al. 2018) formed within transcriptionally-active gene-dense nuclear interior, where the reverse transcription is completed (Francis et al. 2020). Within nuclear speckles-associated domains region, the core catalytic domain (CCD) of the HIV-1 IN interacts with the integrase binding domain (IBD) located on the C-terminal end of the lens epithelium-derived growth factor (LEDGF)/p75 (Figure 2A) (G. Maertens et al. 2003; Vanegas et al. 2005; Cherepanov et al. 2004, 2003), although the N-terminal zinc binding domain of HIV-1 IN is also involved by increasing the affinity to LEDGF/p75 (G. Maertens et al. 2003). The LEDGF/p75 binds to the di- and trimethylated lysine 36 on the histone H3 (H3K36me_{2/3}), a histone modification associated with transcribed gene bodies with a peak towards 3' end of the gene (Bannister et al. 2005; Barski et al. 2007). In the absence of LEDGF/p75, the HIV IN binds to hepatoma-derived growth factor (HDGF) family-related protein 2 (HRP-2), which also contains the IBD, the HIV then integrates into promoters and near CpG islands with higher frequency (Shun et al. 2007). However, given integration site preferences of HIV-1, LEDGF/p75 is the main determinant of HIV-1 targeting. The MLV preferentially integrates into active transcription start sites (TSSs)/promoters and enhancers while disfavoring regions marked with the modifications associated with repressed chromatin (Mitchell et al. 2004; De Ravin et al. 2014). The highly conserved unstructured C-terminal tail (Ct) of the IN C-terminal domain (CTD) interacts with extra-terminal (ET) domain of bromodomain and ET domain (BET) proteins (Figure 2B) (Gupta et al. 2013; De Rijck et al. 2013; De Ravin et al. 2014). The BET proteins bind to acetylated histones through its bromodomain and facilitate tethering of γRV PIC to the sites enriched with the histone modification.

Almost random is the integration of the ASLV. The CTD of ASLV IN binds to the subunits SSRP1 and Spt16 of the facilitates chromatin transcription (FACT) complex (Winans et al. 2017), chromatin-remodeling complex (Orphanides et al. 1998). Both subunits are necessary

for effective integration, since SSRP1 stabilizes Spt16 (Safina et al. 2013). As a result, ASLV exhibits slight preferences for active chromatin (Mitchell et al. 2004). The δ RV HTLV-1 displays integration preferences similar to the ASLV (Derse et al. 2007). Its IN interacts with the B' subunit of serine/threonine protein phosphatase 2A (PP2A) (Maertens 2016), but the exact role of interaction between HTLV-1 IN and PP2A in the δ -retroviral life cycle remains unclear.

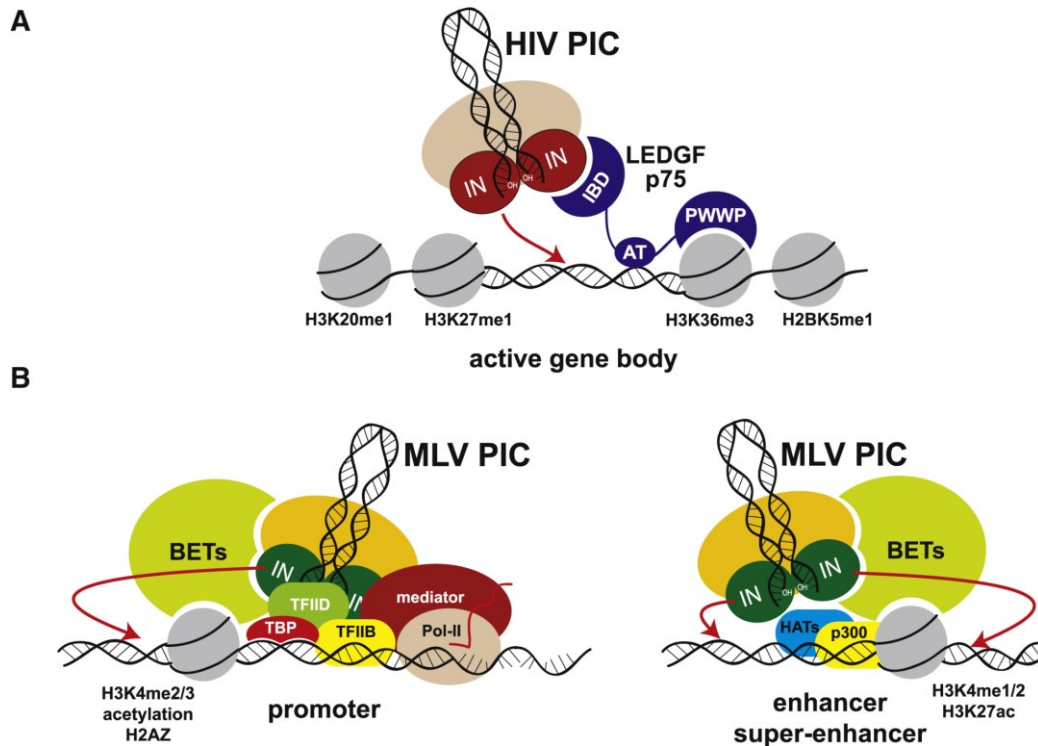


Figure 2. HIV and MLV IN binding partners involved in targeting of PIC. A) The HIV PIC by its IN binds to LEDGF/p75. In turn LEDGF/p75 binds to H3K36me3, histone modification enriched at gene bodies of transcriptionally active genes. Resulting interactions lead to preferential integration of HIV into transcription units of active genes. B) The MLV IN binds to BET proteins, which interact with acetylated histones. As a result, the MLV preferentially integrates into promoters and enhancers of active genes. The figure was adopted from Poletti and Mavilio (2018).

1.1.5 Retargeting of proviral integrations

Integration of the provirus may lead to cell-cycle dysregulation by insertional mutagenesis (reviewed in Bushman 2020), which is mediated by several mechanisms: gene activation by retroviral enhancer activity, gene activation by transcription regulated from retroviral promoter, gene activation by providing platform for termination of transcription, and gene inactivation by proviral integration.

Retargeting of proviruses from preferred transcriptionally active chromatin may provide protection against adverse effects of retroviral infections. With the knowledge of interactions involved in tethering of viral complexes within the infected cells, these interactions may be abrogated. Retargeting approaches involving the IN modifications or IN and non-IN inhibitors were exploited.

First, IN modifications involve fusion of various DNA binding domains to the N or C termini of the INs. The first manipulated IN was the HIV-1 IN when lambda repressor was fused to IN and integrations were dragged into lambda operators-containing target domains (Bushman 1994). Similarly a fusion of HIV-1 IN to the DNA binding domain of the zinc-finger protein zif268 led to increased frequency of integrations near zif268 binding sites (Bushman and Miller 1997).

Next, amino acids which are critical for the interactions between HIV and MLV INs and LEDGF/p75 and BET proteins, respectively, were identified. The replacement or deletion of these critical amino acids leads to altered integration profiles.

Identification of the MLV IN Ct domain as a BET protein binding domain (amino acids 381-408) (De Rijck et al. 2013) and the amino acid tryptophan 390 (W390) as a critical amino acid for interaction (El Ashkar et al. 2014) led to development of BET-independent (Bin) vectors (Figure 3). The INs with W390 replaced by alanine (IN^{W390A}) and deleted Ct (IN^{dCt}) both exhibited ~2-fold decrease in integration frequency near TSSs and CpG islands in comparison to the wild-type (wt) IN (IN^{WT}) (El Ashkar et al. 2014). Although disruption of interaction between the MLV IN and BET proteins leads to a more random pattern of integration, integrations still preferentially occur in open and transcriptionally active chromatin while disfavoring transcriptionally silent chromatin. Of note, even with differences in integration preferences, viruses produced with IN^{WT} or IN^{W390A} showed no differences in *in vivo* infectivity and exhibited similar pathology by developing T-cell lymphoma in mice (Nombela et al. 2022).

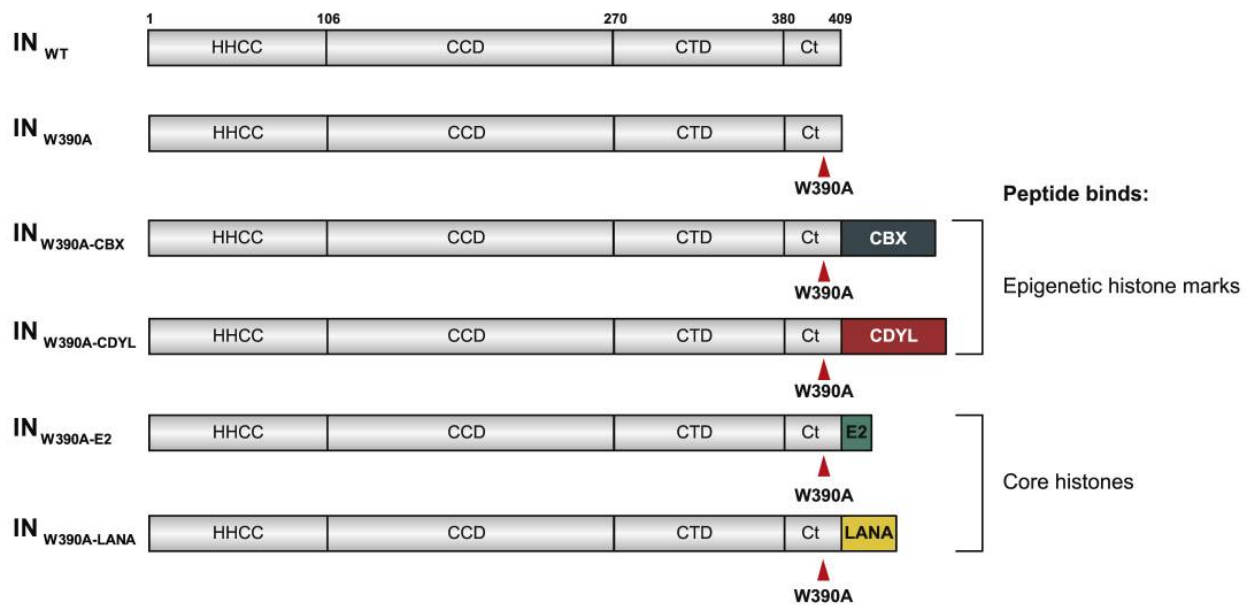


Figure 3. Schematic depiction of wild-type and retargeting, BET-independent, integrases. Mutation in MLV IN Ct leading to tryptophan replacement to alanine (W390A) is enough to disrupt integration between IN and BET proteins and result in more random integration pattern. Additional fusion of different domains to IN^{W390A} leads to further retargeting of integration sites. CBX (chromodomain of heterochromatin-binding protein 1 β) and CDYL (chromodomain of Y-like protein) domains bind to heterochromatin protein 1, E2 (tethering domain of the human papilloma virus E2 protein) and LANA (N-terminal end of Kaposi sarcoma's latency associated nuclear antigen) domains bind to histones 2A and 2B. The figure was adapted from El Ashkar et al (2017).

To further shift the integration preference from the natural integration sites, the retargeting approach by combining amino acid replacement and peptide fusion were utilized (El Ashkar et al. 2017). The IN^{W390A} was additionally fused with the chromodomain of heterochromatin-binding protein 1 β (IN^{CBX}) and chromodomain of Y-like protein (IN^{CDYL}), binding to heterochromatin protein 1, and the tethering domain of the human papilloma virus E2 protein (IN^{E2}) and the N-terminal end of Kaposi sarcoma's latency associated nuclear antigen (IN^{LANA}), binding histones 2A and 2B. While transduction with IN^{CDYL} and IN^{E2} vectors led to integration pattern similar to the IN^{W390} vector, the transduction with IN^{CBX} and IN^{LANA} vectors led to the ~4-fold and ~2-fold decrease of integrations near TSSs and CpG islands in comparison to IN^{WT} and IN^{W390}, respectively.

Mutation of the HIV-1 IN leading to single replacement K258R was shown to disrupt the HIV-1 natural integration preference with increased frequency of integrations into centromeric alpha satellite repeat sequences (Winans et al. 2022). It is noteworthy that similar integration pattern was observed after treatment with GSK878 inhibitor (Marquis et al. 2024), which binds the same CA binding site as inhibitor PF-74 (Gillis et al. 2023), competing with CPSF6 (Bhattacharya et al. 2014) and destabilizing the viral capsid which leads to premature uncoating (Bhattacharya et al. 2014; Shi et al. 2011).

It is noteworthy that all modifications might have adverse effects on the retroviral life cycle. Fusion of peptides to MLV IN Ct did not affect vector production or transduction (El Ashkar et al. 2017), but truncation of MLV IN, even though displaying integration retargeting, negatively affected the levels of viral production, transduction, and integration (El Ashkar et al. 2014). Apart from integration, IN binding to viral RNA is important for correct assembly of viral particles (Kessl et al. 2016; Elliott et al. 2020), truncated IN results in defective morphology and maturation of viral particles (Quillent et al. 1996), and interaction between IN and RT was shown to be important for reverse transcription (Wu et al. 1999).

Retargeting of proviral integrations may be also performed with inhibitors. In early stages of infection, interaction between HIV IN and LEDGF/p75 is inhibited by small allosteric IN inhibitors (ALLINIs), which are also called LEDGF/p75-IN inhibitors (LEDGINs), IN-LEDGF/p75 allosteric inhibitors (INLAIs), or multimeric IN inhibitors (MINIs) (Figure 4). Not only does the binding of inhibitor lead to retargeting of HIV (Vranckx et al. 2016), but it also causes conformational change of IN leading to blockage of catalytic activity of IN (Christ et al. 2012).

HIV proviruses may enter the latency stage, when replication-competent proviruses get transcriptionally silenced, but are prone to reactivation and cause re-emergence of detectable levels of viremia, causing hindrance to viral clearance in HIV-positive patients. ALLINIs-induced retargeted proviruses integrate with higher frequency into sites which showed decreased reactivation from latency after stimulation (Figure 4) (Vranckx et al. 2016). Retargeting of proviruses into integration sites resistant to reactivation from latency provide functional cure of infection by so-called block-and-lock strategy (Debyser et al. 2018), an alternative to shock-and-kill strategy, when reactivated proviruses in latently infected cells are dealt with by immune system and antiretroviral therapy (reviewed in Darcis et al. 2017).

In later stages of infection, ALLINIs enhanced IN multimerization leading to morphologically defective viral particles (Figure 4) (Desimmie et al. 2013; Jurado et al. 2013; Balakrishnan et al. 2013; Le Rouzic et al. 2013).

Other inhibitors utilization involves the BET inhibitors. Disruption of interaction between BET proteins and acetylated histones by BET inhibitors, by competing for acetylated histones-binding pocket on BET proteins, led to a drop of integration from 46% to 5% of integrations near to TSSs (Gupta et al. 2013).

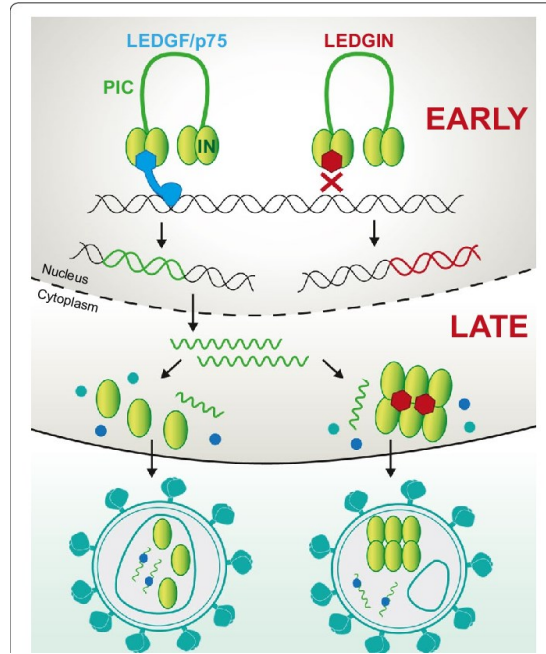


Figure 4. Schematic mechanisms of LEDGIN-mediated retargeting of HIV integration sites (early) and LEDGIN-mediated disruption of morphology of viral particles (late). In the early stage of infection, the HIV IN binding to LEDGF/p75 (blue) results in preferential integration of viral DNA into gene bodies of active genes (green provirus). LEDGIN (red) binds to LEDGF/p75 binding site on HIV IN, which leads to retargeting of HIV integrations (red provirus). Provirus are retargeted into reactivation resistant regions of the genome and thus remain silent. In the late stage of infection, LEDGINS bind multimers of IN together, which leads to impairment of viral particles morphology. The figure was adopted from Vansant et al (2019).

1.1.6 Targeted knock-in

With IN or IN-interacting proteins modification-retargeting approaches, there is still relatively low control over a precise site of integration. Integrations occur near a site of tethering. Current approaches of targeted insertions utilize double-strand breaks or single-strand nicks generation followed by DNA repair-mediated insertion of donor vectors.

Breaks and nicks might be generated by Cre recombinase at pre-inserted loxP sites (Sauer and Henderson 1990), the zinc-finger protein fused to the FokI nuclease (ZFNs) (Kim et al. 1996), transcription activator-like effector nucleases (TALENs) linked to the catalytic domain of FokI (Miller et al. 2011), the clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated protein 9 (Cas9) (Jinek et al. 2013), or the pair-nicking endonucleases (Chen et al. 2017).

Classical IN-based gene transfer of retroviral vectors face variegation of LTR-driven expression and safety issues. The transcriptionally permissive chromatin environment of active genes provides a platform for stable expression from donor vectors (Lombardo et al. 2007; Eyquem et al. 2017). Moreover, it allows insertion of promoterless vectors in open reading frame with target locus, when expression of the donor is under control of the target sites, resulting in physiological levels of transcription. Targeted insertions in active genes were reported for T-cell receptor α constant (*TRAC*) locus (Eyquem et al. 2017), CC motif chemokine receptor 5 (*CCR5*) locus (Lombardo et al. 2007, 2011; Rothemejer et al. 2023), IL-

2 receptor common γ -chain gene (*IL2RG*) locus (Lombardo et al. 2007), adeno-associated integration site 1 (*AAVS1*) locus (Lombardo et al. 2007; Eyquem et al. 2017), or octamer-binding transcription factor 4 (*OCT4*) locus (Hockemeyer et al. 2009).

Several applications were reported. Chimeric antigen receptor (CAR) was inserted into *TRAC* and *CCR5* loci (Eyquem et al. 2017; Rothemejer et al. 2023). CAR T-cell therapy involving modification of T cells leads to elimination of desired cells. Targeted insertion of the promoterless vector into *TRAC* locus led to homogenic expression of CAR and tumor eradication (Eyquem et al. 2017). Targeted insertion of CAR vector *CCR5* was utilized to engineer HIV-resistant CAR T cells resulting in lysis of HIV-infected T cells (Rothemejer et al. 2023).

Other applications include insertion of reporter gene into *OCT4* locus of human embryonic stem cells to monitor the state of pluripotency or insertion of donor vector into *AAVS1* site for overexpression of transfer gene (Hockemeyer et al. 2009).

HIV-1 LTR-equipped expression cassette was inserted by CRISPR-Cas9-induced homologous recombination-mediated targeted knock-in into HIV-1 sites with observed intact latent proviruses to study latent proviruses and their reactivation (Teixeira et al. 2024).

Depending on the donor vectors and target sites, the insertion may influence transcription of target and nearby genes by remodeling of the epigenetic features (Lombardo et al. 2011). Moreover, the insert may carry a cryptic splice site resulting in termination of transcription at the exogenous polyA (Lombardo et al. 2011). Therefore, it is necessary to carefully design vectors to ensure efficiency and safety of inserted vectors.

1.1.7 Integration site-dependent proviral expression

The ASLV exhibits almost random integration with slight preference for active chromatin (Mitchell et al. 2004). The ASLV is efficiently silenced in mammalian cells (Hejnar et al. 1994, 1999). The integration site plays a role in proviral transcriptional silencing (Figure 5) (Senigl et al. 2012). Proviruses integrated in gene bodies of active genes are methylated, mainly by the activity of *de novo* DNA methyltransferase 3B, and silenced. Proviruses integrated in intergenic regions are also silenced, but by the mechanism independent of DNA methylation. Only the proviruses which are integrated close to active TSSs, marked with tri-methylation of lysine 4 on histone H3, remain transcriptionally stable in mammalian cells. There is a correlation between the level of silencing and increased distance from TSSs of active genes. Insertion of Sp1 binding sites, which protect CpG islands against methylation, into ASLV LTRs strengthens the expression stability of ASLV resulting in maintaining stable proviral expression further from TSSs (Hejnar et al. 2001; Senigl et al. 2008; Šenigl et al. 2017).

The MLV is efficiently silenced in embryonic stem/carcinoma cells. The primer-binding site of the MLV is specifically recognized by Krüppel-associated box (KRAB) zinc-finger protein (ZFP) family, which is a huge family of proteins specifically binding to endogenous proviruses (Rowe et al. 2013). The KRAB-ZFP recognizing the MLV is called ZFP809 (Wolf and Goff 2009). The ZFP809 then binds to TRIM28 (tripartite motif containing 28), also called KAP1, TIF1, and KRIP1 (Friedman et al. 1996; Moosmann et al. 1996; Kim et al. 1996). TRIM28 is a scaffold protein that interacts with other proteins regulating gene silencing (Schultz et al. 2002; Sripathy et al. 2006; Karimi et al. 2011; Leung et al. 2011; Zuo et al. 2012).

The expression levels of ZFP809 are similar in embryonic and differentiated cells. However, in the differentiated cells, the ZFP809 is quickly ubiquitinated and degraded (Wang and Goff 2017). The MLV-derived proviruses after transduction with the IN^{WT} vector are transcriptionally stable in the K562 cell line for a long time (Miklík et al. 2018).

1.2 Polyomaviruses

1.2.1 Introduction of polyomaviruses

Viruses of the *Polyomaviridae* family are non-enveloped DNA viruses. A single copy of circular dsDNA is deposited in a capsid with icosahedral symmetry made of pentameric capsomers. After the cell entry, the capsid is partially disassembled in the cytosol, transported into the nucleus, where it is further disassembled. The viral genome remains in the cell as an episome, a small circular extrachromosomal molecule. The genome carries two transcriptional regions, the early and late, which are transcribed in the stages of viral infection they are referring to. The early region encodes regulatory proteins, so-called T antigens, and the late region typically encodes structural proteins. Transcription and replication are controlled from the non-coding regulatory region (NCRR), which is interspersed between the early and late region and contains the early and late promoters, the enhancer, and also the origin of replication.

1.2.2 Classification and representatives of polyomaviruses

According to International Committee on Taxonomy of Viruses the *Polyomaviridae* family is divided into eight genera: *Alpha-* (α), *Beta-* (β), *Delta-*, *Epsilon-*, *Eta-*, *Gamma-*, *Theta-*, *Zeta-polyomavirus* (PyV).

The best-characterized human polyomaviruses are the α -PyV Merkel cell polyomavirus (MCPyV; human polyomavirus 5), which is contributing to Merkel cell carcinoma (MCC), a highly aggressive human skin cancer, the β -PyV BK polyomavirus (BKPyV; human polyomavirus 1) associated with nephropathy, and the β -PyV JC polyomavirus (JCPyV; human polyomavirus 2) associated with progressive multifocal leukoencephalopathy.

In the human population, seroprevalence of various polyomaviruses is high. The first exposure to infection is usually during childhood. Most infections are asymptomatic. Pathologies tend to develop in elderly or immunocompromised individuals (Kean et al. 2009; Gossai et al. 2016). Another broadly studied PyVs are model α -PyV murine PyV and β -PyV simian vacuolating virus 40 (SV40; *Macaca mulatta* polyomavirus 1). The SV40 virus, which is of rhesus monkey origin, was first described by B. H. Sweet and M. R. Hilleman in 1960 as a contamination of poliovirus vaccines (Sweet and Hilleman 1960; Gerber 1962). Immunosuppressed monkeys develop pathologies in kidney or brain tissues (Horvath et al. 1992).

1.2.3 Large T antigen

The early region of polyomavirus genome gives rise to the T antigens, which are the results of alternative splicing and have regulatory functions, e.g. SV40 encodes the large T antigen (LTAg), the small T antigen, and the 17kT antigen. The LTAg contains several functional domains which were reported to be involved in deregulation of the cell cycle, inducing DNA damage response (DDR), viral replication, regulation of early and late transcription, or viral assembly.

The MCPyV and SV40 infections, by the C-terminus of LTAg, activate DDR and cell cycle arrest (Li et al. 2013; Hein et al. 2009), which were shown to be important for efficient replication. Activated DDR kinases prevent the formation of concatemers or linear dsDNA products by utilizing homology-directed DNA repair mechanisms (Sowd et al. 2014; Sowd et al. 2013).

During replication, hexamer of LTAg binds to the origin of DNA replication LTAg-binding sites and initiates structural changes. Due to helicase activity of LTAg, the replication bubble is formed (Borowiec and Hurwitz 1988b, 1988a; Peng and Acheson 1998), and then through the

interactions with LTA_g, the cellular replication machinery assembles (Tsurimoto et al. 1990; Collins and Kelly 1991; Simmons et al. 1996; Waga et al. 1994).

Although activation of DDR was shown to be necessary for efficient viral replication, DDR activated kinases also phosphorylate LTA_g which leads to inhibition of LTA_g helicase activity and hence disruption of viral replication (Homiski et al. 2024).

While LTA_g activates DDR and cell cycle arrest, the LTA_gs also bind to tumor-suppressor proteins p53 (McCormick and Harlow 1980) and retinoblastoma protein (pRb) (Dyson et al. 1990; DeCaprio et al. 1988), inhibit their functions (Dobbelstein and Roth 1998), and deregulate proliferation and apoptosis.

The helicase domain of LTA_g binds to the DNA binding domain of p53, leading to a conformational change of p53 and blocking DNA binding domain of p53, which blocks p53-induced transcription (Bargonetti et al. 1992; Lilyestrom et al. 2006; Kellogg et al. 2022). At the same time, the helicase activity of LTA_g hexamer is blocked by p53 binding (Lilyestrom et al. 2006).

It is noteworthy that early after SV40 infection, p53 may bind to T antigen promoter, competing with a transcription factor Sp1, and preventing viral transcription from early expression region (Drayman et al. 2016).

The pRb inhibition lies in binding and hence blocking transcription factor E2F (Chellappan et al. 1991), and interacting with histone deacetylases (Magnaghi-Jaulin et al. 1998). The LTA_g disrupts interaction between pRb and E2F (Chellappan et al. 1992).

During infection, there is a switch from early transcription region to late transcription region. This switch is ensured by several steps. First, in early stages of infection, the transcription from the late promoter is blocked by transcription repressors (Wiley et al. 1993). Then the LTA_g binds to LTA_g binding sites located within early expression part and blocking early region transcription (Hansen et al. 1981), and LTA_g activates the expression of the late genes (Keller and Alwine 1984), competing with the transcription repressor. Moreover, polyomaviruses encode microRNAs which negatively regulate expression of early transcripts (Sullivan et al. 2009; Seo et al. 2009).

1.2.4 Polyomavirus-induced tumorigenesis

Truncated LTA_g without helicase activity, therefore replication incompetent, is a signature of MCC (Shuda et al. 2008). Interaction with pRB is important for tumorigenesis (Spurgeon et al. 2022; Hesbacher et al. 2016). Not only is binding of the LTA_g to pRB necessary for cell growth (Houben et al. 2012), but also the cells with truncated LTA_g are growing faster than cells with full-length LTA_g (Cheng et al. 2013). The truncated LTA_g do not induce DDR and cell cycle arrest (Li et al. 2013). Apart from MCPyV-positive patient samples with MCC (Cheng et al. 2013), the truncated LTA_g was observed in chronic lymphocytic leukemia (Pantulu et al. 2010).

LTA_g is not the only T antigen involved. Small T antigen stabilizes LTA_g through stabilization domain (Kwun et al. 2013). Moreover, small T antigen of MMC activates non-canonical NF- κ B signaling necessary in MCPyV tumorigenesis to avoid p53-induced senescence (Zhao et al. 2020).

Another feature of tumors is integration of MCPyV genome. The process of integration might lead to LTA_g truncation (Pantulu et al. 2010; Feng et al. 2008). However, apolipoprotein B mRNA editing enzyme, catalytic polypeptide (APOBEC) cytidine deaminase was shown to contribute to mutations of LTA_g sequence and subsequent formation of premature stop codons (Soikkeli et al. 2022).

Another PyV SV40 is clearly tumorigenic in lab animals and causing immortalization of cell lines (reviewed in Pipas 2009). Although the SV40 was associated with brain or lung tumors, non-Hodgkin's lymphoma or malignant pleural mesothelioma, the controversy over the role of SV40 in human tumors still persists (reviewed in Rotondo et al. 2019).

1.3 AID-mediated somatic hypermutations

During the B cell development, immunoglobulin (*Ig*) genes undergo several genetic modifications. After maturation, B cells migrate from bone marrow to the secondary lymphoid organs. After stimulation by antigens, *Ig* genes undergo class switch recombination in the switch region, from *Ig* isotype *IgM* to *IgG*, *IgE*, *IgD*, or *IgA*, for specific immune response and somatic hypermutations (SHM), introducing a substantial number of point mutations in the variable region to produce high affinity antibodies. Class switch recombination and SHM require the activity of the activation-induced cytidine deaminase (AID) (Figure 6) (Muramatsu et al. 1999, 2000; Arakawa et al. 2002).

The AID, expressed from *AICDA* gene, is a member of AID/APOBEC family (Conticello et al. 2005). The AID deaminates deoxycytidine (dC) residues on single-strand DNA (Petersen-Mahrt et al. 2002; Pham et al. 2003). The uracils are then removed by the uracil DNA glycosylase (UNG). Within abasic sites in the *Ig* class switch recombination region, single-strand breaks occur by the activity of apurinic/apyrimidinic endonucleases (APE) 1 and 2. Single-strand breaks in close proximity on the opposite strands result in double-strand breaks. When single-strand breaks are distal, then mismatch repair is deployed to generate double-strand breaks (Guikema et al. 2007; Schrader et al. 2005; Imai et al. 2003; Rada et al. 2002; Schrader et al. 2009).

SHMs are also AID-induced. Hypermutations occur in transcriptionally active genes (Yoshikawa et al. 2002), when AID has an access to single-strand DNA. In case the deaminated dC is not repaired, deoxyuridine (dU) is recognized as deoxythymidine (dT) during replication resulting in C to T transition mutation. If the dU is excised by UNG and abasic site is not repaired before replication, then the resulting abasic site may be filled with any nucleoside and transitions or transversions occur (reviewed in Keim et al. 2013). The dU/dG pairs may also be processed by Error-Prone Mismatch Repair Pathway with A/T mutations at dU-flanking sites (Phung et al. 1998; Wiesendanger et al. 2000; Unniraman and Schatz 2007). Beside *Ig* loci, some non-*Ig* gene loci are mutated with lower frequency. The study of *Ig* light chain locus in the chicken DT40 B cell line lead to an identification of diversification activator (DIVAC) element with SHM targeting properties (Blagodatski et al. 2009) containing transcription factor binding sites (TFBSs) (Buerstedde et al. 2014). However, susceptibility of loci to SHM is closely related to the overall architecture of topologically associated domains (Senigl et al. 2019).

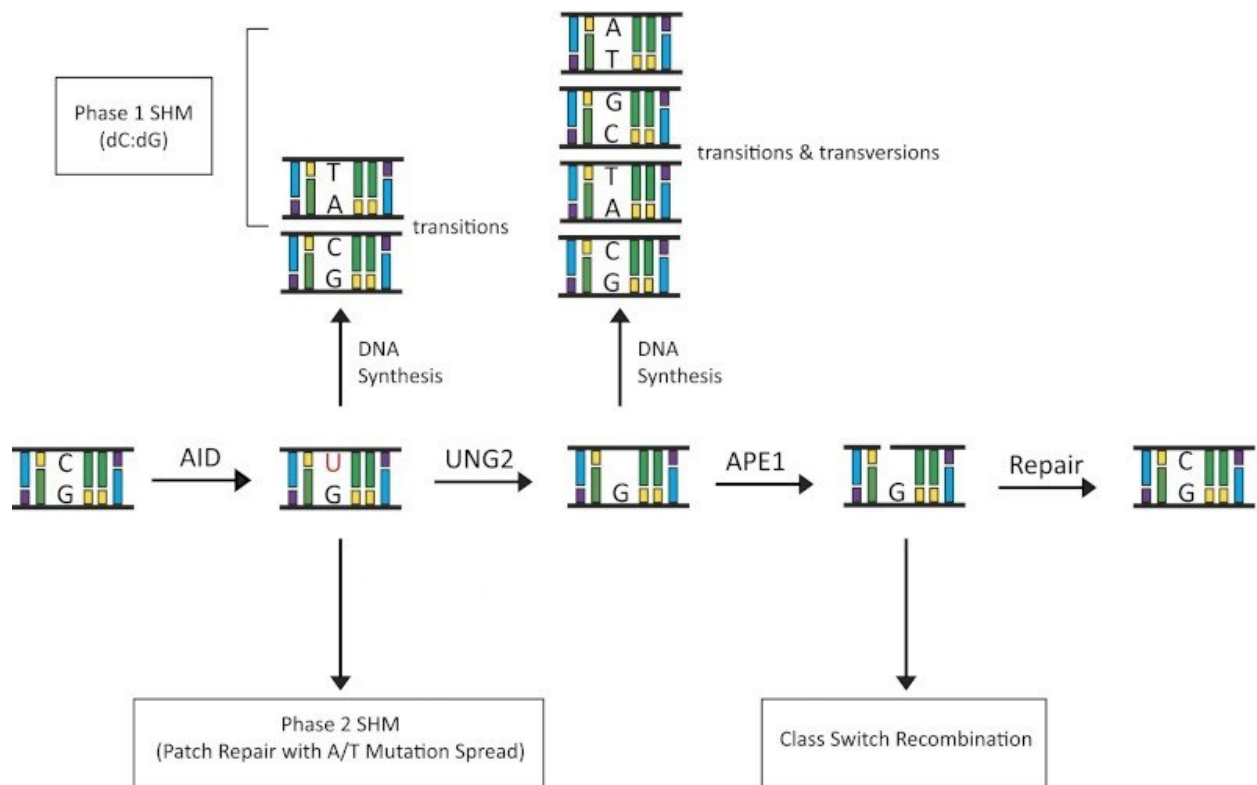


Figure 6. AID-mediated somatic hypermutations and class switch recombination. First, AID (activation-induced cytosine deaminase) deaminates C. The resulting U then can be recognized as T (Phase 1 SHM) or the lesion might be repaired by error-prone mismatch repair pathway, when neighboring nucleotides might get mutated (Phase 2 SHM). The U can also be excised by UNG (uracil DNA glycosylase). If the abasic site is not repaired before replication then any nucleotide might be inserted. At the abasic sites, single-strand breaks occur by the activity of APE (apurinic/apyrimidinic endonucleases). If the DNA breaks are in close proximity on the opposite strands, then the double-strand breaks are made (class switch recombination). The figure was adapted from Keim et al (2013).

1.4 Seizing control of viral properties

With better understanding of viruses, their life cycles and genomes, they were utilized to develop therapeutic and research tools.

DNA viruses, such as adenoviruses, adeno-associated viruses, herpesviruses, and poxviruses, and RNA viruses, such as retroviruses, alphaviruses, paramyxoviruses, and picornaviruses were used in gene transfer (reviewed in Bin Umair et al. 2022). They were turned into high-capacity vectors, for example to treat neurodegenerative disease Friedreich's ataxia (Ventosa et al. 2017), Duchenne muscular dystrophy (Counsell et al. 2017), or X-linked chronic granulomatous disease (Kohn et al. 2020), to develop vaccines against SARS-CoV-2 (Chiuppesi et al. 2020; Hassan et al. 2021), Ebola (Ledgerwood et al. 2017; Lázaro-Frías et al. 2018), rabies virus (Stading et al. 2017), Zika virus (Erasmus et al. 2018), or Lassa virus (M. Wang et al. 2018), or to combat cancer by oncolytic viruses (Studebaker et al. 2017; Porter et al. 2020) or lentiviruses in chimeric antigen receptor T-cell therapy (Jamali et al. 2019).

The MLV-based vectors were utilized for the treatment of hereditary disease, X-linked severe combined immunodeficiency. Originally successful gene therapy trials were shown to be unsuitable for the treatment (Cavazzana-Calvo et al. 2000; Hacein-Bey-Abina et al. 2002). Many patients who underwent the treatment developed leukemia (Hacein-Bey-Abina et al. 2003). Development of leukemia was caused by insertional mutagenesis after retroviral vector integration into proto-oncogene proximity (Hacein-Bey-Abina et al. 2003; Hacein-Bey-Abina et al. 2008; Howe et al. 2008). This adverse event in the treatment, led to preferential usage of self-inactivating (SIN) vectors (reviewed in Fischer and Hacein-Bey-Abina 2020).

The SIN vectors have deletion in the U3/R region of the LTR. Promoter-enhancer activity is then provided from an internal promoter (Zufferey et al. 1998). The SIN vectors possess lower risk of insertional mutagenesis, while preserving long-term stability of transgene expression (Zychlinski et al. 2008). However, the SIN vectors with strong internal enhancer derived from SFFV still cause insertional mutagenesis by enhancer-mediated oncogenesis (Modlich et al. 2006).

Retroviral vectors provide vehicles for transcriptionally stable expression of transgenes. Among others, they were used to study AID-mediated SHM (Senigl et al. 2019), and, in this thesis, for studying the SHM targeting activity of SV40.

Although the involvement of SV40 in cancer in human is still under discussion. The SV40 was used for establishment of immortalized cell lines (Kierstead and Tevethia 1993; Nobre et al. 2010; Tsao et al. 2002).

1.5 Functional organization of the genome

Primary DNA sequence and proteins form chromatin. Apart from the primary role in condensation of genome into more compact structures, chromatin is functionally differentiated. It plays a role in regulation of gene activity, DNA repair, or replication.

Key protein components of chromatin are histones. Histone tails are post-translationally modified. Certain histone modifications together with characteristics, obtained by epigenomic profiling methods, e.g. ChIP-seq (chromatin immunoprecipitation sequencing), DNase-seq (DNaseI sequencing), FAIRE-seq (formaldehyde-assisted isolation of regulatory elements sequencing), or RNA-seq (RNA sequencing), are associated with various functional chromatin states (Ernst and Kellis 2010; Hoffman et al. 2013).

On the higher-scale level, chromatin is organized into so-called topologically associated domains (Dixon et al. 2012), functionally interacting regions. The topologically associated domains are further arranged into larger units. These larger units are divided into A compartments, euchromatin, and B compartments, heterochromatin (Lieberman-Aiden et al. 2009). With higher resolution, more chromatin contacts obtained by chromosome conformation capture (Hi-C) maps, the compartments are divided into so-called subcompartments (Rao et al. 2014). The A compartment is divided into A1 and A2 subcompartments. Both A subcompartments are gene dense and contain highly expressed genes. The B compartment is divided into B1, B2, and B3 subcompartments. B1 subcompartments are enriched for facultative heterochromatin. B2 subcompartments are enriched for pericentromeric heterochromatin and the nuclear lamina. B3 subcompartments are enriched in the nuclear lamina.

Another type of domain is lamina-associated domains (LADs), which are in close contact with the nuclear lamina. LADs can be divided into constitutive LADs, which are highly conserved among cells of different origin, and facultative LADs, which are cell-type specific (Meuleman et al. 2013). Generally, LADs are less gene dense and are associated with transcriptionally repressive environment (Guelen et al. 2008; Peric-Hupkes et al. 2010).

2. Aims

The integration site is one of the key determinants of proviral expression stability. The MLV is transcriptionally stable in naturally occurring integration sites, promoters and enhancers of active genes, even after long-term culture.

- Is MLV LTR-driven transcriptional stability preserved in proviruses retargeted from naturally occurring integration sites?
- Is transcriptional stability a specific characteristic of the MLV or is it a general feature of γ RVs?

MCPyV *LTag* mutations triggered by APOBEC-mediated deamination contribute to LTag truncation observed in human tumors. Similar SV40 LTag truncations have been found in transformed human cells, although SV40 infection has not yet been conclusively linked to human cancer.

- Are the SV40 LTag truncations caused by cytosine deamination? Does SV40 enhancer bear SHM targeting activity allowing AID-induced *LTag* mutations leading to truncation of the protein?

3. Materials and Methods

3.1 Benchtop methods

Bacterial competent cells strains, cultivation and transformation

In case of unstable DNA vectors prone to recombination, e.g. retroviral vectors, we used NEB[®] Stable Competent *E. coli* (Genotype: F' *proA*⁺*B*⁺ *lacI*^q $\Delta(lacZ)M15$ *zzf::Tn10* (Tet^R)/ $\Delta(ara-leu)$ 7697 *araD139 fhuA* $\Delta lacX74$ *galK16 galE15 e14-* $\Phi80\Delta lacZ\Delta M15$ *recA1 relA1 endA1 nupG rpsL* (Str^R) *rph spoT1* $\Delta(mrr-hsdRMS-mcrBC)$; NEB; Cat. n.: C3040H). Otherwise, we used XL-1-Blue competent cells (Genotype: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F *proAB lacIqZ* $\Delta M15$ Tn10 (Tetr)]).

To prepare the competent bacteria, the cells were cultivated in LB medium (LB Miller Broth - 20 g/l; Media Facility IMG) until they reached the optic density 0.8 at 600 nm (Spectrophotometer S-200, Boeco). Then, the cells were centrifuged (2500 rpm, 7 min, 4 °C). The cell pellets from 50 ml cell suspension were resuspended in 15 ml TFB I (pH 5.8; 100 mM RbCl, 50 mM MnCl₂, 10 mM CaCl₂, 30 mM KAc and 15% glycerol) by rotating the vial on ice. The cells were then centrifuged again and resuspended in 2 ml TFB II (pH 6.8; 10 mM MOPS pH 0.7, 10 mM RbCl, 75 mM CaCl₂, 15% glycerol; MOPS - pH 7.0, 1M NaOH) by rotating on ice. 100 μ l aliquots were frozen on dry ice with 96% ethanol (WWR) and stored at -80 °C. The competition was verified by the number of grown colonies after transformation with 1 μ g DNA. Although both strains of competent cells might be cultivated at 37 °C, NEBStable competent cells were cultivated at 30 °C to prevent vector instability. Transformed bacterial cells were cultivated in LB medium or on agarose plates (Media Facility IMG) supplemented with antibiotics. When cultivated in LB media, the bacterial suspensions were incubated in shaker incubators IS-971 (Lab. Companion) or Bühler Incubator Hood TH 15 (Edmund Bühler GmbH). Agarose plates were cultivated in a heated incubator (MIR-262, Sanyo).

Prior to transformation, bacterial cells were mixed with a plasmid cloning reaction and cultivated on ice for 30 min. Bacterial cells were transformed by heat shock method in a stirred water bath WB-4MS (Biosan) at 42 °C for 25 s or 1 min, NEBStable or XL-1, respectively. Immediately after heat shock, we added 500 μ l of LB media and cultivated in a shaker incubator for ~1.5 h. Then, we evenly smeared 150 μ l of bacterial cell suspension on agarose plates supplemented with antibiotics and cultivated overnight in a heated incubator.

For isolation of plasmid from transformed cells, we used an overnight grown culture. When using QIAprep Spin Miniprep Kit (QIAGEN), we inoculated 2 ml of LB media supplemented with antibiotics, and when using QIAfilter Plasmid Midi Kit (QIAGEN), we inoculated 50 ml (XL-1) - 80 ml (NEBStable) of LB media supplemented with antibiotics (100 or 200 μ g antibiotic/ml LB media, respectively) with 60-80 μ l of bacterial suspension following the manufacturer's protocol.

For long-term storage of transformed bacteria, 1 ml bacterial suspension was well mixed with 500 μ l glycerol and stored at -80 °C freezer.

Plasmids and molecular cloning

In wt and Bin MLV vectors related experiments, we worked with pLG vector, the MLV-derived vector with the 5' Moloney murine sarcoma virus (MoMSV) long terminal repeat (LTR) and the 3' Moloney MLV (MoMLV) LTR expressing enhanced green fluorescent protein (EGFP) (Kalina et al. 2007).

When comparing expression stability of different gammaretroviral vectors, first, we replaced the *EGFP* of pLG vector by the destabilized *GFP*, *GFP* fused to the destabilization domain

(d2GFP), by In-Fusion[®] HD Cloning Kit (Takara) following the user guide instructions when the *d2GFP* fragment amplified by Phusion Hot Start DNA Polymerase (ThermoFisher Scientific) with LdG_IF primers (5' – CCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAG – 3' and 5' – ATGGCCTAACGAATTACTACACATTGATCCTAGCAGAAGC – 3') was inserted into EcoRI and NcoI-HF (NEB) restriction sites. The resulting vector pL2dG was then utilized to amplify the vector backbone by gLTR_OUT_IF primers (5' – CATTGGGGGGCTCGTCCGGGATC – 3' and 5' – CATTCCCCCCTTTTGTCAAAGAAACGGAAGGGCCG – 3') and the *d2GFP*-containing fragment by gLTR_IN_IF primers (5' – CATTGGGGGGCTCGTCCGG – 3' and 5' – CATTCCCCCCTTTTCTGGAGAC – 3'). Both amplicons were used in combination with amplified or synthesized γ RV-LTRs to create various γ RV-derived vectors expressing d2GFP by In-Fusion[®] HD Cloning Kit (Takara). MoMLV LTR was amplified from pL2G vector by MoMLV_LTR primers, CrERV LTR was amplified from pCrERV-5 vector (Fábryová et al. 2015) by CrERV_LTR primers. FeLV (GenBank NC_001940 (Chen et al. 1998)), SNV (Shimotohno et al. 1980), and KoRV (GenBank NC_039228 (Hanger et al. 2000)) LTR fragments were synthesized as gBlocks Gene Fragments (IDT) (Oligo Table 1).

The α RV SIN vector pAlpha.SIN.SF.EGFP.wPRE (Suerth et al. 2012) had *EGFP* replaced by *d2GFP* in NcoI and SpeI (NEB) restriction sites. First, *d2GFP* was amplified with pAS_GFP primers and 3' end of the vector was amplified with pAS_PRE primers. Original vector backbone and two amplicons were cloned together with In-Fusion[®] HD Cloning Kit (Takara). Next, the original U3 SFFV promoter was exchanged with other U3 γ RV promoters in XcmI and AscI (NEB) restriction sites. U3 parts and flanking sequences were amplified with primers listed at Oligo Table 2.

For targeted knock-in we used vectors derived from pLG (MoMLV, FeLV, KoRV) and pAG3 (Kalina et al. 2007). First, we replaced *d2GFP* with *mCherry* by using In-Fusion[®] HD Cloning Kit (Takara) when *mCherry* amplified by Phusion Hot Start DNA Polymerase (ThermoFisher Scientific) with gamma_mCherry_InFu (5' – GACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGG – 3' and 5' – AGGCGGCCGCTTGTGCGACGGCCGCAGATTACTTGTACAGC – 3') or AG3_mCherry_InFu (5' – TGCAGCCGACCACCATGGTGAGCAAGGGCGAGGAGG – 3' and 5' – GCTTGGCTGCAGGTGCGACGGCCGCAGATTACTTGTACAGC – 3') primers was inserted into NcoI and Sall (NEB) restriction sites. Final knock-in vectors were constructed from four segments each; EcoRV-NdeI (NEB) pAG3 fragment as a backbone, MfeI-EcoRV (NEB) gamma-*mCherry* or NaeI-EcoRV (NEB) pAGE3-*mCherry* fragment as provirus containing part, and homologous arms amplified by Hot Start DNA Polymerase (ThermoFisher Scientific) with homologous arms' cloning primers (Oligo Table 6) as provirus-flanking sequences for homologous recombination repair.

Variants of SV40 enhancer were amplified by Q5 High-Fidelity DNA Polymerase (NEB) and genomic enhancers were amplified by Phusion High-Fidelity DNA Polymerase (NEB) (Oligo Table 3). Amplicons were cloned into NheI and SpeI (NEB) restriction sites of GFP2 (Blagodatski et al. 2009) and GFP4 (Buerstedde et al. 2014) vectors and HpaI restriction site of GFP7 vector (Senigl et al. 2019) with In-Fusion cloning kit (Takara). The mutated SV40 2x72bp enhancers were prepared with QuickChange site-directed mutagenesis kit (Stratagene) or In-Fusion site-directed mutagenesis following the manufacturer's instruction. The correct assembly of vectors was verified by restriction enzymes and/or sequencing.

gRNA design, preparation of CRISPR vector pX458, testing of CRISPR-Cas9 cutting efficiency, and homologous arm design

For the design of guide RNAs (gRNAs), we used Crispor software (Concordet and Haeussler 2018) (<http://crispor.tefor.net/>). From the suggested oligos, we chose oligos with a high specificity score and a low number of off-targets. Used pSpCas9(BB)-2A-GFP (pX458) (#48138, Addgene (Ran et al. 2013)) vector obtains U6 promoter, therefore, the sense oligos without 5'-G had G added and antisense oligos had C added on the 3' end. Then, gRNA oligos had CACC added on the 5' end of sense oligo (A) and AAAC on the 5' end of antisense oligo (B) (Oligo Table 4). Phosphorylated annealed oligos were then inserted into the BbsI cloning site of the pX458 vector as follows.

First, we simultaneously phosphorylated and annealed oligos. We mixed the reaction: 10 μ M oligo A, 10 μ M oligo B, 1X T4 Ligation Buffer (NEB), 5U T4 polynucleotide kinase (NEB) and water to 10 μ l final volume. Then, we incubated the reaction as follows: 37 °C for 30 min, 95 °C for 5 min and then ramps down to 25 °C at 5 °C/min (0.1 °C/s) rate. After annealing, the annealed oligos were diluted in water 1:250 and stored at -20 °C.

Next, we set up digestion-ligation reaction: 100 ng of backbone vector (pX458), 2 μ l of diluted annealed oligos from the previous step, 1X FastDigest Buffer (ThermoFisher Scientific), 1 mM DTT, 1 mM ATP, 1 μ l FastDigest BbsI (ThermoFisher Scientific; or 20 U BbsI-HF, NEB), 1500 U T7 DNA Ligase (NEB) and water to 20 μ l of final volume. Then, we incubated the reaction as follows: 6 cycles of 37 °C for 5 min and 23 °C for 5 min, and held at 4 °C.

The ligation reaction was treated with PlasmidSafe exonuclease: 11 μ l of digestion-ligation reaction, 1X PlasmidSafe Buffer, 1 mM ATP, 10 U PlasmidSafe exonuclease (Biosearch Technologies). The reaction was incubated at 37 °C for 30 min and then 5 μ l of this reaction was used to transform XL-1 competent cells.

Vectors with correctly inserted oligos in the BbsI cloning site were verified by sequencing with U6_promoter_FW primer 5'-ACGATACAAGGCTGTTAGAGAGA-3'.

Before nucleofection of pX458-gRNA vector, the target sites were sequenced (Oligo Table 5) to verify complementarity between gRNA and host DNA and to obtain a negative control of CRISPR-Cas9 cutting efficiency.

The cells were nucleofected by using Amaxa nucleofector I/II/2b (Lonza) following the manufacturer's protocol. Briefly, after discarding cultivation medium, recommended number of cells (K562: 1×10^6) were resuspended in 100 μ l of Cell Line Nucleofector® Solution (K562: Solution V) with 3 μ g of pX458 with cloned gRNA (pX458-gRNA) and transferred into a cuvette. After the pulse in the nucleofector, 500 μ l of warm cultivating media was added. Then, the cells were seeded on a 6-well plate and cultivated in a CO₂ incubator (Sanyo).

Next, 2 days after nucleofection, we sorted 3,000 GFP-positive cells into 100 μ l of lysis buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.2 mM CaCl₂, 0.001% Triton X-100, 0.001% SDS, 1 mg/ml proteinase K) per sample. After incubation of cells in the lysis buffer at 58 °C for 1 hour and 95 °C for 10 min, the target locus was amplified and sequenced.

From the lysate or purified DNA from wild-type cells, we amplified the targeted locus.

Reaction: 1x NEB LongAmp buffer, 250 μ M dNTPs each, 1 U of TDV polymerase (Taq polymerase with Deep Vent polymerase 200:1 U, both NEB) per reaction, 450 nM forward and reverse primers, and 1 μ l of lysate in water.

Cycling conditions: 94 °C for 2 min, 35 cycles of 94 °C for 30 s, 60 °C for 30 s, and 68 °C for 3 min, then 68 °C for 10 min. The samples were held at 4 °C.

ICE CRISPR Analysis Tool (Synthego Performance Analysis, ICE Analysis. 2019. v3.0. Synthego; [2022]) was used to analyze the CRISPR-Cas9 cutting efficiency. The program requires gRNA sequences, non-treated sample sequences' files and CRISPR-Cas9 treated

sample sequences' files. We used the samples with observed indel profiles to design homologous arms for individual target sites (Oligo Table 6). Homologous arms were designed to immediately flank the indels. The length of homologous arms was ranging between 300-1000 bp. The length of homologous arms was often influenced by designing sufficiently specific primers to amplify the fragments for cloning. The shorter homologous arms are more suitable for subsequent verification of targeted insertion by sequencing.

Targeted knock-in of gammaretroviral vectors verification

Randomly selected clones were tested for targeted insertion by amplification of upstream and downstream genome-insert junctions with provirus-specific and homologous arms-adjacent sites primers by BioTherm DNA polymerase (GeneCraft) (Oligo Table 7). Only the clones amplified on both junctions were further analyzed by droplet digital PCR (ddPCR) for *mCherry* copy number ($CN < 1.5$ = single copy per genome, $CN \geq 2.5$ = > 2 copies per genome; BioRad).

Phenol-chloroform DNA isolation

We added 1 ml of the lysis buffer (0.25 M EDTA, 1% SDS) supplemented with freshly added proteinase K (final concentration 400 µg/ml) to cell pellets. Samples were rotated overnight at 55 °C. Then, we added RNase K (final concentration 300 µg/ml) and rotated samples at 55 °C for 1 hour. We added 800 µl of phenol-chloroform and shook vigorously. We let the samples stand for a while, and then we centrifuged the samples at a table-top centrifuge. All centrifuge steps were done at room temperature, at max. g for 5 minutes. We took the water phase into a new vial and repeated the phenol-chloroform step. Then, we added 1 ml of chloroform, shook vigorously, let it stand for a while, and then spun for 5 min. We added 1 volume of cold 96% EtOH to the water phase and shook the samples. After centrifugation, we carefully removed EtOH and washed the DNA pellet with 1 ml of cold 80% EtOH. Again, we discarded EtOH, air-dried the samples, and diluted them in a convenient volume of the TE buffer (10 mM Tris, 0.1 mM EDTA) based on the size of the DNA pellet after precipitation, usually 30-70 µl. The concentration of DNA in dissolved samples was measured on a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific).

DNA library preparation

For DNA library preparation, we isolated DNA by phenol-chloroform DNA isolation. The high salt isolation (Aljanabi and Martinez 1997) inhibited the subsequent fragmentation step even with prolonged incubation time. The DNA concentrations of the samples were repeatedly measured on the Qubit Fluorometer (Invitrogen) throughout the protocol.

First, we enzymatically fragmented DNA by NEBNext® dsDNA Fragmentase (NEB). We digested 4 µg DNA in 1x fragmentase buffer, 1x BSA, NEBNext® dsDNA Fragmentase® (1 µl/10 µl of a reaction; NEB), and water up to 80 µl reaction volume at 37 °C. Fragmentation time, ranging from 5 to 25 minutes due to DNA isolation differences among the samples, was experimentally determined for every sample individually. The reaction was stopped by adding 0.5 M EDTA (0.1 M final concentration). The size of fragments was checked on a 1% agarose gel. The fragments should be 100-2,000 bp long with the main density between 500-1,000 bp. Next, we selected 300-2,000 bp long fragments by SPRI Magnetic beads (Canvax). SPRI selection was pre-tested on the 100 bp marker (NEB) and tested on fragmented DNA. First, we mixed DNA with well-resuspended SPRI beads in the ratio 1:1, which worked the best for our samples. After incubation at room temperature for 10 min, the beads with bound DNA were collected on a magnetic stand and washed twice with freshly prepared 75% EtOH. Then,

the samples were air-dried and resuspended in 30 µl of 10 mM Tris-Acetate (pH 8.0). After incubation at room temperature for 5 min, the SPRI beads were removed from the samples. The concentrations of the samples were measured (Qubit).

The selected fragments had the ends repaired and added 3'polyA overhangs. The reactions (25 µl of selected fragments reaction with 1x NEB 2.1 buffer, 0.625 mM ATP, 0.25 mM dNTPs, 3 U T4 DNA Polymerase per µg DNA, 10 U T4 Polynucleotide Kinase (3' phosphatase minus) per µg DNA, 5 U Taq DNA Polymerase per µg DNA; all enzymes NEB) were incubated at 25 °C for 20 min, then at 72 °C for 20 min, and hold at 4 °C. After the end-repair reaction, the samples were purified on columns by using High Pure PCR Cleanup Micro Kit (Roche) following the manufacturer's protocol. Samples were eluted in 30 µl of the provided elution buffer. The concentration of the samples was determined (Qubit) and the amount of annealed oligos for ligation was calculated with the 30x excess of the linkers.

Annealed oligos (Oligo Table 8) were mixed with DNA and 1 µl of Quick Ligase (NEB). Reaction was incubated at 25 °C for 10 min.

After purification on columns (Roche), the concentrations of the samples were measured (Qubit) and 150 µg of DNA was added to be amplified in a PCR reaction with adaptor-specific and provirus-specific primers. Blocking oligos containing Locked Nucleic Acid (LNA), which greatly increase melting temperature of the oligo, were added to the reaction to prevent amplification of inner proviral sequences.

First, we ran a linear PCR reaction: 1x Phusion HF buffer, 200 µM dNTPs each, 800 nM MLV-LTR1_F_Bio (5' – Bio-TTCCATGCCTTGCAAAATGGCGT – 3'), 1 µM MLV_blocking2 (5' – {T}{T}CC{C}CCCTT{T}TT{C}T{G}GA{G}AC{T}AAA{T}AAA{A}TCAAAA-(A) – 3'; LNA = {}) and 1 U Phusion® High-Fidelity DNA Polymerase (NEB). The reaction was run as follows: 98 °C for 2 min, then 80 cycles of 98 °C for 15 s, 62 °C for 30 s and 72 °C for 1 min, and then 72 °C for 10 min. After the run, the reactions were immediately placed on ice and 0.5 M EDTA was added (final conc. ~80 mM).

The linear PCR reactions were purified on SPRI beads (1:1) and eluted in 30 µl of elution buffer. Then, the biotinylated products were selected by magnetic Streptavidin beads, Dynabeads MyOne Streptavidin C1 (Invitrogen), following the manufacturer's protocol. Washed coated beads were resuspended in the second PCR reaction.

The second PCR reaction was run with the adaptor-specific primer and the provirus-specific primers containing a barcode (MID) and the sequencing adaptor sequence (IonA). The reaction was mixed as follows: 1x Phusion HF buffer, 200 µM dNTPs each, 400 pM IonA_MIDx_mlvLTR2 (Oligo Table 9), 400 pM Ion_trP1 (5' – CCTCTCTATGGGCAGTCGGTGAT – 3'), 1 µM MLV_blocking2 and 1 U Phusion® High-Fidelity DNA Polymerase (NEB). The reaction was run as follows: 98 °C for 2 min, then 10 cycles of 98 °C for 15 s, 68 °C for 30 s and 72 °C for 2 min, and then 72 °C for 10 min; from 72 °C immediately placed on ice; added 0.5 M EDTA (final conc. ~80 mM).

After the run, the samples were loaded on an 2% agarose gel and 300-500 bp long fragments were isolated. After purification on the columns (Roche), the concentration of the samples was determined by the KAPA Library Quantification Kit for IonTorrent Platforms (Roche). The samples were then pooled in equimolar ratio and mixed with Ion Sphere Particles (Thermo Fisher Scientific). The sequencing was done on the IonTorrent platform (PGM sequencer; Thermo Fisher Scientific).

Apart from clonal populations of infected cells, we also processed the DNA from a clone with a verified single integration site as a control of bioinformatic analysis.

Crosslinking chromatin immunoprecipitation (ChIP)

First, medium from the adherent cells grown to 85-100% of confluency on a P100 plate was discarded and 10 ml of PBS was added. Then, we added the formaldehyde buffer (11% formaldehyde, 50 mM HEPES (pH 7.5), 140 mM NaCl, 1 mM EDTA, 500 mM EGTA) to get 1% formaldehyde. The cells were incubated with formaldehyde for 12 min at room temperature on a rocking platform. The glycine (final conc. 125 mM) was added and after incubation for 5 min at room temperature on a rocking platform, the buffer was discarded and we placed a dish on ice. We washed the cells with 10 ml of ice-cold PBS twice, added PBS supplemented with inhibitors (1x complete protease inhibitor (Roche), 5 mM NaF (Penta chemicals), and 2 mM Na₃VO₄ (Sigma-Aldrich)) and scraped the cells from a dish into a centrifuge tube. Then we centrifuged the tubes and discarded the supernatant.

After the cross-link, the cells were lysed and chromatin was sheared in the Bioruptor sonicator (Diagenode). First, we resuspended a cell pellet from the previous step in 150 µl of ice-cold SDS lysis buffer (50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 1% SDS) supplemented with inhibitors. We incubated the samples on ice for 20 min and diluted the lysate 1:4 with ChIP dilution buffer (16.4 mM Tris-HCl (pH 8.0), 167 mM NaCl, 1.2 mM EDTA, 0.5 mM EGTA, 0.01% SDS, 1.1% Triton X-100) supplemented with inhibitors (600 µl in total). We split the samples into two equal aliquots and sonicated in a sonicator (Bioruptor, Diagenode). The setting of the sonication program was experimentally determined for every sample. We spun the lysates at 13000 rpm for 10 min at 4 °C and combined the lysates.

We took 50 µl lysate aliquots to check the fragment sizes. We added 250 mM NaCl and incubated overnight at 65 °C. Then, we added 12.5 µg RNase A and incubated for 30 min at 37 °C. After RNase A treatment, we diluted the samples with H₂O (1:1) and added 2 µl 0.5M EDTA, 4 µl 1M Tris (pH 6.5), and 1 µl proteinase K (10 mg/ml). We incubated the reactions at 45 °C for 1 h. The samples were purified on the columns by High Pure PCR Product Purification Kit (Roche) and eluted in 50 µl of elution buffer. The purified samples were loaded on a 1.5% agarose gel. The sonicated samples with the fragments with the largest density between 100-1000 bp were used in the following immunoprecipitation (IP) step.

From the pool of two sonicated lysates (1,100 µl in total) we took 30 µl aliquot as an input control and 150 µl aliquots for IP. The IP aliquots were diluted 1:1.5 with ChIP dilution buffer supplemented with inhibitors. We added 1-5 µg of antibody (Anti-Histone H3 [ab1791], Anti-Human IgG [ab2410], and Anti-Lamin B1 [ab133741]; abcam) and incubated the samples overnight at 4 °C in a Stuart SB3 rotator (Bibby Scientific).

Simultaneously, we washed 30 µl of magnetic DynabeadsTM Protein G (Invitrogen) per IP reaction with 1 ml BSA/PBS (250 mg BSA / 50 mL PBS) and then blocked them with BSA/PBS supplemented with 1% Gelatin from cold fish skin and rotated them overnight at 4 °C.

After we washed blocked beads with an ice-cold ChIP dilution buffer twice and resuspended them in a ChIP dilution buffer to the original volume, we added 30 µl of blocked beads to each IP sample and incubated them for 2 h at 4 °C in a rotator.

Then, we collected the beads on a magnetic stand and discarded the supernatant. The beads were washed with several buffers: twice with 750 µl of Low Salt Wash Buffer (20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), three times with 750 µl of ChIP Wash Buffer (50 mM HEPES (pH 7.6), 1 mM EDTA, 0.7% Na-deoxycholate, 1% NP-40/Igepal CA-630, 500 mM LiCl), once with 750 µl of TE buffer supplemented with NaCl (50mM), and once with T₁₀E_{0.1} (10 mM Tris pH 7.6, 0.1 mM EDTA pH 8.0). During every wash step, the samples were incubated for 5 min on ice. After the final wash, we discarded the supernatant and centrifuged the samples (1000 x g for 1 min). We collected the beads on a magnetic stand and discarded buffer. We added 100 µl of ChIP elution buffer (50 mM Tris-HCl

(pH 8.0), 10 mM EDTA, 1% SDS) to the beads and incubated them at 65 °C for 15 minutes with occasional mixing of the samples. Then, we collected the beads on a magnetic stand and transferred the supernatants into clean tubes.

The final step is the de-crosslink of IP and input samples. We diluted IP samples by adding 100 µl of TE and input samples by adding 77 µl of H₂O and 93 µl of ChIP dilution buffer. Then, we added 120 mM NaCl and incubated the samples overnight at 65 °C. Next, we added 50 µg RNase A and incubated for 30 min at 30 °C, then 16 mM EDTA, 66 mM Tris (pH 6.5), and 20 µg Proteinase K and incubated for 1 h at 45 °C. After the de-crosslink, we purified the samples with High Pure PCR Product Purification Kit (Roche), eluted them in 100 µl of provided elution buffer, and analyzed them by Digital Droplet PCR (Biorad).

Droplet digital PCR (ddPCR)

After isolation of DNA, the samples were diluted with water to concentrations 10 ng/µl. Reactions were prepared with ddPCR™ Supermix (BioRad), 10 ng DNA, 200 nM primers, and/or 100 nM probes (Oligo Table 10). Reactions were run in duplicates.

Droplets were prepared from 20 µl of reaction mix and 70 µl of oil in a QX200 Droplet Generator (BioRad). Then, 40 µl of processed sample was carefully transferred into a 96-well plate to avoid damaging generated droplets.

Reactions were amplified in a cycler (Biorad) as follows: (PROBES) 95 °C for 5 min; 40 cycles of 94 °C for 30 s, 58 °C for 30 s, 72 °C 30 s; then 98 °C for 10 min. (PRIMERS) 95 °C for 5 min; 40 cycles of 95 °C for 30 s, 60 °C for 1 min; then 4 °C for 5 min and 90 °C for 5 min. ChIP Reactions were run in QX200™ ddPCR™ EvaGreen Supermix with ChIP primers in the following program: 95 °C 5 min; 40 cycles of 95 °C 30 s, 59 °C 20 s and 72 °C 30 s; 4 °C 5 min; 90 °C 5 min. All programs were run with a ramp rate of 2 °C/s. Amplified samples were read by a QX200 droplet reader (BioRad).

Proviral copy number quantification was measured with ddPCR™ Supermix for Probes (BioRad), primers targeting proviral sequence, probes with FAM against provirus, primers targeting endogenous *RPP30* sequence, and probes with HEX against *RPP30*. q5GFP primers/probes were used for quantification of *GFP*-encoding proviruses, and q3mCh primers/probes were used to quantify copy number of *mCherry*-encoding proviruses. Droplet Generation Oil for Probes or QX200™ Droplet Generation Oil for EvaGreen were used depending on the reaction conditions.

Copy number of provirus per cell was determined as a copy number of FAM-positive target / 2x HEX-positive target.

Protein extraction and western blot (DT40 cells)

Approximately 1x 10⁶ cells were centrifuged at 500xg for 5 min at 4 °C. The cell pellet was resuspended in 500 µl of PBS and then centrifuged at 2000xg for 3 min at 4 °C. Again, the cell pellet was resuspended in 100 µl of ice-cold RIPA buffer (50 mM HEPES (pH 7.6), 1 mM EDTA, 0.7% Na-deoxycholate, 1% NP-40/Igepal CA-630, 500 mM LiCl) with 1x complete protease mix and incubated on ice for 1 h. Then, the lysate was centrifuged at 10.000xg for 10 min at 4 °C. Supernatant was transferred into a clean 1.5 ml tube and mixed with 33.3 µl of 4x NuPage LDS Sample buffer (Invitrogen™) supplemented with 0.2 M DTT, and then incubated for 10 min at 70 °C. The samples could be stored at -80 °C.

Next, 10 µl of lysate with 5 µl of Spectra BR marker (Thermo Fisher Scientific) were loaded on 4-12% NuPage Bis-Tris gel with MOPS buffer in Xcell system (Thermo Fisher Scientific) and run for 1 h at 180 V. Then, the proteins were moved on a nitrocellulose membrane by Xcell blotting system following the manufacturer's protocol. After the transfer, the membrane was

blocked in 5% BSA-TBS for 2-3 h at room temperature. Blocked membranes were washed for 5 min with TBS and incubated with primary antibodies 1.000 x diluted α AID 30F12 Rabbit (Cell Signaling) and 100 x diluted α SV40-LT pAb419 Mouse (Genetex) overnight at 4 °C while rocking and with 10.000 x diluted α GAPDH 5G4 Mouse (Hytest) for 15 min at room temperature while rocking. Again, the membranes were washed with TBS for 5 min. Then, with the secondary antibodies 2.500 x diluted IRdye α Rabbit 680RD, for detection of AID, and 5.000 x diluted IRdye α Mouse 800CW, for detection of large T antigen expression, the incubations were performed 2-3 h at room temperature while rocking. While incubation with 10.000 x diluted IRdye α Mouse 680RD, for detection of GAPDH, was done for 15 min at room temperature while rocking. The incubations with secondary antibodies were performed at dark. All antibodies were diluted in 5% BSA-TBS. Before the imaging by Odyssey Fc (LICOR), the membranes were washed with TBS for 5 min.

Protein extraction and western blot (Ramos and UO-31)

The cells collected from 60 mm dish (VWR) were collected and washed with ice-cold PBS. The cell pellet, approximately 5×10^6 cells, was resuspended in 0.5 ml 1x Laemmli sample buffer. The cell lysate was boiled for 5 min and sheared by repeatedly passing the sample through a needle. Then, 5 μ l of lysate from Ramos or AID-/- cells or 30 μ l of lysate from UO31 with 5 μ l of color protein standard marker (NEB) were loaded on 12% (LTA) or 13% (AID) SDS-PAGE Gel with Elfo buffer in Xcell system and run for 10 min at 100V and then 1 hour at 150V. The the proteins were moved on a PVDF membrane, prewashed with 100% methanol and incubated in blotting buffer for 10 min, by trans-blot SD semi dry transfer system (Biorad) for 30 min at 15V at room temperature. Then, the membranes were blocked with 5% BSA-TBS, for detection of AID, and 5% milk, for the detection of large T antigen, for 1 hour at room temperature. The blocked membranes were washed with TBS for 5 min and incubated with primary antibodies 1.000 x diluted α AID Monoclonal Antibody ZA001 Mouse (Thermo Fisher Scientific), 100 x diluted α SV40-Tag pAb419 Mouse (Genetex), and 2.000 x diluted α GAPDH Mouse (Thermo Fisher Scientific) overnight at 4 °C while rocking. After a wash with TBS for 5 min, the incubations with secondary antibodies 3.000 x diluted α Mouse IgG HRP linked (Cell Signaling) and 3.000 x diluted goat α Mouse IgG HRP linked (NEB) we performed for 1 h at room temperature. The incubations with secondary antibodies were performed at dark. All antibodies were diluted in 5% BSA-TBS. After incubation with secondary antibodies, the membranes were washed with TBS for 5 min. Then, the membranes were treated with chemiluminescent Detection Substrate-West Pico Plus (Thermo Fisher Scientific) and documented by UVITEC Cambridge imaging system (Uvitec).

RNA isolation, reverse transcription and quantitative PCR

Medium from the cells was discarded and 1 ml of RNeasy lysis reagent (Qiagen) was added. The adherent cells were scraped from a dish and transferred into a 2 ml tube. Then, 400 μ l of water was added to samples and incubated for 15 min at room temperature. The samples were centrifuged at 16.000xg for 15 min at room temperature. The supernatant with RNA was transferred into a clean 2 ml tube and 75% ethanol (Penta chemicals) was added. Then, the samples were incubated for 10 min at room temperature and centrifuged at 12.000xg for 10 min at 15 °C. Next, the RNA pellets were washed twice with 75% ethanol. Air-dried pellets were dissolved in water for 10 min at 55 °C.

Isolated RNA was measured on Nanodrop and 1 μ g of RNA was added into DNase treatment reaction by RQ1 RNase-free Dnase kit (Promega) following the manufacturer's protocol. After inactivation of Dnase, the complementary DNA was prepared with Protoscript II Reverse

Transcriptase kit (NEB) following the provider's protocol. After reverse transcription, the sample was diluted with water 1:1.

For the quantitative reaction, 2 µl of cDNA were mixed with 1x MESA GREEN qPCR MasterMix Plus (Eurogentec) and primers for amplification of AID (5'-AATTCAAAAATGTCCGCTGGGC*T-3' and 5'-AGCGGAGGAAGAGCAATTC*C-3') or GAPDH (5'-AGGGCTGCTTTTAACTCTGG*T-3' and 5'-CCCCACTTGATTTTGGAGGG*A-3'). The reactions were run in triplicates in Bio-Rad CFX96™ Real-Time cycler (BioRad) as follows: at 95 °C for 3 min, then 40 cycles of 95 °C for 15 s, 60 °C for 20 s and 72 °C for 20 s, and 72 °C for 10 min. For the standard curve, the serial dilution (5x) of Ramos wt sample was used. Cycles of quantification values were calculated by the CFX Manager software (BioRad).

3.2 Cell culture methods

Cell culture

The cell line K562 was cultivated in RPMI 1640 medium (Sigma) supplemented with 5% fetal calf serum, 5% calf serum (both Gibco), and antibiotics (#A5955, 100x Antibiotic Antimycotic Solution, Stabilized; 10 000 U penicillin, 10 mg streptomycin, and 25 µg amphotericin B per ml; Sigma) at 37 °C and in a 5% CO₂ atmosphere.

The cell line HEK293T was cultivated in DMEM:F12 medium (Sigma) supplemented with 5% fetal calf serum (Gibco), 5% newborn calf serum (Gibco) and 1x antibiotics (Sigma) at 37 °C and in a 5% CO₂ atmosphere.

The DT40 cell line was cultivated in RPMI 1640 HEPES modification medium (Sigma) supplemented with 10% FBS (HyClone), 1% newborn calf serum (Biowest), 1x penicillin-streptomycin antibiotics (Gibco), 1X GlutaMAX (Gibco), and 50 µM β-mercaptoethanol (Sigma-Aldrich) at 40 °C and 5% CO₂.

The Ramos, UO-31, and H28 cell lines were cultivated in RPMI 1640 HEPES modification (Sigma) supplemented with 10% FBS (Gibco), 1x penicillin-streptomycin antibiotic (Gibco), and 1x GlutaMAX (Gibco) at 37 °C and 5% CO₂.

The 293T, U20S, and MDA-MB-231 cell lines were cultivated in Dulbecco's Modified Eagle's Medium - high glucose (Sigma) supplemented with 10% FBS (Gibco) and 1x penicillin-streptomycin (Sigma) at 37 °C and 5% CO₂.

All cell lines were passaged when reaching approximately 90% confluence. The adherent cells were first washed with 1x PBS (Media Facility IMG), and then with 0.125% trypsin, 0.01% EDTA in 1x PBS (Media Facility IMG). After incubation for 5 min at 37 °C, the cells were resuspended in fresh medium. The volumes were adjusted to the size of cultivating plates and dishes. All work with the cell cultures was done in a biohazard box (Schoeller Instruments).

The stock cells were kept at -196 °C in RPMI or DMEM:F12 medium supplemented with 10% DMSO (99%; Serva) and 10% fetal calf serum. The cells were frozen in CryoTube Vials (Thermo Scientific) in a freezing container (Nalgene) filled with isopropanol (Lach Ner) placed in -80 °C freezer to ensure slow freezing 1 °C/min.

Viral vector production and transduction

Gammaretroviral and derived vectors were produced in the HEK293T cell line by co-transfection. A day before a transfection, the cells were seeded on the poly-L-lysine-coated plates/dishes. When using X-Treme Gene HP Transfection Reagent (Roche) following the provider's protocol, and for the calcium phosphate method, 2.5-3 x 10⁶ cells were seeded on a P100 plate. The following day, the cells were co-transfected with the viral genome, Gag-Pol

(Schambach et al. 2006; Suerth et al. 2012; El Ashkar et al. 2017), and pVSV-G (Clontech) vectors in 6:3:1 weight ratios.

First, we prepared mix A containing 30 µg of total DNA in water (up to 1 ml) and 135 µl of 2M CaCl₂·2H₂O. The mix A was then dropwise added to mix B (1.12 ml of HBS and 22 µl of 100x conc. PO₄). The transfection mix was added to the cells, and after incubation for 5 h, the cells were washed with HBS with glycerol and PBS. Then cultivating medium was added.

One day after transfection, the medium was changed with fresh cultivating medium, and two days after transfection, we collected the medium with the virus. To increase viral concentration in stocks, we filtered collected medium through a 0.45 µm syringe filter (Corning) and ultracentrifuged the samples at 24,000 RPM for 2.5 h at 4 °C in the SW40-Ti rotor (Beckman). After discarding the supernatant, the pellet was dissolved in RPMI 1640 medium (Sigma) by overnight shaking at 4 °C. The final viral stocks were kept at -80 °C.

On the day of transduction, the K562 cells were seeded in suitable density on a cultivation dish/plate in fresh cultivating medium. Then, we added virus-containing medium to the cells and incubated for 24 h at 37 °C in a 5% CO₂ atmosphere. A day after transduction, the medium was replaced with fresh cultivating medium.

HIV-derived vector for transduction of the Ramos cell line was produced in the 293T cell line by co-transfection by X-treme HP (Roche). In the transfection reaction, we mixed 1 µg of GFP7 vector, 1 µg of psPAX2 (Addgene; #12260) and 1 µg of pVSV-G (Clontech). Collected virus was filtered through a 0.45 µm SFCA filter and stored at -80 °C.

Targeted knock-in of gammaretroviral vectors

K562 cells were co-nucleofected with 3 µg of each vector, pX458 with cloned gRNA and mCherry-expressing proviral vector flanked with homologous arms corresponding with the gRNA-recognized site, by using Nucleofector™ I/II/2b Device (Lonza). Vector carrying proviral vector for insertion during homologous recombination repair was not linearized prior (restriction enzymes) or during (CRISPR-Cas9) co-nucleofection. Two days post nucleofection, we sorted the population of GFP-positive cells, and 9 days post nucleofection, we sorted the mCherry-positive cells in a single-cell mode into 96-well plates. After expansion, more than 30 days post nucleofection, mCherry expression was analyzed on a flow cytometer LSRII (BD Biosciences) and clonal DNA was isolated on the columns by using DNA Blood and Tissue Kit (QIAGEN).

GFP loss assays

The DT40 *UNG*^{-/-} *AICDA*^{R/puro} cells (Buerstedde et al. 2014) were transfected with GFP2 and GFP4 vectors with cloned SV40 enhancers by electroporation, when 12 x 10⁶ cells were transfected with 40 µg of linearized plasmid with a Gene Pulser Xcell (BioRad): 0.7 kV, 200 Ω, and 25 µF. After the transfection, the cells were seeded on a 96-well plates in growing medium supplemented with extra 5% FCS. One day after transfection, the transfected cells were selected with blasticidin (final concentration 15 µg/ml). After seven days, the primary selected clones were selected for targeted integration with puromycin (final concentration 1 µg/ml). Double selected clones were then subcloned by limiting dilution. After another 12 days, the cells were measured by flow cytometer Accuri C6 cytometer (BD Bioscience) and Novocyte cytometer (Agilent).

The Ramos, UO31, U20S and MDA-MB-231 cells were transduced with GFP7 reporter vector, when 4 x 10⁶ cells were infected with the low multiplicity of infection < 0.01. Viral suspension was added to the cells and cultivated for 40 min at room temperature. Then, cultivation medium was added to the cells and the dish was placed into a cell incubator. Two days after

transduction, the cells were selected with blasticidin (final concentration 5 µg/ml) for two days. Then 7 days after transduction, the GFP-positive cells were sorted in a single-cell mode by an Influx cell sorter (Becton Dickinson). Expanded clones were selected with blasticidin (final concentration 15 µg/ml) 17 days after sorting. After three days of second selection, expression of GFP was measured on LSRII flow cytometer (Becton Dickinson).

Measured data were analyzed by FlowJo software and GraphPad Prism 9 software.

Luciferase assay

SV40 enhancers were cloned into the Sall and BamHI (both NEB) restriction sites of pGL4.23 vector (Promega). The DT40 UNG^{-/-} AID^{R/puro} cells were co-transfected with 20 µg of cloned vectors with 2.5 µg of pGL4.75 Renilla luciferase control vector (Promega). Transfection was done by using the Amaxa Nucleofector Kit V program B-023 (Lonza) or Xcell PBS protocol (BioRad). The relative activity of firefly luciferase to Renilla luciferase was measured with the Dual-Glo Luciferase Assay System (Promega) following the manufacturer's protocol.

Ectopic AID expression

The cells were transduced with the retroviral MSCV-derived vector (gift from Rashu B. Seth) containing human, chicken, or mice AID fused with mCherry (AID-mCherry) (Le and Maizels 2015), IRES and gene for hygromycin resistance. These vectors were transduced into the cell line clones carrying GFP loss assay vectors. Two days after transduction, the cells were selected with hygromycin. GFP expression was measured on Symphony or LSRII flow cytometers (both Becton Dickinson) three weeks after transduction. Measured data were analyzed by FlowJo software.

Large T antigen coding sequence knock-in

SV40 large T antigen region was amplified by Q5 High-Fidelity DNA Polymerase (NEB), for GFP2 vectors, or by Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Scientific), for GFP7 vectors, from SV40 genome WT4 plasmid, which was a kind gift from James DeCaprio. The PCR products were cloned into NheI and SpeI (both NEB) restriction sites of the GFP2 vector or HpaI (NEB) site of the GFP7 vector by In-Fusion cloning kit (Takara) following the manufacturer's protocol.

The DT40 IgL(-) cells were transfected with the GFP2 vector with cloned SV40 large T antigen region, following the transfection protocol of GFP loss assay. One day after transfection, the transfected cells were selected with blasticidin (final concentration 20 µg/ml). The cells were single-cell sorted and the cell clones were selected with puromycin (final concentration 1 µg/ml). Then, the clones were cultivated for 12 weeks and genomic DNA and proteins were isolated by gDNA miniprep kit (Zymo Research) and protein extraction protocol, respectively. The Ramos wt cell line and UO-31 cell line were transduced with lentiviral GFP7 vectors with low multiplicity of infection. Three days after transduction, the cells were sorted for the expression of GFP. After 12 weeks of cultivation, genomic DNA and proteins were isolated by gDNA miniprep kit (Zymo Research) and protein extraction protocol, respectively.

Flow cytometry

Cells analyzed by flow cytometer LSRII (BD Biosciences) were pipetted on a 96-well plate and centrifuged at 1800 RPM for 10 min. After discarding cultivating media, the cells were resuspended in PBS with Hoechst 33258 (1000 x diluted).

For sorting, cells were centrifuged at 200x g for 10 min. After discarding cultivating media, the cells were resuspended in medium without serum. The cells were sorted on BD Influx or FACS Aria IIu (both BD Biosciences) into cultivating medium or lysis buffer.

Immunohistochemistry and Leica SP8 confocal microscopy

The Ramos wt, Ramos AID-/- and UO-31 cells were seeded on microscope poly-L-lysine-coated cover glasses (approximately 3×10^5 cells per 6-well plate). One day after seeding, the cells were fixed with 4% paraformaldehyde for 15 min, then washed with PBS three times, permeabilized with PBS supplemented with 0.1% Triton and 1% BSA for 10 min, and again washed with PBS three times. Then the samples were blocked with PBS supplemented with 2% BSA, 0.15% Glycine, and 10% FCS for 45 min at room temperature and washed with PBS three times. Next, the samples were incubated with 20 x diluted (in PBS supplemented with 0.1% BSA and 0.4% Tween20) Mouse Monoclonal Ab anti-AID antibody (ZA001, Invitrogen) for 1 h at room temperature. After a wash, the samples were incubated with the secondary antibodies 500 x diluted anti-mouse-Cy3 (Jackson ImmunoResearch) for 1 hour in room temperature.

Before the microscopy on Leica SP8 FLIM confocal microscope, the samples were treated with Hoechst 33258 for 15 min at room temperature and washed with PBS three times. Images were deconvoluted by Huygens software, the contrast was enhanced by ImageJ software, and deconvoluted images were further analyzed by Fiji J software.

3.3 Bioinformatic methods

Integration sites

During the DNA library preparation, we used primers with barcodes for individual samples. So, first, reads in the FASTQ format were sorted based on their barcodes using the *cutadapt -g ^B--overlap 8 --discard-untrimmed* command (Martin 2011), where B stands for the barcode sequence, and the read names were modified to contain the name of the barcode group. Reads with LTR-specific primer sequences were selected with *cutadapt -g ^GCTTGCCAAACCTACAGGTG --overlap 20 --discard-untrimmed* command. Next, reads with the minimal length of 15 bp and the last 9 nucleotides of the LTR sequence were selected with *cutadapt -g ^GGTCTTTCA --overlap 8 --discard-untrimmed --minimum-length 15* command. Then adapter sequences were removed from the reads with *cutadapt -a ACCACTAGTGTGCGAC --overlap 10 --minimum-length 15* command. Amplified inner proviral sequences were removed using the *cutadapt -g TTCCCCCCTT --overlap 10 --discard-trimmed* command. Trimmed FASTQ reads were mapped to hg19 and hg38 human genome assemblies by the Bowtie 2 (Langmead and Salzberg 2012) with *bowtie2 -p 20 -q -x hgX* command, where the hgX stands for the name of the assembly. Reads mapped from the start of the read ("MD:Z:0" reads) with a single hit in the genome ("XS:i:" reads) were selected and converted to BED file by *samtools view -S -b* (Danecek et al. 2021) and *bedtools bamtobed -cigar -i* (Quinlan and Hall 2010) commands. Each integration site is read as one genomic LTR-proximal position. ISs represented by at least 5 reads were selected. With ISs appearing within the distance of 5 bp, only the IS with the most reads was selected. The two integration site sets (IS_hg19.bed and IS_hg38.bed) were used independently against the annotations of respective genomic assemblies.

The set of random genomic integration sites was created by the concatenation of three integration site files and subsequent usage of bedtools shuffle command.

Distance to genomic and epigenomic features

The following annotations were obtained from publicly available databases: the genomic segments (Ernst and Kellis 2010) from the University of California Santa Cruz (UCSC) genome annotation database (<https://hgdownload.cse.ucsc.edu/goldenpath/hg19/database/wgEncodeAwgSegmentationChromHMMK562.txt>); the chromatin subcompartments (Rao et al. 2014) from gene expression omnibus (SE63525_K562_Arrowhead_domainlist.txt.gz); the lamina-associated domain (LAD) genomic coordinates from 4D Nucleome Data Portal (Dekker et al. 2017; Reiff et al. 2022) (data accession 4DNFIV776O7C). Distance to the features was calculated by *bedtools closest -d* command.

Large T antigen mutation analysis

The SV40 large T region from the DT40 cells was amplified by Q5 High-Fidelity DNA Polymerase (NEB) and cloned into BamHI and HindIII (both NEB) restriction sites of pUC19 vector (Addgene) by In-Fusion cloning kit (Takara). The cloned plasmids were sequenced and analyzed with SnapGene software.

The SV40 large T region from the Ramos and UO-31 cell lines was divided into three sequencing subregions: 16-810 bp, 694-1521 bp, and 1467-2456 bp. Single-base substitutions, insertions and deletions against reference genome were analyzed. For mutation frequency calculations, the 200-bp bins of large T antigen were utilized, except for the last 273-bp bin.

Software and statistical analysis

Measured cytometric data were measured by FlowJo™ v 10.10 Software (BD Life Sciences). In the γRV experiments, gated populations were exported to csv files, concatenated and analyzed with custom R code (R Core Team 2022; <https://www.r-project.org/>).

γRV statistical tests and plots were created in R. Figures were produced with the ggplot2 package (Wickham 2016). The impact effect-size analysis was performed with the ImpactEffectsize package (Lötsch and Ultsch 2020).

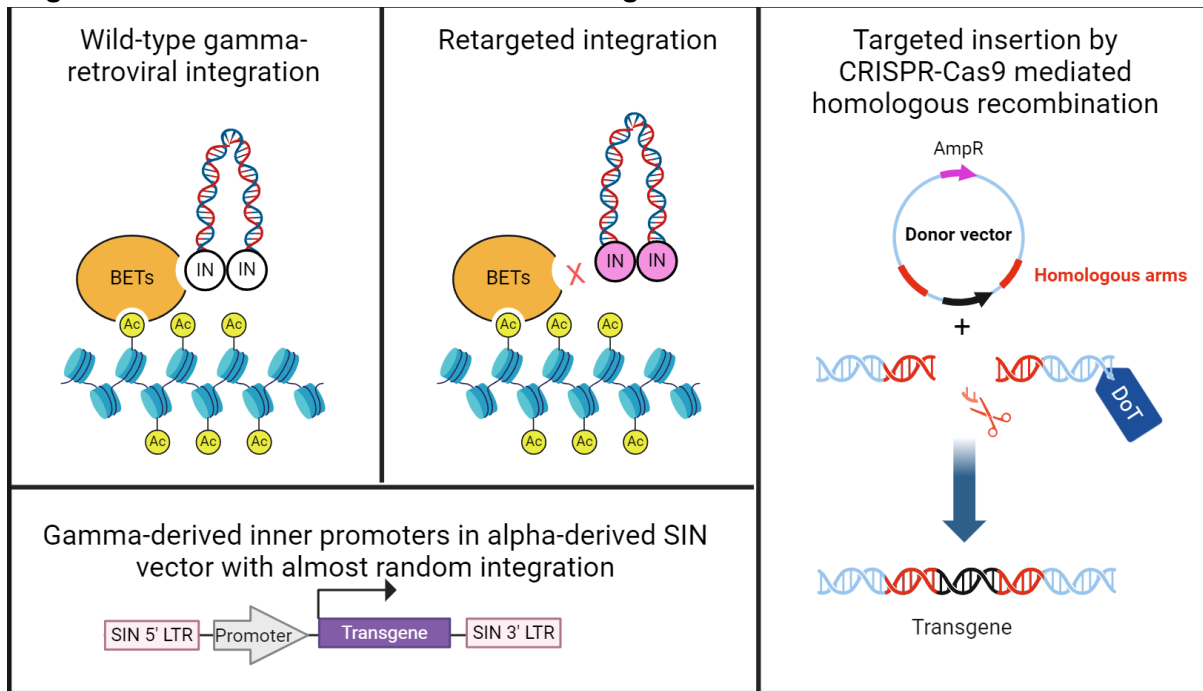
Analysis of ChIP data was performed using GraphPad Prism version 5.2 for Windows (GraphPad Software, San Diego California USA, www.graphpad.com).

Statistical analysis of SV40 experiments was calculated in GraphPad Prism 9 software. For statistical significance of differences in the medians of GFP fluorescence loss and MFI, the Mann-Whitney U test was utilized. Statistical significance of relative luciferase activity was calculated with Unpaired t-test. Statistical significance of WRC mutations compared to overall C targeting mutations was determined with Fisher's exact test.

The schemes in the figures were created with BioRender (<https://www.biorender.com/>).

4. Results

4.1 Gammaretroviral vectors remain transcriptionally stable even after retargeting and targeted insertion in the cells of different origin



4.1.1 Stable expression of MLV-derived vectors with retargeting Bin INs

Previously, the MLV-derived vector was shown to remain stably expressed for a long time in the K562 cell line (Miklík et al. 2018). Even the MLV-derived SIN vectors with internal promoters derived from SFFV LTR or the eukaryotic translation elongation factor 1 alpha were transcriptionally stable after retargeting with IN^{W390A} and IN^{CBX} (Van Looveren et al. 2021). Here we used previously described MLV-derived vector expressing EGFP (Kalina et al. 2007) in combination with IN^{WT} , IN^{W390A} , or IN^{CBX} to assess the transcriptional stability of the retargeting vectors in comparison to the wt vector in the K562 cell line (Figure 7A).

After transduction with low multiplicity of infection (MOI), we measured the EGFP expression on the flow cytometer at 3 days post infection (dpi) (Figure 7B). All vectors exhibited comparable EGFP expression with no significant differences in the intensity. EGFP expression in these transduced cell populations was followed in several time points by flow cytometry. Even at 30 dpi, proviruses were stably expressed in every IN group (Figure 7C).

To observe smaller changes in expression, we followed the expression stability of proviruses in EGFP-positive population sorted at 3 dpi (Figure 7D). In the least stable population transduced with the IN^{WT} vector, EGFP expression dropped to more than 75% of EGFP-positive cells at 14 dpi. However, this EGFP-positive population seemed stable as we did not observe further decrease of expression at 23 dpi.

Next, we shifted from polyclonal populations to clonal populations which were generated by sorting EGFP-positive cells at 3 dpi in a single-cell mode (Figure 7E). The majority of 200 expanded clones in every group exhibited more than 90% EGFP-positive cells at 30 dpi. We observed significant silencing (less than 90% EGFP-positive cells per clone) only in 11 clones (5.5%) transduced with the IN^{WT} vector. We did not observe differences in mean fluorescence intensity distribution among IN groups (Figure 7F).

The MLV retained transcriptional stability of expression after long-term culture. We observed significant silencing only in clonal populations after transduction with the IN^{WT} vector.

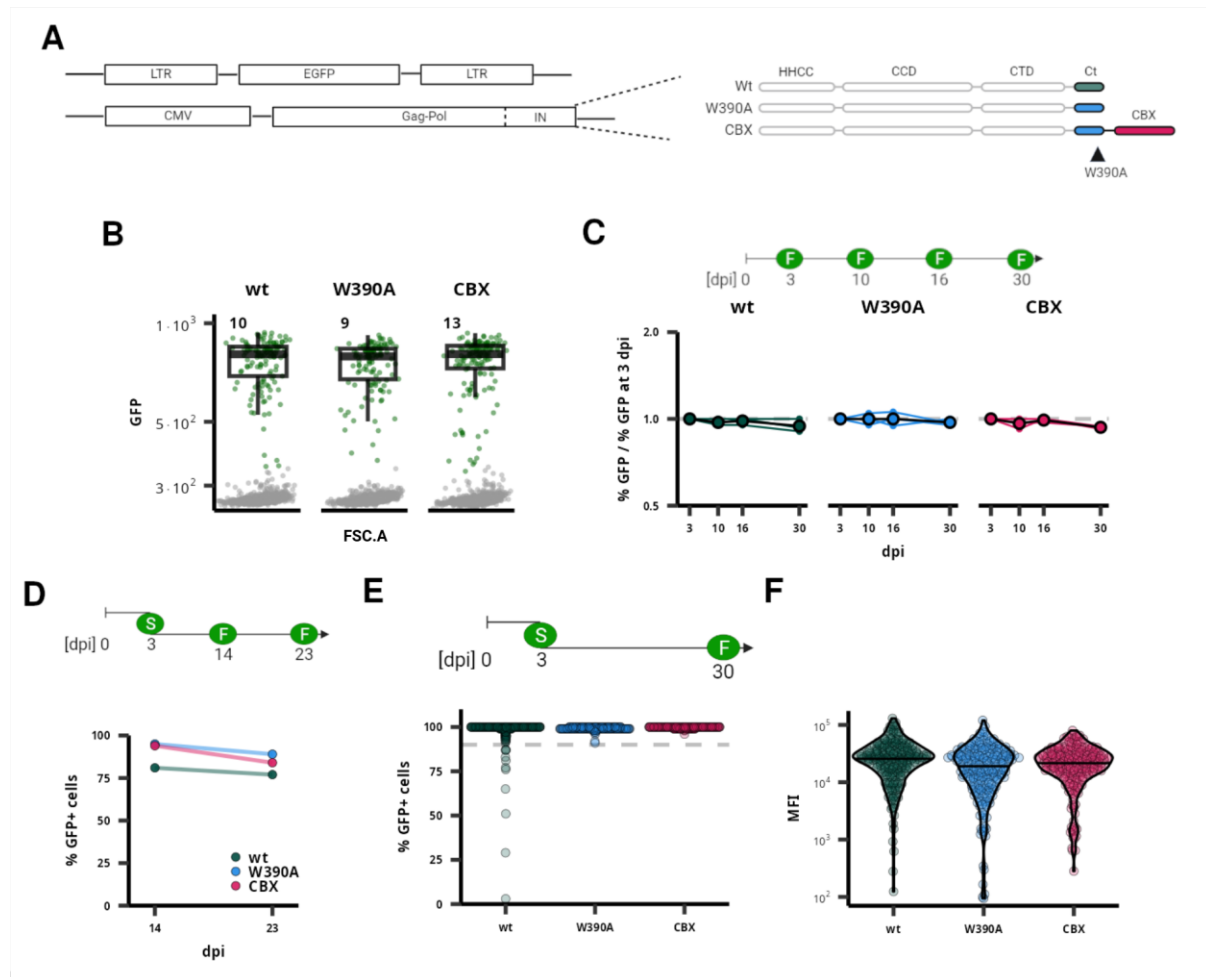


Figure 7. MLV-derived vectors remain transcriptionally stable after retargeting. A) A schematic representation of vectors used for transduction. B) A dot plot of transduced cells at 3 dpi (2,400 live cells for every integrase group). The numbers show the percentage of EGFP-positive cells. C) Expression stability of proviruses in polyclonal populations. Fold change in the percentage of EGFP-positive cells in different time points as compared to the percentage of EGFP-positive cells at 3 dpi. Y-axis is in the log₂ scale. D) The percentage of EGFP-positive cells in the populations sorted for the expression of EGFP at 3 dpi. E), F) Expression of EGFP in clones expanded from EGFP-positive cells single-cell sorted at 3 dpi. E) Fraction of EGFP-positive cells at 30 dpi. The dashed line marks 90% of EGFP-positive cells per clone. F) The mean fluorescence intensity (MFI) in the clones from panel E. (Timelines: selection (S) and flow cytometry (F) at the time indicated under.)

4.1.2 Retargeted integration sites are present in stably expressed population

It was shown that the Bin INs, IN^{W390A} and IN^{CBX} , channel integrations further from active promoters and enhancers (El Ashkar et al. 2017). The shift from IN^{WT} -preferred integration sites towards repressive chromatin was maintained in the populations of transcriptionally active SIN MLV-derived proviruses with internal promoters (Van Looveren et al. 2021). The polyclonal populations transduced with the IN^{WT} and Bin INs vectors showed no expression differences (Figure 7D). So, we scrutinized the integration sites distribution differences of MLV-derived proviruses in the populations of EGFP-positive cells sorted at 3 dpi.

First, we determined the distance of the integration sites to the nearest chromatin segment (Ernst and Kellis 2010). We used Impact effect-size analysis (Lötsch and Ultsch 2020) to assess the significance of differences in integration sites distribution (Figure 8). With the expected highest differences between wt and CBX integration sites datasets, we calculated Impact for these two datasets. Higher Impact, above 0.5, was observed for active transcription start sites, Tss and TssF (F = flanking), enhancers, Enh and EnhF, and Gen5, which is a chromatin segment marked with tri-methylation of lysine 36 on histone H3 and methylation of lysine 4 on histone H3 (Hoffman et al. 2013), histone modifications associated with 5' end of transcriptionally active gene bodies. More specifically, the median distance to the Tss was 0.8 kb, 11.4 kb, and 22.6 kb in the IN^{wt} , IN^{W390A} , and IN^{CBX} integration sites, respectively, and to the Enh was 1.8 kb, 8.5 kb and 15.6 kb in the IN^{wt} , IN^{W390A} , and IN^{CBX} integration sites, respectively (Figure 9). In general, proviral integration sites were distributed closer to segments associated with active regulatory elements after transduction with the IN^{WT} vector, while after transduction with the IN^{CBX} vector, we observed the opposite effect with integration sites distributed closer to segments associated with transcriptionally silent chromatin.

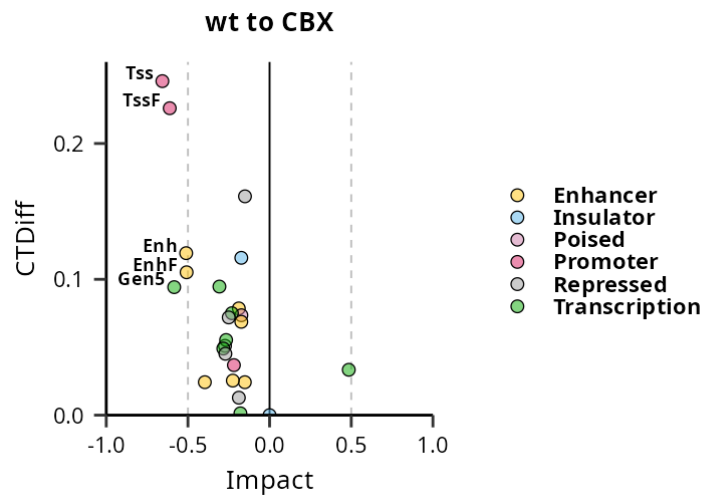


Figure 8. Impact effect-size analysis (Lötsch and Ultsch 2020) of integration sites distances to chromatin segments between IN^{WT} and IN^{CBX} . Grouped segments are differentiated by colors. CTDiff = Difference of the central tendency of distribution.

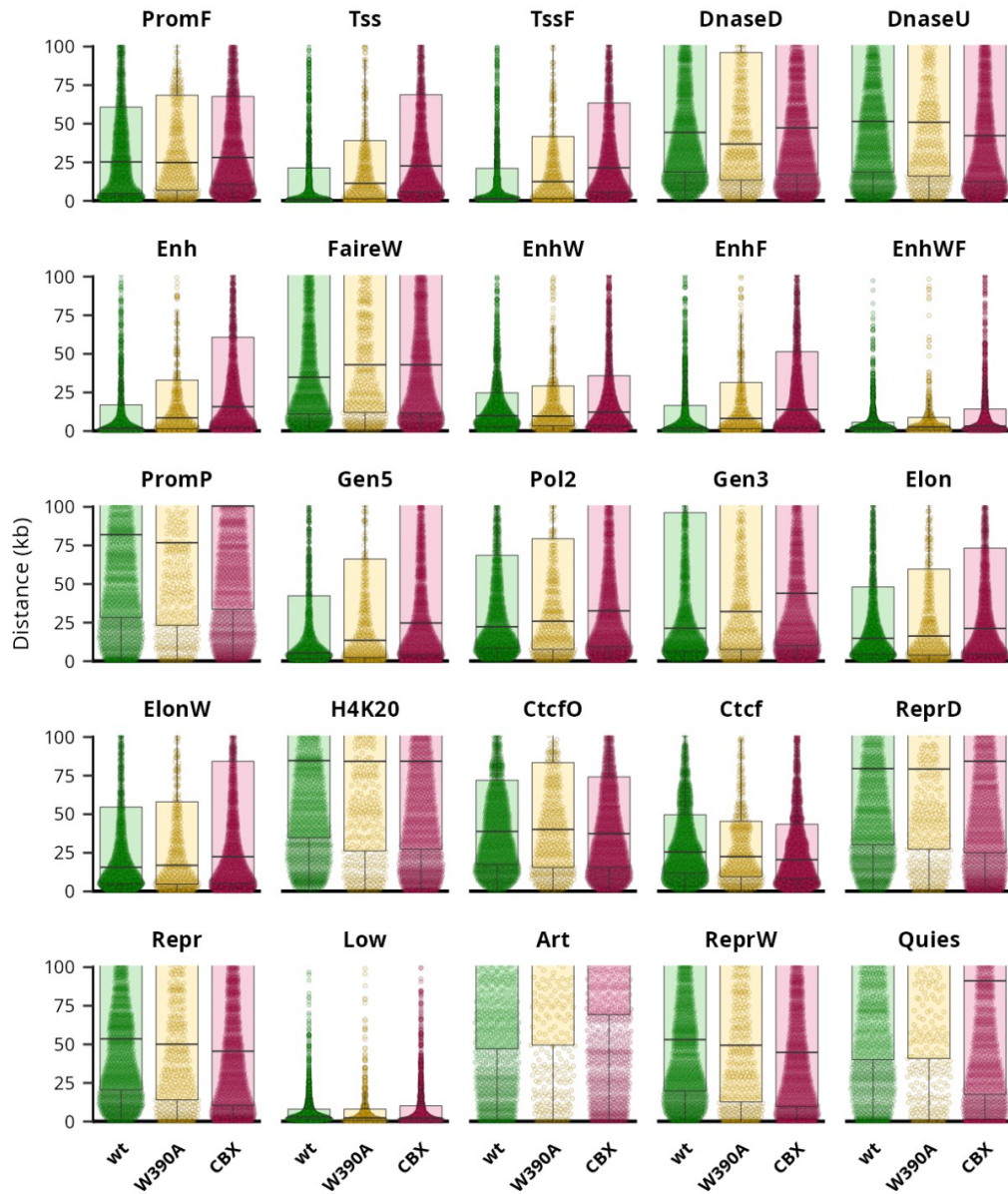


Figure 9. Integration sites distances of integration sites to chromatin segments. Medians and quartile range of the distance distributions are shown. PromF = promoter flanking. Tss = transcriptional start site. TssF = transcriptional start site flanking. DnaseD = DNaseI sensitivity data from Duke University. DnaseU = DNaseI sensitivity data from University of Washington. Enh = enhancer. FaireW = formaldehyde-assisted isolation of regulatory elements. EnhW = weak enhancer. EnhF = strong enhancer. EnhWF = weak enhancer flanking. PromP = poised promoter. Gen5 = 5' end of active gene bodies. Pol2 = RNA polymerase II. Gen3 = 3' end of active gene bodies. Elon = transcriptional elongation. ElonW = weak transcriptional elongation. H4K20 = enrichment of histone modification on lysine 20 on histone H4, associated with transcriptional activation. Ctcf = Ctcf binding sites. CtcfO = Ctcf binding sites, also DNaseI-sensitive sites (= open). Repr = repressed state by Polycomb repressor complex 2. ReprD = repressed state, also DNaseI-sensitive sites (DnaseD). ReprW = weak repression. Low = domains with low activity near active elements. Quies = quiescent, with very little signal of any feature. Art = artifacts.

While evaluating integration sites distribution in relation to the nearest chromatin segments is a common practice, many integration sites might not be functionally related to these segments. Thus, we analyzed distribution of the integration sites within chromatin A/B subcompartments and LADs, which are associated with silent chromatin (Figure 10A). After transduction with the IN^{WT} vector, 68% and 17% of integration sites occurred in transcriptionally active A1 and A2 subcompartments, respectively. Even after transduction with the IN^{CBX} vector, the majority of integration sites were still present in A1 and A2 subcompartments. More specifically, 56% and 21% of proviruses were located in A1 and A2 subcompartments, respectively. The IN^{CBX} vector showed higher frequency of integration into B subcompartments (more than 2-fold) and LADs (more than 3-fold) in comparison to the IN^{WT} vector (Figure 10B). However, even with the retargeting INs, the integration sites of active proviruses exhibited non-random patterns (Figure 10C). Almost 20% of random integration sites were distributed in the A1 subcompartment and almost 50% in LADs.

Integration sites or retargeted proviruses still occurred mainly in transcriptionally active chromatin of A subcompartments. The biggest shift was observed in the population with the IN^{CBX} vector into LADs.

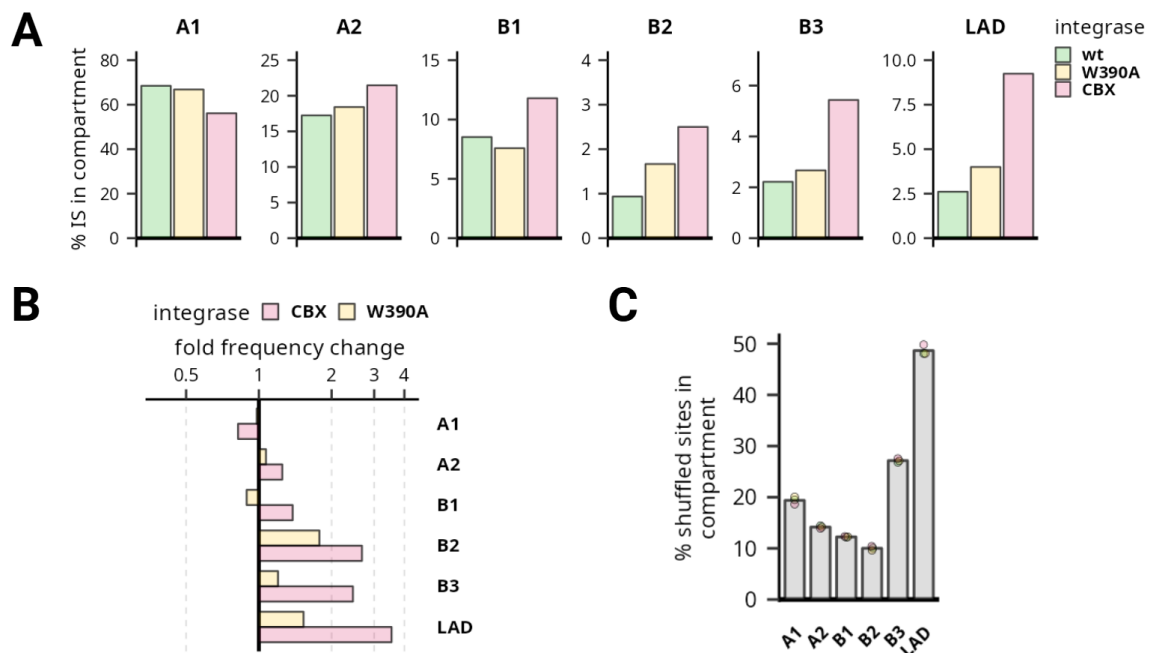


Figure 10. Integration sites analysis of active MLV-derived proviruses after retargeting. A) Frequency of integration site in A/B subcompartments and lamina-associated domains (LADs). B) Fold frequency changes of integrations into A/B subcompartments and LADs with IN^{W390A} and IN^{CBX} in comparison to IN^{WT} . X-axis is in the log₂ scale. C) Frequency of computed random integration sites in A/B subcompartments and LADs.

4.1.3 Stable expression after retargeted integration is a general feature of gammaretroviruses

The unstructured Ct of IN^{WT} is highly conserved among different γ RVs, resulting in the identical integration preference (El Ashkar et al. 2014; Gupta et al. 2013; Gilroy et al. 2016). However, expression stability of other non-MLV γ RVs is poorly studied. We then asked whether previously observed high transcriptional stability is specific to MLV or a general feature of γ RVs. For that purpose, we constructed vectors derived from the pLG vector with LTRs derived from MoMLV, FeLV, SNV, KoRV, and CrERV expressing EGFP fused with the destabilization domain (d2GFP) (Figure 11A).

At 3 dpi, we observed that the vectors equipped with FeLV and SNV LTRs achieved the levels of intensity comparable to the MoMLV-derived vector. After transduction with vectors equipped with MoMLV and SNV LTRs and the IN^{CBX} vector showed slight decrease in d2GFP intensity in comparison to the IN^{WT} vector, while FeLV LTRs with the IN^{CBX} vector exhibited increased d2GFP intensity. In contrast to MoMLV, FeLV and SNV LTRs, we observed only few transduced cells with low intensity of d2GFP with KoRV LTR-equipped vector and failed to effectively transduce the cells with the CrERV LTR-equipped vector (Figure 11B).

We followed the proviral expression of polyclonal populations for two weeks (Figure 11C). While MoMLV and FeLV did not display decrease in number of d2GFP-positive cells, we observed a drop of d2GFP-positive cells in the SNV-transduced population. We also observed a high fluctuation of d2GFP-positive cells proportion in the population transduced with KoRV LTR-equipped vectors at different time points. This fluctuation might be given by a low number of initially transduced cells.

Then, we determined the copy number (CN) of d2GFP per genome (CN of provirus/2x CN of *RPP30*) in the individual populations by ddPCR. Normalizing the d2GFP CN to the proportion of d2GFP-positive cells measured by flow cytometry, we estimated the ratios of active proviral genomes in the given populations (Figure 11D). The highest fraction of transcriptionally active proviruses was observed in the MoMLV LTR-equipped vector with the IN^{WT} vector group with nearly 80% of active proviruses. The MoMLV LTR-equipped vector with the Bin vectors exhibited a decrease in active population proportion to approximately 60%. FeLV, SNV, and KoRV LTR-equipped populations did not exhibit such variability among IN variants. Based on the estimated proportions of active proviruses (less than 80%), we suggest that a significant population of silenced proviruses was present in transduced cells irrespective of the used vectors.

The MLV-based virus-producing system might be used to produce FeLV and SNV LTR-equipped vectors. We revealed previously unnoticed fractions of transcriptionally silent proviruses. Provirus which manage to establish expression remained mostly transcriptionally stable even after retargeting.

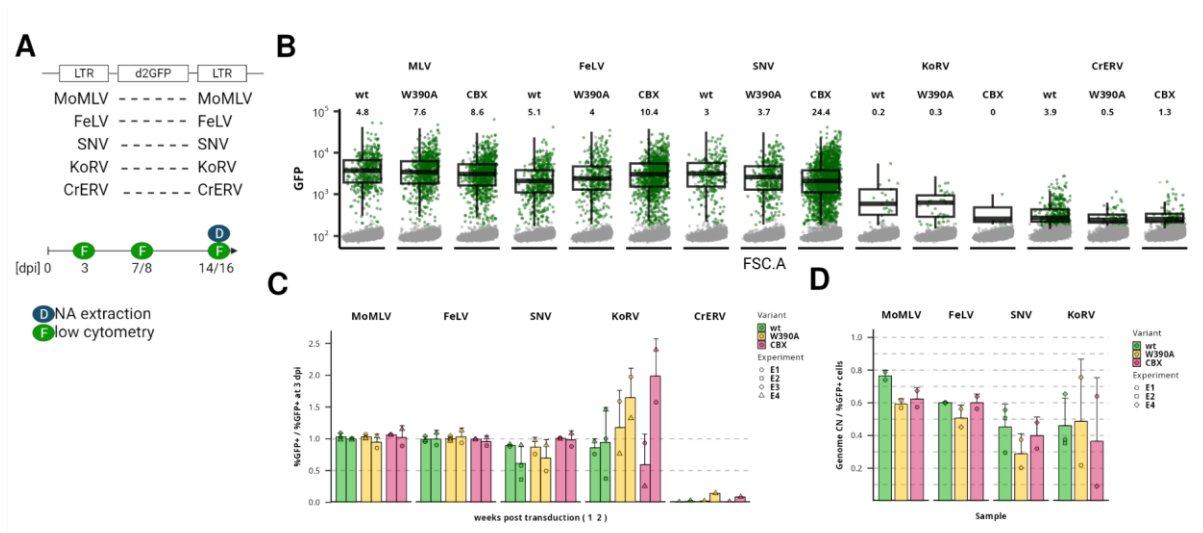


Figure 11. Stability of gammaretroviral LTR-driven expression after retargeting. A) A schematic representation of constructed vectors. B) A dot plot of transduced cells (10,000 cells per sample) with the percentage of d2GFP-positive cells at 3 dpi. The median and quartile range of d2GFP intensity for the d2GFP-positive population are shown. Area of the forward scatter (FSC.A) signal is shown on the x-axis. Both axes are in the log₁₀ scale. C) Fold change in proportion of d2GFP-positive cells, relative to the proportion of d2GFP-positive cells at 3 dpi. Two time points marking one and two weeks after transduction are shown. D) Estimated ratios of active proviral genomes. The estimation was calculated as a ratio of the percentage of d2GFP-positive cells two weeks after transduction measured by flow cytometry to a copy number of *d2GFP* copies per 100 cells.

4.1.4 Stability of expression almost randomly distributed γ RV promoters in the context of α RV SIN vectors

Previously, we tested transcriptional stability of the vectors equipped with full-length γ RV LTRs. In the next step, we decided to test the expression stability regulated by the 3' unique (U3) part of LTRs, acting as an inner promoter, in ASLV-derived SIN vectors (AS) (Suerth et al. 2012). The AS vectors have almost random integration preferences (Mitchell et al. 2004). The AS. γ RV.d2GFP vector originally equipped with strong internal γ RV SFFV promoter was modified with MoMLV, FeLV, SNV, KoRV, or CrERV LTR-derived U3/R regions (Figure 12A). The K562 cell line was transduced with respective AS-derived vectors and first measured by flow cytometry at 3 dpi (Figure 12B). The d2GFP-positive cells in AS.MoMLV-, AS.SFFV-, and AS.FeLV-transduced populations exhibited similar levels of expression intensities. The AS.KoRV-transduced cells reached the lowest expression intensity of d2GFP with median of fluorescence intensity ~5-fold lower than the median fluorescence intensity of AS.MoMLV-transduced cells. The AS.SNV- and AS.CrERV-transduced cells reached the median of d2GFP fluorescence intensity ~2-fold lower than the median fluorescence intensity of AS.MoMLV-transduced cells. The experiment was performed five-times with different MOI, the measured cells by flow cytometry in dot blots of transduced cells represent one of the MOI. Then, we followed the fold change in proportion of d2GFP-positive cells, relative to the proportion of d2GFP-positive cells at 3 dpi, in samples with different MOI and at different time points (Figure 12C). First, we observed slight increase of d2GFP expression in AS.MoMLV, AS.SFFV, and AS.FeLV vector populations in comparison to d2GFP expression at 3 dpi. We observed stable levels of d2GFP-expressing cells in AS.KoRV-transduced cells, while in AS.CrERV-transduced population, we observed slow decrease of d2GFP-positive cells starting at 17 dpi and reaching 83.5% at 31 dpi. The AS.SNV-transduced population exhibited fast decrease of d2GFP-expressing cells falling to 37.5% of d2GFP-positive cells at 31 dpi. Next, we estimated the proviral CN. We performed ddPCR on the genomic DNA collected at 14 dpi from the populations transduced with low MOI. While the levels of d2GFP-positive cells were similar in all samples at 14 dpi, there were differences in proviral CN (Figure 12D). Estimated ratios of active genomes (Figure 12E) were highest in the AS.SFFV and AS.FeLV-transduced populations reaching 80-90% of active proviruses. The population transduced with the AS.MoMLV vector consisted of 60% of active proviruses, which is comparable to the MLV-derived vector with retargeting Bin INs (Figure 11D). The lowest proportion of active proviruses was observed in the population transduced with the AS.SNV vector. γ RVs retain high stability of expression as internal promoters in vectors with almost random integration pattern. We estimated the highest proportion of active proviruses (80-90%) in populations transduced with the AS.SFFV and AS.FeLV vectors, and the lowest proportion (~40%) in populations transduced with the AS.SNV vector.

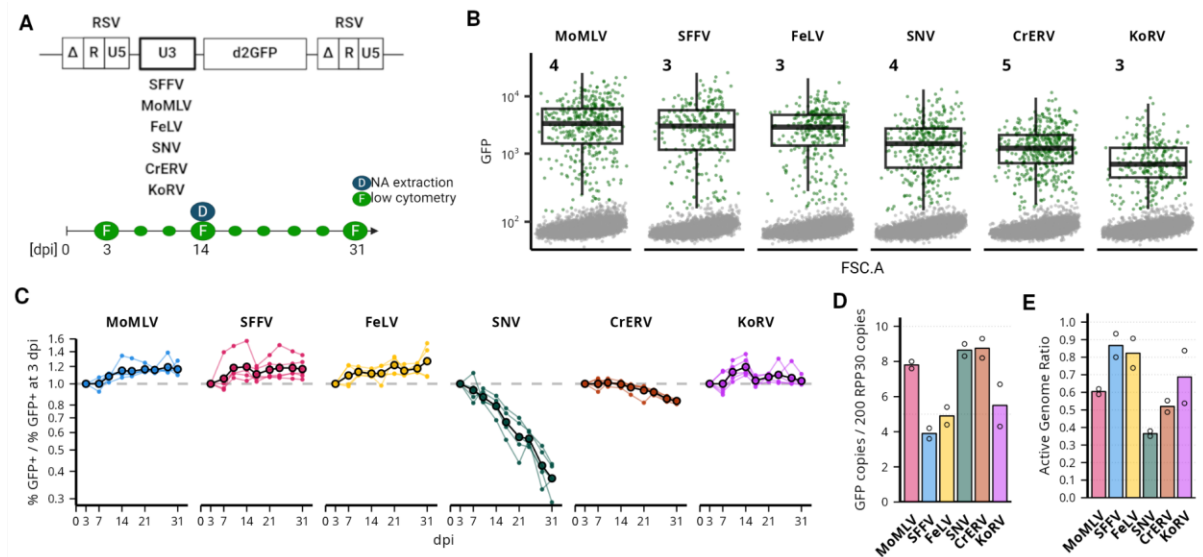


Figure 12. Transcriptional stability of αRV vectors carrying γRV LTRs as internal promoters. A) A schematic representation of αRV SIN vectors with internal promoters derived from the U3 part of studied LTRs and the timeline of the experiment. B) Intensity of d2GFP expression (10,000 cells) and the percentage of d2GFP-positive live cells at 3 dpi. C) Fold change of expression stability during the long culture of transduced cells. Light lines and points represent individual experiments with different multiplicities of infection. Black-outlined points represent the average values. D) *d2GFP* copies per 200 copies of *RPP30* (reference gene) by the droplet digital PCR from genomic DNA collected at 14 dpi (panel C). E) Estimated ratios of active proviral genomes. Calculated as a ratio of the percentage of d2GFP+ cells per *d2GFP* copy number per 100 cells.

4.1.5 Gammaretroviruses display stable expression even after targeted insertion into LADs

In the previous transduction experiment, we observed higher frequency of transcriptionally active proviruses in LADs in the MLV IN^{CBX}-transduced population in comparison to the MLV IN^{WT}-transduced population (Figure 10A). We decided to directly test expression stability of γ RV-derived vectors in the restrictive environment of LADs by CRISPR-Cas9-induced homologous recombination-mediated insertion in the K562 cell line.

First, we selected inter-genic LAD B2/B3 subcompartments and designed guide sequences for the CRISPR-Cas9 system. Five domains were selected randomly (domains 1-5) and two domains were previously hit by the retargeted and stably expressed MLV-derived vectors (domains 6 and 7). We also tested the domain with previously reported targeted insertion (reg9) (Tasan et al. 2018) and used the previously published guide sequence into the gene body of transcriptionally active gene *IFT20* (Kato et al. 2017), which served as a positive control of proviral expression after targeted insertion.

Cutting efficiency of designed gRNAs was tested. Out of the total 23 Domain Target (*DoT*) sites, we obtained four *DoT* sites with observed indels (the percentage of amplified target sites with indels in the CRISPR-Cas9 treated populations): *DoT*:3.1 (14%), *DoT*:6.2 (81%), *DoT*:6.3 (72-74%), and *DoT*:7.3 (34%) (Figure 13A). The first number in the name of *DoT* represents the domain, the second number represents the number of tested gRNA within the domain.

To further evaluate and confirm the LADs-associated features of four *DoT* sites, we checked the ploidy by ddPCR (Figure 13B) and the levels of Lamin B by ChIP-ddPCR (Figure 13C). Apart from *DoT*:7.3, we estimated that other target sites have the same CN as a *RPP30* (reference gene). The LAD *DoTs* exhibited increased levels of Lamin B in comparison to the active genes *IFT20* and *RPP30*. The differences between samples (1) and (2) are given by the different length of fragments before IP.

To insert an expression vector into selected sites, we co-nucleofected the pX458 and the γ RV-derived vector. The particular pX458 vectors carried the sequences for expression of EGFP, Cas9, and gRNAs specific to the selected *DoTs* and *IFT20* gene sites. The γ RV-derived vectors expressed mCherry under control of retroviral LTRs derived from MoMLV, FeLV, KoRV, and ASLV. We observed high expression after transduction with the MoMLV and FeLV-derived vectors and low expression after transduction with the KoRV-derived vector (Figure 11B). The ASLV is expected to be transcriptionally active in mammalian cells only near H3K4me3-marked active promoters (Senigl et al. 2012). The RV LTRs-equipped expression cassettes were flanked with target sites-specific homologous arms for homologous recombination-mediated insertions of expression cassettes into the host cell genome.

We examined 301 randomly selected clones for targeted insertions by amplification of the upstream and downstream genome-insert junctions. We verified targeted insertion in 106 γ RV LTR-containing clones and only in five ASLV LTR-containing clones.

The insertion-verified clones were further analyzed for mCherry CN by ddPCR (Figure 14C). The median mCherry CN per genome (*RPP30* sequence) in clones with *DoT* sites was ~ 2 , while in the *IFT20*-targeted clones it was 3.5. We obtained 30 clones with a single-targeted site ($CN < 1.5$) and 35 clones with two targeted sites ($1.5 \leq CN < 2.5$). We obtained single-copy clones with all 3 γ RV LTR-containing vectors only for *DoT*:7.3. The majority of γ RV LTR-containing clones with a single-targeted site showed stable expression of mCherry more than a month after nucleofection (Figure 14D).

We successfully inserted RV LTR-equipped vectors into B2/B3 subcompartments of LADs. Even in such a hostile environment to transcription, we observed long-term transcriptionally stable insertions in clones targeted with γ RV LTR-containing cassettes. As expected, we observed transcriptionally silent ASLV LTR-containing cassettes in *DoTs* and an exon of active gene *IFT20*.

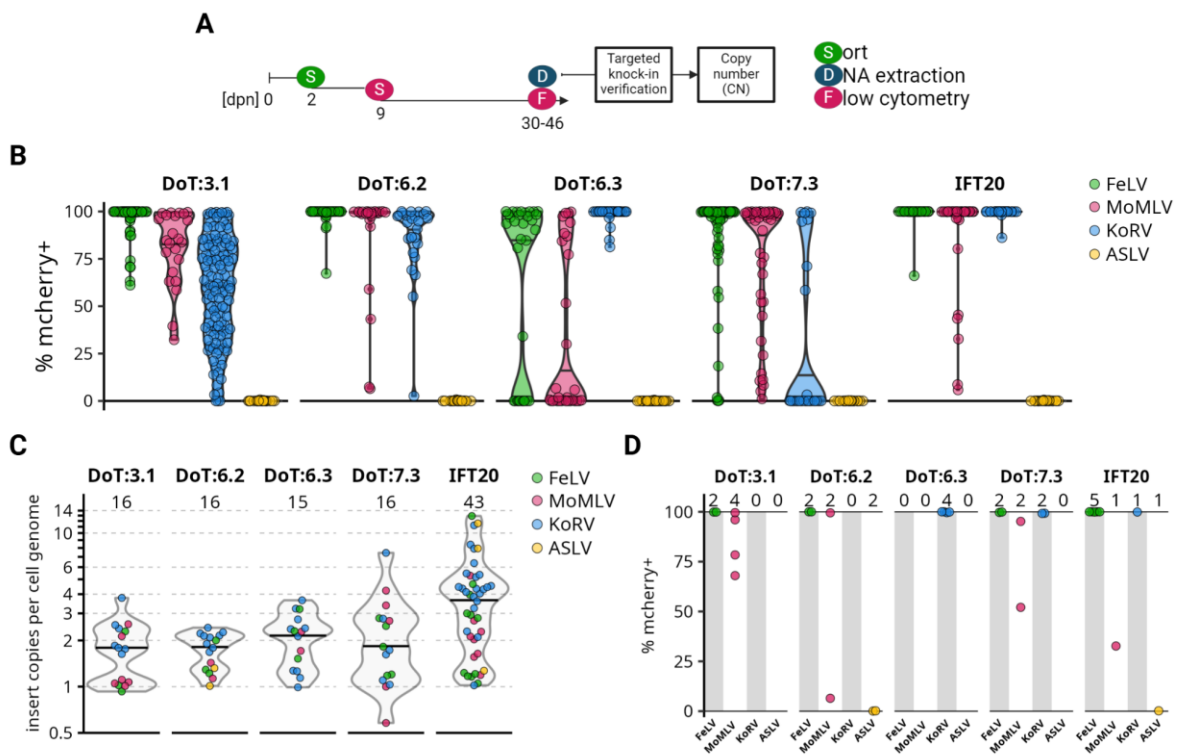


Figure 14. Targeted insertions in the K562 cell line. A) The scheme of the experimental workflow. At two days post nucleofection (dpm), the EGFP-positive population (CRISPR-Cas9 treated) was sorted, then at 9 dpm, the mCherry-positive cells (LTR-controlled expression) were sorted on a 96-well plates in a single-cell mode. After expansion, the clones were measured on a flow cytometer and clonal DNA was isolated. Clones were tested for targeted insertions by amplification of insert-genome junctions and clones with verified insertions had the copy number of mCherry determined. B) % of mCherry-positive cells (y-axis) per clone after expansion measured at 30-46 dpm. Every dot represents a single clone. C) Copy number (CN) of mCherry in clones with verified targeted insertions by PCR amplification and sequencing. Ratio of mCherry CN to 2 copies *RPP30*. D) % of mCherry in clones with a verified single insertion. The numbers in panels C) and D) indicate the number of clones.

4.1.6 FeLV-derived vector's targeted knock-in into the Jurkat cell line

After successful insertions in the K562 cell line, we decided to use the same approach to insert FeLV LTR-equipped expression cassettes into B2/B3 subcompartments of LADs in the widely-used and well-characterized T cell line Jurkat.

From the *DoTs* utilized in targeted knock-in in the K562 cell line, only *DoT:7.3* is predicted as an intra-LAD site. However, we did not obtain the insertions at *DoT:7.3* site in the Jurkat cell line.

Next, we designed and tested gRNAs into other two randomly-selected intergenic intra-LAD domains (*DoT:8* and *DoT:9*). Out of the six tested gRNAs, we observed indels with 4 of them: *DoT:8.1* (36%), *DoT:8.2* (2%), *DoT:8.4* (74-77%), and *DoT:9.2* (61-75%).

Then, we followed the same protocol of targeted knock-in as described with the K562 cell line (Figure 14A). At 36 dpn, we measured the expression of mCherry in 163 expanded clones in total (Figure 15A). We obtained 47, 46 and 61% of stable clones ($\geq 90\%$ mCherry-positive cells per clone) after targeting *DoT:8.1*, *DoT:8.2* and *DoT:8.4*, respectively, and 69% of stable clones after targeting *DoT:9.2*.

We managed to verify targeted insertions by genome-insert junctions' amplification only in *DoT:8.2* and *DoT:9.2* sites in 4 and 21 clones, respectively. In other *DoTs*, we did not amplify the insert-genome junctions or manage to sequence the products of amplification.

Next, in clones with verified targeted insertions, we determined mCherry CN (Figure 15B). For *DoT:8.2*, the median was ~ 1.5 , while with *DoT:9.2*, the median was more than 2. We obtained two and nine clones at *DoT:8.2* and *DoT:9.2*, respectively, with a single insertion per cell (Figure 15C). While the expression cassettes at *DoT:8.2* were not transcriptionally stable ($\geq 90\%$ mCherry+ cells per clone), the majority of LTR-driven expression cassettes at *DoT:9.2* were transcriptionally stable.

We managed to repeat targeted insertions of FeLV LTR-controlled expression cassettes into LADs in the Jurkat cell line. Again, we observed single-copy transcriptionally active insertions in LADs.

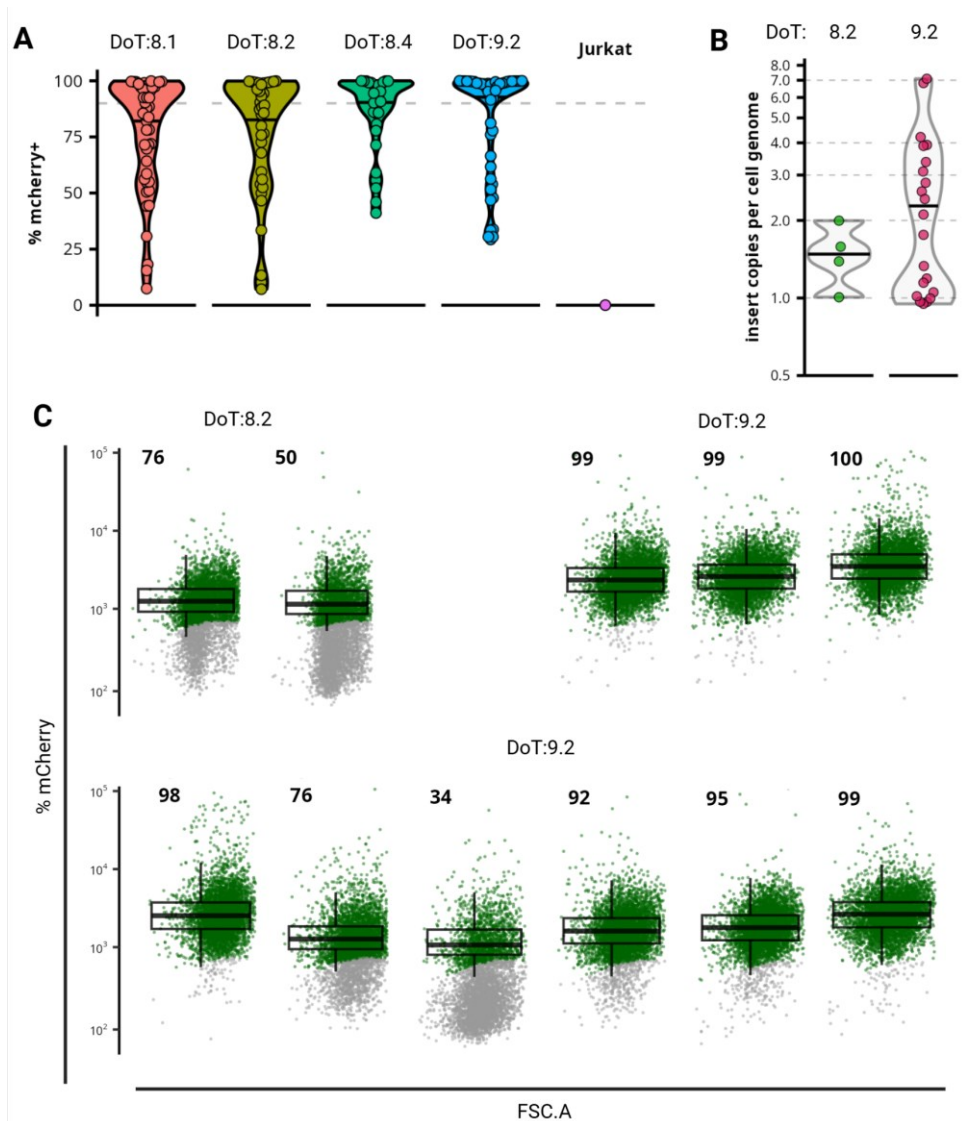
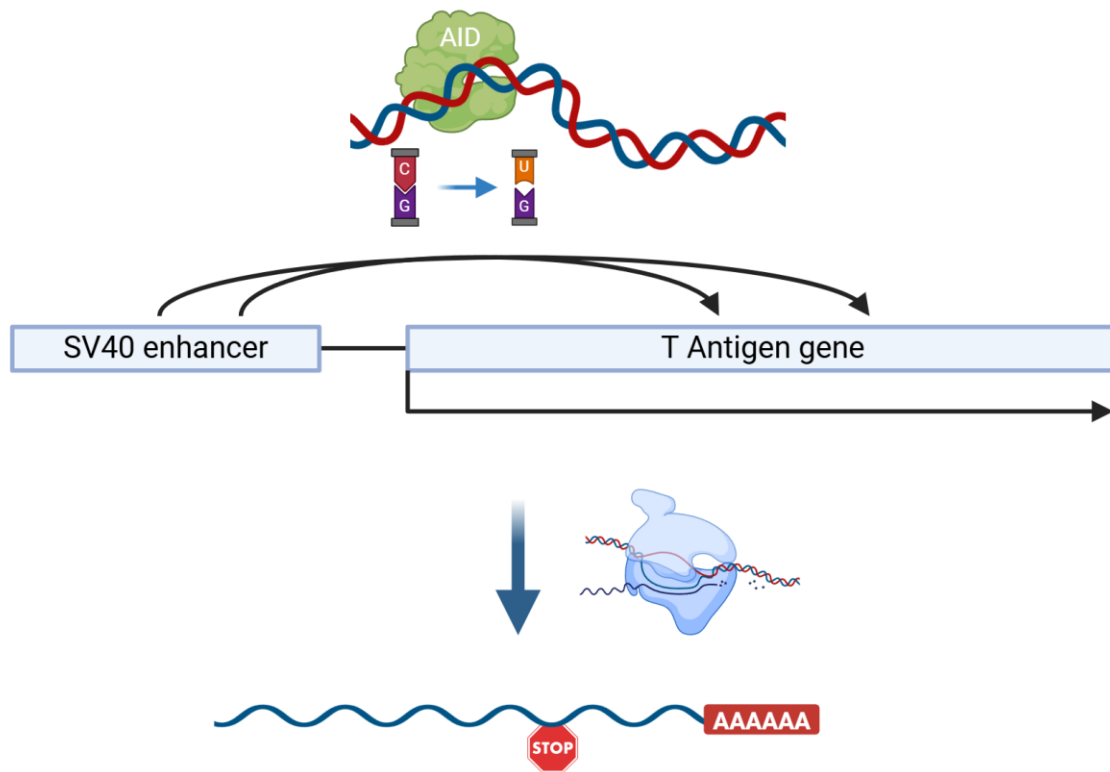


Figure 15. Targeted insertions in the Jurkat cell line. A) The percentage of mCherry-positive cells in expanded clones and negative the Jurkat cell line at 36 dpn. Every dot represents a single clone. B) mCherry copy number per genome (2x CN *RPP30*) in clones with verified targeted insertions. C) Intensity of mCherry expression (10,000 cells) and the percentage of mCherry-positive live cells at 36 dpn in clones with a single insertion. Two DoT:8.2 clones at the top left, nine Dot:9.2 clones at the top right and bottom line. Numbers above dot blots represent the percentage of mCherry-positive cells.

4.2 The SV40 enhancer as an AID-mediated somatic hypermutations targeting element



4.2.1 SV40 enhancer possesses SHM targeting activity in B cell lines

The enhancers and enhance-like sequences of *Ig* genes were shown to target SHM on *Ig* genes during the B cell development to produce high-affinity antibodies. They contain several TFBSs (Figure 16A), which were shown to be important for targeting of SHM in B cells (Kohler et al. 2012; McDonald et al. 2013; Buerstedde et al. 2014; Dinesh et al. 2020). Since the *LTag* of MCPyV was shown to be targeted by cytidine deaminases (Soikkeli et al. 2022), here we tested whether SV40 enhancer (SV40e), which is within SV40 NCRR (Figure 16B) possesses SHM targeting properties.

The TFBSs present in the *Ig* enhancers are also in the SV40e (Figure 16C). Isolated viral strains may contain one or two repeats of 72bp sequence (Lednicky et al. 1998), and therefore are marked as 1x72bp SV40e or 2x72bp SV40e, respectively, in this work. Apart from 1x72bp and 2x72bp, we also cloned duplicated 2x72bp SV40e (= two 2x72 bp SV40e) (Figure 16C), and various 2x72bp TFBS mutants to test the effects of the SV40e TFBSs on SHM activity (Figure 16D). The mutations were introduced individually, E-box1 (mut E1), E-box2 (mut E2), and NFkB (mut NFkB), or in combination, E-box1 with E-box2 (mut E1+E2) and IRF-Ets together with E-boxes, and NFkB (IPEN). Together with the SV40e, we also tested two human, BLK and CD83, and two mouse, Pou2af1 and Argef1, enhancers for their SHM targeting activity as a control. All SV40 and genomic enhancers were cloned into GFP fluorescence loss assay vectors (Blagodatski et al. 2009; Buerstedde et al. 2014; Senigl et al. 2019). GFP4 and GFP7 vectors contain hypermutation target sequences fused with GFP, which is designed to sensitively detect AID activity by generating stop codons upon cytidine mutations (Figure 16E).

sequence with marked TFBS. Boxes represent 72bp repeats. Nucleotides in mutated TFBSs are indicated under the original sequence. E) GFP reporters used in GFP loss assay with marked insertion site of tested enhancers. RSV (Rous sarcoma virus promoter), SIN LTR (self-inactivating long terminal repeats), CMV (cytomegalovirus promoter), HTS (hypermutation target sequence), T2A (self-cleaving T2A peptide), IRES (internal ribosome entry site), BsR (blasticidin resistance), WPRE (woodchuck hepatitis virus posttranscriptional regulatory element).

First, the DT40 cells were transfected with the GFP4 loss assay vectors (Figure 17A). The 2x72bp SV40e and two 2x72bp SV40e increased the GFP fluorescence loss 25- and 39-fold, respectively, when compared to the empty reporter vector. The human Ig λ and chicken Ig λ enhancers similarly increased SHM GFP loss by 10- and 35-fold, respectively. The mutation of E-box2 led to similar GFP fluorescence loss compared to 2x72bp SV40e. Mutations of E-box1, combined mutations of E-box1 and 2, and mutations of NF κ B binding site exhibited significantly decreased SHM activity compared to the 2x72bp SV40e. The combination of mutations in the IPEN element resulted in a complete loss of GFP fluorescence loss. Similarly, the tested genomic enhancers did not exhibit SHM targeting activity. Next, in the DT40 with the GFP2 loss assay vectors, there was a 5-fold decrease with 1x72bp SV40e in comparison to two 2x72bp SV40e (Figure 17B).

When performing GFP7 assay in the Ramos cell line (Figure 17C), two 2x72bp, 2x72bp and 1x72bp SV40e exhibited 88-fold, 43-fold, and 8.5-fold increase of GFP fluorescence loss compared to the empty vector. The two 2x72bp SV40e in comparison to 2x72bp SV40e exhibited 2-fold increase of SHM targeting activity, and 2x72bp SV40e in comparison to 1x72bp SV40e exhibited 5-fold increase.

Measuring the expression levels with luciferase assay in DT40 cells, we observed comparable levels between 1x72bp and 2x72bp SV40e vectors. While gene expression with 2x72bp IPEN SV40e was greatly reduced (Figure 17D).

By utilization of GFP fluorescence assays, we revealed that SV40e possesses SHM targeting activity in the B cell lines while underscoring the importance of TFBSs. In contrast, selected genomic enhancers lack SHM targeting activity.

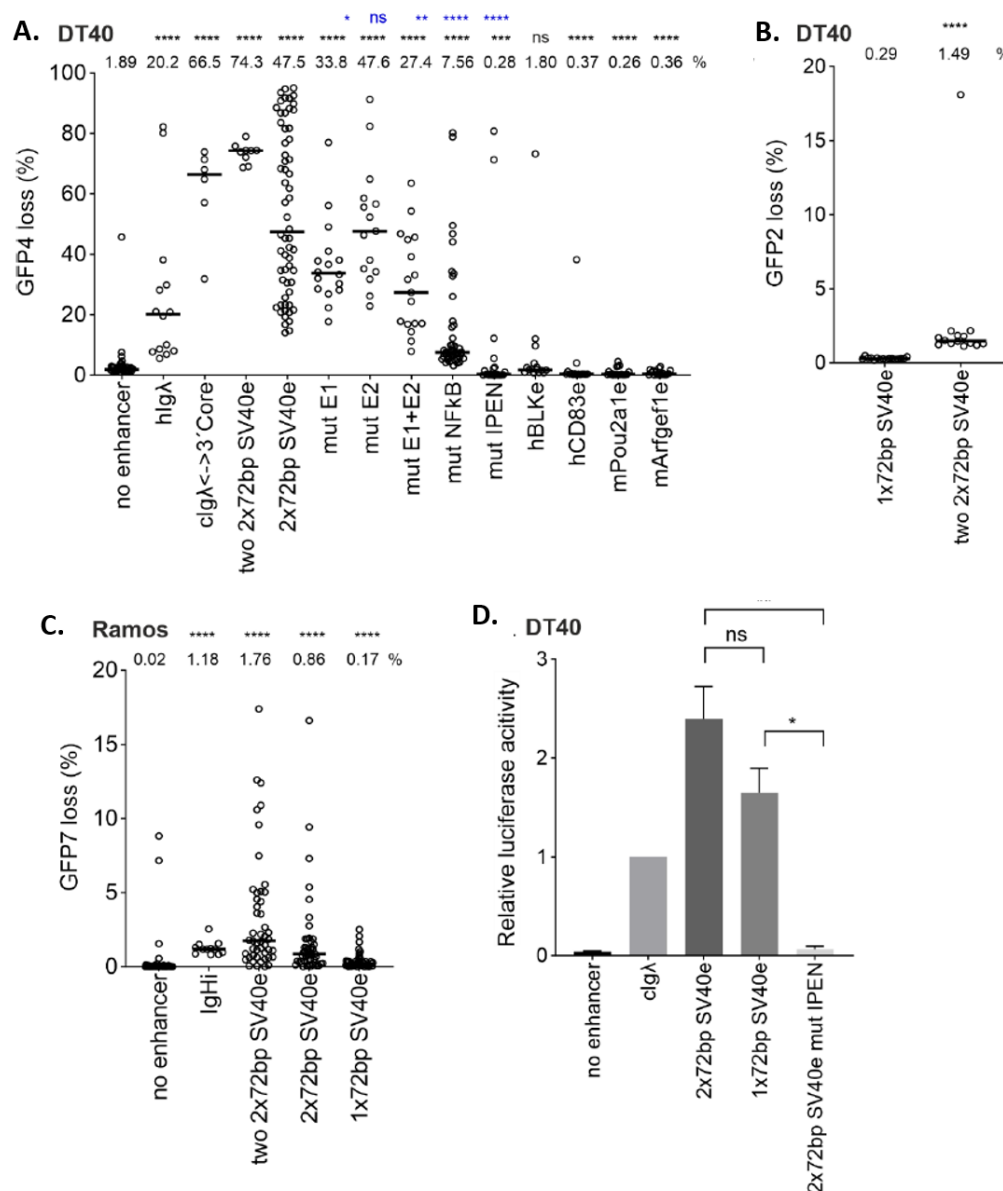


Figure 17. SV40e SHM targeting activity in B cells. A) SHM targeting activity of human Igλ (hlgλ) enhancer, chicken Igλ enhancer 3' core (clgλ<->3'Core), human genomic enhancers BLK and CD83 (hBLKe and hCD83e0, mouse genomic enhancers Arfgef1 and Pou2af1 (mArfgef1e and mPou2af1e), and SV40e variants measured in GFP4 assay in DT40 cells. Statistical significance to negative control is indicated in black and to 2x72bp enhancer in blue (Mann-Whitney U-test). The numbers indicate % median GFP loss. Mut E1 (mutation in E-box1), mut E2 (mutation of E-box2), mut E1 + E2 (combined mutation of E-boxes), mut NFkB (mutation of NFkB binding site), mut IPEN (mutation of IRF-Ets, E-box and NFkB binding sites). B) SHM targeting activity measured by GFP2 assay in DT40 cells. Statistical significance calculated by Mann-Whitney U-test. The numbers indicate % median GFP loss. C) SHM targeting activity measured by GFP7 assay in Ramos cells with human Ig heavy intronic (IgHi) enhancer, 1x72bp SV40e, 2x72bp SV40e and two 2x72bp SV40e. Statistical significance calculated by Mann-Whitney U-test. The numbers indicate % median GFP loss. D) Enhancer activity of selected enhancers measured by luciferase assay in duplicate. Statistical significance was determined by Unpaired t-test.

4.2.2 Enhancer targeted SHM activity is retained in non-B cells

Apart from the DT40 and Ramos cell lines, the non-B-cell UO-31 cell line's AID expression was confirmed by Western blot, quantitative PCR, and immunohistochemistry (Figure 18). After transduction with the GFP7 vector and the two 2x72bp SV40e, we observed 5-fold GFP fluorescence loss in comparison to the empty vector with no enhancer (Figure 19A).

For the lack of cell lines expressing detectable levels of AID, we also transduced the NIH3T3, 293T, U2OS, MDA-MB-231, and Ramos AID^{-/-} (Senigl et al. 2019), as a control, cell lines with the GFP7 vectors without enhancer or with the two 2x72bp SV40e (Figure 19B). We did not observe any GFP fluorescence loss. GFP-negative cells in the 293 cell line were measured due to lower sensitivity of the cells to blasticidin selection.

However, after transduction of the cell lines with the retroviral vector expressing AID fused with mCherry, GFP fluorescence loss was observed (Figure 19C). In the 293 and MDA-MB-231 cell lines, there was a 2- and 18-fold increase of GFP fluorescence loss with the two 2x72bp SV40e in comparison to the vector with no enhancer. It is noteworthy that in contrast to the B cells, the Ig enhancer did not possess SHM targeting activity in the non-B cells.

Here, we revealed that SV40e possesses SHM targeting activity also in the non-B cell lines. In contrast to the B cells, Ig enhancer is not able to target AID-mediated SHM in the non-B cells.

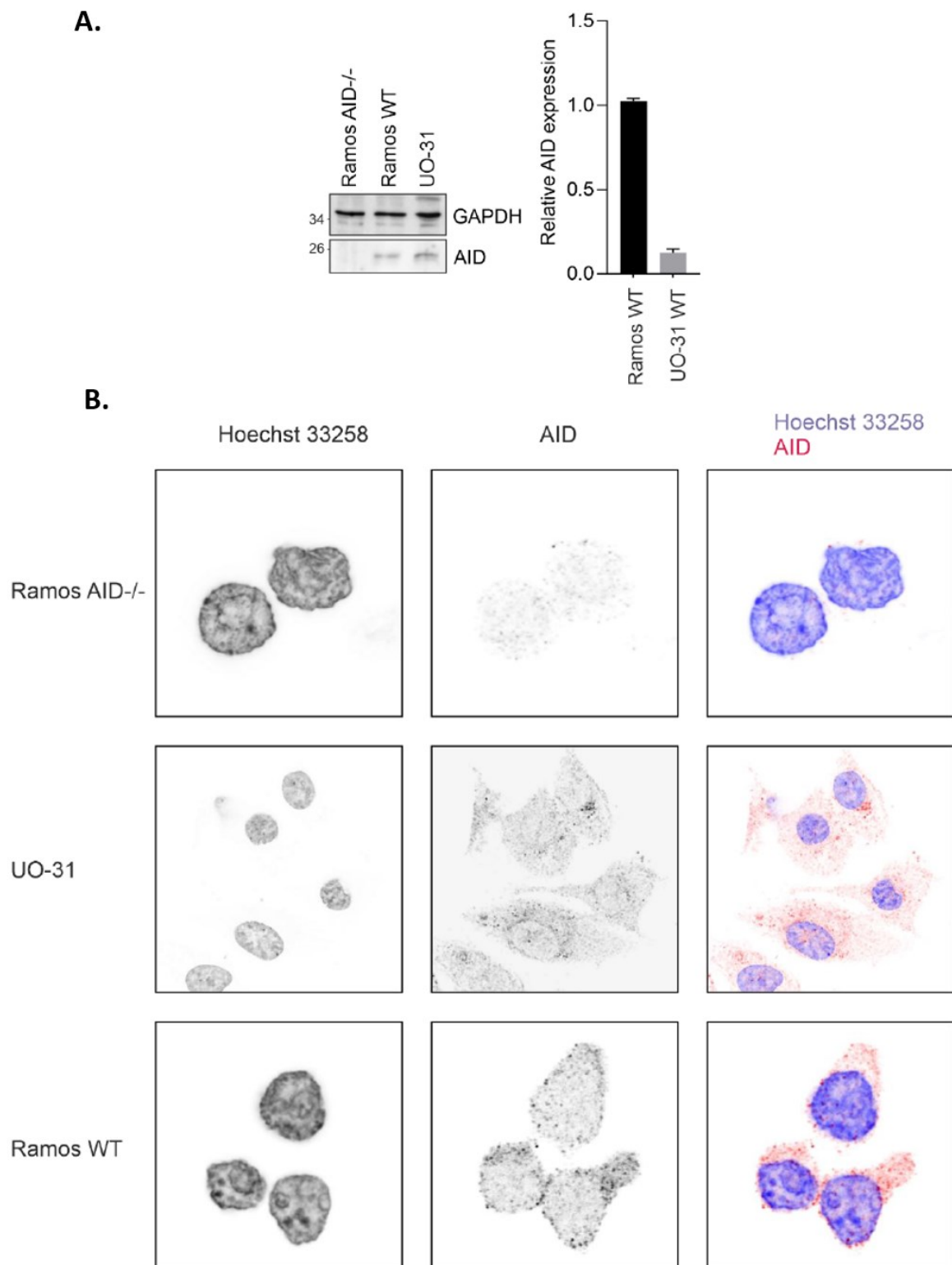


Figure 18. Expression of activation-induced deaminase (AID) in UO-31 cell line. A) Endogenous expression of AID in Ramos WT, UO-31 WT, and Ramos AID^{-/-} as a control. Western blot (left), RT-qPCR (right). B) Immunohistochemistry detection of AID in Ramos WT, UO-31 WT and Ramos AID^{-/-} as a control.

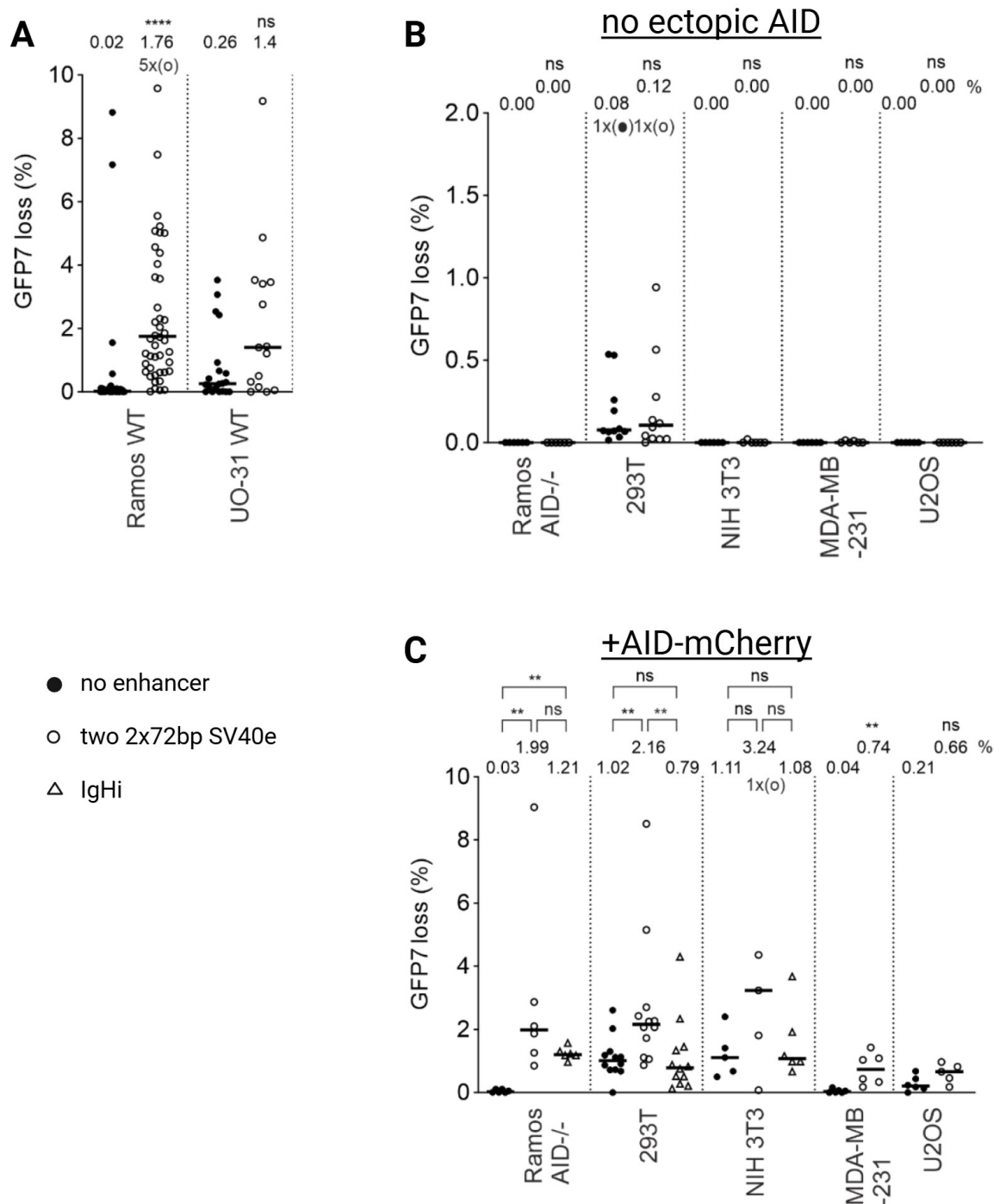


Figure 19. SV40e SHM targeting activity in non-B cells. A) GFP7 loss assay vector with two 2x72 SV40e (open circles). GFP7 vector without an enhancer (black circles) is used as a negative control. B) GFP7 assay in non-B cells without endogenous AID expression and Ramos AID^{-/-} as a control. Vector without an enhancer (black circles) and with two 2x72bp SV40e (open circles). C) GFP7 assay in non-B cells with exogenous AID expression from AID-mCherry expression vector. Ramos AID^{-/-} cells serve as a negative control. GFP7 without enhancer (black circles), two 2x72bp SV40e (open circles), and human Ig heavy intronic (IgHi; open triangles) enhancer. In all panels, statistical significance was calculated by Mann-Whitney U-test. The numbers indicate % median GFP loss.

4.2.3 Truncated large T antigen of SV40 as a result of mutations in SV40 *LT* region in the B and non-B cell lines

To test SV40e SHM targeting activity in a more biologically relevant environment, we utilized the SV40e in combination with the *LTA*g transcription unit in the Ramos and UO-31 cells. The vector also carries upstream SV40 agnoprotein transcribed in the reverse orientation, which is regulated from the same NCRR as the SV40 T antigen (Figure 20A). Twelve weeks after transduction, we sequenced the LT antigen region to assess accumulated mutations in the A/T rich reference sequence (Figure 20B).

In the Ramos cell line, the mutations were distributed along the SV40 *LTA*g region (Figure 20C left). Five premature STOP codons were introduced into the coding sequence (Figure 20D left). 19.14% of all mutations were at AID hotspots WRC (Figure 20E left). From the 47 obtained mutations in total, 53.3% of them were linked to Cs. We observed 25.6% transversions, 68.1% transitions, and 6.4% indels (Figure 20F left).

In the UO-31 cell line, we again sequenced the *LTA*g transcription unit and found mutations along the *LTA*g sequence (Figure 20C right). Five premature STOP codons were introduced into the coding sequences (Figure 20D right). Out of the 25 mutations in total, 40% of them were linked to Cs at AID WRC hotspots (Figure 20E right). We observed 52% of transitions, 36% of transversions, and 12% of indels (Figure 20F right). In general, the mutation frequency in the UO-31 cell line was lower than mutation frequency in the Ramos cells.

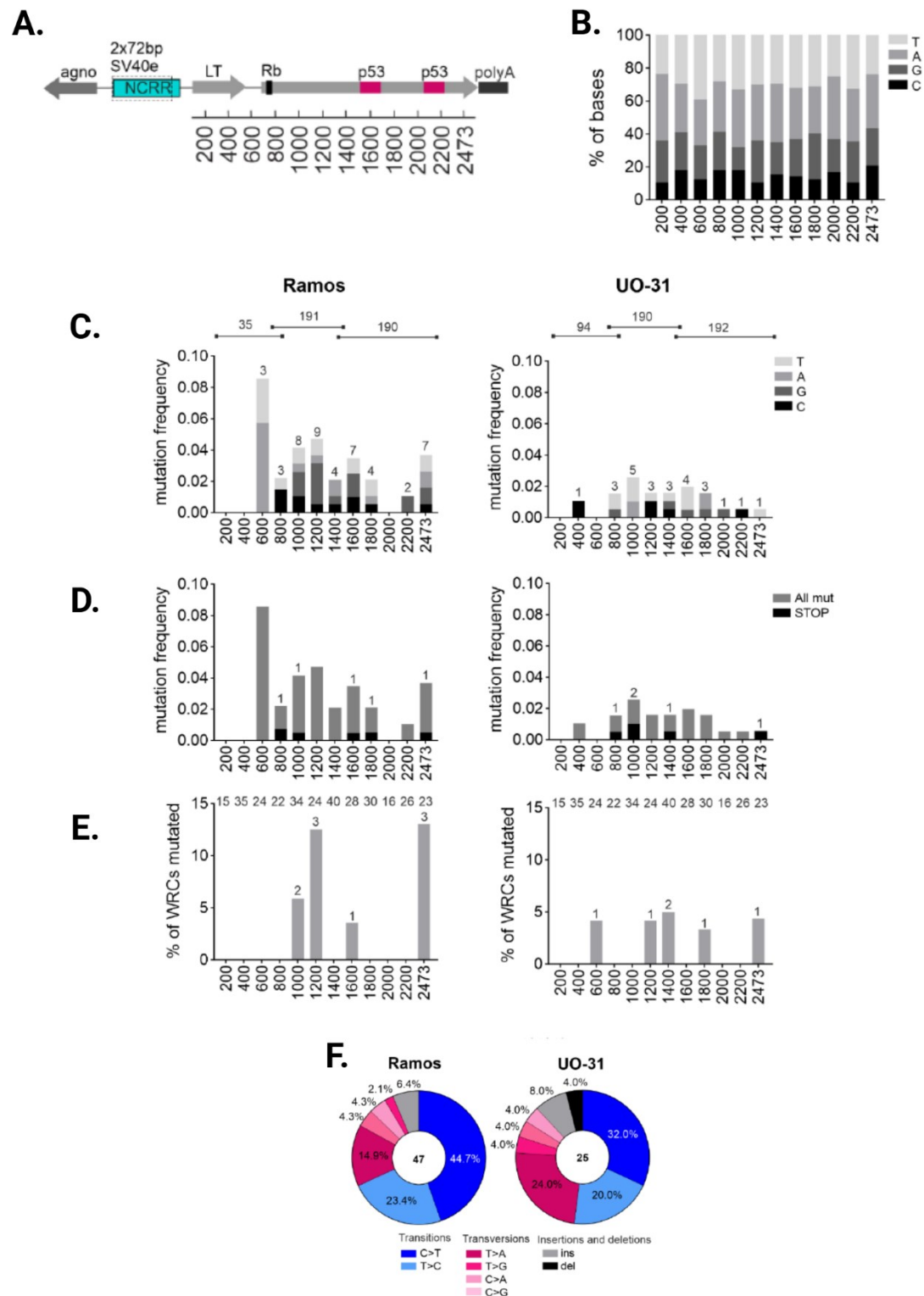


Figure 20. Mutation analysis of SV40 large T antigen. A) A schematic depiction of inserted SV40 *LTA*g transcription unit with indicated 200-bp bins. B) Base content of each bin. C) Mutation distribution along the *LTA*g region with the number of mutations in the bins. Number of obtained sequences is indicated above the graphs. D) Overall and STOP-codon mutations along *LT* region. Y-axis shows the percentage of STOP-codon mutations with absolute

numbers indicated on the top of columns. E) The percentage of mutations targeting AID hotspots WRC along *LT* region. Absolute numbers of WRC mutations (above the columns) from possible WRC motifs present in the SV40 *LTag* unit (above the graph). F) Type of mutations in *LTag*. Color code of substitutions: transitions (blue) and transversions (red). Insertion (gray) and deletions (black).

The verification of *LTag* truncation was done in the DT40 cell line. We obtained a clone with a single nucleotide deletion leading to generation of a STOP codon immediately after the deletion (Figure 21A). As a result of premature STOP codon, we detected truncated *LTag* by Western blot (Figure 21B).

In summary, AID-mediated SHM targeting of SV40 may lead to introduction of premature STOP codons into the *LTag* sequence and truncation of *LTag* in the B and also non-B cell lines.

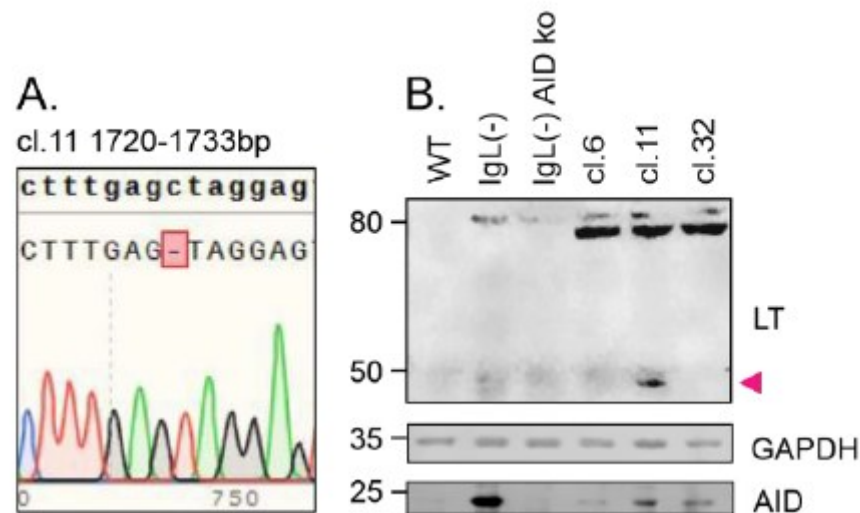


Figure 21. Truncated large T antigen (LTag) in the DT40 cells. A) Sequencing chromatogram from clone 11 with deletion resulting in generation of stop codon after mutation. B) Expression of SV40 *LTag*, GAPDH and AID in DT40 cells by western blot. Clones 6, 11, and 32 have SV40 *LTag* expression cassette in Ig light chain locus. The arrow indicates the truncated *LTag*.

5. Discussion

Chromatin environment at the site of integration plays a role in regulation of proviral expression (Senigl et al 2012; Vansant et al. 2020). γ RVs preferentially integrate into promoters and enhancers of active genes (Mitchell et al. 2004; De Ravin et al. 2014), regulatory elements reported to preserve stable expression of proviruses (Senigl et al. 2012; Miklík et al. 2018; Vansant et al. 2020). Thus, we examined whether γ RV LTR-regulated stability of expression depends on a chromatin environment at the preferred integration sites.

We showed that γ RV LTR promoter activity allows establishment and long-term maintenance of stable proviral expression after retargeting from naturally preferred integration sites. Moreover, the γ RV LTR-equipped vectors establish stable expression even after insertion into the restrictive environment of LADs.

First, we evaluated long-term expression stability of MLV LTR promoter activity. During the long-term culture, we observed only limited silencing of MLV LTR-driven transcription. This long-term stability of expression is in agreement with the previous work (Miklík et al. 2018). The variegation of MLV expression was observed only after transduction with the IN^{WT} vector in the clonal population.

We observed that integration sites obtained from the transduced cells selected for stably expressed proviruses at 3 dpi retained the retargeted integration profile. Van Looveren et al. (2021) also showed only minor differences between non-selected and selected integration sites populations after transduction with MLV SIN vector with SFFV as an internal promoter. However, we showed that even after transduction with Bin vectors the majority of proviruses were still integrated in A compartments, transcriptionally active chromatin. It shows that open chromatin is preferred over more packed chromatin even during disruption of interaction between IN and BET proteins.

Next, we observed that long-term stability of LTR-driven transcription is a general characteristic of γ RV. However, we estimated a significant population of silenced proviruses in the populations transduced with γ RV LTR-equipped vectors. Silenced proviruses result from epigenetic silencing or genetic defects of integrated vectors (Xu et al. 1989; Lorincz et al. 2000). Proviruses were silenced early after infection before the first measurement of proviral expression at 3 dpi by flow cytometry. Moreover, after transduction with MLV LTR-equipped vector with IN^{WT} vector, 20% of proviruses were silent, while after transduction with MLV LTR-equipped vector with Bin vectors, the proportion of transcriptionally proviruses doubled. We did not estimate this increase of silencing between IN^{WT} and Bin vectors groups after transduction with other non-MLV LTR-equipped vectors. Since the only difference among the vectors is the LTR, we thus suggest that the difference results from different strength of promoters and enhancers of individual LTRs and the existence of MLV-specific naturally occurring integration sites which are permissive for MLV LTR-driven expression, while expression from other non-MLV LTR-equipped vectors is restricted. The characteristics of such MLV-specific permissive integration sites remain to be determined.

The γ RV LTR-equipped vectors were produced using the MLV-producing system. While we managed to effectively transduce cells with FeLV and SNV LTR-equipped vectors, we only produced a limited titer of KoRV LTR-equipped vectors suggesting limited compatibility of non-MLV LTR with the MLV-producing system.

In the previous experiments we examined the stability of expression driven from full-length LTR. Next, we evaluated the stability of expression driven from internal γ RV LTR-derived promoters in the vector with an almost random integration pattern. We observed strong silencing after 3 dpi only after transduction with SNV LTR-derived promoter. The RU5 part of the SNV LTR, which is missing in our vector, was reported to contain cis-acting element

important for RNA export from the nucleus, accumulation of unspliced RNAs in the cytoplasm and translation (Butsch et al. 1999; Roberts and Boris-Lawrie 2003). Our results suggest that the RU5 part of the SNV LTR plays a role in long-term stability of proviral expression.

The first generation of MLV-based therapeutic vectors led to development of leukemia due to insertional mutagenesis (Cavazzana-Calvo et al. 2000; Hacein-Bey-Abina et al. 2002). Retargeted vectors thus could provide a platform for safer and, as we showed, long-term stable expression. However, during MLV retargeting attempts using the IN^{dCt} or Bin vectors, the transduction still led to mutagenesis (Loyola et al. 2019; Nombela et al. 2022). Similarly, lentiviral SIN vectors equipped with strong internal SFFV-derived enhancers-promoters were mutagenic (Zychlinski et al. 2008). Therefore, targeted integration may increase safety of the therapeutic vectors.

We utilized CRISPR-Cas9-induced homologous recombination-mediated insertion approach into selected intra-LAD intergenic B2/B3 subcompartments. We showed that γ RV LTRs-equipped vectors are silencing resistant and a single insertion is sufficient for long-term stable expression of transgene in the transcriptionally restrictive environment of LADs.

First, we searched for target sites efficiently cut by the CRISPR-Cas9 system in the LADs in the K562 cell line. Although it was shown that CRISPR-Cas9 cutting efficiency is significantly reduced in heterochromatin regions (Chen et al. 2016), we managed to identify target sites reaching up to 81% efficiency of generating indels by DNA repair mechanism after CRISPR-Cas9 generation of double-strand breaks. Since the nuclease-induced insertion, similarly as IN, requires access to DNA, it is likely that we selected partially open chromatin.

For further evaluation of *DoTs*, we performed ChIP. The size of fragments gained by sonication influences the level of signal. In general, the fragments ranging between 100 to 1000 bp are recommended for IP. We performed the IP with two different samples. The first sonicated sample contained fragments with the main density ranging from 100 to 300 bp on an agarose gel, while the second sonicated sample contained fragments with the main density ranging from 100 to 600 bp on an agarose gel. After performing IP on shorter fragments, the signal should be more target site specific, while after performing IP on longer fragments, it should be more domain specific. Differences between two sonicated samples for the same *DoTs* could be explained either by partially open chromatin at *DoTs* and therefore decreased levels of Lamin B or dilution of antibody by binding shorter fragments without amplified sequence. *RPP30* as active chromatin exhibited similarly lower levels of Lamin B at both samples. Transcriptionally active *IFT20* locus exhibited higher levels of Lamin B than *RPP30* suggesting differences at local chromatin at those two sites.

Recommendation on the length of homologous arms for homologous recombination-mediated insertions greatly varied. During designing arms homologous to sequences flanking our target domains, we were limited by repetitive sequences in *DoTs* to design specific primers for homologous arms amplification. Resulting homologous arms ranged from 310 to 1050 bp. We managed to obtain targeted insertions with any tested lengths of homologous arms, verified by amplification of genome-insert junctions. However, shorter homologous arms were more convenient for insert verification by sequencing.

We performed the targeted knock-in only in the pre-selected target sites with observed indels caused by CRISPR-Cas9-induced double strand breaks error-prone DNA repair pathway. Differences in the CRISPR efficiency in CRISPR-expressing cells might be given different accessibility of the target sites and kinetics of DNA repair pathway at the site of cleavage, with the indels accumulating over time (Brinkman et al. 2018). Despite the different ratios of modified sequences, ranging from 14 to 81%, we managed to insert the vectors into every tested *DoTs*.

We managed to insert MoMLV, FeLV, and KoRV LTRs-equipped vectors into LADs by homologous recombination. However, homologous recombination DNA repair pathway was reported to be employed to repair double-strand breaks in active chromatin (Aymard et al. 2014), while in LADs, employment of non-homologous end joining was preferred (Lemaître et al. 2014), which is in agreement with reported targeted insertion into LADs in the HCT116 cell line (Tasan et al. 2018). We suggest that the discrepancy between reported preferential DNA repair mechanism and our observed insertions might be due to cell-type or DoTs specific differences given by the local chromatin at the site of targeting.

We observed that different target sites differently supported expression from different vectors. The local chromatin environment of LADs plays a role in regulation of expression (Leemans et al. 2019). We thus suggest that the observed differences in level of transcriptional stability might be given by different sensitivity of promoter/enhancer of LTRs to LAD chromatin.

Next, we estimated the higher number of inserted vectors in the majority of clones. Apart from off-targets, which we did not detect by standard amplification methods, the higher number of inserted vectors might be explained by the formation of concatemers at the target sites (Lombardo et al. 2007, 2011; Suchy et al. 2024). In order to obtain targeted insertions, we co-transfected high amounts of plasmids. Either downscaling the plasmid amount or changing the vector delivery, e.g. from transfection to transduction, might potentially lead to the decreased CN. Moreover, the higher copy number was observed after targeting active gene *IFT20* in comparison to targeting intra-LAD *DoTs*. Transcriptionally silent sites could thus provide a platform with a better control of targeted insertions.

After targeting insertions into intra-LAD B2/B3 compartments in the K562 cell line. We decided to repeat the targeted knock-in in the Jurkat cell line. Although LADs are generally considered highly conserved, LADs might differ in various cell lines or under different conditions (Robson et al. 2017; Das, Martin, and McCord 2023). In the Jurkat cell line, from the previously used *DoTs*, only the *DoT:7.3* is predicted as a LAD. We even observed similar CRISPR-Cas9 activity. However, we did not manage to verify targeted insertions at that *DoT* suggesting possible differences at the local chromatin between the Jurkat and K565 cell lines. Efficiency of targeted insertions should thus be experimentally verified for every target site in the modified cells.

We observed long-term stable expression from inserted γ RV LTR-equipped vectors in intra-LAD B2/B3 subcompartments, which are generally considered as a transcriptionally repressive environment. Transcriptionally silent intra-LAD regions are omitted in the search for safe harbors, safe sites in the genome for transgene without disrupting expression of endogenous genes (Shrestha et al. 2022; Autio et al. 2024). Instead, targeted insertions were performed into active chromatin, into *TRAC*, *CCR5*, *IL2RG*, *AAVS1*, and *OCT4* loci (Eyquem et al. 2017; Lombardo et al. 2007, 2011; Rothemejer et al. 2023; Hockemeyer et al. 2009). However, these target sites in active chromatin are not true safe harbors, without influence on the level of transcription of nearby genes. We advocate for intra-LAD to potentially be such safe harbors providing that intra-LAD *DoTs* would pass necessary safety tests eliminating the risk of transcriptional deregulation of endogenous genes. Moreover, all the epigenetic features of integration or insertion sites refer to the state before vector delivery. The effect of insertion on a target site remains to be elucidated.

Regarding the second project, SV40 related MCPyV enhancer possesses SHM targeting activity (Soikkeli et al. 2022). Mutations in the *LTA*g sequence may lead to introduction of premature STOP codons and truncation of *LTA*g. Thus, we tested SHM targeting activity of SV40 enhancer and its role in truncation of *LTA*g.

AID-mediated SHM is cell-type specific with a different set of genes mutated in different cell lines (Qian et al. 2014). So far, several requirements for AID-mediated SHM were identified. AID-mediated SHM occurs in the transcriptionally active loci of the genome which hold SHM targeting elements containing TFBSs and their overall domain architecture is susceptible to SHM. Susceptibility of the domains was characterized by enrichment for the cohesin loader NIPBL and markers of paused RNA polymerase II (Yoshikawa et al. 2002; Buerstedde et al. 2014; Senigl et al. 2019). However, the complete mechanism of AID-mediated SHM is still unclear.

First, we observed SHM targeting activity of SV40e in B and also non-B cells by utilization of GFP loss assays. Ig enhancers which have SHM targeting activity in B cells (Blagodatski et al. 2009), did not exhibit such activity in non-B cells. Similarly, genomic enhancers did not possess SHM targeting activity in B cells suggesting requirement of cell-type- and enhancer-specific factors. Nevertheless, we observed that SHM targeting activity of SV40e is retained in avian, mice, and human cells suggesting evolutionary conserved mechanism of SHM targeting. The differences between Ig enhancers, genomic enhancers, and SV40e remain to be researched. Enhancers targeting SHM activity into nearby sequences contain a number of TFBSs. We observed the cooperative function of TFBS on SHM, which is in an agreement with the previous report (Buerstedde et al. 2014).

The SV40e of the early region may contain several 72bp repeats (Lednický et al. 1998). The number of 72bp repeats influence the level of transcription. The number of repeats correlates with the expression level (Kumar et al. 1986). After co-nucleofection, the duplicated 72bp SV40e provided the growth advantage over the single 72bp SV40e. Apart from higher expression and faster replication, we showed that two 2x72bp SV40e have higher AID-mediated SHM targeting activity than 2x72bp SV40e (2-fold) and 1x72bp SV40e (10-fold). Nevertheless, the single 72bp SV40e might have an advantage in decreased immunogenicity. Next, we analyzed the AID-mediated SHM in the context of the *LTA*g transcription unit. We observed generation of mutations leading to introduction of STOP codons and truncation of *LTA*g. Although the role of SV40 in human tumorigenesis is still unclear (Rotondo et al. 2019), the truncated SV40 *LTA*g was also observed in transformed human cell lines (Gish and Botchan 1987; Kao et al. 1993).

Apart from SV40 potential role in tumorigenesis by deregulation of cell cycle by *LTA*g, we suggest another mechanism of tumorigenesis. If integrated into the host genome the SV40e could target AID-mediated SHM into the nearby sequences such as tumor-suppressor genes and proto-oncogenes.

6. Conclusions

The integrated MLV has previously been reported to be transcriptionally stable over the long term. Here we show that this long-term transcriptional stability is not specific to the MLV but is a general property of γ RVs. The high transcriptional stability of γ RVs is maintained in retargeted proviruses in non- IN^{WT} -preferred integration sites. γ RVs may remain transcriptionally stable even in the restrictive environment of LADs.

Transcriptionally silent LADs could therefore represent a safe platform for gene transfer, so called safe-harbor. However, safety and cell cycle dysregulation after insertion remains to be validated. Moreover, we observed different support of expression from inserts in different genomic and epigenomic contexts. For efficient expression from transgene, the level of expression given by the combination of used vector and specific target domain should be experimentally verified.

We also tested SV40e SHM targeting activity and its role in SV40 LTA γ truncation. By GFP fluorescence assays, we revealed that SV40e possesses SHM targeting activity in B and also non-B cell lines while underscoring the importance of TFBSs. In contrast, selected genomic enhancers lack AID-mediated SHM targeting activity in the B cells, and Ig enhancer lack AID-mediated SHM targeting activity in the non-B cells suggesting differences in expression of specific factors for AID-mediated SHM activity.

AID-mediated SHM may lead to generation of mutations in *LTA γ* sequence or, after integration into the genome, into tumor-suppressor genes and proto-oncogenes and thus possibly contribute to tumorigenesis.

Contribution to the publications

First-author publication

Miklík, Dalibor, Martina Slavková, Dana Kučerová, Chahrazed Mekadim, Jakub Mrázek, and Jiří Hejnar. 2024. "Long Terminal Repeats of Gammaretroviruses Retain Stable Expression after Integration Retargeting." *Viruses* 16 (10). <https://doi.org/10.3390/v16101518>.

- Preparation of DNA library (results 4.1.2)
- Targeted knock-ins (results 4.1.5 and 4.1.6) – design of the vectors, characterization of target sites, generating of clones with insertions and their verification, chromatin immunoprecipitation
- Schemes preparation in Biorender software
- Writing of the draft and review

Contributing-author publication

Šenigl, Filip, Anni Soikkeli, Salomé Prost, David G. Schatz, Martina Slavková, Jiří Hejnar, and Jukka Alinikula. 2024. "The SV40 Virus Enhancer Functions as a Somatic Hypermutation-Targeting Element with Potential Tumorigenic Activity." *Tumour Virus Research*, October, 200293.

- Isolation of RNA, reverse transcription and analysis of AID expression levels by RT-qPCR

Oligo Tables

MoMLV_LTR_F	5' – AAAAAGGGGGGAATGAAAGACCCAC – 3'
MoMLV_LTR_R	5' – ACGAGCCCCCAAATGAAAGACCC – 3'
CrERV_LTR_F	5' – AAAAAGGGGGGAATGTGTAGGGAGAACAACGGAATGTAG – 3'
CrERV_LTR_R	5' – CATTTGGGGGCTCGTTGTTATGCCCGATGTCCGAATC – 3'
KoRV_LTR	5' – <u>AAAAAGGGGGGAATGAAGGAGGCAGAAATCATGAGGCAGAAATCATT</u> <u>CCGTGGAGTATGGAACTACCCGGAGGGCCCAAGGTTTAGGGACAGGT</u> <u>GCAGCCAGGCACAGTAAAAGGTCAGAGCAAGAAAAACAAGGAAGATT</u> <u>GGAGTGCCAAACGGGATATCTGTGGTCATGCACCTGAGTCCCCACCCG</u> <u>GACTTATGCAACAATTCCAGAAATAGCTGAGCTCATAACAGTTTCTAG</u> <u>GGTGCCCTCAGCAGTTTCTAGAACCCTCTCGTGACCGGAGTTTTTATTC</u> <u>AACTAACCAATGATCTTGCTCCTCGCTTCCGTACCCGCGCTTTTTTGCTA</u> <u>TAAATGAGACCAAGAATCCACCTGGCGCGCCAGTCCCTCTAGGTGAC</u> <u>TGAGTCGCCCCGAGTACTCGTAAGTTCAATAAACCTCTTGCTATTTGCATC</u> <u>CGGAGTTGTGTTGCGGTTGATCCTGGGAGGGTTTCTCAAGGTCGGAGG</u> <u>ACTACCCGAACATTGGGGTCTTTCA</u> <u>TTTGGGGGCTCGT – 3'</u>
FeLV_LTR	5' – <u>AAAAAGGGGGGAATGAAAGACCCCTACCCCAAATTTAGCCAGCTACT</u> <u>GCAGTGGTGCCATTTCAAGGCATGGAAAATTACTCAAGTATGTTCCC</u> <u>ATGAGATACAAGGAAGTTAGAGGCTAAAACAGGATATCTGTGGTTAAG</u> <u>CACCTGGGCCCCGGCTTGAGGCCAAGAACAGTTAAACCCCGGATATAGC</u> <u>TGAAACAGCAGAAGTTTCAAGGCCGCTACCAGCAGTCTCCAGGCTCCCC</u> <u>AGTTGACCAGGGTTCGACCTTCCGCCTCATTTAACTAACCAATCCCCAC</u> <u>GCCTCTCGCTTCTGTGCGCGCGCTTCTGCTATAAAACGAGCCATCAGCC</u> <u>CCCAACGGGCGCGCAAGTCTTTGCTGAGACTTGACCGCCCCGGGTACCC</u> <u>GTGTACGAATAAACCTCTTGCTGATTGCATCTGACTCGTGCTCGGTGT</u> <u>TCTGTGGGCGCGGGGTCTCATCGCCGAGGAAGACCTAGTTCAAGGGTC</u> <u>TTTCATTTGGGGGCTCGT – 3'</u>
SNV_LTR	5' – <u>AAAAAGGGGGGAATGTGGGAGGGAGCTCTGGGGGAAATAGCGCTGGC</u> <u>TCGCAACTGCTATATTAGCTTCTGTACTCATGCTTGCTTGCCTGGCCACTA</u> <u>ACCGCCATATTAGCTTCTGTACACATGCTTGCTTGCCGTAGCCGCCATTG</u> <u>TACTTGATATGCCATTTCTCGGAATCGGCATCAAGTTTCGTTCTCGAGA</u> <u>GCAAGCCCACAAACCACAAAAGGAAACGCGCACC GAAGGCAAGCATCA</u> <u>GACCACTTGCGCCATCCAATCATGAACGGACACGAGATCGGACTATCAT</u> <u>ACTGGAGCCAATGGTTGTAAGGGCAGATGCTACTCTCCAATGAGGGA</u> <u>AAATGTCATGTAACACCCTGTAAGCTGTAAGCGGCTATATAAGCCGGGT</u> <u>ACATCTCTTGCTCGGGGTCGCCGTCCTGCACATTGTTGTTGTGACGTGCG</u> <u>GCCCAGATTGCAATCTGTAATAAACTTTTCTTCTGAATCCTCAGATTG</u> <u>GCAGTGAGAGGAGATTTTGTTCGTGGTGTTGGCTGGCCTACTGGGTGG</u> <u>GCGCAGGGATCCGGACTGAATCCGTAGTACTTCGGTACAACATTTGGGG</u> <u>GCTCGT – 3'</u>

Oligo Table 1. Primers and fragments for LTR cloning into pL2dG vector. LTR sequence underlined.

pAS_XcmI_F	5' – TGGTCGGCCACAGACGG – 3'
pAS-GFP_InFu_F	5' – CCGGTCGCCACCATGGTGAGCAAGGGCGAGG – 3'
pAS-d2GFP_InFu_R	5' – GACGCGGCCGCTTTACACATTGATCCTAGCAGAAGCACAGG – 3'
pAS_5intP_R	5' – GCTTGATCCGCAGGCCGAC – 3'
pAS_FeU3_InFu_F	5' – GCCTGCGGATCAAGCACCCCCTACCCCAAAATTTAGC – 3'
pAS_AscI-FeU3_InFu_R	5' – CGGAGGACTGGCGCGCCCGTTGGGGGC – 3'
pAS_SNU3_InFu_F	5' – GCCTGCGGATCAAGCTGTGGGAGGGAGCTCTGGG – 3'
pAS_AscI-SNU3_InFu_R	5' – CGGAGGACTGGCGCGCCGAGCAAGAGATGTACCCG – 3'
pAS_KoU3_InFu_F	5' – GCCTGCGGATCAAGCTGAAGGAGGCAGAAATCATGAGGC – 3'
pAS_AscI-KoU3_InFu_R	5' – CGGAGGACTGGCGCGCCAGGTGGATTCTTTGG – 3'
pAS_MoU3_InFu_F	5' – GCCTGCGGATCAAGCTGAAAGACCCACCTGTAGGTTTG – 3'
pAS_AscI-MoU3_InFu_R	5' – CGGAGGACTGGCGCGCCCGAGTGAGGGGTTG – 3'
pAS_CrU3_InFu_F	5' – GCCTGCGGATCAAGCTGTGTAGGGAGAACAAACGGAATGTAG – 3'
pAS_AscI-CrU3_InFu_R	5' – CGGAGGACTGGCGCGCCCGAGCCCGGTTTTC – 3'

Oligo Table 2. pAS_yRV_d2GFP cloning primers

Name Coordinates (GRCm38)	Primers
hBLKe chr8:11,536,697-11,539,424	5' – TCACGGATCCACTAGTTGCCTGGCCAATATTCTTGACTTTA – 3' 5' – CAGGGAGCAGGCTAGCATGTAAGTGGCCAGCAATTGAAGA – 3'
hCD83e chr6:14,093,401-14,096,308	5' – TCACGGATCCACTAGTAAATAATGCCCCAAGTCTCACTGG – 3' 5' – CAGGGAGCAGGCTAGCTCTGCCTTCTCTTGACAACAAACA – 3'
mArfgef1e chr7:24,903,876-24,904,597	5' – TCACGGATCCACTAGTCGAGCGCATGAGTGTTTTATGCTT – 3' 5' – CAGGGAGCAGGCTAGCTACAACAGGCGCTCACTAATTCTG – 3'
mPou2af1e chr9:51,225,574-51,226,089	5' – TCACGGATCCACTAGTCTCTCTCTTTATATATCTCTAGAGACAGC – 3' 5' – CAGGGAGCAGGCTAGCTCCAAAAAGAAAATGGTGGTTT – 3'
SV40 enhancer	5' – ATGTAACTGCAGGGCCTGAAATAACCTCT – 3' 5' – ATGTTAACCTAGTGGGACTATGGTTGCTGA – 3'

Oligo Table 3. SV40 enhancers and genomic enhancers cloning primers

gRNA_DoT:1.1A	5' – CACCGTTGACCCACTCTTTGCACTGGG – 3'
gRNA_DoT:1.1B	5' – AAACCCAGTGCAAAGAGTGGGTCAAC – 3'
gRNA_DoT:1.2A	5' – CACCGAGATACCACCGAGTAAATGAGG – 3'
gRNA_DoT:1.2B	5' – AAACCCTCATTTACTCGGTGGTATCTC – 3'
gRNA_DoT:1.3A	5' – CACCGAATAGGAGCATTGCTACATTGG – 3'
gRNA_DoT:1.3B	5' – AAACCCAATGTAGCAATGCTCCTATTC – 3'
gRNA_DoT:2.1A	5' – CACCGCTACTGATAGACCACATCGAGG – 3'
gRNA_DoT:2.1B	5' – AAACCCTCGATGTGGTCTATCAGTAGC – 3'
gRNA_DoT:2.2A	5' – CACCGATCCCCAAGACAACGGAGGAGG – 3'
gRNA_DoT:2.2B	5' – AAACCCTCCTCCGTTGTCTTGGGGATC – 3'
gRNA_DoT:3.1A	5' – CACCGTAATCCTCTAGGGATGCCGTGG – 3'
gRNA_DoT:3.1B	5' – AAACCCACGGCATCCCTAGAGGATTAC – 3'
gRNA_DoT:3.2A	5' – CACCGTCACTCAGCTATGCATAACTGG – 3'
gRNA_DoT:3.2B	5' – AAACCCAGTTATGCATAGCTGAGTGAC – 3'
gRNA_DoT:3.3A	5' – CACCGTGCTCCCAGGTAGCCTAGTGGG – 3'
gRNA_DoT:3.3B	5' – AAACCCCACTAGGCTACCTGGGAGCAC – 3'
gRNA_DoT:4.1A	5' – CACCGTGTCCTTGATATGAATCAGAGG – 3'
gRNA_DoT:4.1B	5' – AAACCCTCTGATTCATATCAAGGACAC – 3'
gRNA_DoT:4.2A	5' – CACCGAGTAGTTGACCGTAGTCATAGG – 3'
gRNA_DoT:4.2B	5' – AAACCCTATGACTACGGTCAACTACTC – 3'
gRNA_DoT:4.3A	5' – CACCGAGTACGACTATACATATCAAGG – 3'
gRNA_DoT:4.3B	5' – AAACCCTTGATATGTATAGTCGTACTC – 3'
gRNA_DoT:5.1A	5' – CACCGCGTTGCTGCTCAGCGACTCTGG – 3'
gRNA_DoT:5.1B	5' – AAACCCAGAGTCGCTGAGCAGCAACGC – 3'
gRNA_DoT:5.2A	5' – CACCGACAGGAAGGTATAGGCCTCTGG – 3'
gRNA_DoT:5.2B	5' – AAACCCAGAGGCCTATACCTTCCTGTC – 3'
gRNA_DoT:5.3A	5' – CACCGTTCCCATTTGTCTAGAATCGGGG – 3'
gRNA_DoT:5.3B	5' – AAACCCCCGATTCTAGACAATGGGAAC – 3'
gRNA_DoT:5.4A	5' – CACCGGCAGAATGTAAGGACCGTGGGG – 3'
gRNA_DoT:5.4B	5' – AAACCCCCACGGTCCTTACATTCTGCC – 3'
gRNA_DoT:5.5A	5' – CACCGTTCACAAGTGTATAAGGACAGG – 3'
gRNA_DoT:5.5B	5' – AAACCCTGTCCTTATACACTTGTGAAC – 3'
gRNA_DoT:6.1A	5' – CACCGATATTAGAGTATCCCGTGAG – 3'
gRNA_DoT:6.1B	5' – AAACCTCACGGGATACTCTAATATC – 3'
gRNA_DoT:6.2A	5' – CACCGAGGGATCTCCTTAGTACCG – 3'
gRNA_DoT:6.2B	5' – AAACCGGTACTAAGGAGATCCCTC – 3'
gRNA_DoT:6.3A	5' – CACCGAATGTCGCTTTCCTACCGT – 3'
gRNA_DoT:6.3B	5' – AAACACGGTAGGAAAGCGACATTC – 3'
gRNA_DoT:7.1A	5' – CACCGCTAAATCCCTTGACAATTGG – 3'
gRNA_DoT:7.1B	5' – AAACCCAATTGTCAAGGGATTTAGC – 3'
gRNA_DoT:7.2A	5' – CACCGTCACTCAGCACCTATGGAC – 3'
gRNA_DoT:7.2B	5' – AAACGTCCATAGGGTGCTGAGTGAC – 3'
gRNA_DoT:7.3A	5' – CACCGTTCTACTTGCTGGCTAACAC – 3'

gRNA_DoT:7.3B	5' – AAACGTGTTAGCCAGCAAGTAGAAC – 3'
gRNA_DoT:8.1A	5' – CACCGCCGGGATTAGCACAATCAGC – 3'
gRNA_DoT:8.1B	5' – AAACGCTGATTGTGCTAATCCCGGC – 3'
gRNA_DoT:8.2A	5' – CACCGAGGATACTACACCACAACG – 3'
gRNA_DoT:8.2B	5' – AAACCGTTGTGGTGTAGTATCCTC – 3'
gRNA_DoT:8.3A	5' – CACCGCAAACGTCCCGATTACGATG – 3'
gRNA_DoT:8.3B	5' – AAACCATCGTAACTGGGACGTTTGC – 3'
gRNA_DoT:8.4A	5' – CACCGATTAATGCGCCTAACTACTC – 3'
gRNA_DoT:8.4B	5' – AAACGAGTAGTTAGGCGCATTATC – 3'
gRNA_DoT:9.1A	5' – CACCGCAAGCCGCTCACTACCGTAG – 3'
gRNA_DoT:9.1B	5' – AAACCTACGGTAGTGAGCGGCTTGC – 3'
gRNA_DoT:9.2A	5' – CACCGTCAATCAGGAGTACCCACG – 3'
gRNA_DoT:9.2B	5' – AAACCGTGGGTACTCCTGATTGAC – 3'
IFT20-gRNA2A	5' – CACCGTGATGAACTGAACAAGCTGA – 3'
IFT20-gRNA2B	5' – AAACCTCAGCTTGTTCACTTCATCAC – 3'
region9A	5' – CACCGCCTATACCTTTACCGATAGC – 3'
region9B	5' – AAACGCTATCGGTAAAGGTATAGGC – 3'

Oligo Table 4. gRNA oligos

DoT:3.1_FW	5' – CCATTGTATGGTTGCCCCATC – 3'
DoT:3.1_RV	5' – GACTCACATATCCAGCCTCAAAC – 3'
DoT:6.2_FW	5' – AACTCTCCAGTGGCTCAGGA – 3'
DoT:6.2_RV	5' – CCCAGGGTCAGGAGGTTAAA – 3'
DoT:6.3_FW	5' – AGGGGCTGAGATAAGGAGAAAT – 3'
DoT:6.3_RV	5' – CAAATCTGGACACGGGGACC – 3'
DoT:7.3_FW	5' – AGGTGAAAAGAGGTTCTTCCCTG – 3'
DoT:7.3_RV	5' – CTGCTGAAGGAAGGTCTCAGT – 3'
DoT:8.1_FW	5' – TGCCACAACAACCTCCCTTT – 3'
DoT:8.1_RV	5' – GTGAGGGGGAAGCCAAACAC – 3'
DoT:8.2_FW	5' – ATCACAGAAACAGTCCGCA – 3'
DoT:8.2_RV	5' – AAAACGAAGGGGTGAGCAGG – 3'
DoT:8.3_FW	5' – TCGTGATGGAGGGCGTGAT – 3'
DoT:8.3_RV	5' – ATTTGGCTGGCCTCTGGTTT – 3'
DoT:8.4_FW	5' – ATGCAGTTTGCCTCCAGCAAT – 3'
DoT:8.4_RV	5' – GCCCATTCCTGGTTTTCT – 3'
DoT:9.1_FW	5' – AGAGAAGGCACCAGCAAGCA – 3'
DoT:9.1_RV	5' – CCCCTCATACTGGCTCCTGA – 3'
DoT:9.2_FW	5' – TGGTAGGTGGGGTGTGTTGT – 3'
DoT:9.2_RV	5' – ACGTGCATTGTTAGAGCAGTCA – 3'
IFT20_FW	5' – GCAGGCATTCCTTGGGTGG – 3'
IFT20_RV	5' – CTGTGGCAGGGCCAAGATGA – 3'

Oligo Table 5. Target site verification primers (wt + CRISPR-Cas9-treated samples)

DoT:3.1_5'ha_NdeI_InFu_FW	5' – ATTCACACCGCATAAACTGGGTTCTCTCCTTTTGTAGT – 3'
DoT:3.1_5'ha_MfeI_InFu_RV (γ)	5' – GAATTGCTAGCAATTTTCTCAATATTTATTGAGTAGGAGAC – 3'
DoT:3.1_5'ha_NaeI_InFu_RV (α)	5' – ACCTGAATGGAAGCCTTCTCAATATTTATTGAGTAGGAGAC – 3'
DoT:3.1_3'ha_NspI_InFu_FW (γ)	5' – TGAAAACCTCTGACATAGGGATGCCGTGGAGGAGTA – 3'
DoT:3.1_3'ha_EcoRV_InFu_FW (α)	5' – GTCGGAATGGACGATTAGGGATGCCGTGGAGGAGTA – 3'
DoT:3.1_3'ha_EcoRV_InFu_RV	5' – GGCCTCTTGCGGGATCCATGCACCCCTGTGAGTG – 3'
DoT:6.2_5'ha_NdeI_InFu_FW	5' – ATTCACACCGCATAAAGGTGCCTGAAGATGGGATGG – 3'
DoT:6.2_5'ha_MfeI_InFu_RV (γ)	5' – GAATTGCTAGCAATTGGAGCTTTGTAACAATGGATTCAT – 3'
DoT:6.2_5'ha_NaeI_InFu_RV (α)	5' – ACCTGAATGGAAGCCGGAGCTTTGTAACAATGGATTCAT – 3'
DoT:6.2_3'ha_NspI_InFu_FW (γ)	5' – TGAAAACCTCTGACAAGATCCCTCTCTGGGTCTGG – 3'
DoT:6.2_3'ha_EcoRV_InFu_FW (α)	5' – GTCGGAATGGACGATAGATCCCTCTCTGGGTCTGG – 3'
DoT:6.2_3'ha_EcoRV_InFu_RV	5' – GGCCTCTTGCGGGATCTACAAGTACCATGTGTGTCCCC – 3'
DoT:6.3_5'ha_NdeI_InFu_FW	5' – ATTCACACCGCATATAACCATGCTCTATCTCCCTCAGC – 3'
DoT:6.3_5'ha_MfeI_InFu_RV (γ)	5' – GAATTGCTAGCAATTTAGGAGAGATGCAGTTCAACGAT – 3'
DoT:6.3_5'ha_NaeI_InFu_RV (α)	5' – ACCTGAATGGAAGCCTAGGAGAGATGCAGTTCAACGAT – 3'
DoT:6.3_3'ha_NspI_InFu_FW (γ)	5' – TGAAAACCTCTGACAGACATTCCTGAGTCGGGTTTC – 3'
DoT:6.3_3'ha_EcoRV_InFu_FW (α)	5' – GTCGGAATGGACGATGACATTCCTGAGTCGGGTTTC – 3'
DoT:6.3_3'ha_EcoRV_InFu_RV	5' – GGCCTCTTGCGGGATGCTAGAGAGAATCACAACCAGTCC – 3'
DoT:7.3_5'ha_NdeI_InFu_FW	5' – ATTCACACCGCATAAAATCCTCTGCACATAGTGGGT – 3'
DoT:7.3_5'ha_MfeI_InFu_RV (γ)	5' – GAATTGCTAGCAATTACACAATAGAAGCTAATTATTCTGTCTGG – 3'
DoT:7.3_5'ha_NaeI_InFu_RV (α)	5' – ACCTGAATGGAAGCCACACAATAGAAGCTAATTATTCTGTCTGG – 3'
DoT:7.3_3'ha_NspI_InFu_FW (γ)	5' – TGAAAACCTCTGACACCAGCAAGTAGAAGCGCAGAC – 3'
DoT:7.3_3'ha_EcoRV_InFu_FW (α)	5' – GTCGGAATGGACGATCCAGCAAGTAGAAGCGCAGAC – 3'
DoT:7.3_3'ha_EcoRV_InFu_RV	5' – GGCCTCTTGCGGGATAGCCAGTAAACACCCAACGG – 3'
DoT:8.1_5'ha_NdeI_InFu_FW (γ)	5' – ATTCACACCGCATAGCAACCGGAGAGAGAGGATGA – 3'
DoT:8.1_5'ha_MfeI_InFu_RV (γ)	5' – GAATTGCTAGCAATTGCAGGTGGCTGGGGAC – 3'
DoT:8.1_3'ha_NspI_InFu_FW (γ)	5' – TGAAAACCTCTGACACCTTGGGCACAGACATGGAGC – 3'
DoT:8.1_3'ha_EcoRV_InFu_RV (γ)	5' – GGCCTCTTGCGGGATCCTTCTGAGATTTGCGGACTGG – 3'
DoT:8.2_5'ha_NdeI_InFu_FW (γ)	5' – ATTCACACCGCATAGCAGTGCTACTGAGGGAAGC – 3'
DoT:8.2_5'ha_MfeI_InFu_RV (γ)	5' – GAATTGCTAGCAATTTCTCCTGCTATTTTTAAAATGTGTCTG – 3'
DoT:8.2_3'ha_NspI_InFu_FW (γ)	5' – TGAAAACCTCTGACACCCATCCTTAAACGTAACCTAAGGCT – 3'

	3'
DoT:8.2_3'ha_EcoRV_InFu_RV (γ)	5' – GGCCTCTTGCGGGATGCCTCCTGTCTCTCATCCGT – 3'
DoT:8.4_5'ha_NdeI_InFu_FW (γ)	5' – ATTCACACCGCATACAAAGGACTCAAGGCCCCAA – 3'
DoT:8.4_5'ha_MfeI_InFu_RV (γ)	5' – GAATTGCTAGCAATTGGAAGATGTTACATTGGGCTGGC – 3'
DoT:8.4_3'ha_NspI_InFu_FW (γ)	5' – TGAAAACCTCTGACATGTTCTGTGTGTCATAATAACACACCAT – 3'
DoT:8.4_3'ha_EcoRV_InFu_RV (γ)	5' – GGCCTCTTGCGGGATGCCCAGACAAGGTGCTACAA – 3'
DoT:9.2_5'ha_NdeI_InFu_FW (γ)	5' – ATTCACACCGCATATGGGTACAACCAGTGAGCTAGA – 3'
DoT:9.2_5'ha_MfeI_InFu_RV (γ)	5' – GAATTGCTAGCAATTCACCTCTGAACACTGACCATAGGT – 3'
DoT:9.2_3'ha_NspI_InFu_FW (γ)	5' – TGAAAACCTCTGACATTGACCCAGTTGGGGCTGT – 3'
DoT:9.2_3'ha_EcoRV_InFu_RV (γ)	5' – GGCCTCTTGCGGGATACCACGTGCATTGTTAGAGCA – 3'
IFT20_5'ha_NdeI_InFu_FW	5' – ATTCACACCGCATATGTCCTCCTCTGGCTCTTTAC – 3'
IFT20_5'ha_MfeI_InFu_RV (γ)	5' – GAATTGCTAGCAATTCTTGTTTCAGTTCATCAAAGTGTAGC – 3'
IFT20_5'ha_NaeI_InFu_RV (α)	5' – ACCTGAATGGAAGCCCTTGTTTCAGTTCATCAAAGTGTAGC – 3'
IFT20_3'ha_NspI_InFu_FW (γ)	5' – TGAAAACCTCTGACAGACCCAGAGGTTACCCAGCAG – 3'
IFT20_3'ha_EcoRV_InFu_FW (α)	5' – GTCGGAATGGACGATGACCCAGAGGTTACCCAGCAG – 3'
IFT20_3'ha_EcoRV_InFu_RV	5' – GGCCTCTTGCGGGATAACAGAGAACCTGCCCTGA – 3'

Oligo Table 6. Primers for homologous arms cloning. (γ) = γRV LTRs-containing vectors. (α) = ASLV LTRs-containing vectors.

CoPr_gamCh_5'LTR_FW (γ)	5' – GCAACTTATCTGTGTCTGTCCGA – 3'
CoPr_gamCh_5'LTR_RV (γ)	5' – TCGGACAGACACAGATAAGTTGC – 3'
CoPr_gamCh_3'LTR_FW (γ)	5' – TTTAGTCTCCAGAAAAAGGGGGG – 3'
CoPr_gamCh_3'LTR_RV (γ)	5' – CCCCCCTTTTCTGGAGACTAAA – 3'
CoPr_AmCh_5'LTR_FW (α)	5' – AGGGAGCTCCAGGGCCCG – 3'
CoPr_AmCh_5'LTR_RV (α)	5' – CCCTGGAGCTCCCTCCGA – 3'
CoPr_AmCh_3'LTR_FW (α)	5' – AGCCAAGCTTATCGATGCGAT – 3'
CoPr_AmCh_3'LTR_RV (α)	5' – ATCGCATCGATAAGCTTGGCT – 3'
DoT:3.1_knock-in_FW	5' – CCGAGTGACAGGGTTGGTATTG – 3'
DoT:3.1_knock-in_RV	5' – GCAGGGTGCAAAGGAATAAGA – 3'
DoT:6.2_knock-in_FW	5' – GATCAACATTTTCAGGCTTCTGC – 3'
DoT:6.2_knock-in_RV	5' – CCCTCTCCTCTCAACCTACGA – 3'
DoT:6.3_knock-in_FW	5' – AGCATCTGTCTAAGTCACTGGGT – 3'
DoT:6.3_knock-in_RV	5' – GCAAAGAGAAGTGGGTCTGTCC – 3'
DoT:7.3_knock-in_FW	5' – CCACAAGTTCGTTCTTACAGC – 3'
DoT:7.3_knock-in_RV	5' – AGCCTTCCTTGGATAAGAGAACA – 3'
DoT:8.1_knock-in_FW	5' – AAGGCATTGTTGGGGACACA – 3'
DoT:8.1_knock-in_RV	5' – TGAGCTTCCCTCAGTAGCAC – 3'
DoT:8.2_knock-in_FW	5' – ATCACAGAAACCAGTCCGCA – 3'
DoT:8.2_knock-in_RV	5' – AAAACGAAGGGGTGAGCAGG – 3'
DoT:8.4_knock-in_FW	5' – ACCAACACAGGTGGGACTGG – 3'
DoT:8.4_knock-in_RV	5' – GCACAGGGTACATGTCAGGC – 3'
DoT:9.2_knock-in_FW	5' – CCTGGAGTCTTGATGCCCAA – 3'
DoT:9.2_knock-in_RV	5' – GGCACTTCGAGTGCAATCAC – 3'
IFT20_knock-in_FW	5' – TCCCAATTCAGGCAAGCATACG – 3'
IFT20_knock-in_RV	5' – ACCAAGCTCTCTCCACAACCT – 3'

Oligo Table 7. Primers for verification of targeted insertions. (γ) = γRV LTRs-containing vectors. (α) = ASLV LTRs-containing vectors.

Ion-Link-A	5' – CGCAGCATCGATGCATCGTAGCAGACCTCTCTATGGGCAGTCGGTGATGTCGACACTAGTGGT – 3'
Ion-Link-Ba (ligation)	5' – (P)-CCACTAGTGTCGACATCACCGACTGAAATA-(A) – 3'
Ion-Link-Bb (blocking)	5' – {G}CTA{C}GATG{C}ATC{G}AT{G}AA-(A) – 3'

Oligo Table 8. Adaptor oligos. {} LNA. (P) phosphate. (A) amino modifier C6.

lonA_MID1_mlvLTR2	5' – CATCTCATCCCTGCGTGTCTCCGACTCAGACGAGTGCCTGCTTGCCAAACCTACAGGTGG – 3'
lonA_MID2_mlvLTR2	5' – CATCTCATCCCTGCGTGTCTCCGACTCAGACGCTCGACAGCTTGCCAAACCTACAGGTGG – 3'
lonA_MID3_mlvLTR2	5' – CATCTCATCCCTGCGTGTCTCCGACTCAGAGACGCACTCGCTTGCCAAACCTACAGGTGG – 3'
lonA_MID4_mlvLTR2	5' – CATCTCATCCCTGCGTGTCTCCGACTCAGAGCACTGTAGGCTTGCCAAACCTACAGGTGG – 3'
lonA_MID5_mlvLTR2	5' – CATCTCATCCCTGCGTGTCTCCGACTCAGATCAGACACGGCTTGCCAAACCTACAGGTGG – 3'
lonA_MID6_mlvLTR2	5' – CATCTCATCCCTGCGTGTCTCCGACTCAGATATCGCGAGGCTTGCCAAACCTACAGGTGG – 3'
lonA_MID7_mlvLTR2	5' – CATCTCATCCCTGCGTGTCTCCGACTCAGCGTGTCTCTAGCTTGCCAAACCTACAGGTGG – 3'
lonA_MID8_mlvLTR2	5' – CATCTCATCCCTGCGTGTCTCCGACTCAGCTCGCGTGTCTGCTTGCCAAACCTACAGGTGG – 3'
lonA_MID9_mlvLTR2	5' – CATCTCATCCCTGCGTGTCTCCGACTCAGTAGTATCAGCGCTTGCCAAACCTACAGGTGG – 3'
lonA_MID10_mlvLTR2	5' – CATCTCATCCCTGCGTGTCTCCGACTCAGTCTCTATGCGGCTTGCCAAACCTACAGGTGG – 3'
lonA_MID11_mlvLTR2	5' – CATCTCATCCCTGCGTGTCTCCGACTCAGCAGCATCAGCTTGCCAAACCTACAGGTGG – 3'
lonA_MID12_mlvLTR2	5' – CATCTCATCCCTGCGTGTCTCCGACTCAGGCATATGCACGCTTGCCAAACCTACAGGTGG – 3'
lonA_MID13_mlvLTR2	5' – CATCTCATCCCTGCGTGTCTCCGACTCAGGAGTACAGTCGCTTGCCAAACCTACAGGTGG – 3'

Oligo Table 9. Provirus-specific primers used in the second PCR reaction. Sequencing adaptor sequence (red), barcode (purple), provirus-specific sequence (green).

q5GFP_MLV_F	5' – TCCTTCTCTAGGCGCCGAATTGATCC – 3'
q5GFP_R_2	5' – GTCGCCGTCCAGCTCGACCAG – 3'
q5GFP_pAS_F1	5' – TGAGTCGGCCGGTCAATCAAGC – 3'
q5GFP_probe	5' – FAM-AGCTGTTACCGGGGTGGTGCCCATC-BHQ-1 – 3'
q3mCh_F	5' – AGTTGGACATCACCTCCCAC – 3'
q3mCh_R	5' – CTTGTACAGCTCGTCCATGC – 3'
q3mCh_probe	5' – FAM-ACTACACCATCGTGGAAACAGTACGAAC-BHQ-1 – 3'
RPP30_F	5' – GATTTGGACCTGCGAGCG – 3'
RPP30_R	5' – GCGGCTGTCTCCACAAGT – 3'
RPP30_probe	5' – HEX-TTCTGACCTGAAGGCTCTGCGC-IB – 3'
dd_DoT:3.1_FW	5' – CCATTGTATGGTTGCCCCATC – 3'
dd_DoT:3.1_RV	5' – GACTCACATATCCAGCCTCAAAC – 3'
dd_DoT:6.2_FW	5' – AACTCTCCAGTGGCTCAGGA – 3'
dd_DoT:6.2_RV	5' – CCCAGGGTCAGGAGGTTAAA – 3'
dd_DoT:6.3_FW	5' – AGGGGCTGAGATAAGGAGAAAT – 3'
dd_DoT:6.3_RV	5' – CAAATCTGGACACGGGGACC – 3'
dd_DoT:7.3_FW	5' – AGGTGAAAAGAGGTTCTTCCCTG – 3'
dd_DoT:7.3_RV	5' – CTGCTGAAGGAAGGTCTCAGT – 3'
dd_IFT20_FW	5' – GCAGGCATTTCTTGGGTGG – 3'
dd_IFT20_RV	5' – CTGTGGCAGGGCCAAGATGA – 3'
dd_DoT:8.2_FW	5' – CCACATGCCTGTCTATTATCCCC – 3'
dd_DoT:8.2_RV	5' – GCAGATTTGACTTTGGCGGC – 3'
dd_DoT:9.2_FW	5' – TGGTAGGTGGGGTGTGTTGT – 3'
dd_DoT:9.2_RV	5' – TGCACATGGACAAAAGTGTCTG – 3'

Oligo Table 10. ddPCR primers and probes.

References

*review

- Achuthan, Vasudevan, Jill M. Perreira, Gregory A. Sowd, Maritza Puray-Chavez, William M. McDougall, Adriana Paulucci-Holthauzen, Xiaolin Wu, et al. 2018. "Capsid-CPSF6 Interaction Licenses Nuclear HIV-1 Trafficking to Sites of Viral DNA Integration." *Cell Host & Microbe* 24 (3): 392–404.e8.
- Aljanabi, S. M., and I. Martinez. 1997. "Universal and Rapid Salt-Extraction of High Quality Genomic DNA for PCR-Based Techniques." *Nucleic Acids Research* 25 (22): 4692–93.
- Arakawa, Hiroshi, Jessica Hauschild, and Jean-Marie Buerstedde. 2002. "Requirement of the Activation-Induced Deaminase (AID) Gene for Immunoglobulin Gene Conversion." *Science* 295 (5558): 1301–6.
- Autio, Matias I., Efthymios Motakis, Arnaud Perrin, Talal Bin Amin, Zenia Tiang, Dang Vinh Do, Jiaxu Wang, et al. 2024. "Computationally Defined and in Vitro Validated Putative Genomic Safe Harbour Loci for Transgene Expression in Human Cells." *eLife* 13 (January). <https://doi.org/10.7554/eLife.79592>.
- Aymard, François, Beatrix Bugler, Christine K. Schmidt, Emmanuelle Guillou, Pierre Caron, Sébastien Briois, Jason S. Iacovoni, et al. 2014. "Transcriptionally Active Chromatin Recruits Homologous Recombination at DNA Double-Strand Breaks." *Nature Structural & Molecular Biology* 21 (4): 366–74.
- Bachmann, Nadine, Chantal von Siebenthal, Valentina Vongrad, Teja Turk, Kathrin Neumann, Niko Beerenwinkel, Jasmina Bogojeska, et al. 2019. "Determinants of HIV-1 Reservoir Size and Long-Term Dynamics during Suppressive ART." *Nature Communications* 10 (1): 3193.
- Balakrishnan, Mini, Stephen R. Yant, Luong Tsai, Christopher O'Sullivan, Rujuta A. Bam, Angela Tsai, Anita Niedziela-Majka, Kirsten M. Stray, Roman Sakowicz, and Tomas Cihlar. 2013. "Non-Catalytic Site HIV-1 Integrase Inhibitors Disrupt Core Maturation and Induce a Reverse Transcription Block in Target Cells." *PloS One* 8 (9): e74163.
- Bannister, Andrew J., Robert Schneider, Fiona A. Myers, Alan W. Thorne, Colyn Crane-Robinson, and Tony Kouzarides. 2005. "Spatial Distribution of Di- and Tri-Methyl Lysine 36 of Histone H3 at Active Genes." *The Journal of Biological Chemistry* 280 (18): 17732–36.
- Bargonetti, J., I. Reynisdóttir, P. N. Friedman, and C. Prives. 1992. "Site-Specific Binding of Wild-Type p53 to Cellular DNA Is Inhibited by SV40 T Antigen and Mutant p53." *Genes & Development* 6 (10): 1886–98.
- Barski, Artem, Suresh Cuddapah, Kairong Cui, Tae-Young Roh, Dustin E. Schones, Zhibin Wang, Gang Wei, Iouri Chepelev, and Keji Zhao. 2007. "High-Resolution Profiling of Histone Methylations in the Human Genome." *Cell* 129 (4): 823–37.
- Battivelli, Emilie, Matthew S. Dahabieh, Mohamed Abdel-Mohsen, J. Peter Svensson, Israel Tojal Da Silva, Lillian B. Cohn, Andrea Gramatica, et al. 2018. "Distinct Chromatin Functional States Correlate with HIV Latency Reactivation in Infected Primary CD4 T Cells." *eLife* 7 (May). <https://doi.org/10.7554/eLife.34655>.
- Bhattacharya, Akash, Steven L. Alam, Thomas Fricke, Kaneil Zadrozny, Jaroslaw Sedzicki, Alexander B. Taylor, Borries Demeler, et al. 2014. "Structural Basis of HIV-1 Capsid Recognition by PF74 and CPSF6." *Proceedings of the National Academy of Sciences* 111 (52): 18625–30.
- *Bin Umair, Musab, Fujimura Nao Akusa, Hadia Kashif, Seerat-E-Fatima, Fatima Butt, Marium Azhar, Iqra Munir, et al. 2022. "Viruses as Tools in Gene Therapy, Vaccine Development, and Cancer Treatment." *Archives of Virology* 167 (6): 1387–1404.
- Blagodatski, Artem, Vera Batrak, Sabine Schmidl, Ulrike Schoetz, Randolph B. Caldwell, Hiroshi Arakawa, and Jean-Marie Buerstedde. 2009. "A Cis-Acting Diversification Activator Both Necessary and Sufficient for AID-Mediated Hypermutation." *PLoS*

- Genetics* 5 (1): e1000332.
- Borowiec, J. A., and J. Hurwitz. 1988a. "ATP Stimulates the Binding of Simian Virus 40 (SV40) Large Tumor Antigen to the SV40 Origin of Replication." *Proceedings of the National Academy of Sciences of the United States of America* 85 (1): 64–68.
- . 1988b. "Localized Melting and Structural Changes in the SV40 Origin of Replication Induced by T-Antigen." *The EMBO Journal* 7 (10): 3149–58.
- Brinkman, Eva K., Tao Chen, Marcel de Haas, Hanna A. Holland, Waseem Akhtar, and Bas van Steensel. 2018. "Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks." *Molecular Cell* 70 (5): 801–13.e6.
- Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop. 1989. "Retroviral Integration: Structure of the Initial Covalent Product and Its Precursor, and a Role for the Viral IN Protein." *Proceedings of the National Academy of Sciences of the United States of America* 86 (8): 2525–29.
- Buerstedde, Jean-Marie, Jukka Alinikula, Hiroshi Arakawa, Jessica J. McDonald, and David G. Schatz. 2014. "Targeting of Somatic Hypermutation by Immunoglobulin Enhancer and Enhancer-like Sequences." *PLoS Biology* 12 (4): e1001831.
- Bushman, F. D. 1994. "Tethering Human Immunodeficiency Virus 1 Integrase to a DNA Site Directs Integration to Nearby Sequences." *Proceedings of the National Academy of Sciences of the United States of America* 91 (20): 9233–37.
- Bushman, F. D., and M. D. Miller. 1997. "Tethering Human Immunodeficiency Virus Type 1 Preintegration Complexes to Target DNA Promotes Integration at Nearby Sites." *Journal of Virology* 71 (1): 458–64.
- *Bushman, Frederic D. 2020. "Retroviral Insertional Mutagenesis in Humans: Evidence for Four Genetic Mechanisms Promoting Expansion of Cell Clones." *Molecular Therapy: The Journal of the American Society of Gene Therapy* 28 (2): 352–56.
- Butsch, M., S. Hull, Y. Wang, T. M. Roberts, and K. Boris-Lawrie. 1999. "The 5' RNA Terminus of Spleen Necrosis Virus Contains a Novel Posttranscriptional Control Element That Facilitates Human Immunodeficiency Virus Rev/RRE-Independent Gag Production." *Journal of Virology* 73 (6): 4847–55.
- Cavazzana-Calvo, M., S. Hacein-Bey, G. de Saint-Basile, F. Le Deist, and A. Fischer. 2000. "[Gene therapy of severe combined immunodeficiencies]." *Transfusion clinique et biologique : journal de la Societe francaise de transfusion sanguine* 7 (3): 259–60.
- Chellappan, S., V. B. Kraus, B. Kroger, K. Munger, P. M. Howley, W. C. Phelps, and J. R. Nevins. 1992. "Adenovirus E1A, Simian Virus 40 Tumor Antigen, and Human Papillomavirus E7 Protein Share the Capacity to Disrupt the Interaction between Transcription Factor E2F and the Retinoblastoma Gene Product." *Proceedings of the National Academy of Sciences of the United States of America* 89 (10): 4549–53.
- Chellappan, S. P., S. Hiebert, M. Mudryj, J. M. Horowitz, and J. R. Nevins. 1991. "The E2F Transcription Factor Is a Cellular Target for the RB Protein." *Cell* 65 (6): 1053–61.
- Cheng, Jingwei, Orit Rozenblatt-Rosen, Kelly G. Paulson, Paul Nghiem, and James A. DeCaprio. 2013. "Merkel Cell Polyomavirus Large T Antigen Has Growth-Promoting and Inhibitory Activities." *Journal of Virology* 87 (11): 6118–26.
- Chen, H., M. K. Bechtel, Y. Shi, A. Phipps, L. E. Mathes, K. A. Hayes, and P. Roy-Burman. 1998. "Pathogenicity Induced by Feline Leukemia Virus, Rickard Strain, Subgroup A Plasmid DNA (pFRA)." *Journal of Virology* 72 (9): 7048–56.
- Chen, Heng-Chang, Javier P. Martinez, Eduard Zorita, Andreas Meyerhans, and Guillaume J. Filion. 2016. "Position Effects Influence HIV Latency Reversal." *Nature Structural & Molecular Biology* 24 (1): 47–54.
- Chen, Xiaoyu, Josephine M. Janssen, Jin Liu, Ignazio Maggio, Anke E. J. 't Jong, Harald M. M. Mikkers, and Manuel A. F. V. Gonçalves. 2017. "In Trans Paired Nicking Triggers Seamless Genome Editing without Double-Stranded DNA Cutting." *Nature Communications* 8 (1): 657.
- Chen, Xiaoyu, Marrit Rinsma, Josephine M. Janssen, Jin Liu, Ignazio Maggio, and Manuel A. F. V. Gonçalves. 2016. "Probing the Impact of Chromatin Conformation on Genome Editing Tools." *Nucleic Acids Research* 44 (13): 6482–92.

- Cherepanov, Peter, Eric Devroe, Pamela A. Silver, and Alan Engelman. 2004. "Identification of an Evolutionarily Conserved Domain in Human Lens Epithelium-Derived Growth Factor/transcriptional Co-Activator p75 (LEDGF/p75) That Binds HIV-1 Integrase." *The Journal of Biological Chemistry* 279 (47): 48883–92.
- Cherepanov, Peter, Goedele Maertens, Paul Proost, Bart Devreese, Jozef Van Beeumen, Yves Engelborghs, Erik De Clercq, and Zeger Debyser. 2003. "HIV-1 Integrase Forms Stable Tetramers and Associates with LEDGF/p75 Protein in Human Cells." *The Journal of Biological Chemistry* 278 (1): 372–81.
- Chiappesi, Flavia, Marcela D'alincourt Salazar, Heidi Contreras, Vu H. Nguyen, Joy Martinez, Yoonsuh Park, Jenny Nguyen, et al. 2020. "Development of a Multi-Antigenic SARS-CoV-2 Vaccine Candidate Using a Synthetic Poxvirus Platform." *Nature Communications* 11 (1): 6121.
- Christ, Frauke, Stephen Shaw, Jonas Demeulemeester, Belete A. Desimmie, Arnaud Marchand, Scott Butler, Wim Smets, et al. 2012. "Small-Molecule Inhibitors of the LEDGF/p75 Binding Site of Integrase Block HIV Replication and Modulate Integrase Multimerization." *Antimicrobial Agents and Chemotherapy* 56 (8): 4365–74.
- Chun, T. W., D. Engel, M. M. Berrey, T. Shea, L. Corey, and A. S. Fauci. 1998. "Early Establishment of a Pool of Latently Infected, Resting CD4(+) T Cells during Primary HIV-1 Infection." *Proceedings of the National Academy of Sciences of the United States of America* 95 (15): 8869–73.
- Collins, K. L., and T. J. Kelly. 1991. "Effects of T Antigen and Replication Protein A on the Initiation of DNA Synthesis by DNA Polymerase Alpha-Primase." *Molecular and Cellular Biology* 11 (4): 2108–15.
- Concordet, Jean-Paul, and Maximilian Haeussler. 2018. "CRISPOR: Intuitive Guide Selection for CRISPR/Cas9 Genome Editing Experiments and Screens." *Nucleic Acids Research* 46 (W1): W242–45.
- Conticello, Silvestro G., Cornelia J. F. Thomas, Svend K. Petersen-Mahrt, and Michael S. Neuberger. 2005. "Evolution of the AID/APOBEC Family of Polynucleotide (deoxy)cytidine Deaminases." *Molecular Biology and Evolution* 22 (2): 367–77.
- Counsell, John R., Zeinab Asgarian, Jinhong Meng, Veronica Ferrer, Conrad A. Vink, Steven J. Howe, Simon N. Waddington, et al. 2017. "Lentiviral Vectors Can Be Used for Full-Length Dystrophin Gene Therapy." *Scientific Reports* 7 (March): 44775.
- Danecek, Petr, James K. Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O. Pollard, Andrew Whitwham, et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience* 10 (2). <https://doi.org/10.1093/gigascience/giab008>.
- *Darcis, Gilles, Benoît Van Driessche, and Carine Van Lint. 2017. "HIV Latency: Should We Shock or Lock?" *Trends in Immunology* 38 (3): 217–28.
- Das, Priyot, Rebeca San Martin, and Rachel Patton McCord. 2023. "Differential Contributions of Nuclear Lamina Association and Genome Compartmentalization to Gene Regulation." *Nucleus*, December. <https://doi.org/10.1080/19491034.2023.2197693>.
- Davies, Peers, Scott Jones, Stephen Dunham, and Rachael E. Tarlinton. 2023. "Associations between Small Ruminant Lentivirus Infection and Total Milk Yield and Somatic Cell Count in a Dairy Sheep Flock." *The Veterinary Record* 192 (11): e2731.
- *Debyser, Zeger, Gerlinde Vansant, Anne Bruggemans, Julie Janssens, and Frauke Christ. 2018. "Insight in HIV Integration Site Selection Provides a Block-and-Lock Strategy for a Functional Cure of HIV Infection." *Viruses* 11 (1). <https://doi.org/10.3390/v11010012>.
- DeCaprio, J. A., J. W. Ludlow, J. Figge, J. Y. Shew, C. M. Huang, W. H. Lee, E. Marsilio, E. Paucha, and D. M. Livingston. 1988. "SV40 Large Tumor Antigen Forms a Specific Complex with the Product of the Retinoblastoma Susceptibility Gene." *Cell* 54 (2): 275–83.
- Dekker, Job, Andrew S. Belmont, Mitchell Guttman, Victor O. Leshyk, John T. Lis, Stavros Lomvardas, Leonid A. Mirny, et al. 2017. "The 4D Nucleome Project." *Nature* 549 (7671): 219–26.
- De Ravin, Suk See, Ling Su, Narda Theobald, Uimook Choi, Janet L. Macpherson, Michael Poidinger, Geoff Symonds, et al. 2014. "Enhancers Are Major Targets for Murine

- Leukemia Virus Vector Integration." *Journal of Virology* 88 (8): 4504–13.
- De Rijck, Jan, Christine de Kogel, Jonas Demeulemeester, Sofie Vets, Sara El Ashkar, Nirav Malani, Frederic D. Bushman, et al. 2013. "The BET Family of Proteins Targets Moloney Murine Leukemia Virus Integration near Transcription Start Sites." *Cell Reports* 5 (4): 886–94.
- Derse, David, Bruce Crise, Yuan Li, Gerald Princler, Nicole Lum, Claudia Stewart, Connor F. McGrath, Stephen H. Hughes, David J. Munroe, and Xiaolin Wu. 2007. "Human T-Cell Leukemia Virus Type 1 Integration Target Sites in the Human Genome: Comparison with Those of Other Retroviruses." *Journal of Virology* 81 (12): 6731–41.
- Desimmie, Belete Ayele, Rik Schrijvers, Jonas Demeulemeester, Doortje Borrenberghs, Caroline Weydert, Wannes Thys, Sofie Vets, et al. 2013. "LEDGINs Inhibit Late Stage HIV-1 Replication by Modulating Integrase Multimerization in the Virions." *Retrovirology* 10 (May):57.
- Dinesh, Ravi K., Benjamin Barnhill, Anoj Ilanges, Lizhen Wu, Daniel A. Michelson, Filip Senigl, Jukka Alinikula, Jeffrey Shabanowitz, Donald F. Hunt, and David G. Schatz. 2020. "Transcription Factor Binding at Ig Enhancers Is Linked to Somatic Hypermutation Targeting." *European Journal of Immunology* 50 (3): 380–95.
- Dixon, Jesse R., Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, and Bing Ren. 2012. "Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions." *Nature* 485 (7398): 376–80.
- Dobbelstein, M., and J. Roth. 1998. "The Large T Antigen of Simian Virus 40 Binds and Inactivates p53 but Not p73." *The Journal of General Virology* 79 (Pt 12) (December):3079–83.
- Drayman, Nir, Orly Ben-Nun-Shaul, Veronika Butin-Israeli, Rohit Srivastava, Ariel M. Rubinstein, Caroline S. Mock, Ela Elyada, Yinon Ben-Neriah, Galit Lahav, and Ariella Oppenheim. 2016. "p53 Elevation in Human Cells Halt SV40 Infection by Inhibiting T-Ag Expression." *Oncotarget* 7 (33): 52643–60.
- Dyson, N., R. Bernards, S. H. Friend, L. R. Gooding, J. A. Hassell, E. O. Major, J. M. Pipas, T. Vandyke, and E. Harlow. 1990. "Large T Antigens of Many Polyomaviruses Are Able to Form Complexes with the Retinoblastoma Protein." *Journal of Virology* 64 (3): 1353–56.
- Einkauf, Kevin B., Guinevere Q. Lee, Ce Gao, Radwa Sharaf, Xiaoming Sun, Stephane Hua, Samantha My Chen, et al. 2019. "Intact HIV-1 Proviruses Accumulate at Distinct Chromosomal Positions during Prolonged Antiretroviral Therapy." *The Journal of Clinical Investigation* 129 (3): 988–98.
- Einkauf, Kevin B., Matthew R. Osborn, Ce Gao, Weiwei Sun, Xiaoming Sun, Xiaodong Lian, Elizabeth M. Parsons, et al. 2022. "Parallel Analysis of Transcription, Integration, and Sequence of Single HIV-1 Proviruses." *Cell* 185 (2): 266–82.e15.
- El Ashkar, Sara, Jan De Rijck, Jonas Demeulemeester, Sofie Vets, Paradise Madlala, Katerina Cermakova, Zeger Debyser, and Rik Gijsbers. 2014. "BET-Independent MLV-Based Vectors Target Away From Promoters and Regulatory Elements." *Molecular Therapy. Nucleic Acids* 3 (7): e179.
- El Ashkar, Sara, Dominique Van Looveren, Franziska Schenk, Lenard S. Vranckx, Jonas Demeulemeester, Jan De Rijck, Zeger Debyser, Ute Modlich, and Rik Gijsbers. 2017. "Engineering Next-Generation BET-Independent MLV Vectors for Safer Gene Therapy." *Molecular Therapy. Nucleic Acids* 7 (June):231–45.
- Elleder, Daniel, Adam Pavlíček, Jan Paces, and Jirí Hejnar. 2002. "Preferential Integration of Human Immunodeficiency Virus Type 1 into Genes, Cytogenetic R Bands and GC-Rich DNA Regions: Insight from the Human Genome Sequence." *FEBS Letters* 517 (1-3): 285–86.
- Elliott, Jennifer L., Jenna E. Eschbach, Pratibha C. Koneru, Wen Li, Maritza Puray-Chavez, Dana Townsend, Dana Q. Lawson, Alan N. Engelman, Mamuka Kvaratskhelia, and Sebla B. Kutluay. 2020. "Integrase-RNA Interactions Underscore the Critical Role of Integrase in HIV-1 Virion Morphogenesis." *eLife* 9 (September). <https://doi.org/10.7554/eLife.54311>.

- Engelman, A., K. Mizuuchi, and R. Craigie. 1991. "HIV-1 DNA Integration: Mechanism of Viral DNA Cleavage and DNA Strand Transfer." *Cell* 67 (6): 1211–21.
- Erasmus, Jesse H., Amit P. Khandhar, Jeff Guderian, Brian Granger, Jacob Archer, Michelle Archer, Emily Gage, et al. 2018. "A Nanostructured Lipid Carrier for Delivery of a Replicating Viral RNA Provides Single, Low-Dose Protection against Zika." *Molecular Therapy: The Journal of the American Society of Gene Therapy* 26 (10): 2507–22.
- Ernst, Jason, and Manolis Kellis. 2010. "Discovery and Characterization of Chromatin States for Systematic Annotation of the Human Genome." *Nature Biotechnology* 28 (8): 817–25.
- Eyquem, Justin, Jorge Mansilla-Soto, Theodoros Giavridis, Sjoukje J. C. van der Stegen, Mohamad Hamieh, Kristen M. Cunanan, Ashlesha Odak, Mithat Gönen, and Michel Sadelain. 2017. "Targeting a CAR to the TRAC Locus with CRISPR/Cas9 Enhances Tumour Rejection." *Nature* 543 (7643): 113–17.
- Fábryová, Helena, Tomáš Hron, Hana Kabíčková, Mary Poss, and Daniel Elleder. 2015. "Induction and Characterization of a Replication Competent Cervid Endogenous Gammaretrovirus (CrERV) from Mule Deer Cells." *Virology* 485 (November):96–103.
- *Fandiño, Sergio, Esperanza Gomez-Lucia, Laura Benítez, and Ana Doménech. 2023. "Avian Leukosis: Will We Be Able to Get Rid of It?" *Animals : An Open Access Journal from MDPI* 13 (14). <https://doi.org/10.3390/ani13142358>.
- Feng, Huichen, Masahiro Shuda, Yuan Chang, and Patrick S. Moore. 2008. "Clonal Integration of a Polyomavirus in Human Merkel Cell Carcinoma." *Science* 319 (5866): 1096–1100.
- *Fischer, Alain, and Salima Hacein-Bey-Abina. 2020. "Gene Therapy for Severe Combined Immunodeficiencies and beyond." *The Journal of Experimental Medicine* 217 (2). <https://doi.org/10.1084/jem.20190607>.
- Francis, Ashwanth C., Mariana Marin, Parmit K. Singh, Vasudevan Achuthan, Mathew J. Prellberg, Kristina Palermينو-Rowland, Shuiyun Lan, et al. 2020. "HIV-1 Replication Complexes Accumulate in Nuclear Speckles and Integrate into Speckle-Associated Genomic Domains." *Nature Communications* 11 (1): 3505.
- Friedman, J. R., W. J. Fredericks, D. E. Jensen, D. W. Speicher, X. P. Huang, E. G. Neilson, and F. J. Rauscher 3rd. 1996. "KAP-1, a Novel Corepressor for the Highly Conserved KRAB Repression Domain." *Genes & Development* 10 (16): 2067–78.
- Gerber, P. 1962. "An Infectious Deoxyribonucleic Acid Derived from Vacuolating Virus (SV40)." *Virology* 16 (January):96–97.
- Gessain, Antoine, and Olivier Cassar. 2012. "Epidemiological Aspects and World Distribution of HTLV-1 Infection." *Frontiers in Microbiology* 3 (November):388.
- Gillis, Eric P., Kyle Parcella, Michael Bowsher, James H. Cook, Christiana Iwuagwu, B. Narasimhulu Naidu, Manoj Patel, et al. 2023. "Potent Long-Acting Inhibitors Targeting the HIV-1 Capsid Based on a Versatile Quinazolin-4-One Scaffold." *Journal of Medicinal Chemistry* 66 (3): 1941–54.
- Gilroy, Kathryn L., Anne Terry, Asif Naseer, Jeroen de Ridder, Amin Allahyar, Weiwei Wang, Eric Carpenter, et al. 2016. "Gamma-Retrovirus Integration Marks Cell Type-Specific Cancer Genes: A Novel Profiling Tool in Cancer Genomics." *PloS One* 11 (4): e0154070.
- Gish, W. R., and M. R. Botchan. 1987. "Simian Virus 40-Transformed Human Cells That Express Large T Antigens Defective for Viral DNA Replication." *Journal of Virology* 61 (9): 2864–76.
- Gossai, Anala, Tim Waterboer, Heather H. Nelson, Angelika Michel, Martina Willhauck-Fleckenstein, Shohreh F. Farzan, Anne G. Hoen, et al. 2016. "Seroepidemiology of Human Polyomaviruses in a US Population." *American Journal of Epidemiology* 183 (1): 61–69.
- Guelen, Lars, Ludo Pagie, Emilie Brasset, Wouter Meuleman, Marius B. Faza, Wendy Talhout, Bert H. Eussen, et al. 2008. "Domain Organization of Human Chromosomes Revealed by Mapping of Nuclear Lamina Interactions." *Nature* 453 (7197): 948–51.
- Guikema, Jeroen E. J., Erin K. Linehan, Daisuke Tsuchimoto, Yusaku Nakabeppu, Phyllis R. Strauss, Janet Stavnezer, and Carol E. Schrader. 2007. "APE1- and APE2-Dependent DNA Breaks in Immunoglobulin Class Switch Recombination." *The Journal of Experimental Medicine* 204 (12): 3017–26.

- Gupta, Saumya Shree, Tobias Maetzig, Goedele N. Maertens, Azar Sharif, Michael Rothe, Magdalena Weidner-Glunde, Melanie Galla, Axel Schambach, Peter Cherepanov, and Thomas F. Schulz. 2013. "Bromo- and Extraterminal Domain Chromatin Regulators Serve as Cofactors for Murine Leukemia Virus Integration." *Journal of Virology* 87 (23): 12721–36.
- Hacein-Bey-Abina, Salima, Alain Fischer, and Marina Cavazzana-Calvo. 2002. "Gene Therapy of X-Linked Severe Combined Immunodeficiency." *International Journal of Hematology* 76 (4): 295–98.
- Hacein-Bey-Abina, Salima, Alexandrine Garrigue, Gary P. Wang, Jean Soulier, Annick Lim, Estelle Morillon, Emmanuelle Clappier, et al. 2008. "Insertional Oncogenesis in 4 Patients after Retrovirus-Mediated Gene Therapy of SCID-X1." *The Journal of Clinical Investigation* 118 (9): 3132–42.
- Hacein-Bey-Abina, Salima, Christof von Kalle, Manfred Schmidt, Françoise Le Deist, Nicolas Wulffraat, Elisabeth McIntyre, Isabelle Radford, et al. 2003. "A Serious Adverse Event after Successful Gene Therapy for X-Linked Severe Combined Immunodeficiency." *The New England Journal of Medicine* 348 (3): 255–56.
- Hacein-Bey-Abina, S., C. Von Kalle, M. Schmidt, M. P. McCormack, N. Wulffraat, P. Leboulch, A. Lim, et al. 2003. "LMO2-Associated Clonal T Cell Proliferation in Two Patients after Gene Therapy for SCID-X1." *Science (New York, N.Y.)* 302 (5644): 415–19.
- Hanger, J. J., L. D. Bromham, J. J. McKee, T. M. O'Brien, and W. F. Robinson. 2000. "The Nucleotide Sequence of Koala (*Phascogale* *Cinereus*) Retrovirus: A Novel Type C Endogenous Virus Related to Gibbon Ape Leukemia Virus." *Journal of Virology* 74 (9): 4264–72.
- Hansen, U., D. G. Tenen, D. M. Livingston, and P. A. Sharp. 1981. "T Antigen Repression of SV40 Early Transcription from Two Promoters." *Cell* 27 (3 Pt 2): 603–13.
- Hassan, Ahmed O., Friederike Feldmann, Haiyan Zhao, David T. Curiel, Atsushi Okumura, Tsing-Lee Tang-Huau, James Brett Case, et al. 2021. "A Single Intranasal Dose of Chimpanzee Adenovirus-Vectored Vaccine Protects against SARS-CoV-2 Infection in Rhesus Macaques." *Cell Reports. Medicine* 2 (4): 100230.
- Hein, Jennifer, Sergei Boichuk, Jiaping Wu, Yuan Cheng, Raimundo Freire, Parmjit S. Jat, Thomas M. Roberts, and Ole V. Gjoerup. 2009. "Simian Virus 40 Large T Antigen Disrupts Genome Integrity and Activates a DNA Damage Response via Bub1 Binding." *Journal of Virology* 83 (1): 117–27.
- Hejnar, J., P. Hájková, J. Plachy, D. Elleder, V. Stepanets, and J. Svoboda. 2001. "CpG Island Protects Rous Sarcoma Virus-Derived Vectors Integrated into Nonpermissive Cells from DNA Methylation and Transcriptional Suppression." *Proceedings of the National Academy of Sciences of the United States of America* 98 (2): 565–69.
- Hejnar, J., J. Plachý, J. Geryk, O. Machon, K. Trejbalová, R. V. Guntaka, and J. Svoboda. 1999. "Inhibition of the Rous Sarcoma Virus Long Terminal Repeat-Driven Transcription by in Vitro Methylation: Different Sensitivity in Permissive Chicken Cells versus Mammalian Cells." *Virology* 255 (1): 171–81.
- Hejnar, J., J. Svoboda, J. Geryk, V. J. Fincham, and R. Hák. 1994. "High Rate of Morphological Reversion in Tumor Cell Line H-19 Associated with Permanent Transcriptional Suppression of the LTR, v-Src, LTR Provirus." *Cell Growth & Differentiation: The Molecular Biology Journal of the American Association for Cancer Research* 5 (3): 277–85.
- Hesbacher, Sonja, Lisa Pfitzer, Katharina Wiedorfer, Sabrina Angermeyer, Andreas Borst, Sebastian Haferkamp, Claus-Jürgen Scholz, Marion Wobser, David Schrama, and Roland Houben. 2016. "RB1 Is the Crucial Target of the Merkel Cell Polyomavirus Large T Antigen in Merkel Cell Carcinoma Cells." *Oncotarget* 7 (22): 32956–68.
- Hockemeyer, Dirk, Frank Soldner, Caroline Beard, Qing Gao, Maisam Mitalipova, Russell C. DeKaveler, George E. Katibah, et al. 2009. "Efficient Targeting of Expressed and Silent Genes in Human ESCs and iPSCs Using Zinc-Finger Nucleases." *Nature Biotechnology* 27 (9): 851–57.
- Hoffman, Michael M., Jason Ernst, Steven P. Wilder, Anshul Kundaje, Robert S. Harris, Max

- Libbrecht, Belinda Giardine, et al. 2013. "Integrative Annotation of Chromatin Elements from ENCODE Data." *Nucleic Acids Research* 41 (2): 827–41.
- Homiski, Caleb, Rama Dey-Rao, Shichen Shen, Jun Qu, and Thomas Melendy. 2024. "DNA Damage-Induced Phosphorylation of a Replicative DNA Helicase Results in Inhibition of DNA Replication through Attenuation of Helicase Function." *Nucleic Acids Research* 52 (17): 10311–28.
- Horvath, C. J., M. A. Simon, D. J. Bergsagel, D. R. Pauley, N. W. King, R. L. Garcea, and D. J. Ringler. 1992. "Simian Virus 40-Induced Disease in Rhesus Monkeys with Simian Acquired Immunodeficiency Syndrome." *The American Journal of Pathology* 140 (6): 1431–40.
- Houben, Roland, Christian Adam, Anne Baeurle, Sonja Hesbacher, Johannes Grimm, Sabrina Angermeyer, Katharina Henzel, et al. 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. Journal International Du Cancer* 130 (4): 847–56.
- Howe, Steven J., Marc R. Mansour, Kerstin Schwarzwaelder, Cynthia Bartholomae, Michael Hubank, Helena Kempinski, Martijn H. Brugman, et al. 2008. "Insertional Mutagenesis Combined with Acquired Somatic Mutations Causes Leukemogenesis Following Gene Therapy of SCID-X1 Patients." *The Journal of Clinical Investigation* 118 (9): 3143–50.
- Ho, Ya-Chi, Liang Shan, Nina N. Hosmane, Jeffrey Wang, Sarah B. Laskey, Daniel I. S. Rosenbloom, Jun Lai, Joel N. Blankson, Janet D. Siliciano, and Robert F. Siliciano. 2013. "Replication-Competent Noninduced Proviruses in the Latent Reservoir Increase Barrier to HIV-1 Cure." *Cell* 155 (3): 540–51.
- Huang, Amy S., Victor Ramos, Thiago Y. Oliveira, Christian Gaebler, Mila Jankovic, Michel C. Nussenzweig, and Lillian B. Cohn. 2021. "Integration Features of Intact Latent HIV-1 in CD4+ T Cell Clones Contribute to Viral Persistence." *The Journal of Experimental Medicine* 218 (12). <https://doi.org/10.1084/jem.20211427>.
- Imai, Kohsuke, Geir Slupphaug, Wen-I Lee, Patrick Revy, Shigeaki Nonoyama, Nadia Catalan, Leman Yel, et al. 2003. "Human Uracil-DNA Glycosylase Deficiency Associated with Profoundly Impaired Immunoglobulin Class-Switch Recombination." *Nature Immunology* 4 (10): 1023–28.
- Jamali, Arezoo, Laura Kapitza, Thomas Schaser, Ian C. D. Johnston, Christian J. Buchholz, and Jessica Hartmann. 2019. "Highly Efficient and Selective CAR-Gene Transfer Using CD4- and CD8-Targeted Lentiviral Vectors." *Molecular Therapy. Methods & Clinical Development* 13 (June):371–79.
- Jiang, Chenyang, Xiaodong Lian, Ce Gao, Xiaoming Sun, Kevin B. Einkauf, Joshua M. Chevalier, Samantha M. Y. Chen, et al. 2020. "Distinct Viral Reservoirs in Individuals with Spontaneous Control of HIV-1." *Nature* 585 (7824): 261–67.
- Jinek, Martin, Alexandra East, Aaron Cheng, Steven Lin, Enbo Ma, and Jennifer Doudna. 2013. "RNA-Programmed Genome Editing in Human Cells." *eLife* 2 (January):e00471.
- Jurado, Kellie A., Hao Wang, Alison Slaughter, Lei Feng, Jacques J. Kessl, Yasuhiro Koh, Weifeng Wang, et al. 2013. "Allosteric Integrase Inhibitor Potency Is Determined through the Inhibition of HIV-1 Particle Maturation." *Proceedings of the National Academy of Sciences of the United States of America* 110 (21): 8690–95.
- Kalina, Jirí, Filip Senigl, Alena Micáková, Jitka Mucksová, Jana Blazková, Haifeng Yan, Martin Poplstein, Jirí Hejnar, and Pavel Trefil. 2007. "Retrovirus-Mediated in Vitro Gene Transfer into Chicken Male Germ Line Cells." *Reproduction* 134 (3): 445–53.
- Kao, C., P. Hauser, W. S. Reznikoff, and C. A. Reznikoff. 1993. "Simian Virus 40 (SV40) T-Antigen Mutations in Tumorigenic Transformation of SV40-Immortalized Human Uroepithelial Cells." *Journal of Virology* 67 (4): 1987–95.
- Karimi, Mohammad M., Preeti Goyal, Irina A. Maksakova, Misha Bilenky, Danny Leung, Jie Xin Tang, Yoichi Shinkai, et al. 2011. "DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs." *Cell Stem Cell* 8 (6): 676–87.
- Katoh, Yohei, Saki Michisaka, Shohei Nozaki, Teruki Funabashi, Tomoaki Hirano, Ryota

- Takei, and Kazuhisa Nakayama. 2017. "Practical Method for Targeted Disruption of Cilia-Related Genes by Using CRISPR/Cas9-Mediated, Homology-Independent Knock-in System." *Molecular Biology of the Cell* 28 (7): 898–906.
- Kean, Jaime M., Suchitra Rao, Michael Wang, and Robert L. Garcea. 2009. "Seroepidemiology of Human Polyomaviruses." *PLoS Pathogens* 5 (3): e1000363.
- *Keim, Celia, David Kazadi, Gerson Rothschild, and Uttiya Basu. 2013. "Regulation of AID, the B-Cell Genome Mutator." *Genes & Development* 27 (1): 1–17.
- Keller, J. M., and J. C. Alwine. 1984. "Activation of the SV40 Late Promoter: Direct Effects of T Antigen in the Absence of Viral DNA Replication." *Cell* 36 (2): 381–89.
- Kellogg, Caroline, Valentina L. Kouznetsova, and Igor F. Tsigelny. 2022. "Interactions of Large T-Antigen (LT) Protein of Polyomaviruses with p53 Unfold Their Cancerogenic Potential." *Journal of Biomolecular Structure & Dynamics* 40 (11): 5243–52.
- Kessl, Jacques J., Sebla B. Kutluay, Dana Townsend, Stephanie Rebensburg, Alison Slaughter, Ross C. Larue, Nikoloz Shkriabai, et al. 2016. "HIV-1 Integrase Binds the Viral RNA Genome and Is Essential during Virion Morphogenesis." *Cell* 166 (5): 1257–68.e12.
- Kierstead, T. D., and M. J. Tevethia. 1993. "Association of p53 Binding and Immortalization of Primary C57BL/6 Mouse Embryo Fibroblasts by Using Simian Virus 40 T-Antigen Mutants Bearing Internal Overlapping Deletion Mutations." *Journal of Virology* 67 (4): 1817–29.
- Kim, S. S., Y. M. Chen, E. O'Leary, R. Witzgall, M. Vidal, and J. V. Bonventre. 1996. "A Novel Member of the RING Finger Family, KRIP-1, Associates with the KRAB-A Transcriptional Repressor Domain of Zinc Finger Proteins." *Proceedings of the National Academy of Sciences of the United States of America* 93 (26): 15299–304.
- Kim, Y. G., J. Cha, and S. Chandrasegaran. 1996. "Hybrid Restriction Enzymes: Zinc Finger Fusions to Fok I Cleavage Domain." *Proceedings of the National Academy of Sciences of the United States of America* 93 (3): 1156–60.
- Kohler, Kristin M., Jessica J. McDonald, Jamie L. Duke, Hiroshi Arakawa, Sally Tan, Steven H. Kleinstein, Jean-Marie Buerstedde, and David G. Schatz. 2012. "Identification of Core DNA Elements That Target Somatic Hypermutation." *Journal of Immunology* 189 (11): 5314–26.
- Kohn, Donald B., Claire Booth, Elizabeth M. Kang, Sung-Yun Pai, Kit L. Shaw, Giorgia Santilli, Myriam Armant, et al. 2020. "Lentiviral Gene Therapy for X-Linked Chronic Granulomatous Disease." *Nature Medicine* 26 (2): 200–206.
- Kumar, R., T. A. Firak, C. T. Schroll, and K. N. Subramanian. 1986. "Activation of Gene Expression Is Adversely Affected at High Multiplicities of Linked Simian Virus 40 Enhancer." *Proceedings of the National Academy of Sciences of the United States of America* 83 (10): 3199–3203.
- Kwon, Hyun Jin, Masahiro Shuda, Huichen Feng, Carlos J. Camacho, Patrick S. Moore, and Yuan Chang. 2013. "Merkel Cell Polyomavirus Small T Antigen Controls Viral Replication and Oncoprotein Expression by Targeting the Cellular Ubiquitin Ligase SCFFbw7." *Cell Host & Microbe* 14 (2): 125–35.
- Langmead, Ben, and Steven L. Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59.
- Lázaro-Frías, Adrián, Sergio Gómez-Medina, Lucas Sánchez-Sampedro, Karl Ljungberg, Mart Ustav, Peter Liljeström, César Muñoz-Fontela, Mariano Esteban, and Juan García-Arriaza. 2018. "Distinct Immunogenicity and Efficacy of Poxvirus-Based Vaccine Candidates against Ebola Virus Expressing GP and VP40 Proteins." *Journal of Virology* 92 (11). <https://doi.org/10.1128/JVI.00363-18>.
- Ledgerwood, Julie E., Adam D. DeZure, Daphne A. Stanley, Emily E. Coates, Laura Novik, Mary E. Enama, Nina M. Berkowitz, et al. 2017. "Chimpanzee Adenovirus Vector Ebola Vaccine." *The New England Journal of Medicine* 376 (10): 928–38.
- Lednicky, J. A., A. S. Arrington, A. R. Stewart, X. M. Dai, C. Wong, S. Jafar, M. Murphey-Corb, and J. S. Butel. 1998. "Natural Isolates of Simian Virus 40 from Immunocompromised Monkeys Display Extensive Genetic Heterogeneity: New Implications for Polyomavirus Disease." *Journal of Virology* 72 (5): 3980–90.
- Leemans, Christ, Marloes C. H. van der Zwalm, Laura Brueckner, Federico Comoglio, Tom

- van Schaik, Ludo Pagie, Joris van Arensbergen, and Bas van Steensel. 2019. "Promoter-Intrinsic and Local Chromatin Features Determine Gene Repression in LADs." *Cell* 177 (4): 852–64.e14.
- Lemaître, Charlène, Anastazja Grabarz, Katerina Tsouroula, Leonid Andronov, Audrey Furst, Tibor Pankotai, Vincent Heyer, et al. 2014. "Nuclear Position Dictates DNA Repair Pathway Choice." *Genes & Development* 28 (22): 2450–63.
- Le, Quy, and Nancy Maizels. 2015. "Cell Cycle Regulates Nuclear Stability of AID and Determines the Cellular Response to AID." *PLoS Genetics* 11 (9): e1005411.
- Le Rouzic, Erwann, Damien Bonnard, Sophie Chasset, Jean-Michel Bruneau, Francis Chevreuil, Frédéric Le Strat, Juliette Nguyen, et al. 2013. "Dual Inhibition of HIV-1 Replication by Integrase-LEDGF Allosteric Inhibitors Is Predominant at the Post-Integration Stage." *Retrovirology* 10 (November):144.
- Leung, Danny C., Kevin B. Dong, Irina A. Maksakova, Preeti Goyal, Ruth Appanah, Sandra Lee, Makoto Tachibana, et al. 2011. "Lysine Methyltransferase G9a Is Required for de Novo DNA Methylation and the Establishment, but Not the Maintenance, of Proviral Silencing." *Proceedings of the National Academy of Sciences of the United States of America* 108 (14): 5718–23.
- Lewinski, Mary K., Masahiro Yamashita, Michael Emernan, Angela Ciuffi, Heather Marshall, Gregory Crawford, Francis Collins, et al. 2006. "Retroviral DNA Integration: Viral and Cellular Determinants of Target-Site Selection." *PLOS Pathogens* 2 (6): e60.
- Lieberman-Aiden, Erez, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, et al. 2009. "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome." *Science (New York, N.Y.)* 326 (5950): 289–93.
- Li, Jing, Xin Wang, Jason Diaz, Sabrina H. Tsang, Christopher B. Buck, and Jianxin You. 2013. "Merkel Cell Polyomavirus Large T Antigen Disrupts Host Genomic Integrity and Inhibits Cellular Proliferation." *Journal of Virology* 87 (16): 9173–88.
- Lilyestrom, Wayne, Michael G. Klein, Rongguang Zhang, Andrzej Joachimiak, and Xiaojiang S. Chen. 2006. "Crystal Structure of SV40 Large T-Antigen Bound to p53: Interplay between a Viral Oncoprotein and a Cellular Tumor Suppressor." *Genes & Development* 20 (17): 2373–82.
- Lombardo, Angelo, Daniela Cesana, Pietro Genovese, Bruno Di Stefano, Elena Provasi, Daniele F. Colombo, Margherita Neri, et al. 2011. "Site-Specific Integration and Tailoring of Cassette Design for Sustainable Gene Transfer." *Nature Methods* 8 (10): 861–69.
- Lombardo, Angelo, Pietro Genovese, Christian M. Beausejour, Silvia Colleoni, Ya-Li Lee, Kenneth A. Kim, Dale Ando, et al. 2007. "Gene Editing in Human Stem Cells Using Zinc Finger Nucleases and Integrase-Defective Lentiviral Vector Delivery." *Nature Biotechnology* 25 (11): 1298–1306.
- Lorincz, M. C., D. Schübeler, S. C. Goeke, M. Walters, M. Groudine, and D. I. Martin. 2000. "Dynamic Analysis of Proviral Induction and De Novo Methylation: Implications for a Histone Deacetylase-Independent, Methylation Density-Dependent Mechanism of Transcriptional Repression." *Molecular and Cellular Biology* 20 (3): 842–50.
- Lötsch, Jörn, and Alfred Ultsch. 2020. "A Non-Parametric Effect-Size Measure Capturing Changes in Central Tendency and Data Distribution Shape." *PloS One* 15 (9): e0239623.
- Loyola, Lorenz, Vasudevan Achuthan, Kathryn Gilroy, Gillian Borland, Anna Kilbey, Nancy Mackay, Margaret Bell, et al. 2019. "Disrupting MLV integrase:BET Protein Interaction Biases Integration into Quiescent Chromatin and Delays but Does Not Eliminate Tumor Activation in a MYC/Runx2 Mouse Model." *PLoS Pathogens* 15 (12): e1008154.
- *Lv, Guanxin, Jianfa Wang, Shuai Lian, Hai Wang, and Rui Wu. 2024. "The Global Epidemiology of Bovine Leukemia Virus: Current Trends and Future Implications." *Animals : An Open Access Journal from MDPI* 14 (2). <https://doi.org/10.3390/ani14020297>.
- Maertens, Goedele, Peter Cherepanov, Wim Pluymers, Katrien Busschots, Erik De Clercq, Zeger Debyser, and Yves Engelborghs. 2003. "LEDGF/p75 Is Essential for Nuclear and Chromosomal Targeting of HIV-1 Integrase in Human Cells." *The Journal of Biological*

- Chemistry* 278 (35): 33528–39.
- Maertens, Goede N. 2016. “B’-Protein Phosphatase 2A Is a Functional Binding Partner of Delta-Retroviral Integrase.” *Nucleic Acids Research* 44 (1): 364–76.
- Magnaghi-Jaulin, L., R. Groisman, I. Naguibneva, P. Robin, S. Lorain, J. P. Le Villain, F. Troalen, D. Trouche, and A. Harel-Bellan. 1998. “Retinoblastoma Protein Represses Transcription by Recruiting a Histone Deacetylase.” *Nature* 391 (6667): 601–5.
- Marquis, Kaitlin A., John Everett, Adrian Cantu, Alexander McFarland, Scott Sherrill-Mix, Mark Krystal, Kyle Parcella, Eric Gillis, Robert A. Fridell, and Frederic D. Bushman. 2024. “The HIV-1 Capsid-Targeted Inhibitor GSK878 Alters Selection of Target Sites for HIV DNA Integration.” *AIDS Research and Human Retroviruses* 40 (2): 114–26.
- Martin, Marcel. 2011. “Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads.” *EMBnet.journal* 17 (1): 10.
- McCormick, F., and E. Harlow. 1980. “Association of a Murine 53,000-Dalton Phosphoprotein with Simian Virus 40 Large-T Antigen in Transformed Cells.” *Journal of Virology* 34 (1): 213–24.
- McDonald, Jessica J., Jukka Alinikula, Jean-Marie Buerstedde, and David G. Schatz. 2013. “A Critical Context-Dependent Role for E Boxes in the Targeting of Somatic Hypermutation.” *Journal of Immunology* 191 (4): 1556–66.
- Meuleman, Wouter, Daan Peric-Hupkes, Jop Kind, Jean-Bernard Beaudry, Ludo Pagie, Manolis Kellis, Marcel Reinders, Lodewyk Wessels, and Bas van Steensel. 2013. “Constitutive Nuclear Lamina-Genome Interactions Are Highly Conserved and Associated with A/T-Rich Sequence.” *Genome Research* 23 (2): 270–80.
- Michieletto, D., M. Lusic, D. Marenduzzo, and E. Orlandini. 2019. “Physical Principles of Retroviral Integration in the Human Genome.” *Nature Communications* 10 (1): 1–11.
- Miklík, Dalibor, Jiří Grim, Daniel Elleder, and Jiří Hejnar. 2023. “Unraveling the Palindromic and Nonpalindromic Motifs of Retroviral Integration Site Sequences by Statistical Mixture Models.” *Genome Research* 33 (8): 1395–1408.
- Miklík, Dalibor, Filip Šenigl, and Jiří Hejnar. 2018. “Proviruses with Long-Term Stable Expression Accumulate in Transcriptionally Active Chromatin Close to the Gene Regulatory Elements: Comparison of ASLV-, HIV- and MLV-Derived Vectors.” *Viruses* 10 (3). <https://doi.org/10.3390/v10030116>.
- Miller, Jeffrey C., Siyuan Tan, Guijuan Qiao, Kyle A. Barlow, Jianbin Wang, Danny F. Xia, Xiangdong Meng, et al. 2011. “A TALE Nuclease Architecture for Efficient Genome Editing.” *Nature Biotechnology* 29 (2): 143–48.
- Mitchell, Rick S., Brett F. Beitzel, Astrid R. W. Schroder, Paul Shinn, Huaming Chen, Charles C. Berry, Joseph R. Ecker, and Frederic D. Bushman. 2004. “Retroviral DNA Integration: ASLV, HIV, and MLV Show Distinct Target Site Preferences.” *PLoS Biology* 2 (8): E234.
- Modlich, Ute, Jens Bohne, Manfred Schmidt, Christof von Kalle, Sabine Knöss, Axel Schambach, and Christopher Baum. 2006. “Cell-Culture Assays Reveal the Importance of Retroviral Vector Design for Insertional Genotoxicity.” *Blood* 108 (8): 2545–53.
- Moosmann, P., O. Georgiev, B. Le Douarin, J. P. Bourquin, and W. Schaffner. 1996. “Transcriptional Repression by RING Finger Protein TIF1 Beta That Interacts with the KRAB Repressor Domain of KOX1.” *Nucleic Acids Research* 24 (24): 4859–67.
- Müller, H. P., and H. E. Varmus. 1994. “DNA Bending Creates Favored Sites for Retroviral Integration: An Explanation for Preferred Insertion Sites in Nucleosomes.” *The EMBO Journal*, October. <https://doi.org/10.1002/j.1460-2075.1994.tb06794.x>.
- Muramatsu, M., K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, and T. Honjo. 2000. “Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme.” *Cell* 102 (5): 553–63.
- Muramatsu, M., V. S. Sankaranand, S. Anant, M. Sugai, K. Kinoshita, N. O. Davidson, and T. Honjo. 1999. “Specific Expression of Activation-Induced Cytidine Deaminase (AID), a Novel Member of the RNA-Editing Deaminase Family in Germinal Center B Cells.” *The Journal of Biological Chemistry* 274 (26): 18470–76.
- Nobre, André, Ieva Kalve, Konstantin Cesnulevicius, Daniela Ragancokova, Andreas Ratzka, Nina Halfer, Maike Wesemann, Klaus Krampfl, Peter Claus, and Claudia Grothe. 2010.

- "Characterization and Differentiation Potential of Rat Ventral Mesencephalic Neuronal Progenitor Cells Immortalized with SV40 Large T Antigen." *Cell and Tissue Research* 340 (1): 29–43.
- Nombela, Ivan, Martine Michiels, Dominique Van Looveren, Lukas Marcelis, Sara El Ashkar, Siska Van Belle, Anne Bruggemans, et al. 2022. "BET-Independent Murine Leukemia Virus Integration Is Retargeted and Selects Distinct Genomic Elements for Lymphomagenesis." *Microbiology Spectrum* 10 (4): e0147822.
- Orphanides, G., G. LeRoy, C. H. Chang, D. S. Luse, and D. Reinberg. 1998. "FACT, a Factor That Facilitates Transcript Elongation through Nucleosomes." *Cell* 92 (1): 105–16.
- Pantulu, N. Deepa, Christian P. Pallasch, Anna Kordelia Kurz, Ahmad Kassem, Lukas Frenzel, Sebastian Sodenkamp, Hans Michael Kvasnicka, Clemens M. Wendtner, and Axel Zur Hausen. 2010. "Detection of a Novel Truncating Merkel Cell Polyomavirus Large T Antigen Deletion in Chronic Lymphocytic Leukemia Cells." *Blood* 116 (24): 5280–84.
- Parisi, Francesca, Niccolò Fonti, Francesca Millanta, Giulia Freer, Mauro Pistello, and Alessandro Poli. 2023. "Exploring the Link between Viruses and Cancer in Companion Animals: A Comprehensive and Comparative Analysis." *Infectious Agents and Cancer* 18 (1): 40.
- Peng, Y. C., and N. H. Acheson. 1998. "Polyomavirus Large T Antigen Binds Cooperatively to Its Multiple Binding Sites in the Viral Origin of DNA Replication." *Journal of Virology* 72 (9): 7330–40.
- Peric-Hupkes, Daan, Wouter Meuleman, Ludo Pagie, Sophia W. M. Bruggeman, Irina Solovei, Wim Brugman, Stefan Gräf, et al. 2010. "Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation." *Molecular Cell* 38 (4): 603–13.
- Petersen-Mahrt, Svend K., Reuben S. Harris, and Michael S. Neuberger. 2002. "AID Mutates E. Coli Suggesting a DNA Deamination Mechanism for Antibody Diversification." *Nature* 418 (6893): 99–103.
- Pham, Phuong, Ronda Bransteitter, John Petruska, and Myron F. Goodman. 2003. "Processive AID-Catalysed Cytosine Deamination on Single-Stranded DNA Simulates Somatic Hypermutation." *Nature* 424 (6944): 103–7.
- Phung, Q. H., D. B. Winter, A. Cranston, R. E. Tarone, V. A. Bohr, R. Fishel, and P. J. Gearhart. 1998. "Increased Hypermutation at G and C Nucleotides in Immunoglobulin Variable Genes from Mice Deficient in the MSH2 Mismatch Repair Protein." *The Journal of Experimental Medicine* 187 (11): 1745–51.
- *Pipas, James M. 2009. "SV40: Cell Transformation and Tumorigenesis." *Virology* 384 (2): 294–303.
- Poletti, Valentina, and Fulvio Mavilio. 2018. "Interactions between Retroviruses and the Host Cell Genome." *Molecular Therapy Methods & Clinical Development* 8 (March): 31–41.
- Porter, Caroline E., Amanda Rosewell Shaw, Youngrock Jung, Tiffany Yip, Patricia D. Castro, Vlad C. Sandulache, Andrew Sikora, et al. 2020. "Oncolytic Adenovirus Armed with BiTE, Cytokine, and Checkpoint Inhibitor Enables CAR T Cells to Control the Growth of Heterogeneous Tumors." *Molecular Therapy: The Journal of the American Society of Gene Therapy* 28 (5): 1251–62.
- Pruss, D., R. Reeves, F. D. Bushman, and A. P. Wolffe. 1994. "The Influence of DNA and Nucleosome Structure on Integration Events Directed by HIV Integrase." *The Journal of Biological Chemistry* 269 (40): 25031–41.
- Qian, Jason, Qiao Wang, Marei Dose, Nathanael Pruett, Kyong-Rim Kieffer-Kwon, Wolfgang Resch, Genqing Liang, et al. 2014. "B Cell Super-Enhancers and Regulatory Clusters Recruit AID Tumorigenic Activity." *Cell* 159 (7): 1524–37.
- Quillent, C., A. M. Borman, S. Paulous, C. Dauguet, and F. Clavel. 1996. "Extensive Regions of Pol Are Required for Efficient Human Immunodeficiency Virus Polyprotein Processing and Particle Maturation." *Virology* 219 (1): 29–36.
- Quinlan, Aaron R., and Ira M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics (Oxford, England)* 26 (6): 841–42.
- Rada, Cristina, Gareth T. Williams, Hilde Nilsen, Deborah E. Barnes, Tomas Lindahl, and

- Michael S. Neuberger. 2002. "Immunoglobulin Isotype Switching Is Inhibited and Somatic Hypermutation Perturbed in UNG-Deficient Mice." *Current Biology: CB* 12 (20): 1748–55.
- Ran, F. Ann, Patrick D. Hsu, Jason Wright, Vineeta Agarwala, David A. Scott, and Feng Zhang. 2013. "Genome Engineering Using the CRISPR-Cas9 System." *Nature Protocols* 8 (11): 2281–2308.
- Rao, Suhas S. P., Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D. Bochkov, James T. Robinson, Adrian L. Sanborn, et al. 2014. "A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping." *Cell* 159 (7): 1665–80.
- Reiff, Sarah B., Andrew J. Schroeder, Koray Kırılı, Andrea Cosolo, Clara Bakker, Luisa Mercado, Soohyun Lee, et al. 2022. "The 4D Nucleome Data Portal as a Resource for Searching and Visualizing Curated Nucleomics Data." *Nature Communications* 13 (1): 2365.
- Roberts, Tiffiney M., and Kathleen Boris-Lawrie. 2003. "Primary Sequence and Secondary Structure Motifs in Spleen Necrosis Virus RU5 Confer Translational Utilization of Unspliced Human Immunodeficiency Virus Type 1 Reporter RNA." *Journal of Virology* 77 (22): 11973–84.
- Robson, Michael I., Jose I. de Las Heras, Rafal Czapiewski, Aishwarya Sivakumar, Alastair R. W. Kerr, and Eric C. Schirmer. 2017. "Constrained Release of Lamina-Associated Enhancers and Genes from the Nuclear Envelope during T-Cell Activation Facilitates Their Association in Chromosome Compartments." *Genome Research* 27 (7): 1126–38.
- Rothemejer, Frederik Holm, Nanna Pi Lauritsen, Anna Karina Juhl, Mariane Høgsbjerg Schleimann, Saskia König, Ole Schmeltz Søgaard, Rasmus O. Bak, and Martin Tolstrup. 2023. "Development of HIV-Resistant CAR T Cells by CRISPR/Cas-Mediated CAR Integration into the Locus." *Viruses* 15 (1). <https://doi.org/10.3390/v15010202>.
- *Rotondo, John Charles, Elisa Mazzoni, Ilaria Bononi, Mauro Tognon, and Fernanda Martini. 2019. "Association Between Simian Virus 40 and Human Tumors." *Frontiers in Oncology* 9 (July): 670.
- Rowe, Helen M., Marc Friedli, Sandra Offner, Sonia Verp, Daniel Mesnard, Julien Marquis, Tugce Aktas, and Didier Trono. 2013. "De Novo DNA Methylation of Endogenous Retroviruses Is Shaped by KRAB-ZFPs/KAP1 and ESET." *Development* 140 (3): 519–29.
- Safina, Alfiya, Henry Garcia, Mairead Commane, Olga Guryanova, Seamus Degan, Kateryna Kolesnikova, and Katerina V. Gurova. 2013. "Complex Mutual Regulation of Facilitates Chromatin Transcription (FACT) Subunits on Both mRNA and Protein Levels in Human Cells." *Cell Cycle* 12 (15): 2423–34.
- Sauer, B., and N. Henderson. 1990. "Targeted Insertion of Exogenous DNA into the Eukaryotic Genome by the Cre Recombinase." *The New Biologist* 2 (5): 441–49.
- Schambach, Axel, Jens Bohne, Saurabh Chandra, Elke Will, Geoffrey P. Margison, David A. Williams, and Christopher Baum. 2006. "Equal Potency of Gammaretroviral and Lentiviral SIN Vectors for Expression of O6-Methylguanine-DNA Methyltransferase in Hematopoietic Cells." *Molecular Therapy : The Journal of the American Society of Gene Therapy* 13 (2): 391–400.
- *Schrader, Carol E., Jeroen E. J. Guikema, Xiaoming Wu, and Janet Stavnezer. 2009. "The Roles of APE1, APE2, DNA Polymerase Beta and Mismatch Repair in Creating S Region DNA Breaks during Antibody Class Switch." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 364 (1517): 645–52.
- Schrader, Carol E., Erin K. Linehan, Sofia N. Mochegova, Robert T. Woodland, and Janet Stavnezer. 2005. "Inducible DNA Breaks in Ig S Regions Are Dependent on AID and UNG." *The Journal of Experimental Medicine* 202 (4): 561–68.
- Schultz, David C., Kasirajan Ayyanathan, Dmitri Negorev, Gerd G. Maul, and Frank J. Rauscher 3rd. 2002. "SETDB1: A Novel KAP-1-Associated Histone H3, Lysine 9-Specific Methyltransferase That Contributes to HP1-Mediated Silencing of Euchromatic Genes by KRAB Zinc-Finger Proteins." *Genes & Development* 16 (8): 919–32.
- Senavirathne, Gayan, James London, Anne Gardner, Richard Fishel, and Kristine E. Yoder. 2023. "DNA Strand Breaks and Gaps Target Retroviral Intasome Binding and Integration."

- Nature Communications* 14 (1): 7072.
- Senigl, Filip, Miroslav Auxt, and Jirí Hejnar. 2012. "Transcriptional Provirus Silencing as a Crosstalk of de Novo DNA Methylation and Epigenomic Features at the Integration Site." *Nucleic Acids Research* 40 (12): 5298–5312.
- Senigl, Filip, Yaakov Maman, Ravi K. Dinesh, Jukka Alinikula, Rashu B. Seth, Lubomira Pecnova, Arina D. Omer, et al. 2019. "Topologically Associated Domains Delineate Susceptibility to Somatic Hypermutation." *Cell Reports* 29 (12): 3902–15.e8.
- Šenigl, Filip, Dalibor Miklík, Miroslav Auxt, and Jirí Hejnar. 2017. "Accumulation of Long-Term Transcriptionally Active Integrated Retroviral Vectors in Active Promoters and Enhancers." *Nucleic Acids Research* 45 (22): 12752–65.
- Senigl, Filip, Jirí Plachý, and Jirí Hejnar. 2008. "The Core Element of a CpG Island Protects Avian Sarcoma and Leukosis Virus-Derived Vectors from Transcriptional Silencing." *Journal of Virology* 82 (16): 7818–27.
- Seo, Gil Ju, Chun Jung Chen, and Christopher S. Sullivan. 2009. "Merkel Cell Polyomavirus Encodes a microRNA with the Ability to Autoregulate Viral Gene Expression." *Virology* 383 (2): 183–87.
- Shi, Jiong, Jing Zhou, Vaibhav B. Shah, Christopher Aiken, and Kevin Whitby. 2011. "Small-Molecule Inhibition of Human Immunodeficiency Virus Type 1 Infection by Virus Capsid Destabilization." *Journal of Virology* 85 (1): 542–49.
- Shimotohno, K., S. Mizutani, and H. M. Temin. 1980. "Sequence of Retrovirus Provirus Resembles that of Bacterial Transposable Elements." *Nature* 285 (5766): 550–54.
- Shrestha, Dewan, Aishee Bag, Ruiqiong Wu, Yeting Zhang, Xing Tang, Qian Qi, Jinchuan Xing, and Yong Cheng. 2022. "Genomics and Epigenetics Guided Identification of Tissue-Specific Genomic Safe Harbors." *Genome Biology* 23 (1): 1–17.
- Shuda, Masahiro, Huichen Feng, Hyun Jin Kwun, Steven T. Rosen, Ole Gjoerup, Patrick S. Moore, and Yuan Chang. 2008. "T Antigen Mutations Are a Human Tumor-Specific Signature for Merkel Cell Polyomavirus." *Proceedings of the National Academy of Sciences of the United States of America* 105 (42): 16272–77.
- Shun, Ming-Chieh, Nidhanapati K. Raghavendra, Nick Vandegraaff, Janet E. Daigle, Siobhan Hughes, Paul Kellam, Peter Cherepanov, and Alan Engelman. 2007. "LEDGF/p75 Functions Downstream from Preintegration Complex Formation to Effect Gene-Specific HIV-1 Integration." *Genes & Development* 21 (14): 1767–78.
- Siliciano, Janet D., Joleen Kajdas, Diana Finzi, Thomas C. Quinn, Karen Chadwick, Joseph B. Margolick, Colin Kovacs, Stephen J. Gange, and Robert F. Siliciano. 2003. "Long-Term Follow-up Studies Confirm the Stability of the Latent Reservoir for HIV-1 in Resting CD4+ T Cells." *Nature Medicine* 9 (6): 727–28.
- Simmons, D. T., T. Melendy, D. Usher, and B. Stillman. 1996. "Simian Virus 40 Large T Antigen Binds to Topoisomerase I." *Virology* 222 (2): 365–74.
- Soikkeli, Anni I., Minna K. Kyläniemi, Harri Sihto, and Jukka Alinikula. 2022. "Oncogenic Merkel Cell Polyomavirus T Antigen Truncating Mutations Are Mediated by APOBEC3 Activity in Merkel Cell Carcinoma." *Cancer Research Communications* 2 (11): 1344–54.
- Sowd, Gregory A., Nancy Yan Li, and Ellen Fanning. 2013. "ATM and ATR Activities Maintain Replication Fork Integrity during SV40 Chromatin Replication." *PLoS Pathogens* 9 (4): e1003283.
- Sowd, Gregory A., Dviti Mody, Joshua Eggold, David Cortez, Katherine L. Friedman, and Ellen Fanning. 2014. "SV40 Utilizes ATM Kinase Activity to Prevent Non-Homologous End Joining of Broken Viral DNA Replication Products." *PLoS Pathogens* 10 (12): e1004536.
- Sowd, Gregory A., Erik Serrao, Hao Wang, Weifeng Wang, Hind J. Fadel, Eric M. Poeschla, and Alan N. Engelman. 2016. "A Critical Role for Alternative Polyadenylation Factor CPSF6 in Targeting HIV-1 Integration to Transcriptionally Active Chromatin." *Proceedings of the National Academy of Sciences* 113 (8): E1054–63.
- Spurgeon, Megan E., Jingwei Cheng, Ella Ward-Shaw, Frederick A. Dick, James A. DeCaprio, and Paul F. Lambert. 2022. "Merkel Cell Polyomavirus Large T Antigen Binding to pRb Promotes Skin Hyperplasia and Tumor Development." *PLoS Pathogens* 18 (5): e1010551.

- Sripathy, Smitha P., Jessica Stevens, and David C. Schultz. 2006. "The KAP1 Corepressor Functions to Coordinate the Assembly of de Novo HP1-Demarcated Microenvironments of Heterochromatin Required for KRAB Zinc Finger Protein-Mediated Transcriptional Repression." *Molecular and Cellular Biology* 26 (22): 8623–38.
- Stading, Ben, James A. Ellison, William C. Carson, Panayampalli Subbian Satheshkumar, Tonie E. Rocke, and Jorge E. Osorio. 2017. "Protection of Bats (*Eptesicus fuscus*) against Rabies Following Topical or Oronasal Exposure to a Recombinant Raccoon Poxvirus Vaccine." *PLoS Neglected Tropical Diseases* 11 (10): e0005958.
- Studebaker, Adam W., Brian J. Hutzen, Christopher R. Pierson, Kellie B. Haworth, Timothy P. Cripe, Eric M. Jackson, and Jeffrey R. Leonard. 2017. "Oncolytic Herpes Virus rRp450 Shows Efficacy in Orthotopic Xenograft Group 3/4 Medulloblastomas and Atypical Teratoid/Rhabdoid Tumors." *Molecular Therapy Oncolytics* 6 (September):22–30.
- Suchy, Fabian P., Daiki Karigane, Yusuke Nakauchi, Maimi Higuchi, Jinyu Zhang, Katja Pekrun, Ian Hsu, et al. 2024. "Genome Engineering with Cas9 and AAV Repair Templates Generates Frequent Concatemeric Insertions of Viral Vectors." *Nature Biotechnology*, April, 1–10.
- Suerth, Julia D., Tobias Maetzig, Martijn H. Brugman, Niels Heinz, Jens-Uwe Appelt, Kerstin B. Kaufmann, Manfred Schmidt, et al. 2012. "Alpharetroviral Self-Inactivating Vectors: Long-Term Transgene Expression in Murine Hematopoietic Cells and Low Genotoxicity." *Molecular Therapy: The Journal of the American Society of Gene Therapy* 20 (5): 1022–32.
- Sullivan, Christopher S., Chang K. Sung, Christopher D. Pack, Adam Grundhoff, Aron E. Lukacher, Thomas L. Benjamin, and Don Ganem. 2009. "Murine Polyomavirus Encodes a microRNA That Cleaves Early RNA Transcripts but Is Not Essential for Experimental Infection." *Virology* 387 (1): 157–67.
- Sweet, B. H., and M. R. Hilleman. 1960. "The Vacuolating Virus, S.V.40." *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine*, November. <https://doi.org/10.3181/00379727-105-26128>.
- Tarlington, Rachael, Joanne Meers, Jon Hanger, and Paul Young. 2005. "Real-Time Reverse Transcriptase PCR for the Endogenous Koala Retrovirus Reveals an Association between Plasma Viral Load and Neoplastic Disease in Koalas." *The Journal of General Virology* 86 (Pt 3): 783–87.
- Tasan, Ipek, Gabriela Sustackova, Liguozhang, Jiah Kim, Mayandi Sivaguru, Mohammad Hamedirad, Yuchuan Wang, et al. 2018. "CRISPR/Cas9-Mediated Knock-in of an Optimized TetO Repeat for Live Cell Imaging of Endogenous Loci." *Nucleic Acids Research* 46 (17): e100.
- Teixeira, Ana Rafaela, Cintia Bittar, Gabriela S. Silva Santos, Thiago Y. Oliveira, Amy S. Huang, Noemi Linden, Isabella A. T. M. Ferreira, et al. 2024. "Transcription of HIV-1 at Sites of Intact Latent Provirus Integration." *The Journal of Experimental Medicine* 221 (9). <https://doi.org/10.1084/jem.20240391>.
- Tsao, Sai Wah, Xianghong Wang, Yu Liu, Yuk Chun Cheung, Huichen Feng, Zhong Zheng, Natalie Wong, et al. 2002. "Establishment of Two Immortalized Nasopharyngeal Epithelial Cell Lines Using SV40 Large T and HPV16E6/E7 Viral Oncogenes." *Biochimica et Biophysica Acta* 1590 (1-3): 150–58.
- Tsurimoto, T., T. Melendy, and B. Stillman. 1990. "Sequential Initiation of Lagging and Leading Strand Synthesis by Two Different Polymerase Complexes at the SV40 DNA Replication Origin." *Nature* 346 (6284): 534–39.
- *Chameettachal, A., Farah Mustafa, Tahir A. Rizvi. 2023. "Understanding Retroviral Life Cycle and Its Genomic RNA Packaging." *Journal of Molecular Biology* 435 (3): 167924.
- Unniraman, Shyam, and David G. Schatz. 2007. "Strand-Biased Spreading of Mutations during Somatic Hypermutation." *Science* 317 (5842): 1227–30.
- Vanegas, Maria, Manuel Llano, Sharon Delgado, Daniah Thompson, Mary Peretz, and Eric Poeschla. 2005. "Identification of the LEDGF/p75 HIV-1 Integrase-Interaction Domain and NLS Reveals NLS-Independent Chromatin Tethering." *Journal of Cell Science* 118 (Pt 8): 1733–43.

- Van Looveren, Dominique, Giorgia Giacomazzi, Irina Thiry, Maurilio Sampaolesi, and Rik Gijbsbers. 2021. "Improved Functionality and Potency of next Generation BinMLV Viral Vectors toward Safer Gene Therapy." *Molecular Therapy. Methods & Clinical Development* 23 (December):51–67.
- Vansant, Gerlinde, Heng-Chang Chen, Eduard Zorita, Katerina Trejbalová, Dalibor Miklík, Guillaume Filion, and Zeger Debyser. 2020. "The Chromatin Landscape at the HIV-1 Provirus Integration Site Determines Viral Expression." *Nucleic Acids Research* 48 (14): 7801–17.
- Vansant, Gerlinde, Lenard S. Vranckx, Irena Zurnic, Dominique Van Looveren, Paulien Van de Velde, Christopher Nobles, Rik Gijbsbers, Frauke Christ, and Zeger Debyser. 2019. "Impact of LEDGIN Treatment during Virus Production on Residual HIV-1 Transcription." *Retrovirology* 16 (1): 8.
- Ventosa, Maria, Zetang Wu, and Filip Lim. 2017. "Sustained FXN Expression in Dorsal Root Ganglia from a Nonreplicative Genomic HSV-1 Vector." *The Journal of Gene Medicine* 19 (11): 376–86.
- Vranckx, Lenard S., Jonas Demeulemeester, Suha Saleh, Annegret Boll, Gerlinde Vansant, Rik Schrijvers, Caroline Weydert, et al. 2016. "LEDGIN-Mediated Inhibition of Integrase-LEDGF/p75 Interaction Reduces Reactivation of Residual Latent HIV." *EBioMedicine* 8 (June):248–64.
- Waga, S., G. Bauer, and B. Stillman. 1994. "Reconstitution of Complete SV40 DNA Replication with Purified Replication Factors." *The Journal of Biological Chemistry* 269 (14): 10923–34.
- Wang, Cheng, and Stephen P. Goff. 2017. "Differential Control of Retrovirus Silencing in Embryonic Cells by Proteasomal Regulation of the ZFP809 Retroviral Repressor." *Proceedings of the National Academy of Sciences of the United States of America* 114 (6): E922–30.
- Wang, Min, Jenny Jokinen, Irina Tretyakova, Peter Pushko, and Igor S. Lukashevich. 2018. "Alphavirus Vector-Based Replicon Particles Expressing Multivalent Cross-Protective Lassa Virus Glycoproteins." *Vaccine* 36 (5): 683–90.
- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer.
- Wiesendanger, M., B. Kneitz, W. Edelmann, and M. D. Scharff. 2000. "Somatic Hypermutation in MutS Homologue (MSH)3-, MSH6-, and MSH3/MSH6-Deficient Mice Reveals a Role for the MSH2-MSH6 Heterodimer in Modulating the Base Substitution Pattern." *The Journal of Experimental Medicine* 191 (3): 579–84.
- Wiley, S. R., R. J. Kraus, F. Zuo, E. E. Murray, K. Loritz, and J. E. Mertz. 1993. "SV40 Early-to-Late Switch Involves Titration of Cellular Transcriptional Repressors." *Genes & Development* 7 (11): 2206–19.
- Winans, Shelby, Ross C. Larue, Carly M. Abraham, Nikolozi Shkriabai, Amelie Skopp, Duane Winkler, Mamuka Kvaratskhelia, and Karen L. Beemon. 2017. "The FACT Complex Promotes Avian Leukosis Virus DNA Integration." *Journal of Virology* 91 (7). <https://doi.org/10.1128/JVI.00082-17>.
- Winans, Shelby, Hyun Jae Yu, Kenia de Los Santos, Gary Z. Wang, Vineet N. KewalRamani, and Stephen P. Goff. 2022. "A Point Mutation in HIV-1 Integrase Redirects Proviral Integration into Centromeric Repeats." *Nature Communications* 13 (1): 1474.
- Wolf, Daniel, and Stephen P. Goff. 2009. "Embryonic Stem Cells Use ZFP809 to Silence Retroviral DNAs." *Nature* 458 (7242): 1201–4.
- Wu, X., H. Liu, H. Xiao, J. A. Conway, E. Hehl, G. V. Kalpana, V. Prasad, and J. C. Kappes. 1999. "Human Immunodeficiency Virus Type 1 Integrase Protein Promotes Reverse Transcription through Specific Interactions with the Nucleoprotein Reverse Transcription Complex." *Journal of Virology* 73 (3): 2126–35.
- Xu, L., J. K. Yee, J. A. Wolff, and T. Friedmann. 1989. "Factors Affecting Long-Term Stability of Moloney Murine Leukemia Virus-Based Vectors." *Virology* 171 (2): 331–41.
- Xu, Wenqin, Cynthia K. Stadler, Kristen Gorman, Nathaniel Jensen, David Kim, Haoqiang Zheng, Shaohua Tang, William M. Switzer, Geoffrey W. Pye, and Maribeth V. Eiden. 2013. "An Exogenous Retrovirus Isolated from Koalas with Malignant Neoplasias in a US

- Zoo." *Proceedings of the National Academy of Sciences of the United States of America* 110 (28): 11547–52.
- Yoshikawa, Kiyotsugu, Il-Mi Okazaki, Tomonori Eto, Kazuo Kinoshita, Masamichi Muramatsu, Hitoshi Nagaoka, and Tasuku Honjo. 2002. "AID Enzyme-Induced Hypermutation in an Actively Transcribed Gene in Fibroblasts." *Science* 296 (5575): 2033–36.
- Zhao, Jiawei, Yuemeng Jia, Shunli Shen, Jiwoong Kim, Xun Wang, Eunice Lee, Isaac Brownell, et al. 2020. "Merkel Cell Polyomavirus Small T Antigen Activates Noncanonical NF- κ B Signaling to Promote Tumorigenesis." *Molecular Cancer Research: MCR* 18 (11): 1623–37.
- Zufferey, R., T. Dull, R. J. Mandel, A. Bukovsky, D. Quiroz, L. Naldini, and D. Trono. 1998. "Self-Inactivating Lentivirus Vector for Safe and Efficient in Vivo Gene Delivery." *Journal of Virology* 72 (12): 9873–80.
- Zuo, Xiaopan, Jipo Sheng, Ho-Tak Lau, Carol M. McDonald, Monica Andrade, Dana E. Cullen, Fong T. Bell, et al. 2012. "Zinc Finger Protein ZFP57 Requires Its Co-Factor to Recruit DNA Methyltransferases and Maintains DNA Methylation Imprint in Embryonic Stem Cells via Its Transcriptional Repression Domain." *The Journal of Biological Chemistry* 287 (3): 2107–18.
- Zychlinski, Daniela, Axel Schambach, Ute Modlich, Tobias Maetzig, Johann Meyer, Elke Grassman, Anjali Mishra, and Christopher Baum. 2008. "Physiological Promoters Reduce the Genotoxic Risk of Integrating Gene Vectors." *Molecular Therapy : The Journal of the American Society of Gene Therapy* 16 (4): 718–25.